Mol. Cells, Vol. 19, No. 3, pp. 365-374

Molecules and Cells

Identification of Essential Genes in *Streptococcus pneumoniae* by Allelic Replacement Mutagenesis

Jae-Hoon Song[1,2,*], Kwan Soo Ko[1], Ji-Young Lee[1], Jin Yang Baek[1], Won Sup Oh[2], Ha Sik Yoon[3], Jin-Yong Jeong[4], and Jongsik Chun[5]

[1] *Asian-Pacific Research Foundation for Infectious Diseases (ARFID), Seoul 135-710, Korea;*

[2] *Division of Infectious Diseases, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 135-710, Korea;*

[3] *LG Life Sciences, Inc., Daejeon 305-850, Korea;*

[4] *Asan Institute for Life Sciences, Seoul 138-736, Korea;*

[5] *School of Biological Sciences, Seoul National University, Seoul 151-742, Korea.*

(Received January 7, 2005; Accepted March 2, 2005)

To find potential targets of novel antimicrobial agents, we identified essential genes of *Streptococcus pneumoniae* using comparative genomics and allelic replacement mutagenesis. We compared the genome of *S. pneumoniae* R6 with those of *Bacillus subtilis*, *Enterococcus faecalis*, *Escherichia coli*, and *Staphylococcus aureus*, and selected 693 candidate target genes with > 40% amino acid sequence identity to the corresponding genes in at least two of the other species. The 693 genes were disrupted and 133 were found to be essential for growth. Of these, 32 encoded proteins of unknown function, and we were able to identify orthologues of 22 of these genes by genomic comparisons. The experimental method used in this study is easy to perform, rapid and efficient for identifying essential genes of bacterial pathogens.

Keywords: Allelic Replacement Mutagenesis; Essential Genes; Genomics; New Antimicrobial Agent; Pneumococci.

Introduction

The emergence of antibiotic resistance in major bacterial pathogens over the past decades poses a growing challenge to public health, and underscores the need for new antimicrobial agents. *Streptococcus pneumoniae* is one of the best examples of the global emergence of resistance in a pathogen (Adam, 2002). However, it is very difficult to discover new classes of antibiotics, the oxazolidinones being the latest example of a successful launch in clinical practice. The discovery of novel antimicrobial agents among natural products or modification of existing antibiotics cannot circumvent the problem of antimicrobial resistance. The recent development of bacterial genomics and the availability of genomic sequences (Fleischmann *et al.*, 1995; Rosamond and Allsop, 2000) allow the identification of potentially novel antibacterial targets. Although the identification of new drug targets does not guarantee the development of new chemical compounds, it may be a first step.

The genomes of bacterial pathogens contain on average more than 2,000 genes, of which less than 25 are targeted by current antibiotics. This suggests that there are many undiscovered drug targets in bacterial genomes. The recent publication of the complete chromosomal DNA sequence of *S. pneumoniae* strains R6 (Hoskins *et al.*, 2001) and TIGR4 (Tettelin *et al.*, 2001), facilitates the genome-wide identification of essential genes encoding potential targets for antibacterial drugs. Recently, 113 conserved essential genes were identified in *S. pneumoniae* by disrupting over 300 candidates using a suicide vector (Thanassi *et al.*, 2002). In another study, 36 essential genes for growth were identified among 144 open reading frames (ORFs) with previously uncharacterized functions (Zalacain *et al.*, 2004). However, these studies have investigated only about 1/5 of the more than 2,000 genes of *S. pneumoniae*.

In the current study, we identified novel essential genes in *S. pneumoniae* using comparative genomics followed by allelic replacement mutagenesis.

* To whom correspondence should be addressed.
Tel: 82-2-3410-0320; Fax: 82-2-3410-0328
E-mail: jhsong@smc.samsung.co.kr

Abbreviation: ORF, open reading frame.

Materials and Methods

Bacterial strains and growth conditions The pneumococcal strain used in this study was the type 2 Avery strain D39, obtained from Dr. James C. Paton of the Molecular Microbiology Unit, Women's and Children's Hospital, North Adelaide, S.A., Australia. It was maintained routinely in Todd-Hewitt broth or agar (Difco, Becton-Dickinson, Sparks, MD) supplemented with 0.5% yeast extract (Difco) (THYE). When needed, 400 μg/ml of kanamycin (Sigma-Aldrich, USA) was added.

Extraction of chromosomal DNA To extract chromosomal DNA, *S. pneumoniae* D39 was grown overnight on blood agar plates at 37°C in 5% CO_2. Chromosomal DNA was extracted by the method of Paton *et al.* (1986). A single colony was picked up with an inoculating loop and resuspended in 20 ml of THYE with 400 μg/ml sodium bicarbonate. The cells were grown at 37°C to $OD_{600\ nm}$ 0.4–0.6, then chilled on ice and harvested by centrifugation at 5,000 rpm for 15 min at 4°C. The pellet was resuspended and washed once with 20 ml of ice-cold TE buffer (10 mM Tris-HCl, pH 8.0 and 1 mM EDTA), centrifuged as above, and the resulting pellet quick frozen at –20°C. The cells were thawed and resuspended in 5 ml TE buffer, and 0.005% sodium deoxycholate and 0.01% SDS added. After incubation at 37°C for 10 min, during which the cells were lysed, 500 μg/ml proteinase K (Sigma Chemical Co., USA) was added and incubated for an additional 10 min. The cell lysate in buffer was gently extracted with an equal volume of phenol/chloroform/isoamyl alchohol (25:24:1) (Gibco-BRL, USA) (Kim, 2003). After centrifugation at 8,000 rpm for 10 min, the upper layer was removed and extracted twice with an equal volume of chloroform/isoamyl alcohol (24:1). The final aqueous extract was made up to 0.3 M sodium acetate, and 2.2 volumes of ethanol were overlayed. The DNA was spooled onto a glass rod and redissolved in 2 ml of TE buffer overnight at 4°C. This preparation was dialyzed against 400 volumes of TE buffer before storing at 4°C. DNA concentrations were determined from the absorbance at 260 nm and adjusted to 0.5 μg/μl.

