Available online at www.sciencedirect.com

SCIENCE DIRECT

Gene 337 (2004) 189–198

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

Sequencing and analysis of the large virulence plasmid pLVPK of *Klebsiella pneumoniae* CG43

Ying-Tsong Chen[a], Hwan-You Chang[b], Yi-Chyi Lai[b], Chao-Chi Pan[b], Shih-Feng Tsai[c,d], Hwei-Ling Peng[a,*]

[a] *Department of Biological Science and Technology, National Chiao Tung University, 75, Po-Ai Street, Hsinchu, Taiwan, ROC*
[b] *Institute of Molecular Medicine, National Tsing Hua University, Hsinchu, Taiwan, ROC*
[c] *Institute of Genetics, National Yang-Ming University, Taipei, Taiwan, ROC*
[d] *Division of Molecular and Genomic Medicine, National Health Research Institutes, Taipei, Taiwan, ROC*

Received 18 February 2004; received in revised form 26 April 2004; accepted 6 May 2004

Received by J.A. Engler

Abstract

We have determined the entire DNA sequence of pLVPK, which is a 219-kb virulence plasmid harbored in a bacteremic isolate of *Klebsiella pneumoniae*. A total of 251 open reading frames (ORFs) were annotated, of which 37% have homologous genes of known function, 31% match the hypothetical genes in the GenBank database, and the remaining 32% are novel sequences. The obvious virulence-associated genes carried by the plasmid are the capsular polysaccharide synthesis regulator *rmpA* and its homolog *rmpA2*, and multiple iron-acquisition systems, including *iucABCDiutA* and *iroBCDN* siderophore gene clusters, *Mesorhizobium loti fepBC* ABC-type transporter, and *Escherichia coli fecIRA*, which encodes a Fur-dependent regulatory system for iron uptake. In addition, several gene clusters homologous with copper, silver, lead, and tellurite resistance genes of other bacteria were also identified. Identification of a replication origin consisting of a *repA* gene lying in between two sets of iterons suggests that the replication of pLVPK is iteron-controlled and the iterons are the binding sites for the *repA* to initiate replication and maintain copy number of the plasmid. Genes homologous with *E. coli sopA/sopB* and *parA/parB* with nearby direct DNA repeats were also identified indicating the presence of an F plasmid-like partitioning system. Finally, the presence of 13 insertion sequences located mostly at the boundaries of the aforementioned gene clusters suggests that pLVPK was derived from a sequential assembly of various horizontally acquired DNA fragments.

Keywords: *rmpA*; Siderophore; Heavy-metal resistance; Tellurite resistance; Iteron

1. Introduction

Klebsiella pneumoniae is an important cause of community-acquired bacterial pneumonia, occurring particularly in chronic alcoholics and commonly results in a high fatality rate if untreated. Nevertheless, the vast majority of *K. pneumoniae* infections are associated with hospitalization. It has been estimated that *K. pneumoniae* causes up to 8% of all nosocomial bacterial infections in developed countries, and its colonization in hospitalized patients appears to be associated with the use of antibiotics (Schaberg et al., 1991). Recently, the prevalence of multiple-drug-resistant *K. pneumoniae* strains has significantly restricted the availability of antibiotics for effective treatment of the bacterial infections.

Despite its significance, our knowledge of the pathogenicity of the bacterium is rather limited. Clinically isolated *K. pneumoniae* usually produces large amounts of capsular polysaccharides (CPS) as reflected by the formation of glistening mucoid colonies. The CPS provides the bacterium an anti-phagocytic ability and prevents the bacteria from being killed by serum bactericidal factors (Simmons-Smit et al., 1986). Additional virulence-associated factors identified so far in *K. pneumoniae* include lipopolysaccharides, several adhesins, and iron-acquisition systems (Simmons-Smit et al., 1986; Nassif and Sansonetti, 1986). The small numbers of

Abbreviations: bp, base pair(s); IS, insertion sequence(s); kb, kilobase(s); LB, Luria–Bertani (medium); ORF, open reading frame; ori, origin of DNA replication.

* Corresponding author. Tel.: +886-3-5742910; fax: +886-3-5742909.
E-mail address: hlpeng@cc.nctu.edu.tw (H.-L. Peng).

0378-1119/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2004.05.008

known virulence-associated factors rather limit the possible targets for drug development, thus making the intervention of bacterial infection rather difficult.

Several strategies including in vivo expression technology, subtractive DNA hybridization, and signature-tagged mutagenesis have been adopted to identify virulence-associated genes in *K. pneumoniae*. These efforts have allowed the identification of many novel genes that might be important for the bacterium to infect humans. For instance, by using the in vivo expression technology, we have identified the presence of a plasmid-borne iron-acquisition gene cluster in *K. pneumoniae* that is primarily expressed in the hosts (Lai et al., 2001). Nevertheless, further investigation of the functional roles of these novel sequences has been significantly hampered by the lack of the complete genome sequence of *K. pneumoniae*.

Most of the blood isolates of *K. pneumoniae* harbor a large plasmid of 200 kb in size (Peng et al., 1991). The plasmid has been demonstrated to contain the aerobactin siderophore biosynthesis genes and curing of the plasmid would result in an avirulent phenotype (Nassif and Sansonetti, 1986). In our laboratory, we also found that the loss of pLVPK, a plasmid of the similar size harbored in *K. pneumoniae* CG43, a highly virulent clinical isolate of K2 serotype (Lai et al., 2003), resulted in a loss of colony mucoidy, the ability to synthesize aerobactin, and a 1000-fold decrease of virulence. It is conceivable that the plasmid is likely to carry many additional virulence-associated genes and complete sequencing of the plasmid would hence be the most straightforward way for their identification. We herein report the 219-kb sequence and annotation of this large virulence plasmid from *K. pneumoniae* CG43.

2. Materials and methods

2.1. Sequencing of pLVPK

The DNA of pLVPK was isolated from *K. pneumoniae* CG43 by using a Qiagen Plasmid Purification kit and fragmented by sonication. The DNA fragments were then resolved on a 0.7% low melting point agarose gel and DNA of size ranging from 2.0 to 3.0 kb were recovered, blunt-repaired by *Bal*31 nuclease, and subsequently cloned into the pUC18 vector. A total of 2304 clones were sequenced from both ends to achieve approximately 11-fold coverage of the plasmid. Sequences were assembled initially using the Phred/Phrap program (Ewing et al., 1998) with optimized parameters and the quality score was set to >20. When all the sequences assembled into 11 major contigs (>20 reads; >2 kb), the Consed program (Gordon et al., 1998) was then used for the final sequence closure (autofinishing). Finally, several gaps among contigs were closed either by primer walking on selected clones, which were identified by analysis on the forward and the reverse links of each of the contigs, or by sequencing the DNA amplicons generated by PCR.

