Polish Journal of Microbiology
2007, Vol. 56, No 2, 83–88

ORIGINAL PAPER

Heterogeneity of *galF* and *gnd* of the *cps* Region for Capsule Synthesis in Clinical Isolates of *Klebsiella pneumoniae*

RAFAŁ GIERCZYŃSKI

Department of Bacteriology, National Institute of Hygiene, Warsaw, Poland

Received 14 March 2007, revised 5 May 2007, accepted 10 May 2007

Abstract

The capsular polysaccharide (CPS) plays important role in *Klebsiella* spp pathogenesis. Capsular types K1 and K2 of *Klebsiella pneumoniae* are considered most virulent for humans. The capsule biosynthesis region flanking genes *galF* and *gnd* from clinical isolates and reference strains of *K. pneumoniae* were screened for polymorphism. Nucleotide sequence analysis of *galF* and *gnd* revealed a high heterogeneity. However, deduced amino acid sequences demonstrated that the majority of mutations were silent implying GalF and Gnd are strongly conserved. This may suggest importance of these *loci* in the CPS biosynthesis and may argue for their potential usefulness in *Klebsiella* genotyping.

Key words: *Klebsiella pneumoniae*, capsule, heterogeneity of CPS region

Introduction

Klebsiella pneumoniae, an important nosocomial pathogen, causes suppurative infection, pneumonia, urinary tract infection and septicaemia in humans, especially immunosuppressed or suffering from underlying diseases like diabetes mellitus (Podschun and Ulmann, 1998; Fung *et al.*, 2002). This bacterium is also known as an etiological agent of community acquired bacterial pneumonia occurring in chronic alcoholics and people of low social status, which is often fatal if untreated (Podschun and Ulmann, 1998).

Generally, clinical isolates of *K. pneumoniae* produce abundant capsular polysaccharide (CPS). The thickness of capsular layer was shown to be an important factor for high virulence, since poorly encapsulated isolates, however appeared into the alveolar epithelial cells, were avirulent in mouse model of pneumonia (Lai *et al.*, 2003; Astorza *et al.*, 2004). Capsule production, is a general prerequisite for virulence since it protects the bacterium from phagocytosis and killing by serum factors. The genomic organisation of the chromosomal *K. pneumoniae cps* region responsible for capsule K2 synthesis in strain Chedid was determined by Arakawa *et al.* (1995). Although this cluster contains 19 open reading frames (*orfs*), 15 were found to be indispensable for capsule production. These *orfs*, numbered from 1 to 15, are transcribed at the same direction from two promoters located upstream of *orf1* (*galF*) and *orf3*. In addition to the nucleotide sequence of the *cps* region of Chedid which is deposited in GenBank (http://www.ncbi.nlm.nih.gov) under accession number D21242, corresponding sequences for K1 capsule biosynthesis pathway of *K. pneumoniae* strains DTS and NTUH-K2044 have been recently added: AY762939 and AB198423 respectively. Moreover, the complete genome sequence of strain MGH78578 *K. pneumoniae* K52 became available at Genome Sequencing Center at Washington University Medical School (http://genome.wustl.edu).

The Southern hybridisation experiments performed by Arakawa *et al.* (1995) shown that upstream section of the *cps* region from *orf1* to *orf3* and downstream section from *orf12* to *orf15* (*gnd*) were present in all tested *Klebsiella* capsular reference strains. The results of PCR screening performed during previous study (Gierczyński *et al.*, 2005a) and results obtained by Brisse *et al.* (2004) generally confirmed findings reported by Arakawa *et al.* (1995), however suggested that the downstream section is likely limited to sole *gnd*. The *galF* and *gnd* genes are present in a homologous *cps* clusters of *K. pneumoniae* K1 strains DTS

Corresponding author: R. Gierczyński, Department of Bacteriology National Institute of Hygiene, 00-791 Chocimska 24, Warsaw, Poland; phone: (48) 22 542 12 44, fax: (48) 22 542 13 07; **e-mail: rgierczynski@pzh.gov.pl**

and NTUH-K2044 as well in K52 strain MGH78578. In contrast, the inner *cps* cluster flanked by *orf4* and *orf14* was shown to reveal genetic heterogeneity that correlated to capsular type of *Klebsiella* reference strains (Brisse *et al.*, 2004).