Selection of target genes by comparative genomics Genome sequence data for *S. pneumoniae* R6 were obtained from a proprietary database (Pathogenome; Genome Therapeutics Corp., USA) and from publicly available data from GenBank and TIGR (The Institute for Genomic Research). Target genes were selected using the Microbial Concordance Tool (Bruccoleri *et al.*, 1998; Thanassi *et al.*, 2002) as follows: the amino acid sequences of 2,046 *S. pneumoniae* ORFs were compared with those of *Bacillus subtilis*, *Enterococcus faecalis*, *Escherichia coli*, and *Staphylococcus aureus*, and genes with > 40% amino acid sequence identity to the corresponding genes in at least two of the other species were selected.

Allelic replacement mutagenesis with the kanamycin resistance cassette The procedure for allelic replacement mutagenesis is presented in Fig. 1 (Chalker *et al.*, 2001; Zalacain *et al.*, 2004). A kanamycin resistance cassette (904 bp) containing Kan^R from *Staphyloccus aureus* ATCC 43300 was amplified with the primer set, Kan-F (5′-AAC AGT GAA TTG GAG TTC GTC TTG TTA TA-3′) and Kan-R (5′-GCT TTT TAG ACA TCT AAA TCT AGG TA-3′) (Pierce *et al.*, 2002; Trieu-Cuot and Courvalin *et al.*, 1983). Two pairs of gene-specific primers, L-F/L-R and R-F/R-R, were used to amplify the left and right flanking regions of each target gene, generating PCR products of 500 to 800 bp in length. Primers L-R and R-F each consisted of 21 nucleotides (5′-GAC GAA CTC CAA TTC ACT GTT-3′ and 5′-AGA TTT AGA TGT CTA AAA AGC-3′, respectively), identical to a segment of the promoter region and the 3′-end of the Kan^R gene, plus 23 nucleotides of target gene-specific sequence.

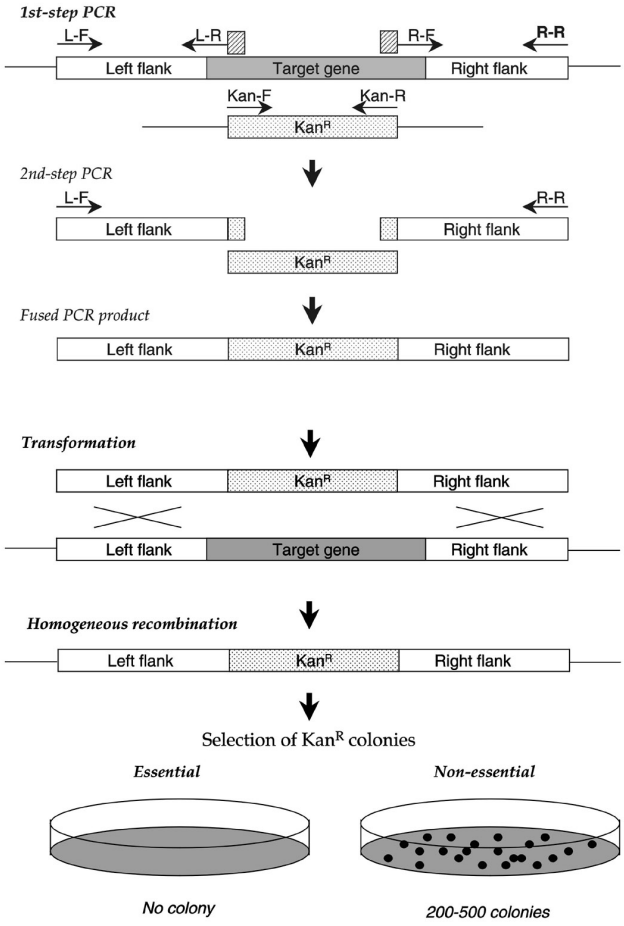


Fig. 1. Overall strategy for allelic replacement mutagenesis and gene knockout by homologous recombination. In the first PCR reaction, the *aph*III gene (Kan^R) and the up- and downstream regions of target genes are amplified. In the second PCR reaction, the up- and downstream fragments are fused to the amplified Kan^R gene using primer L-F and R-R. The fused PCR product is introduced into *S. pneumoniae* strain by transformation, and homologous recombination replaces the target gene with Kan^R. If the target gene is essential, no Kan^R colonies are obtained.

In order to minimize potential polar effects of the mutants, primers were designed so that the flanking genes and intergenic regions including potential promoters would remain intact. In addition, transcription termination signals were removed from Kan^R and the cassettes were designed to integrate in the same orientation as the target genes to ensure transcription of the downstream region. PCR amplifications were run in a 96-well plate format in the following conditions: 30 cycles of 94°C for 1 min, 55°C for 1 min, and 72°C for 1 min 30 s, and final extension of 72°C for 10 min. Each PCR product was purified using a Core-One PCR purification kit (Corebio Systems Co., Korea). A template mixture of the amplified Kan^R gene and the two PCR products flanking the target gene was then subjected to PCR amplification to produce a linear fused product using primers L-F and R-R. The PCR was carried out in 50 μl containing 2 μl of each of the left and right flanking PCR products and the Kan^R gene cassette, 5 μl of 10× buffer, 1 μl of each primer (L-F and R-R) (25 pmol/μl), 5 μl of dNTP mix (25 mM each), and 1 unit of *Taq* polymerase. The PCR condition was as follows: 30 cycles of 94°C for 40 s, 50°C for 40 s, and 72°C for 2 min 30 s, and a final extension of 72°C for 10 min.