2.2. Gene prediction and annotation

GLIMMER 2.02 (Delcher et al., 1999), a program that searches for protein coding regions, was used to identify those ORFs possessing more than 30 codons. Overlapping and closely clustered ORFs were manually inspected. The predicted polypeptide sequences were used to search the protein database with the BLAST (NCBI database), and the clusters of orthologous groups (COGs) of proteins database was used to identify families to which the predicted proteins were related. Mobile elements and repetitive sequences were identified using pairwise comparison with the known insertion sequences. The presence of tRNA sequences was identified by the program tRNAscan-SE (Lowe and Todd, 1997). The G+C nucleotide composition analysis was made by GCWin of the G-Language package (Arakawa et al., 2003).

2.3. Drug susceptibility assay

Tellurite, copper, silver, and lead susceptibility for the strains were determined essentially as described (Menoharan et al., 2003). *E. coli, K. pneumoniae* CG43, and its derivatives were propagated at 37 °C in Luria–Bertani (LB) broth. The overnight-grown cells were spread onto LB plates and the 3MM paper discs (5-mm diameter) impregnated with aliquots of a serial dilution of K_2TeO_3, $CuSO_4$, $AgNO_3$, and $Pb(NO_3)_2$ solutions were placed on top of each of the plates. The plates were then incubated at 37 °C for another 12 h and the inhibition zone was measured. Iron-acquisition activity was assayed using iron-deprived M9 plates (with 200 μM 2,2′-dipyridyl) and the paper discs were impregnated with a serial dilution of $FeCl_3$ solution. After spreading the overnight-grown bacteria onto the plates, the iron-loaded discs were then placed on top of each of the plates. The plates were incubated at 37 °C for 12 h, and the growth zones around the paper discs were measured.

2.4. Nucleotide sequence accession number

The nucleotide sequences reported in this paper have been submitted to GenBank under the accession no. AY378100.

3. Results and discussion

3.1. General overview

The entire DNA sequence consists of 219,385 bp forming a circular plasmid (Fig. 1). The size and the predicted restriction enzyme cutting sites are consistent with the ex-

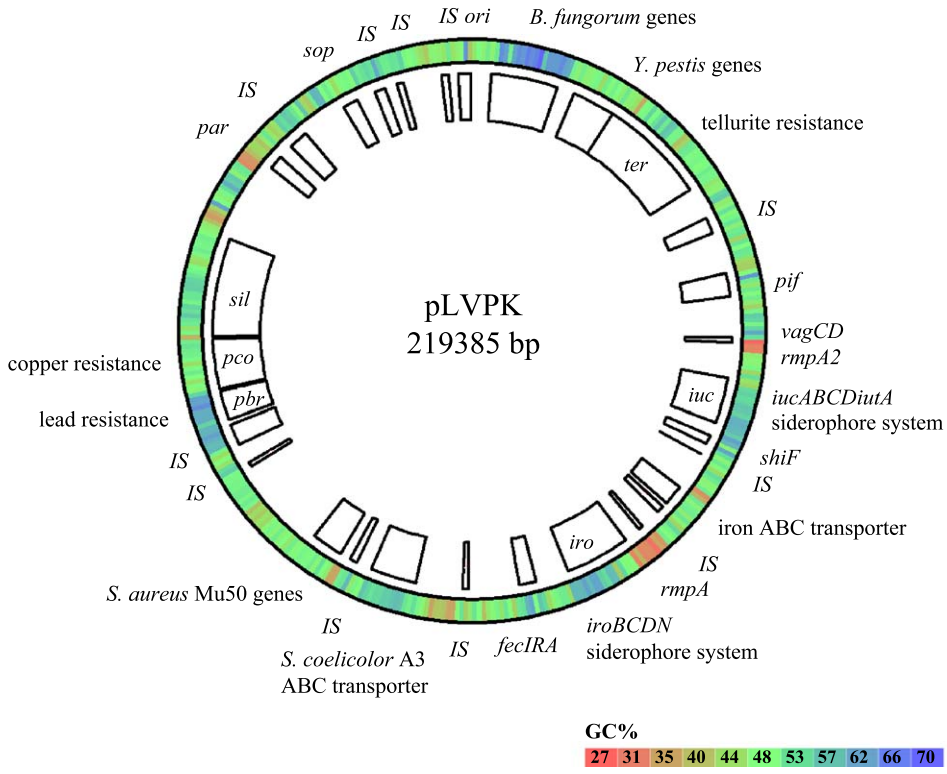

Fig. 1. Map of the pLVPK plasmid. The circle map shows the major features of the plasmid including the replication origins (*ori*), partition regions (*par, sop*), major gene clusters, and *IS* elements (*IS*). The positions of some of the ORFs and gene clusters are depicted in box and their contents were labeled. Gene clusters, which are homologous unknown gene clusters reported from other species, are labeled with the species name respectively. The G + C% contents along the plasmid are shown in colors from low% (red) to high% (purple).

perimental findings using pulse-field gel electrophoresis. The plasmid contains 251 ORFs, as determined by the Glimmer program. The possible functions of these ORFs were subsequently analyzed by comparing the sequence to the current nonredundant protein database of the National Center for Biotechnology Information using BLAST software through the Internet. Approximately 37% of the 251 ORFs have significant amino acid sequence similarity (>60%) with the genes of known function in GenBank or with protein domains or motifs in protein databases. Despite their lack of homology to the known genes, the deduced amino acid sequences of 31% of the ORFs matched the hypothetical genes in the database. The remaining 32% had lower or no significant sequence similarities (< 20%) with those in the database and their functions could not be assigned.

The average G + C content of the plasmid is 50.35%, which is somewhat lower than that of the *K. pneumoniae* MGH78578 genome (G + C = ~ 55%). The G + C content plotted along the pLVPK sequence with a window size of 1000 bp is shown in Fig. 2. Four regions (Box 1–4) with a significant high G + C content in comparison with the average of the whole plasmid sequence were identified. The Box 1 consists of 9 ORFs showing 56–90% sequence similarity to an unknown gene cluster in *Burkholderia fungorum* genome. The second and third high G + C regions contain two iron-acquisition systems: *iut* and *iro* genes, respectively. The fourth box covered the lead-resistant *pbr* gene cluster and its nearby transposase gene. Two low G + C content regions are also marked in Fig. 2, which include the two mucoidy regulator encoding genes, *rmpA* (34.6%) and *rmpA2* (31.9%). The values of G + C at the third codon are even lower with 29.2% for *rmpA* and 28% for *rmpA2*.