Studies on genetic diversity of genes flanking the *cps* region in *Klebsiella* are rather limited. Most of them were conducted on a limited number of laboratory strains and their mutants. Previously a single strand conformation polymorphism (SSCP) in selected fragments of *galF* and *gnd* was tested in 56 clinical strains isolated from infants in six nosocomial outbreaks (Gierczyński *et al.*, 2005b). Eighteen and ten patterns were distinguished for *galF* and *gnd* respectively, that suggested a high nucleotide polymorphism.

The main purpose of this study was to determine the nucleotide and amino acid diversity of *galF* and *gnd* fragments previously tested by the SSCP and to check whether these genes may be useful for genotyping and phylogenetic analyses of *K. pneumoniae.* In addition, a comparative phylogenetic analysis of the *galF* and *gnd* gene variants found in tested strains and data from GenBank was performed to trace eventual co-evolution.

Experimental

Materials and Methods

Bacterial isolates and database nucleotide sequences. Twelve strains *of K. pneumoniae* representing different SSCP haplotypes of *galF* and *gnd* described previously (Gierczyński *et al.*, 2005b) were examined (Table I). Biochemical properties (biogroup) of tested strains were determined as described by Kałużewski (1965). To eliminate occurrence of delayed results of biochemical tests, development period was reduced to 24 hours. The O-antigen serotyping was performed as described by Kałużewski (1968) using unencapsulated variants of tested strains. Moreover, database nucleotide sequences of *galF* and *gnd* from strains Chedid (Genbank accession no.: D21242), DTS (AY762939) and NTUH-K2044 (AB198423), MGH78578 (http://genome.wustl.edu) and their homologues from *Escherichia coli* strain E69 K30 (AF503613) were used for phylogenetic analysis.

Table I
Characteristic of tested strains of *K. pneumoniae*

Strain	Sample type (localisation)	Phenotype		SSCP-genotypes[a]		
		O-type	Bio-group	*galF* 355	*galF* 366	*gnd*
A5054	Reference O1:K1	O1	3	A1	A2	A2
B5055	Reference O1:K2	O1	10	A	A	A
30	Blood (Warszawa)	O2	13	B	A1	B
478/02	Blood (Łódź)	O3	11	A2	A1	B1
234	Blood (Poznań)	O1	8	B	C	B3
57	CSF (Warszawa)	O1	8	A	B	A
11	Lung (Warszawa)	O1	1	A1	A	C
145/02	Nose (Zamość)	NA	18	B1	C1	C1
222	Stool (Warszawa)	O3	14	B1	D1	A2
5	Stool (Warszawa)	O1	2	A1	D	C
216	Stool (Warszawa)	O3	3	B	D1	C2
273	Stool (Poznań)	NA	11	B2	E	A1

[a] – SSCP genotypes distinguished in previous study (Gierczyński *et al.*, 2005b),
NA – not assigned

PCR analysis. Genomic DNA templates were prepared as described by Gierczyński *et al.* (2004) but the lysosyme treating was omitted. Primers listed in Table II and *Taq* DNA Polymerase (Fermentas, Lithuania) in $(NH_4)_2SO_4$ reaction buffer with 1.5 mM $MgCl_2$ supplied by the manufacturer were used for amplification. PCR was performed in 25 ml volume using Mastercycler 5333 (Eppendorf, Germany) as follows: 3 min at 94°C and 35 cycles for 30 seconds of each denaturation at 94°C, annealing at 58°C and elongation at 72°C. DNA synthesis was completed at 72°C for 3 min.

DNA sequencing. PCR products were subjected for sequence analysis using automated fluorescent DNA sequencer 377 and BigDye Terminator v3.1. (Applied Biosystems, USA) in accordance to manufacturer's instructions. Both DNA strands were sequenced. The determined sequences of tested loci were compared to sequences deposited in GenBank database using BLASTN (Altschul *et al.*, 1997) software utility (National Center for Biotechnology Information, USA)

Table II
List of PCR primers

Locus	Primer signature	Nucleotide sequence		Amplicon size	Amplified region
		Forward primer	Reward primer		
galF	or1355	gctgccgatcgttgataagcc	gttatagcgcagcgggtcagc	355	90 – 444
galF	or1366	tggtcctgccggatatcatcc	actgcttcttcgccagttcgg	366	389 – 754
gnd	or15343	gttgaatccctcgagacacc	caccgatataggtcacacacg	343	181 – 523

The amplified regions were counted from the initial nucleotide of the coding sequence of appropriate gene of strain Chedid (D21242). The amplicon size was indicated in base pairs.