Preparation of competent *S. pneumoniae* cells A colony was picked from a fresh blood agar plate and resuspended in 1.5 ml of THYE. One hundred μl of the suspension was used to inoculate 50 ml of the same medium and grown at 37°C overnight. Five ml of the culture was added to 45 ml fresh medium and grown at 37°C to OD_{600} for 4 to 5 h. Sterile glycerol was added to a final concentration of 10%, and the cells were aliquoted in 1 ml volumes, frozen in a dry ice-ethanol bath, and stored at −80°C.

Transformation and gene disruption The linear fused product was introduced into the chromosomal genome of *S. pneumoniae* D39 by transformation and homologous recombination. As a result the Kan^R gene cassette replaced the chromosomal copy of the target gene, thereby creating a gene knockout. Pneumococcal transformation was carried out under the following conditions (Havarstein *et al.*, 1995): 1 μg DNA and 200 μl *S. pneumoniae* D39 competent cells were diluted 1:10 in competence medium containing THYE, 0.2% bovine serum albumin, 0.01% $CaCl_2$, and 100 ng/ml peptide pheromone CSP (H-Glu-Met-Arg-Leu-Ser-Lys-Phe-Arg-Asp-Phe-Ile-Leu-Gln-Arg-Lys-Lys-OH). CSP was synthesized at Takara Korea (Korea) (Havarstein *et al.*, 1995). The cells were incubated at 37°C for 2.5 to 3 h without shaking and plated on THYE plus 400 μg/ml kanamycin, a concentration shown in a preliminary study not to permit the growth of any kanamycin resistant derivatives, and grown at 37°C for 24 h in a CO_2 incubator. If no Kan^R colonies were obtained, the transformation was repeated at least twice. Genes were regarded as essential if no colony appeared in any of the three transformations. If one or more Kan^R colonies were obtained, the target genes were considered non-essential. In all transformation experiments, THYE with 5% lysed sheep blood was used for growth of bacteria and preparation of competent cells.

Confirmation of gene replacement Gene replacement in mutant clones was confirmed by PCR. Genomic DNAs of mutant and wild-type strains were used as templates for PCR amplification with primers L-F and R-R to verify correct incorporation of the fused construct into the mutant genome. The PCR reaction was carried out under the same conditions as the fusion of the three PCR products (30 cycles of 95°C for 40 s, 50°C for 40 s, and 72°C for 2 min 30 s). Correct incorporation of the fused construct gives rise to a larger or smaller PCR product in a mutant than in the wild type, depending on the target gene.

Results

Screening of target genes by comparative genomics By comparing the genome of *S. pneumoniae* R6 with those of *B. subtilis*, *E. faecalis*, *E. coli*, and *S. aureus*, a total of 693 candidate target genes were selected by the criterion of 40% global amino acid sequence identity with the corresponding gene in at least two bacterial species. Genes that had already been reported (Thanassi *et al.*, 2002) were excluded from subsequent analysis.

Identification of essential genes by allelic replacement mutagenesis Our procedure for identifying essential genes by allelic replacement mutagenesis was evaluated using an essential and a non-essential *S. pneumoniae* gene identified previously (Thanassi *et al.*, 2002). Knockout of the non-essential gene (spr0746) typically produced 300 to 500 colonies, while knockout of the essential gene (spr0478) produced no colonies. Mutant strains generated by homologous recombination yielded larger or smaller PCR fragments than the wild type (Fig. 2).

Our method was also evaluated with regard to polar effects. spr0004 and spr0005 (SP0004 and SP0005 in TIGR4) had been reported to be essential (Thanassi *et al.*, 2002). However, we identified it as non-essential using Kan^R cassette without polarity (Fig. 3). Thus, allelic replacement mutagenesis can determine if genes of both monocistronic and polycistronic types are essential.

Essential genes of *S. pneumoniae* The essential genes of *S. pneumoniae* identified in this study are listed in Table 1. Of 693 disrupted candidate genes tested, 133 were identified as essential. Of these, 32 encode proteins of unknown function. In addition to the 113 essential genes reported by Thanassi *et al.* (2002), the functions of a total of 246 essential genes are summarized in Table 2. The genes are classified on the basis of their clusters-of-orthologous-groups (COGs) functional categories (Tatusov *et al.*, 1997). Our study identified novel essential genes related to carbohydrate transport and metabolism (G). Essential genes related to cell motility and secretion (N) and signal transduction (T) were not identified in this study, and genes related to secondary metabolite biosynthesis, trans-