3.2. Virulence-associated genes

The BLAST search revealed an 18-kb region, which is highly similar to the SHI-2 pathogenicity island (PAI) of *Shigella flexneri* (Moss et al., 1999). The SHI-2-like region includes the iron-acquisition genes *iucABCDiutA*, *vagCD*, the unknown function ORF *shiF*, and *rmpA2*, a known virulence-associated gene in *K. pneumoniae* (Lai et al., 2003). Elsewhere the PAI-like region, a *rmpA2* homolog, *rmpA*, and two additional gene clusters associated with iron metabolism were also found.

One interesting finding in pLVPK is the presence of *rmpA* and *rmpA2*, two genes encoding regulatory proteins for CPS synthesis in *K. pneumoniae*. CPS has been known to be a major virulence factor in *K. pneumoniae* that protects the bacterium from the bactericidal activity of serum complements and macrophages (Simmons-Smit et al., 1986). The gene *rmpA* was first identified in *K. pneumoniae* as a determinant controlling the CPS biosynthesis (Nassif et al.,

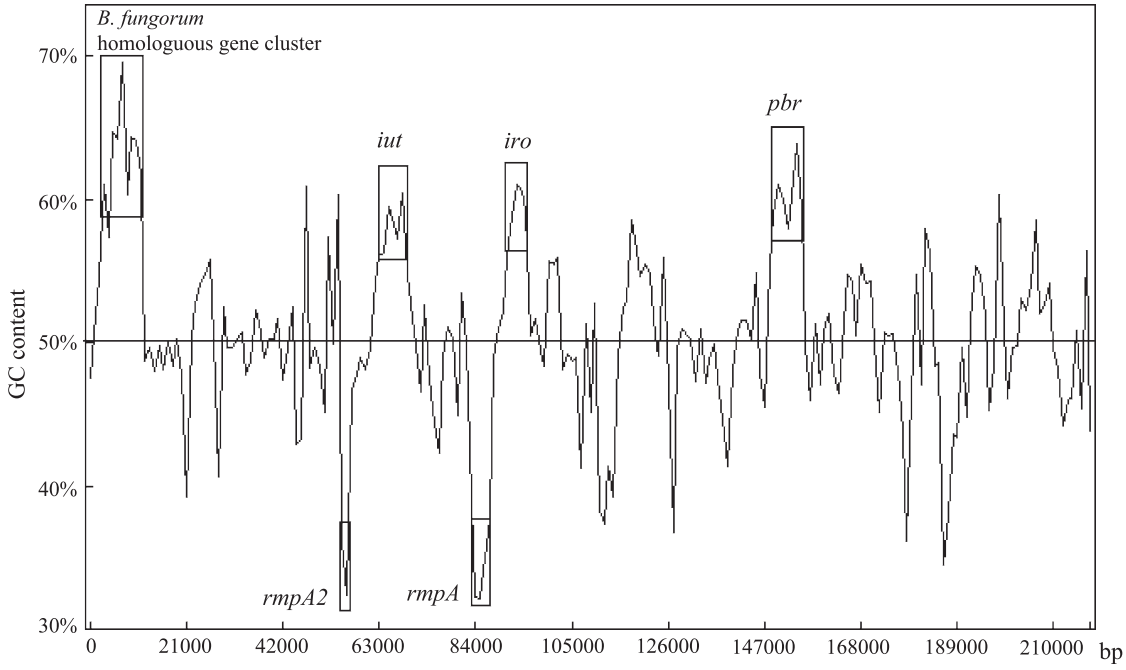

Fig. 2. Base composition of pLVPK. The G+C content along the sequence of pLVPK is calculated with a window of 1000 bp. The horizontal line indicates 50% G+C content, and the selected ORFs are shown as open boxes drawn to the exact scale. The region that contained the genes similar to that of the *B. fungorum* gene cluster is labeled. The *iut* and *iro* siderophore gene clusters, the lead-resistance gene clusters (*pbr*), and the region of *rmpA* and *rmpA2* genes, respectively, are also indicated.

1989). The gene *rmpA2*, which was named because of its high similarity with *rmpA*, was identified later (Wacharotayankun et al., 1993). Since the major difference between these two gene products is that the RmpA2 has an extended N-terminal region, it has been generally thought that *rmpA* and *rmpA2* are the same gene, and the *rmpA* reported earlier by Nassif et al. was a truncated form of *rmpA2*. Our sequencing result shows that *rmpA* and *rmpA2*, which share 81% nucleotide sequence homology (78% identity in the 194 comparable amino acids), are actually two independent loci 29 kb apart (Fig. 3a). Southern hybridization analysis of the plasmid using an *rmpA2* probe also confirmed the presence of two copies of the gene (Fig. 3b). The finding not only clarified that *rmpA* is not a part of *rmpA2*, but also demonstrated that both the genes are plasmid-borne. Our laboratory has recently found that RmpA2 protein directly interacts with the promoters of the K2 CPS biosynthesis genes through its carboxyl terminal helix-turn-helix motif-containing portion (Lai et al., 2003). Thus, we believe that RmpA could also interact with the *cps* gene promoter, although how it activates the *cps* gene expression and the interplay between these two Rmp proteins remain to be investigated.

The *K. pneumoniae vagCD* products exhibit 94% and 84% amino acid sequence identities with that of the VagC and VagD on pR64 of *Salmonella enterica* serovar Dublin. Like the *vagCD* of pR64, the two genes are also overlapped by one nucleotide. It has been proposed that VagC and VagD might be involved in the coordination of plasmid replication and cell division, and disruption of the *vagC* locus would reduce the bacterial virulence (Pullinger and Lax, 1992). The high sequence similarity suggests that *vagCD* genes on the pLVPK also participate in the maintenance of the plasmid stability. Interestingly, the G+C content of the *vagCD* genes (~ 70%) is significantly higher than that of the *rmpA2* (31.9%), which is located only 1.1 kb away, implying that *rmpA2* and *vagCD* were recruited onto pLVPK independently.

3.3. Iron-acquisition systems

The capability of iron acquisition is generally a prerequisite for a pathogen to establish infection when entering the hosts. In pLVPK, two siderophore-mediated iron-acquisition systems, *iucABCDiutA* and *iroBCDN*, were identified. The *iucABCDiutA* operon, which was first reported on pCoIV-K30 in *E. coli* (Ambrozic et al., 1998), consists of five genes responsible for synthesis and transport of the hydroxymate siderophore aerobactin. The presence of the aerobactin synthesis and utilization genes has also been reported for *Salmonella* and *Shigella* spp., indicating that the genes are freely transferable within the Enterobacteriaceae. This notion is also consistent with the finding that the *iucABCDiutA* gene cluster is flanked by two transposable elements, IS*630* and IS*3*, and 3′ sequences of *E. coli* K12 $tRNA^{Lys}$ and $tRNA^{Trp}$, which have been proposed to play a role in the horizontal transfer of PAIs between bacterial pathogens (Hou, 1999).