Evolutionary trees. Phylogenetic and molecular evolutionary analyses were conducted using MEGA software version 2.1. (Kumar *et al.*, 2001). The unweighed pair group method with arithmetic means (UPGMA) was used to generate dendrograms of genetic diversity of tested strains. Neighbour-Joining method with p-distance parameter was applied to generate dendrogram reflecting phylogenetic relatedness of tested strains based upon the deduced amino acid sequences of the analysed genes. Kimura 2 parameter model was applied to both UPGMA and Neighbour-Joining methods.

Results

All tested strains yielded PCR-products of selected fragments of *galF* and *gnd*. The size of amplicons was in agreement to expected size calculated from the nucleotide sequence of the *cps* region of strain Chedid. The nucleotide sequencing of *galF* and *gnd* fragments previously tested by Multitemperature-SSCP confirmed genetic diversity of the SSCP genotypes (Gierczyński *et al.*, 2005b). Obtained nucleotide sequences were compared to the reference sequence of strain Chedid. Nucleotide substitutions within analysed fragments of *galF* and *gnd* from tested strains versus strain Chedid are shown in Table III and IV respectively.

When compared to strain Chedid, a number of 61 single nucleotide polymprphisms (SNPs) was detected in tested strains for both analysed genes. It is noteworthy the nucleotide polymorphism in *gnd* was about four times higher than in *galF* (Tables III and IV). In this case, 43 SNPs were found in 343 base pairs long DNA fragment, whereas seven and eleven SNPs were detected in *galF* fragments 355 and 366 respectively. In spite of the high nucleotide sequence diversity of *gnd*, only two amino acid substitutions were found. Three amino acid substitutions were observed in *galF* fragment 366, whereas no one was detected in *galF* fragment 355.

The nucleotide sequences of *galF* and *gnd* determined in this study for selected representatives of tested and reference strains were compared to the homologous database sequences of strains Chedid, DTS, NTUH-K2044, MGH78578 and *E. coli* E69 to determine phylogenetic relatedness of tested isolates (Fig. 1). Dendrogram generated by the UPGMA based upon combined sequences of *galF* and *gnd*

Table III
The nucleotide sequences of SSCP-genotypes of the *galF* gene fragment 355 bp (A) and 366 bp (B) compared to the *galF* sequence of *K. pneumoniae* strain Chedid

A

Strain (SSCP)	Position of single nucleotide polymorphisms						
	153	210	213	261	297	345	381
Chedid (NA)	c	c	a	g	g	c	g
A5054 (A1)	–	–	c	–	–	–	–
B5055 (A)	–	t	c	–	–	–	–
478/02 (A2)	t	–	c	–	–	–	–
234 (B)	–	t	c	–	a	–	–
273 (B2)	–	–	–	–	–	–	a
222 (B1)	–	–	c	a	–	t	–

B

Strain (SSCP)	Position of single nucleotide polymorphisms										
	504	533	537	551	585	609	618	633	709	717	729
Chedid (NA)	c	t	c	c	c	g	t	g	c	g	t
A5054 (A2)	–	–	–	–	t	–	a	–	T[c]	–	c
B5055 (A)	–	–	–	–	–	–	c	–	–	–	c
478/02 (A1)	t	–	t	–	–	a	c	–	–	a	c
145/02 (C1)	–	–	–	–	t	a	c	–	–	a	a
234 (C)	–	–	–	A[b]	–	–	c	–	–	a	c
216 (D1)	–	–	–	–	–	–	c	a	–	a	c
57 (B)	–	C[a]	–	–	–	–	c	–	–	–	c
5 (D)	t	–	–	–	t	a	c	–	–	a	c

Position of a single nucleotide polymorphism (SNP) was indicated as the number of nucleotides counted from the initial nucleotide of *galF* coding sequence of strain Chedid (D21242). Nucleotides identical to sequence D21242 were indicated by dashes. Amino acids substitutions are shown in capitals. [a] – Alanine instead Valine; [b] – Glutamine instead Proline; [c] – Cysteine instead Arginine; NA – not assigned.