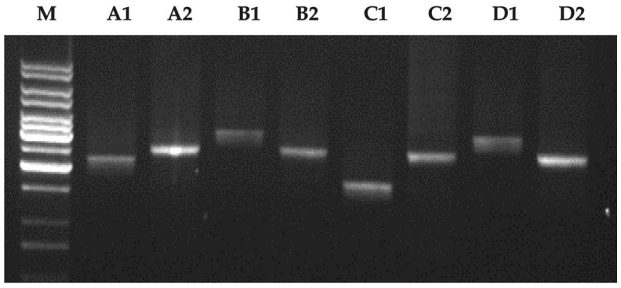
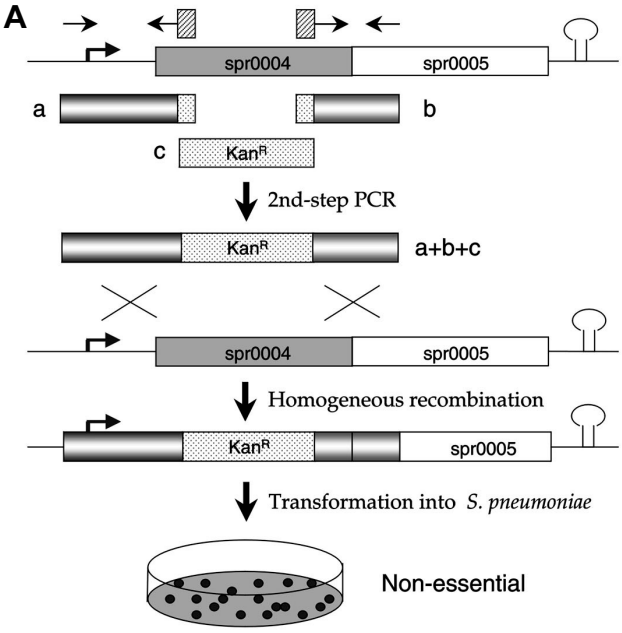


Fig. 2. Confirmation of gene replacement. Lane M, 1 kb ladder marker; lane A1, spr0147 wild type (wt) (2265 bp); lane A2, spr0147 mutant (2542 bp); lane B1, spr0232 wt (3258 bp); lane B2, spr0232 mutant (2600 bp); lane C1, spr0746 wt (1718 bp); lane C2, spr0746 mutant (2448 bp); lane D1, spr1153 wt (3302 bp); lane D2, spr1153 mutant (2414 bp).

port, and catabolism (Q) were not identified in either study (Table 2).

Discussion

The identification of essential genes in pathogenic bacteria is of considerable relevance because they are potential targets for novel antimicrobial agents (Zhang *et al.*, 2004). Several studies based on genome-driven, target-based approaches have provided a valuable inventory of essential genes that can be used to select and validate antimicrobial agents (Akerley *et al.*, 2002; Arigoni *et al.*, 1998; Forsyth *et al.*, 2002; Hutchison *et al.*, 1999; Ji *et al.*, 2001; Thanassi *et al.*, 2002). Peptide deformylase (PDF) inhibitors are a good example of the genomics-driven, target-based approach to discovery of novel antimicrobial agents. Several methods have been used to identify conserved essential gene in a few bacterial species; random insertional mutagenesis by transposons or plasmids (Akerley *et al.*, 2002; Hutchison *et al.*, 1999; Kobayashi *et al.*, 2003; Sassetti *et al.*, 2001), inducible expression of antisense RNA molecules (Forsyth *et al.*, 2002; Ji *et al.*, 2001), and systematic gene inactivation (Chalker *et al.*, 2001; Thanassi *et al.*, 2002). In this study, we used a very efficient method to identify essential genes in *S. pneumoniae*, namely comparative genomics combined with allelic replacement mutagenesis (Zalacain *et al.*, 2004). Compared with previous methods (Akerley *et al.*, 2002; Forsyth *et al.*, 2002; Ji *et al.*, 2001; Kobayashi *et al.*, 2003; Sassetti *et al.*, 2001; Thanassi *et al.*, 2002), this method has some advantages for identifying essential genes of pathogenic bacteria. First, stepwise filtering of ORFs through genome comparison with other species based on simple criteria can reduce the number of genes to be tested (Bruccoleri *et al.*, 1998; Thanassi *et al.*, 2002). Second, *a priori* knowledge of target genes makes it unnecessary to sequence *a posteriori* for identification.



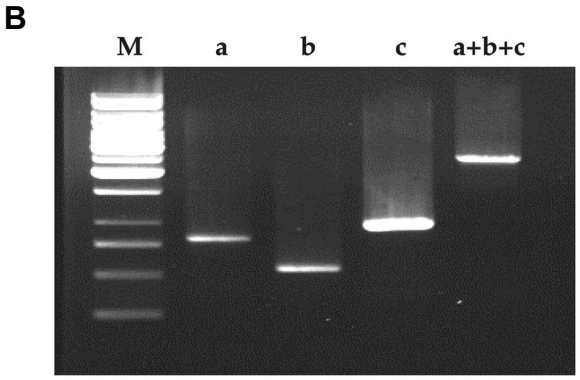


Fig. 3. Knockout of spr0004 to confirm the absence of a polar effect. **A.** In the first PCR reaction, the Kan^R gene lacking the transcriptional termination signal and the up- and downstream regions of spr0004 were amplified. As a result, the upstream (a, 813 bp) and downstream (b, 558 bp) regions and Kan^R (c, 904 bp) were obtained and a roughly 2.3 kb fused PCR product (a + b + c, 2274 bp) was produced by the second PCR. There were hundreds of colonies on the THYE agar plate containing kanamycin, indicating that spr0004 is non-essential. **B.** Lane M, 1 kb ladder marker; lane a, spr0004 left fragment (813 bp); lane b, spr0004 right fragment (558 bp); lane c, Kan^R cassette (904 bp); lanes a + b + c, second PCR fragment (2274 bp).

Third, allelic replacement mutagenesis by two-step PCR does not require a vector for recombination. Fourth, the method minimizes polar effects and can be applied to both monocistronic and polycistronic genes.