The *iroBCDEN* gene cluster, first described in *S. enterica*, is known to participate in the uptake of catecholate-type

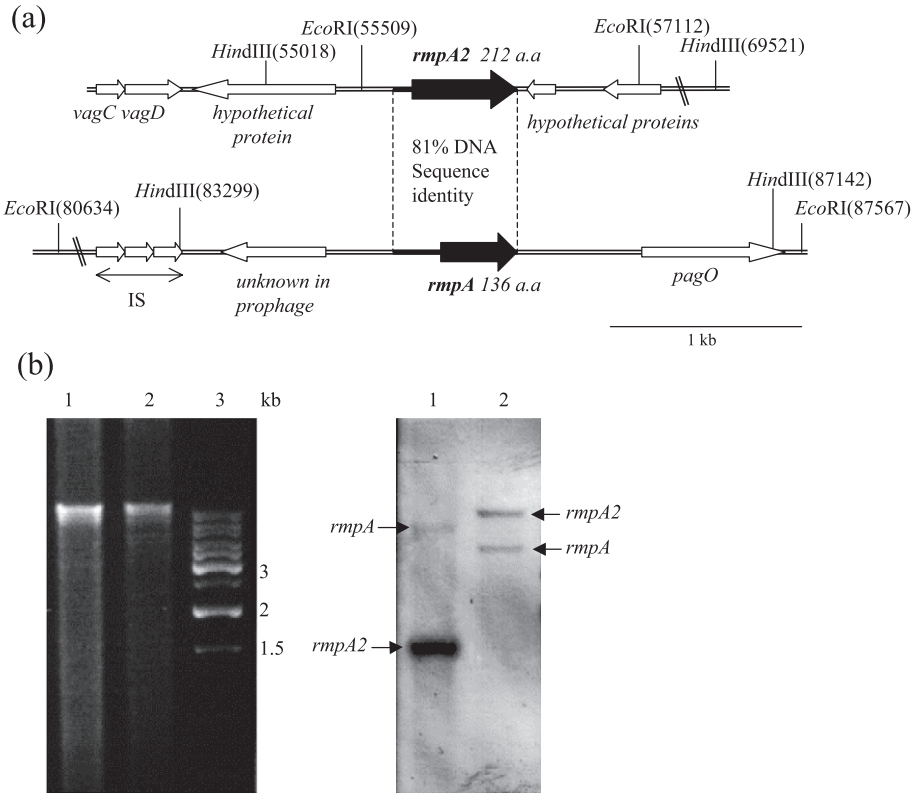

Fig. 3. Comparison maps of the *rmpA* and *rmpA2* and their neighboring genes. (a) The restriction map of *rmpA2* and *rmpA*. (b) Southern analysis of pLVPK (lane 2). Southern hybridization with a probe prepared from the PCR product of the 636-bp *rmpA2* coding sequence is shown at the right. Two *Eco*RI fragments of 6.9 and 1.6 kb, respectively, representing the fragment containing *rmpA* and *rmpA2* are detected. The 3.8- and 14.6-kb fragments in lane 2 represent the *rmpA*- and *rmpA2*-containing *Hin*dIII fragments.

siderophores. Recently, similar gene cluster contained in a PAI was also found either on the chromosome or a transmissible plasmid in the uropathogenic *E. coli* (Sorsa et al., 2003). It should be mentioned that the *iro* gene cluster in pLVPK lacks *iroE* gene. Nevertheless, the absence of *iroE* gene probably would not affect the utilization of catecholate siderophore by the bacterium since it has been demonstrated in *E. coli* that an *iroE* mutation does not hinder the siderophore utilization activity (Sorsa et al., 2003).

A two-gene operon that encodes an ABC-type transporter related to *M. loti* FepBC was noted on pLVPK at nucleotide positions 77,450–80,256. The identity between the pLVPK genes and FepBC is 38% and 44%, respectively. These genes also share significant homology with many ABC transporters mediating translocation of iron, siderophores, and heme (Koster, 2001). Although the contribution of this putative ABC transporter in the uptake of iron remains unclear, it is undoubtedly advantageous for the bacteria to have multiple iron-acquisition systems in order to obtain iron from the frequently changing environment.

Finally, a gene cluster similar to *E. coli fecIRA*, which is responsible for regulating the uptake of ferric citrate in a Fe^{2+}-Fur-dependent manner, was identified approximately 3 kb upstream of the *iroBCDN*. In *E. coli, fecIR* genes are within a large gene cluster with *fecABCDE* that are the structural genes for iron citrate uptake and are thought to be the target of FecIR regulatory system (Braun et al., 2003). However, there are no observable *fecABCDE* homologs in pLVPK. This phenomenon is not that unusual. As shown in Fig. 4, the homologs of *fecIRA*, but not *fecBCDE*, have been identified experimentally in *Bordetella* spp. as well as in several other bacterial species. It is not clear what the target genes are for these FecIRA-like regulatory systems in these bacteria (Braun et al., 2003). One possibility is that a *fecABCDE* gene cluster could be located on *K. pneumoniae* chromosome. Alternatively, the FepBC-like ABC-type iron transporter encoding genes on pLVPK could be the target gene of the FecIRA regulators.

It should be pointed out here that the pLVPK *fecR* open reading frame is disrupted by an in-frame termination codon. FecR is an inner membrane protein that senses whether FecA, the outer membrane ferric citrate receptor, is bound to the substrate and, in response, activates FecI, which is known as a transcription factor. Deletion analysis of the *fecR* in *E. coli* has shown that a minimum of 59 amino acids in length of the FecR N-terminal derivative is still able to activate the FecI and, subsequently, a constitutive expression of the downstream target genes (Ochs et al., 1995). Thus, despite the presence of an internal stop codon, the *fecR* of pLVPK may still be capable of encoding a truncated but functional product and may result in a constitutive iron-acquisition phenotype in *K. pneumoniae* CG43.