Table IV
The nucleotide sequences of SSCP-genotypes of the *gnd* gene fragment 343 bp compared to the *gnd* sequence of *K. pneumoniae* strain Chedid

Strain (SSCP)	Position of single nucleotide polymorphisms														
	204	207	233	237	240	243	252	261	273	276	282	300	303	306	309
Chedid (NA)	t	t	c	c	c	c	t	g	t	t	c	c	c	c	c
A5054 (A2)	c	–	–	–	–	–	c	–	c	–	–	–	–	–	–
B5055 (A)	c	–	–	–	–	–	c	–	c	–	–	–	–	–	–
478/02 (B1)	c	–	–	–	–	–	c	–	c	c	t	–	t	–	–
145/02 (C1)	c	–	–	–	–	–	c	–	c	–	–	t	–	–	–
30 (B)	c	–	T[a]	–	–	t	c	t	–	–	–	–	–	t	–
5 (C)	–	c	–	t	a	t	c	–	c	–	–	–	–	–	t
	324	339	342	345	351	375	381	387	390	396	399	417	426	432	435
Chedid (NA)	c	t	a	c	c	c	t	c	c	g	c	c	t	c	g
A5054 (A2)	–	–	g	a	–	–	–	t	t	a	g	–	–	g	–
B5055 (A)	–	–	g	a	–	–	–	t	t	a	g	–	–	g	–
478/02 (B1)	–	–	–	g	a	–	–	t	t	–	–	t	–	t	–
145/02 (C1)	–	c	–	g	–	a	–	t	t	–	g	–	a	t	a
30 (B)	–	–	g	g	–	–	–	t	t	–	–	t	–	t	–
5 (C)	t	–	–	–	–	–	a	t	t	–	g	–	a	–	–
	450	451	453	456	459	462	471	474	477	480	486	489	498		
Chedid (NA)	a	c	g	a	a	a	g	g	t	g	g	t	c		
A5054 (A2)	g	t	a	t	t	g	–	–	–	t	t	–	t		
B5055 (A)	g	t	a	t	t	g	–	T[b]	–	t	t	–	t		
478/02 (B1)	g	t	a	t	t	g	–	–	–	t	t	–	t		
145/02 (C1)	–	–	–	t	t	–	–	–	c	–	t	a	t		
30 (B)	g	t	a	t	t	g	–	T[b]	–	t	t	–	t		
5 (C)	–	–	–	t	t	–	a	–	c	–	t	a	t		

Position of single nucleotide polymorphism (SNP) was indicated by the number of nucleotides counted from the initial nucleotide of *gnd* coding sequence of strain Chedid (D21242). Nucleotides identical to sequence D21242 were indicated by dashes. Amino acid substitutions are shown in capitals. [a] – Valine instead Alanine; [b] – Histidine instead Glutamine; NA – not assigned.

divided tested strains into two main clusters. However, these clusters appear closely related when compared to strains Chedid and MGH78578. The phylogenetic reconstruction (Fig. 1.B) exhibited that strains MGH78578 and *E. coli* E69 are significantly distant from the other strains tested. Moreover, two topologically significant branches represented by strain B5055 and Chedid were distinguished. The distant position of strains MGH78578 and E69 was also reflected by dendrogram based upon the deduced amino acid sequences of tested strains (Fig. 1.C). The majority of tested strains together with strain Chedid were placed at the same branch while strains (30 and 57) isolated from fatal cases were grouped closely to strain B5055.

Discussion

In this study we determined the single nucleotide polymorphism (SNP) of *galF* and *gnd* in clinical strains of *K. pneumoniae* belonging to three O-antigen serogroups. Although a relatively high polymorphism was observed, neither specific genotype nor cluster of genotypes was found to be exclusive for epidemic strains. Nevertheless, the high nucleotide diversity found in analysed fragments of *galF* and *gnd* may suggest that these loci are potentially useful targets for the multilocus sequence typing (MLST).

The significant disproportion of the SNPs in *galF* and *gnd* observed for clinical and reference strains implies that these loci are exposed to different evolutionary pressure. Nelson and Selander (1994) found that the nucleotide diversity in *gnd* (orf15) is surprisingly high in both *K. pneumoniae* and *E. coli*. In the last species recombination at *gnd* has occurred with such high frequency that the indicated evolutionary relationships among strains are not congruent with those estimated for other housekeeping genes. This was attributed to a horizontal co-transfer of *gnd* with adjacent locus *rfb* mediating O-antigen synthesis, whose activities are subject to diversifying selection because of the host immune response (Nelson and Selander, 1994). Moreover, Rahn *et al.* (1999) observed exchange of large portions of the *cps* cluster between various strains of *Klebsiella* as well between *Klebsiella* and *E. coli* K1 group strains. In this light,