We identified 133 genes in *S. pneumoniae* that are essential for growth. We excluded essential genes already identified by Thanassi *et al*. (2002). Our data included genes present in eukaryotic cells, whereas Thanassi *et al.*

Table 1. List of new essential genes of *S. pneumoniae**.

R6 gene No.	Gene name	Gene description
Information storage and processing		
Translation, ribosomal structure and biogenesis		
spr0243	proS	Prolyl-tRNA synthetase
spr0372	serS	Seryl-tRNA synthetase
spr0393	gatB	Glutamyl tRNA-Gln amidotransferase subunit B
spr0395	gatC`	Glutamyl tRNA-Gln amidotransferase, subunit C
spr0626	lysS	Lysyl-tRNA synthetase (lysine--tRNA ligase) (LYSRS)
spr0670	rheB	ATP-dependent RNA helicase
spr0861	infC	Translation initiation factor IF-3
spr0992	map	Methionine aminopeptidase
spr1310	fms	Peptide deformylase, N-formylmethionylaminoacyl-tRNA deformylase
spr1329	glyQ	Glycyl-tRNA synthetase alpha chain
spr1413	cca	tRNA nucleotidyltransferase
spr1472	thrS	Threonyl-tRNA synthetase 1
spr1502	ileS	Isoleucyl-tRNA synthetase
spr1580	fmt	Methionyl-tRNA formyltransferase
spr1853	rnpA	Ribonuclease P - protein component
spr1881	gltX	Glutamyl-tRNA synthetase (glutamate--tRNA ligase)
spr1910	tyrS	Tyrosyl-tRNA synthetase 1
spr2034	trpS	Tryptophanyl-tRNA synthetase
Transcription		
spr0478	nusA	Transcription termination
spr0670	rheB	ATP-dependent RNA helicase
spr0979	rpoD	RNA polymerase sigma factor 70
spr1439	codV	Transcriptional pleiotropic repressor
spr1776	rpoC	DNA-dependent RNA polymerase
spr2013	comFA	Involved in transformation; required for DNA uptake but not for binding; Related to ATP-dependent RNA/DNA helicase
DNA replication, recombination and repair		
spr0670	rheB	ATP-dependent RNA helicase
spr0756	parE	Topoisomerase IV subunit B
spr0978	dnaG	DNA primase
spr0996	radC	DNA repair protein
spr1024	ligA	DNA ligase
spr1555	dnaI	Primosome component (helicase loader)
spr2013	comFA	Involved in transformation; required for DNA uptake but not for binding; Related to ATP-dependent RNA/DNA helicase
Cellular processes		
Cell division and chromosome partitioning		
spr0666	ftsE	ABC transporter ATP-binding protein & cell division protein
spr1510	ftsZ	Cell division protein FtsZ
spr1511	ftsA	Cell division protein FtsA
Post-translational modification, protein turnover, chaperones		
spr0131	gcp	Secreted metalloendopeptidase Gcp
spr1064	nrdH	Glutaredoxin-like protein involved in electron transport system for ribonucleotide reductase system NrdEF
spr1312	trxB	Thioredoxin reductase
spr1427	clpX	ATP-dependent Clp protease ATP-binding subunit (class III heat-shock protein)

(Continued)

R6 gene No.	Gene name	Gene description
Cell envelope biogenesis, outer membrane		
spr0139	ugd	UDP-glucose dehydrogenase
spr0304	pbpX	Penicillin-binding protein 2X
spr0603	murD	UDP-N-acetylmuramoyl-L-alanine--D-glutamate ligase
spr0604	murG	Undecaprenyl-PP-MurNAc-pentapeptide-UDPGlcNAc GlcNAc transferase
spr1247	murB	UDP-N-acetylenolpyruvoylglucosamine reductase
spr1373	murC	UDP-N-acetylmuramate-alanine ligase
spr1460	galE	UDP-glucose 4-epimerase
spr1514	murF	UDP-N-acetylmuramoylalanine-D-glutamyl-lysine-D-alanyl-D-alanine ligase
spr1515	ddl	D-alanine-D-alanine ligase
spr1540	alr	Alanine racemase
spr1647	galE	UDP-glucose 4-epimerase
spr1696	murI	Glutamate racemase
Inorganic ion transport and metabolism		
spr0173	arsC	Arsenate reductase, putative
spr0906	lmb	Lipoprotein
spr1410	pacL	P-type ATPase - calcium transporter
spr1430	dpr	DNA binding protein starved cells-like peroxide resistance protein
Metabolism		
Energy production and conversion		
spr1359	atpC	Proton-translocating ATPase, F1 sector, epsilon-subunit
spr1360	atpD	Proton-translocating ATPase, F1 sector, beta-subunit
spr1361	atpG	Proton-translocating ATPase, F1 sector, gamma-subunit
spr1362	atpA	Proton-translocating ATPase, F1 sector, alpha-subunit
spr1364	atpF	Proton-translocating ATPase, F0 sector, subunit b
spr1458	fer	Ferredoxin
spr1963	adh2	Probable alcohol dehydrogenase
Carbohydrate transport and metabolism		
spr0081	ABC-MSP	ABC transporter membrane-spanning permease - sugar transport
spr0356	mtlA	Mannitol PTS EII
spr0359	mtlD	Mannitol-1-phosphate 5-dehydrogenase
spr0530	fba	Fructose-bisphosphate aldolase
spr0646	bgl-truncation	Phospho-beta-gluco or galactosidase, truncation
spr0796	pfkA	6-phosphofructokinase I
spr0797	pykF	Pyruvate kinase I; fructose-stimulated
spr0856	celA	Competence protein
spr1036	eno	Enolase
spr1063	ptsH	Histidine-containing phosphocarrier protein of the PTS
spr1432	tpi	Triose phosphate isomerase
spr1617	sacA	Sucrose-6-phosphate hydrolase
spr1797	rpe	Pentose-5-phosphate-3-epimerase
spr1825	gapA	Glyceraldehyde 3-phosphate dehydrogenase (phosphorylating)
spr1882	gpi	Glucose-6-phosphate isomerase
spr1971	fucU	Fucose pathway protein, function unknown
spr1973	fcsK	Fucose kinase
Amino acid transport and metabolism		
spr0882	pepB	Group B oligopeptidase
spr1001	nifS	Pyridoxal-phosphate dependent aminotransferase
spr1434	metA	Homoserine O-succinyltransferase
spr1553	serA	Phosphoglycerate dehydrogenase
spr1632	trpB	Tryptophan synthase beta chain