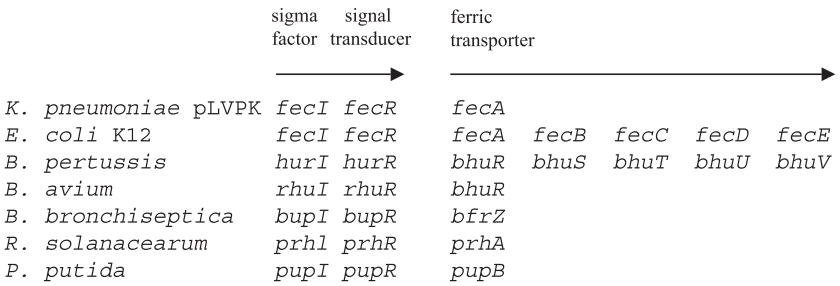

Fig. 4. Comparison of the organization of *fecIRA* genes of pLVPK with similar iron transport and signaling systems. The arrows indicate the transcription orientation of the genes. The genes in the *fecIRA* gene cluster of the pLVPK are very similar to that of the *E. coli fec* operon, however, lacking the *fecBCDE* genes.

The hydroxamate bioassay with the aerobactin indicator strain *E. coli* LG1522 showed that the plasmid-cured strain, CG43-101, loses the aerobactin activity in comparison with its parental strain CG43. In addition, the iron-acquisition activity assay revealed that CG43-101 apparently has a smaller growth zone around the iron-loaded disc. These results indicated that the iron-acquisition capability of the bacteria could mostly be attributed to the plasmid pLVPK.

3.4. Genes related to metal resistance

Heavy metals at certain concentrations in the cell may form unspecific complex compounds leading to a toxic effect. Many genes for the maintenance of the heavy-metal ion homeostasis have been identified in bacteria. Three physically linked gene clusters, as shown in Fig. 5 (152,306–177,234 bp), were identified in the pLVPK that are related to metal resistance phenotype in *K. pneumoniae*. These gene clusters include homologs of the lead-resistance genes *pbrRSABC* of *Ralstonia metallidurans* CH34 (Borremans et al., 2001), the copper-resistance genes *pcoEABCDRS* of *E. coli* plasmid pRJ1004 (Brown et al., 1995), and the silver-resistance gene cluster *silCBAPsilRSE* of *S. enterica* serovar Typhimurium (Gupta et al., 1999). By using disk diffusion assay, we have found that the resistance against silver and copper ions between *K. pneumoniae* CG43 and a plasmid-cured strain, CG43-101, remains the same.

A putative lead-resistance gene cluster, *pbrRABC*, showed a 63–71% deduced amino acid sequence identity with that of the *R. metallidurans pbrTRABCD* genes. The *R. metallidurans* lead-resistance operon, carried on a large plasmid, pMOL30, contains *pbrT* for Pb^{2+} uptake, *pbrA* for Pb^{2+} efflux, *pbrB* for a putative integral membrane

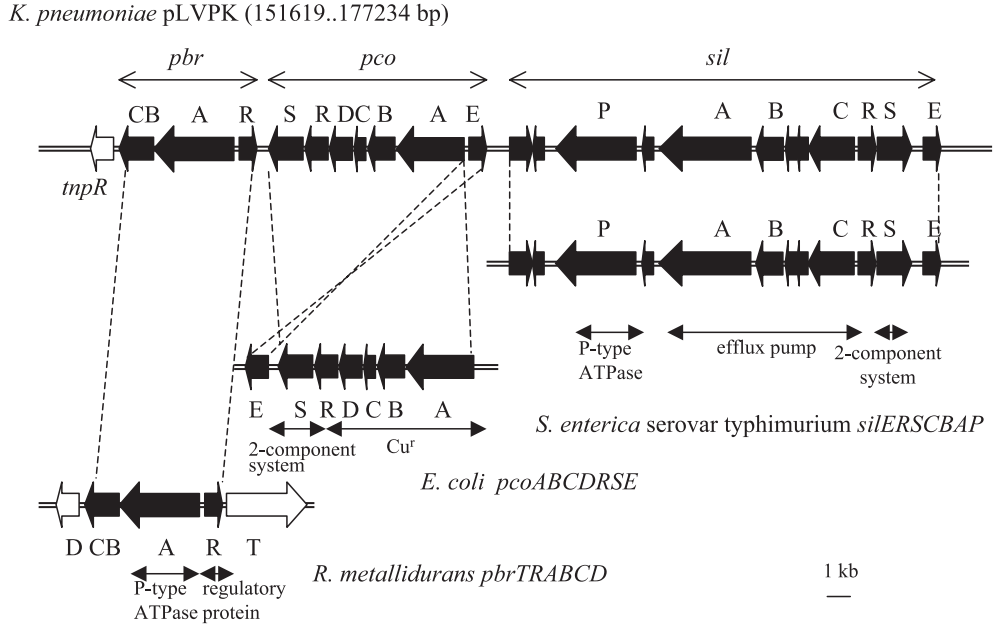

Fig. 5. The heavy-metal-resistant gene clusters. The silver-resistance *sil* gene cluster, in comparison with that of the *S. enterica* serovar Typhimurium pMG101 (Gupta et al., 1999), the copper-resistance *pco* gene cluster, in comparison with that of the *E. coli* pRJ1004 (Brown et al., 1995), and the lead-resistance *pbr* gene cluster, in comparison with that of the *R. metallidurans* pMOL30, are shown (Borremans et al., 2001). The homologous gene clusters are depicted in solid black and the transcriptional orientation of the individual ORFs are shown by arrows.

protein, *pbrC* for a putative prolipoprotein signal peptidase, *pbrD* that confers lead sequestration, and *pbrR* that regulates the transcription of *pbrABCD* (Borremans et al., 2001). Unlike that of the *R. metallidurans*, the *pbr* gene clusters of pLVPK contain only the efflux system (*pbrABC*) and regulator encoding genes (*pbrR*) (Fig. 5), which suggest a simple lead-efflux mechanism similar to that of the CadA ATPase of *Staphylococcus aureus* and the ZntA ATPase of *E. coli* (Rensing et al., 1998). In contrast to the indifference of copper and silver ion resistance, the lead susceptibility increased in the disk diffusion assay after curing of the plasmid. The *pbr* genes in the pLVPK may contribute to the adaptation of *K. pneumoniae* in lead-polluted human inhabitants.

A gene cluster encoding *E. coli terZABCDE* homolog was also identified. The *terZABCDE* has been shown previously to be a part of a PAI, which also contains integrase, prophage, and urease genes in *E. coli* EDL933 (Taylor et al., 2002). This gene cluster also provides the resistance to bacteriophage infection as well as resistance to pore-forming colicins. Although *terBCDE* are sufficient for the tellurite resistance property, the functions of each of these genes are unknown. The 14.7-kb region (19,890–34,588 bp) containing *terZABCDE* genes and 12 putative ORFs of pLVPK are comparable to the *ter* genes-containing region in the *E. coli* O157 genome. The homology is interrupted downstream of the *terZABCDE* region by an *E. coli* pTE53 tellurite resistance *terF* homolog and IS*903* gene (Fig. 6a). A recent study suggests that the Te^r-containing pathogenicity island in enterohemorrhagic *E. coli* isolates was acquired from plasmid. With considerable degree of sequence homology (75–98% amino acid sequence similarity respectively with that of the *E. coli* O157 *terZABCDE*), the *ter* genes of the pLVPK are likely horizontally acquired. It has been speculated that the *ter* system most likely plays other functional roles such as protection against host defenses so as to be stably maintained in the bacterium (Taylor et al., 2002).