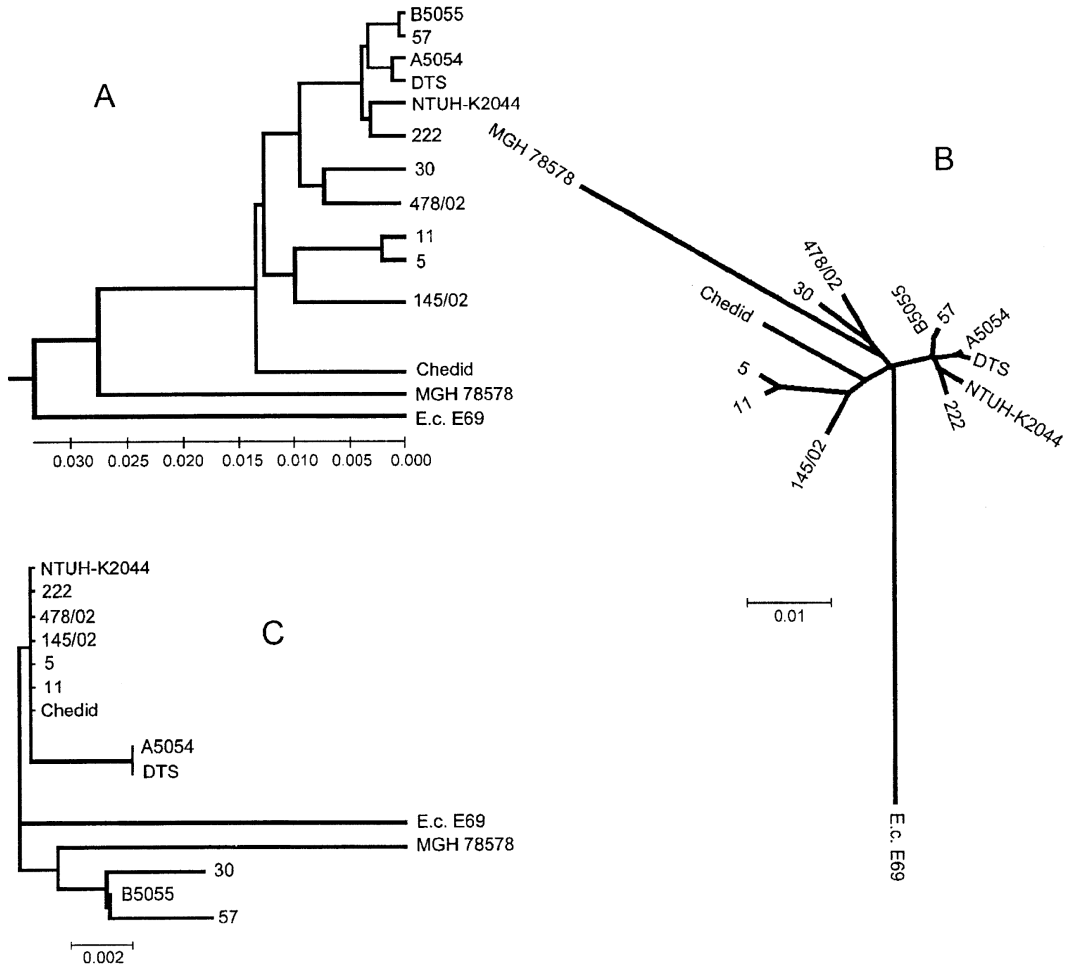

Figure 1. Results of the phylogenetic analysis of the nucleotide and deduced amino acid sequences of *galF* and *gnd* from tested strains and selected database sequences. A – genetic diversity shown by UPGMA; B – based on nucleotide sequence unrooted phylogenetic tree of tested strains; C – based on amino acid sequences phylogenetic tree of tested strains.

application of the sole *gnd* for the phylogenetic analyses of *K. pneumoniae* may lead to false conclusions. On the other hand, the amino acid sequences of both GalF and Gnd were strongly conserved in tested strains that together with ubiquitous presence of these genes in *K. pneumoniae* (Gierczyński *et al.*, 2005a) may argue for their important role in capsule synthesis. This thesis is supported by results of mutagenesis experiments performed by Arakawa *et al.* (1995). Insertion of Tn5 in *gnd* abolished capsular synthesis in strain Chedid, while the lack of *galF* resulted in unstable CPS production characterised as the mixed capsule phenotype. Taken together results presented herein and conclusions made by Nelson and Selander (1994), the high level of nucleotide diversity of *galF* and *gnd* appears to be protected from evolutionary bias by the strong evolutionary pressure. Therefore, these loci may be useful markers for strain distinguishing or subtyping.