(Continued)

R6 gene No.	Gene name	Gene description
Nucleotide transport and metabolism		
spr0047	purF	Amidophosphoribosyl transferase
spr0049	purN	5-phosphoribosylglycinamide transformylase 1
spr0210	adk	Adenylate kinase (ATP-AMP transphosphorylase)
spr0585	thyA	Thymidylate synthase
spr0845	pyrH	UMP kinase
spr1065	nrdE	Ribonucleoside-diphosphate reductase (major subunit)
Coenzyme metabolism		
spr0266	sulA	Dihydropteroate synthase
spr0267	sulB	Dihydrofolate synthetase
spr0268	sulC	GTP cyclohydrolase
spr0269	sulD	Aldolase-pyrophosphokinase
spr0671	metK	S-adenosylmethionine synthetase
spr0729	folD	Fold bifunctional protein; includes: methylenetetrahydrofolate dehydrogenase, methenyltetrahydrofolate cyclohydrolase.
spr1111	dfp N-terminous	Similar to N-terminal region of the Dfp protein
spr1429	dfr	Dihydrofolate reductase
spr1715	birA	Biotin--(acetyl-CoA-carboxylase) ligase
Lipid metabolism		
spr0338	mvk	Mevalonate kinase
spr1541	acpS	Acyl Carrier protein synthase
spr1571	mvaS	3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) synthase
spr2027	pgsA	Phosphotidylglycerophosphate synthase
Poorly characterized		
General function prediction only		
spr0881	coiA	Competence protein
spr2012	comFC	Involved in transformation (competence for DNA uptake)
Function unknown		
spr0003		Hypothetical protein
spr0091		Conserved hypothetical protein
spr0129		Conserved hypothetical protein
spr0145		Hypothetical protein
spr0156		Conserved hypothetical protein
spr0328		Conserved hypothetical protein
spr0755		Conserved hypothetical protein
spr0788		Hypothetical protein
spr0965		Hypothetical protein
spr0972		Conserved hypothetical protein
spr0982		Conserved Hypothetical protein
spr1081		Hypothetical protein
spr1150		Conserved hypothetical protein
spr1158		Conserved hypothetical protein
spr1282		Hypothetical protein
spr1409		Conserved hypothetical protein
spr1411		Conserved hypothetical protein
spr1419		Conserved hypothetical protein
spr1433		Conserved hypothetical protein
spr1468		Conserved hypothetical protein
spr1535		Hypothetical protein
spr1603		Conserved hypothetical protein

(Continued)

R6 gene No.	Gene name	Gene description
spr1642		Hypothetical protein
spr1666		Conserved hypothetical protein
spr1671		Conserved hypothetical protein
spr1674		Conserved hypothetical protein
spr1697		Hypothetical protein
spr1804		Conserved hypothetical protein
spr1856		Hypothetical protein
spr1914		Hypothetical protein
spr1950		Conserved hypothetical protein
spr1987		Hypothetical protein, truncation

* Patents are pending for the newly identified essential genes of *S. pneumoniae*.

Table 2. Functional classification of the 246 essential genes of *S. pneumoniae*.

Cellular role[a]	Subtotal[b]	Total[b]
Information storage and processing		52 (28)
Translation, ribosomal structure, and biogenesis (J)	28 (17)	
Transcription (K)	9 (4)	
DNA replication, recombination, and repair (L)	15 (7)	
Cellular processes		48 (23)
Cell division and chromosome partitioning (D)	6 (3)	
Posttranslational modification, protein turnover, chaperones (O)	5 (4)	
Cell envelope biogenesis, outer membrane (M)	26 (12)	
Cell motility and secretion (N)	4 (-)	
Inorganic ion transport and metabolism (P)	5 (4)	
Signal transduction mechanisms (T)	2 (-)	
Metabolism		77 (48)
Energy production and conversion (C)	10 (7)	
Carbohydrate transport and metabolism (G)	20 (17)	
Amino acid transport and metabolism (E)	11 (5)	
Nucleotide transport and metabolism (F)	9 (6)	
Coenzyme metabolism (H)	13 (9)	
Lipid metabolism (I)	14 (4)	
Secondary metabolites biosynthesis, transport, and catabolism (Q)	-	
Poorly characterized		69 (34)
General function prediction only (R)	18 (2)	
Function unknown (S)	51 (32)	
Total		246 (133)

[a] Classification was based on COG (Clusters of Orthologous Groups of proteins) functional categories of the NCBI (http://www.ncbi.nlm.nih.gov/COG/).