A chromosomally located ORF which showed 77% amino acid sequence identity with the *E. coli* tellurite-resistant gene *tehB* (Taylor et al., 2002) has also been recently isolated in our laboratory from *K. pneumoniae* CG43. Deletion of the *tehB*-like gene had no apparent effect on tellurite resistance of the bacteria (Fig. 6b), suggesting that the tellurite resistance of the bacteria is determined by the *ter* gene cluster of pLVPK rather than the *tehB* homolog.

3.5. Replication and plasmid maintenance

DNA sequence analysis also revealed a single plasmid replication region of 1756 bp (217,448–219,203 bp), which consists of *repA* and sequence elements with characteristics of plasmid replicons that employ an iteron-based

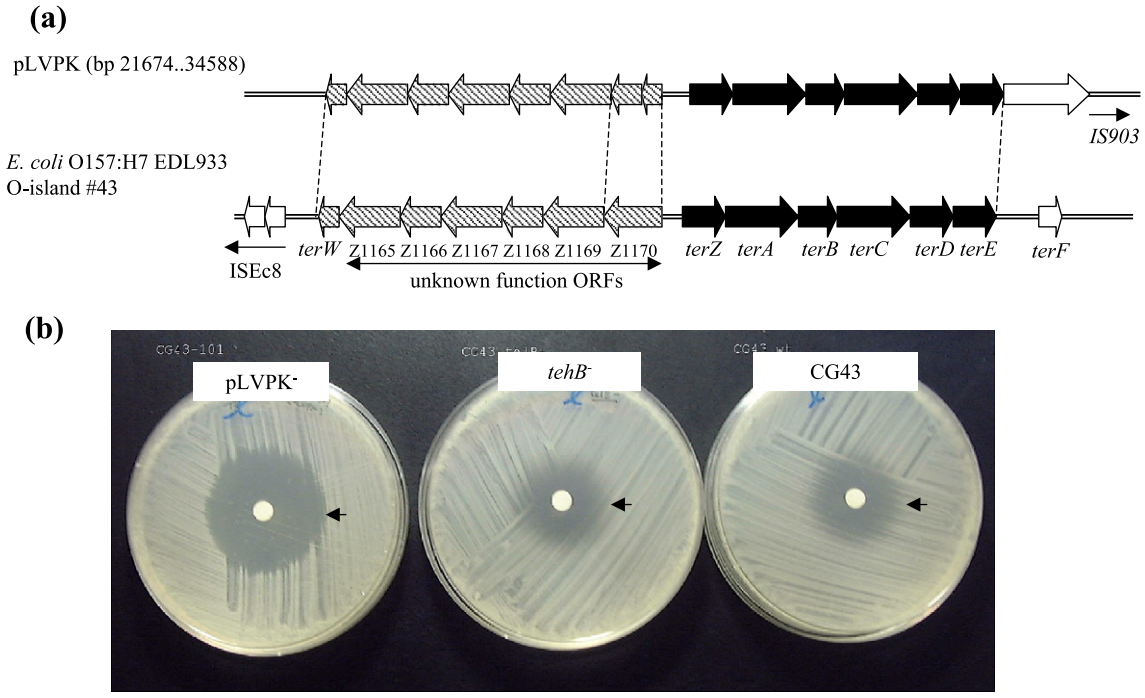

Fig. 6. The tellurite resistance genes in *K. pneumoniae* CG43. (a) The *ter* gene cluster of pLVPK similar to *terZABCDE* of *E. coli* O157:H7 EDL933 O-island #43, a region of the EDL933 chromosome not homologous to *E. coli* K-12 MG1655. The homologous regions are shown in hatched and solid arrow. (b) Effects of the *tehB* deletion and curing of pLVPK on the tellurite resistance of the bacteria. Tellurite resistance assays of *K. pneumoniae* CG43S3, the pLVPK-cured bacteria (CG-101), and the *tehB* mutant CG43S3T1 using the discs immersed, respectively, with 20 and 40 μg K_2TeO_3 are shown. The arrows indicate the margins of the inhibition zone.

replication initiation and control mechanism (Chattoraj, 2000). The *repA* product showed a high sequence similarity to a number of plasmid replication initiation proteins, including RepFIB of *S. enterica* serovar Typhi R27 plasmid (60% identity), RepFIB of *E. coli* O103:H2 (43% identity), RepA of *Yersinia pestis* KIM plasmid pMT-1 (42% identity), and RepA of *S. enterica* serovar Typhi plasmid pHCM2 (42% identity). As shown in the multiple sequence alignment in Fig. 7a, RepA appears to be an initiator for plasmid replication, which is able to bind the flanking repeated sequences through its DNA binding structures, a winged-helix domain, and a leucine-zipper motif (Chattoraj, 2000). We have also found two sets of iterons, 4 21-bp and 13 42-bp direct repeats, located, respectively, at the upstream and downstream of the *repA* locus (Fig. 7b). The sequences are most likely the specific binding sites for the RepA protein to initiate replication of the plasmid and also control the plasmid copy number (Chattoraj, 2000).

3.6. Segregation control machineries

A region (203,493–203,994 bp) consisting of 11 copies of a 43-bp repeat (5′ -gggaccacggtcccacctgcatcgtcgtttaggtttt-cagcct-3′) is believed to be required for segregation control of the plasmid. Next to the 43-bp direct repeat pattern, positioned are the genes encoding *sopA* and *sopB* homologous. The organization is comparable to that of the *sop* operon, which governs the partition of the F plasmid (Yates et al., 1999). In addition to *sopAB*, genes showing sequence similarity with *parAB* of *E. coli* P1 phage were identified. It has been shown previously that the corresponding partitioning site in the P1 *parAB* system is composed of direct or inverted repeats (Davis and Austin, 1988). We also noted