Interestingly, in spite of the apparent nucleotide diversity the majority of tested strains revealed amino acid sequence of GalF and Gnd typical for strain Chedid. Hence, analysis of both the nucleotide and amino acid sequences demonstrated that two main branches represented by strains B5055 and Chedid can be distinguished among *K. pneumoniae* K2 isolates. This finding is in accordance with results of the restriction fragment length polymorphism (RFLP) of the *cps* inner region obtained by Brisse *et al.* (2004) who distinguished five RFLP profiles (C-patterns) for *K. pneumoniae* K2 isolates. In contrast, a single profile was observed for all tested K1 isolates.

Acknowledgements

This work was supported by grants-in-aid for scientific research (3P05D-00225 and 3P05D-05825) from the Ministry of Science and Higher Education of Poland.

Literature

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

Arakawa Y., R. Wacharotayankun, T. Nagatsuka, H. Ito, N. Kato and M. Ohta. 1995. Genomic organization of the *Klebsiella pneumoniae cps* region responsible for serotype K2 capsular

polysaccharide synthesis in the virulent strain Chedid. *Infect. Immun.* 177: 1788–1796.

Astorza B., G. Cortes, C. Crespi, C. Saus, J.M. Rojo and S. Alberti. 2004. C3 promotes clearance of *Klebsiella pneumoniae* by A549 epithelial cells. *Infect. Immun.* 72: 1767–1774.

Brisse S., S. Issenhuth-Jeanjean and P.A.D. Grimont. 2004. Molecular serotyping of *Klebsiella* species isolates by restriction of the amplified capsular antigen gene cluster. *J. Clin. Microbiol.* 42: 3388–3398.

Fung C.-P., F.-Y. Chang, S.-C. Lee, B.-S. Hu, B. I.-T. Kuo, C.-Y. Liu, M. Ho and L.K. Siu. 2002. A global emerging disease of *Klebsiella pneumoniae* liver abscess: is serotype K1 an important factor for complicated endophthalmitis? *Gut* 50: 420–424.

Gierczyński R., S. Kałużewski, A. Rakin, M. Jagielski, A. Zasada, A. Jakubczak, B. Borkowska-Opacka and W. Rastawicki. 2004. Intriguing diversity of *Bacillus anthracis* in eastern Poland – the molecular echoes of the past outbreaks. *FEMS Microbiol. Lett.* 239: 235–240.

Gierczyński R., S. Kałużewski, A.A. Zasada, W. Rastawicki and M. Jagielski. 2005a. Occurrence of selected loci of the *cps* region for capsule K1 or K2 synthesis in epidemic strains of *Klebsiella pneumoniae* isolated from infants in Poland. (in Polish) *Med. Dośw. Mikrobiol.* 57: 51–63

Gierczyński R., S. Kałużewski, A.A. Zasada, W. Rastawicki and M. Jagielski. 2005b. Single strand conformation polymorphism (SSCP) of selected loci of the *cps* region for capsule synthesis in epidemic and occasionaly isolated strains of *Klebsiella pneumoniae* in Poland (in Polish). *Med. Dośw. Mikrobiol.* 57: 153–161.

Kałużewski S. 1965. Biochemical properties of *Klebsiella pneumoniae* applied to strain typing. *Exp. Med. Microbiol.* 17: 1–8.

Kałużewski S. 1968. Some partial agents of unencapsulated variants of group O2 *Klebsiella.* I. Characteristics of strains and antigen O preparations. *Exp. Med. Microbiol.* 20:16–32.

Kumar S., K. Tamura, I.B. Jakobsen and M. Nei. 2001. MEGA2: Molecular Evolutionary Genetics Analysis software. *Bioinformatics* 17: 1244–1245.

Lai, Y.-Ch., H.-L. Peng and H.-Y. Chang. 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.

Nelson K. and R.K. Selander. 1994. Intergeneric transfer and recombination of the 6-phosphogluconate dehydrogenase gene (*gnd*) in enteric bacteria. *Proc. Natl. Acad. Sci. USA* 91: 10227–10231.

Podschun R. and U. Ulmann. 1998. *Klebsiella* spp. as nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors. *Clin. Microbiol. Rev.* 11: 589–603.

Rahn A., J. Drummelsmith and C. Whitfield. 1999. Conserved organisation of the *cps* gene clusters for expression of *Escherichia coli* group 1K antigens: relationship to colanic acid biosynthesis locus and the *cps* genes from *Klebsiella pneumoniae. J. Bacteriol.* 181: 2307–2313.