[b] Data include essential genes reported by Thanassi *et al.* (2002). Numbers of essential genes identified in this study are indicated in parenthesis.

excluded those genes from their list. Examples of the newly found essential genes include the peptide deformylase (PDF) gene, shown to be essential in *Escherichia coli* (Waller and Clements, 2002). This gene could be a target for a broad-spectrum antimicrobial agent (Apfel *et al.*, 2001; Mills, 2003; Waller and Clements, 2002). Genes involved in the biosynthesis of UDP-N-acetyl muramyl pentapeptide such as *murD*, *murG*, *murB*, *murC*, *murF*, and *murI* were also found to be essential in this study. Since these genes are responsible for the biosynthesis of UDP-N-acetyl muramyl pentapeptide in cell wall formation they have been assessed as targets for new antimicrobial agents (El Zoeiby *et al.*, 2003).

Thirty two genes whose function is unknown were identified as essential. Although the usefulness of such genes for drug development is uncertain, they may pro-

Table 3. Comparison of essential genes with unknown function in R6 with a comprehensive microbial resource (www.tigr.org) using gene vs. all alignment.

R6 gene no.	*Streptococcus mitis* NCTC 12261		Putative identification in another bacterial genome[a]
	Identity (%)	P-value	
spr0003	96.9	6.7 e^{-32}	
spr0091	None		Glycosyl transferse in *Lactococcus lactis* IL1403
spr0129	97.4	1.3 e^{-108}	Glycoprotease family subfamily
spr0145	None		
spr0156	95.6	4.7 e^{-79}	*Nrdl* protein (imported)
spr0328	None		Similar to BL0464 of *Bifidobacterium longum* NCC 2705
spr0755	94.8	5.6 e^{-99}	Putative membrane protein
spr0788	96.6	6.2 e^{-77}	Putative esterase superfamily
spr0965	36.0	8.4 e^{-08}	Tetrahydrodipicolinate acetyltransferase in *Staphylococcus aureus*
spr0972	None		Transcriptional regulator, Cro/CI family in *Streptococcus agalactiae* 2603V/R
spr0982	96.6	6.2 e^{-219}	Glycosyl transferase in *S. pneumoniae* TIGR4 and other bacterial species
spr1081	93.7	3.6 e^{-33}	Probable PTS system IIA component
spr1150	98.2	3.2 e^{-231}	Putative polysaccharide biosynthesis protein
spr1158	96.7	1.5 e^{-80}	
spr1282	None		
spr1409	None		Putative glutathione S-transferase in *S. pneumoniae* TIGR4 and other bacterial species
spr1411	96.5	39 e^{-178}	Cation efflux family protein in S. pneumoniae TIGR4 and other bacterial species
spr1419	88.1	2.4 e^{-130}	
spr1433	96.9	2.5 e^{-112}	DnaD-like domain protein
spr1468	None		
spr1535	26.5	2.8 e^{-10}	
spr1603	83.8	2.2 e^{-58}	Putative GTP cyclohydrolase I subfamily
spr1642	None		
spr1666	96.3	4.4 e^{-67}	
spr1671	94.9	8.8 e^{-46}	Transcriptional regulator, MerR family in *S. pneumoniae* TIGR4 and other bacterial species
spr1674	56.4	4.3 e^{-76}	Putative nicotinamide mononucleotide transporter PnuC
spr1697	99.0	1.7 e^{-35}	
spr1804	95.2	8.1 e^{-91}	Primase-related protein in *S. pneumoniae* TIGR4 and *S. mitis* NCTC12261
spr1856	85.2	1.7 e^{-51}	None in *S. pneumoniae* TIGR4, putative methyltransferase
spr1914	None		
spr1950	95.2	1.8 e^{-134}	ROK family protein in *S. pneumoniae* TIGR4 and other bacterial species
spr1987	None		

[a] Putatively identified in *Streptococcus mitis* NCTC12261 if not otherwise specified.

vide completely new opportunities (Akerley *et al.*, 2002; Forsyth *et al.*, 2002; Ji *et al.*, 2001; Kobayashi *et al.*, 2003; Thanassi *et al.*, 2002; Waller and Clements, 2002). As they are likely to encode components of basic cellular processes (Akerley *et al.*, 2002; Jordan *et al.*, 2002), their study is important for understanding bacterial cell biology as well as for use as targets of narrow-spectrum antimicrobial agents. A comprehensive microbial resource (http://www.tigr.org) was used to analyze the genes of *S. pneumoniae* R6 whose function was unknown. Using the option "gene vs. all alignment", we compared sequence identities and P-values for the R6 strain with another *S. pneumoniae* strain, TIGR4, and with other species including *Streptococcus mitis* NCTC 12261 (Table 3) and were able to identify orthologues of 22 of these essential genes with unknown function. spr1856 had no similar gene in *S. pneumoniae* TIGR4, but it had an orthologue in *S. mitis* NCTC 12261. Some essential genes are shared by other bacterial species (Kobayashi *et al.*, 2003), while other genes are not always essential in other species (Akerley *et al.*, 2002). Conserved essential genes present in many species could be targets of broad-spectrum agents, whereas essential genes unique to particular species could be targets of narrow-spectrum agents.

Interestingly, two genes in the fucose pathway (*fucU* and *fucK*) were identified as essential. However an experiment using the fucose promoter to regulate gene expression in *S. pneumoniae* (Chan *et al.*, 2003), indicated

that the fucose pathway was inactive in pneumococci. Therefore, their function in *S. pneumoniae* needs to be further investigated.

In summary, we have identified 133 new essential genes in *S. pneumoniae* using comparative genomics and allelic replacement mutagenesis via two-step PCR. The experimental method used in this study is simple and efficient method for identifying essential genes in bacterial pathogens.