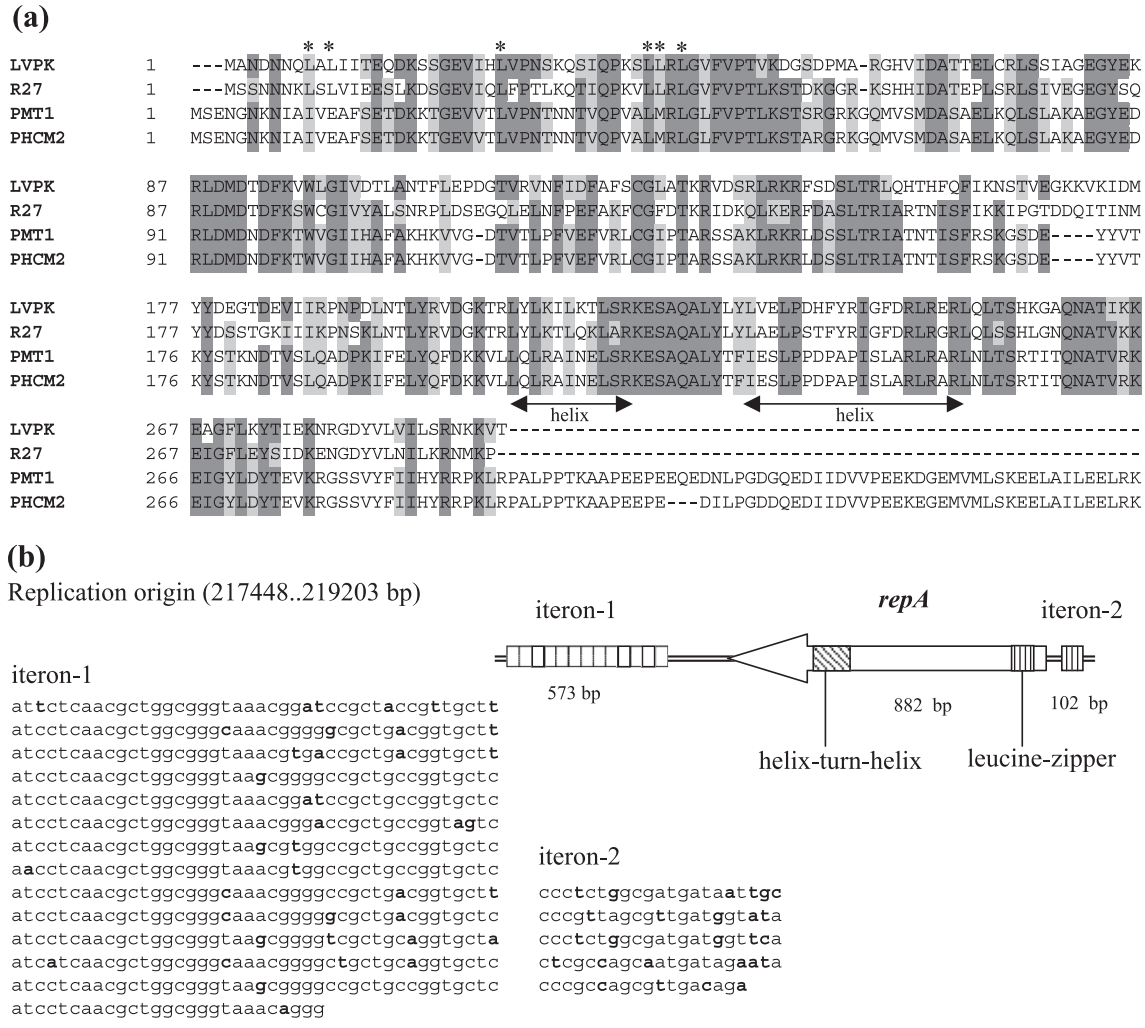

Fig. 7. The *repA* and the iteron sequences at the replication origin. (a) Sequence alignment of the *repA* from pLVPK (LVPK), R27 of *S. enterica* serovar Typhi (R27), pHMT-1 of *Y. pestis* (PMT1), and pHCM2 of *S. enterica* serovar Typhi (PHCM2). Leucine residues near the N-terminus are marked with asterisks, and the two-helix regions corresponding to the helix-turn-helix DNA-binding domain are shown. (b) The replication origin including the *repA* structural gene and two iteron sequences nearby of the pLVPK. The sequences of the two sets of iterons, iteron-1 (217,448–218,020 bp) and iteron-2 (219,102–219,203 bp), are shown with each row, respectively, representative for each of the adjacent repeat units.

that a 66-bp direct repeat upstream of the *parAB* homologs is found, which indicates that they also contribute to the partitioning control of the pLVPK. It is reasonable that such a large plasmid has meticulous maintenance systems. Nevertheless, how these two partitioning systems contribute to the maintenance of pLVPK remains to be confirmed.

3.7. Heterogeneity

Pathogenic bacteria have obtained a significant proportion of their genetic diversity by acquisition of DNA from other organisms. Many of the gene clusters identified in pLVPK are homologous to the unknown gene clusters in the other organisms. Although with unknown functions, the homologs of the gene clusters contained in the 9-kb region from nucleotide 2522 to 11,618 and the 5.9-kb region from nucleotide 13,997 to 19,886 were found, respectively, in the genome of *B. fungorum* and *Y. pestis* KIM. A gene cluster which encodes a putative ABC transporter system (117,432–113,670 bp) is also identified for which the deduced amino acid sequences are similar to those of the putative ABC transporter system of *Streptomyces coelicolor* A3. A region (46,979–51,336 bp) comparable to the phage infection inhibition *pif* region of *E. coli* F plasmid was also identified. The boundary sequences of these gene clusters, as well as that of the PAI-like region, are mobile elements including insertion sequences and short pieces of 3′ -sequences of tRNA genes. With the involvement of the transposons and the tRNA sequences, horizontal gene transfers have made possible these gene clusters to be introduced into the plasmid and hence affect the ecological and pathological characteristics of bacteria.

4. Conclusions

- The 219-kb plasmid contains 251 ORFs, of which 37% have homologous genes of known function, 31% match the hypothetical genes in GenBank, and the remaining 32% are novel sequences.
- Virulence-associated genes identified include the CPS synthesis regulator gene *rmpA2* and its homolog *rmpA* and multiple iron-acquisition system genes *iucABC-DiutA*, *iroBCDN*, *fepBC*, and *fecIRA*.
- Gene clusters homologous with copper silver, lead, and tellurite-resistant genes were identified. After curing of the plasmid, resistance of the bacteria to either lead or tellurite was found to decrease. However, no apparent effect was observed on the bacterial susceptibility to silver and copper.
- Identification of a *repA* gene lying in between two sets of iterons suggests that the replication of the plasmid is iteron-controlled. Furthermore, the presence of two regions similar to the *sop* and *par* genes indicates the presence of an F plasmid-like partitioning system.

Acknowledgements

This work was supported in part by National Science Council of the Republic of China (NSC92-2311-B009-001 to HLP and 92-2311-B009-001 to HYC). We are grateful to N. Venkatesan and J. Vatsyayan for critical reading of the manuscript.

References

Ambrozic, J., Ostroversnik, A., Starcic, M., Kuhar, I., Grabnar, M., Zgur-Bertok, D., 1998. *Escherichia coli* CoIV plasmid pRK100: genetic organization, stability and conjugal transfer. Microbiology 144, 343–352.

Arakawa, K., Mori, K., Ikeda, K., Matsuzaki, T., Kobayashi, Y., Tomita, M., 2003. G-language genome analysis environment: a workbench for nucleotide sequence data mining. Bioinformatics 19, 305–306.

Borremans, B., Hobman, J.L., Provoost, A., Brown, N.L., van der Lelie, 2001. Cloning and functional analysis of the *pbr* lead resistance determinant of *Ralstonia metallidurans* CH34. J. Bacteriol. 183, 5651–5658.

Braun, V., Mahren, S., Ogierman, M., 2003. Regulation of the FecI-type ECF factor by transmembrane signaling. Curr. Opin. Microbiol. 6, 173–180.

Brown, N.L., Barrett, S.R., Camakaris, J., Lee, B.T., Rouch, D.A., 1995. Molecular genetics and transport analysis of the copper-resistance determinant (*pco*) from *Escherichia coli* plasmid pRJ1004. Mol. Microbiol. 17, 1153–1166.

Chattoraj, D.K., 2000. Control of plasmid DNA replication by iterons: no longer paradoxical. Mol. Microbiol. 37, 467–476.

Davis, M.A., Austin, S.J., 1988. Recognition of the P1 plasmid centromere analog involves binding of the ParB protein and is modified by a specific host factor. EMBO J. 7, 1881–1888.

Delcher, A.L., Harmon, D., Kasif, S., White, O., Salzberg, S.L., 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. 27, 4636–4641.

Ewing, B., Hillier, L., Wendl, M.C., Green, P., 1998. Base-calling of automated sequencer traces using Phred: I. Accuracy assessment. Genome Res. 8, 175–185.

Gordon, D., Abajian, C., Green, P., 1998. Consed: a graphical tool for sequence finishing. Genome Res. 8, 195–202.

Gupta, A., Matsui, K., Lo, J.F., Silver, S., 1999. Molecular basis for resistance to silver cations in *Salmonella*. Nat. Med. 5, 183–188.

Hou, Y.M., 1999. Transfer RNAs and pathogenicity islands. TIBS 24, 295–298.

Koster, W., 2001. ABC transporter-mediated uptake of iron, siderophores, heme and vitamin B12. Res. Microbiol. 152, 291–301.

Lai, Y.C., Peng, H.L., Chang, H.Y., 2001. Identification of genes induced in vivo during *Klebsiella pneumoniae* CG43 infection. Infect. Immun. 69, 7140–7145.

Lai, Y.C., Peng, H.L., Chang, H.Y., 2003. RmpA2, an activator of capsule biosynthesis in *Klebsiella pneumoniae* CG43, regulates K2 cps gene expression at the transcriptional level. J. Bacteriol. 185, 788–800.

Lowe, T.M., Todd, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 955–964.

Menoharan, A., Pai, R., Shankar, V., Thomas, K., Lalitha, M.K., 2003. Comparison of disc diffusion and E test methods with agar dilution for antimicrobial susceptibility testing of *Haemophilus influenzae*. Indian J. Med. Res. 117, 81–87.

Moss, J.E., Cardozo, T.J., Zychlinsky, A., Groisman, E.A., 1999. The *selC*-associated SHI-2 island of *Shigella flexneri*. Mol. Microbiol. 33, 74–83.

Nassif, X., Sansonetti, P.J., 1986. Correlation of the virulence of *Klebsiella pneumoniae* K1 and K2 with the presence of a plasmid encoding aerobactin. Infect. Immun. 54, 603–608.

Nassif, X., Fournier, J.M., Arondel, J., Sansonetti, P.J., 1989. Mucoid phenotype of *Klebsiella pneumoniae* is a plasmid-encoded virulence factor. Infect. Immun. 57, 546–552.

Ochs, M., Veitinger, S., Kim, I., Welz, D., Angerer, A., Braun, V., 1995. Regulation of citrate-dependent iron transport of *Escherichia coli*: *fecR* is required for transcription activation by FecI. Mol. Microbiol. 15, 119–132.

Peng, H.L., Wang, P.Y., Wu, J.L., Chiu, C.T., Chang, H.Y., 1991. Molecular epidemiology of *Klebsiella pneumoniae*. Zhonghua Min guo Wei sheng wu Ji Mian yi xue Za zhi 24, 264–271.

Pullinger, G.D., Lax, A.J., 1992. A *Salmonella dublin* virulence plasmid locus that affects bacterial growth under nutrient-limited conditions. Mol. Microbiol. 6, 1631–1643.

Rensing, C., Sun, Y., Mitra, B., Rosen, B., 1998. Pb(II)-translocating P-type ATPases. J. Biol. Chem. 273, 32614–32617.

Schaberg, D.R., Culver, D.H., Gaynes, R.P., 1991. Major trends in the microbial etiology of nosocomial infection. Am. J. Med. 91, 72S–75S.

Simmons-Smit, A.M., Verweij-van Vught, A.M., MacLaren, D.M., 1986. The role of K antigens as virulence factors in *Klebsiella*. J. Med. Microbiol. 21, 133–137.

Sorsa, L.J., Dufke, S., Heesemann, J., Schubert, S., 2003. Characterization of an *iroBCDEN* gene cluster on a transmissible plasmid of uropathogenic *Escherichia coli*: evidence for horizontal transfer of a chromosomal virulence factor. Infect. Immun. 71, 3285–3293.

Taylor, D.E., Rooker, M., Keelan, M., Ng, L.K., Martin, I., Perna, N.T., Burland, N.T.V., Blattner, F.R., 2002. Genomic variability of O islands encoding tellurite resistance in *Enterohemorrhagic Escherichia coli* O157:H7 isolates. J. Bacteriol. 184, 4690–4698.

Wacharotayankun, R., Arakawa, Y., Ohta, M., Tanaka, K., Akashi, T., Mori, M., Kato, N., 1993. Enhancement of extracapsular polysaccharide synthesis in *Klebsiella pneumoniae* by RmpA2, which shows homology to NtrC and FixJ. Infect. Immun. 61, 3164–3174.

Yates, P., Lane, D., Biek, D.P., 1999. The F plasmid centromere, *sopC*, is required for full repression of the *sopAB* operon. J. Mol. Biol. 290, 627–638.