Acknowledgment This study was supported by the IMT-2000 project of the Ministry of Commerce, Industry and Energy, Republic of Korea.

References

Adam, D. (2002) Global antibiotic resistance in *Streptococcus pneumoniae. J. Antimicrob. Chemother.* **50**, (Topic T1), 1–5.

Akerley, B. J., Rubin, E. J., Novick, V. L., Amaya, K., Judson, N., *et al.* (2002) A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenzae*. *Proc. Natl. Acad. Sci. USA* **99**, 966–971.

Apfel, C. M., Locher, H., Evers, S., Takács, B., Hubschwerlen, C., *et al.* (2001) Peptide deformylase as an antibacterial drug target: target validation and resistance development. *Antimicrob. Agents Chemother.* **45**, 1058–1064.

Arigoni, F., Talabot, F., Peitsch, M., Degerton, M. D., Meldrum, E., *et al.* (1998) A genome-based approach for the identification of essential bacterial genes. *Nat. Biotechnol.* **16**, 851–858.

Bruccoleri, R. E., Dougherty, T. J., and Davison, D. B. (1998) Concordance analysis of microbial genomes. *Nucleic Acids Res.* **16**, 4482–4486.

Chalker, A. F., Minehart, H. W., Hughes, N. J., Koretke, K. K., Lonetto, M. A., *et al.* (2001) Systematic identification of selective essential genes in *Helicobacter pylori* by genome prioritization and allelic replacement mutagenesis. *J. Bacteriol.* **183**, 1259–1268.

Chan, P. F., O'Dwyer, K. M., Palmer, L. M., Ambrad, J. D., Ingraham, K. A., *et al.* (2003) Characterization of a novel fucose-regulated promoter (P_{fcsK}) suitable for gene essentiality and antibacterial mode-of -action studies in *Streptococcus pneumoniae*. *J. Bacteriol.* **185**, 2051–2058.

El Zoeiby, A., Sanschagrin, F., and Levesque, R. C. (2003) Structure and function of the Mur enzymes: development of novel inhibitors. *Mol. Microbiol.* **47**, 1–12.

Forsyth, R. A., Haselbeck, R. J., Ohlsen, K. L., Yamamoto, R. T., Xu, H., *et al.* (2002) A genome-wide strategy for the identification of essential genes in *Staphylococcus aureus*. *Mol. Microbiol.* **43**, 1387–1400.

Havarstein, L., Coomaraswamy, G., and Morrison, D. A. (1995) An unmodified heptadecaptide pheromone induces competence for genetic transformation in *Streptococcus pneumoniae. Proc. Natl. Acad. Sci. USA* **92**, 11140–11144.

Hoskins, J., Alborn, W. E. Jr, Arnold, J., Blaszczak, L. C., Burgett, S., *et al.* (2001) Genome of the bacterium *Streptococcus pneumoniae* strain R6. *J. Bacteriol.* **183**, 5709–5717.

Hutchison, C. A., Pterson, S. N., Gill, S. R., Cline, R. T., White, O., *et al.* (1999) Global transposon mutagenesis and a minimal *Mycoplasma* genome. *Science* **286**, 2165–2169.

Ji, Y., Zhang, B., Van Horn, S. F., Warren, P., Woodnutt, G., *et al.* (2001) Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. *Science* **293**, 2266–2269.

Jordan, I. K., Rogozin, I. B., Wolf, Y. I., and Koonin, E. V. (2002) Essential genes are more evolutionary conserved than are nonessential genes in bacteria. *Genome Res.* **12**, 962–968.

Kim, C.-H. (2003) A *Salmonella typhimurium rfaE* mutant recovers invasiveness for human epithelial cells when complemented by wild type *rfaE* (controlling, biosynthesis of ADP-L-glycero-D-manno-heptose-containing lipopolysaccharide). *Mol. Cells* **15**, 226–232.

Kobayashi, K., Ehrlich, S. D., Albertini, A., Amati, G., Andersen, K. K., *et al.* (2003) Essential *Bacillus subtilis* genes. *Proc. Natl. Acad. Sci. USA* **100**, 4678–4683.

Mills, S. D. (2003) The role of genomics in antimicrobial discovery. *J. Antimicrob. Chemother.* **51**, 749–752.

Paton, J. C., Berry, A. M., Lock, R. A., Hansman, D., and Manning, P. A. (1986) Cloning and expression in *Escherichia coli* of *the Streptococcus pneumoniae* gene encoding pneumolysin. *Infect. Immun*. **54**, 50–55.

Pierce, B. J., Ianelli, F., and Pozzi, F. (2002) Construction of new unencapsulated (rough) strains of *Streptococcus pneumoniae Res. Microbiol.* **153**, 243–247.

Sassetti, C. M., Boyd, D. H., and Rubin, E. J. (2001) Comprehensive identification of conditionally essential genes in mycobacteria. *Proc. Natl. Acad. Sci. USA* **98**, 12712–12717.

Tatusov, R. L., Koonin, E. V., and Lipman, D. J. (1997) A genomic perspective on protein families. *Science* **278**, 631–637.

Tettelin, H., Nelson, K. E., Paulsen, I. T., Eisen, J. A., Read, T. D., *et al.* (2001) Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* **293**, 498–506.

Thanassi, J. A., Hartman-Neumann, S. L., Dougherty, T. J., Dougherty, B. A., and Pucci, M. J. (2002) Identification of 113 conserved essential genes using a high-throughput gene disruption system in *Streptococcus pneumoniae*. *Nucleic Acids Res.* **30**, 3152–3162.

Trieu-Cuot, P. and Courvalin, P. (1983) Nucleotide sequence of the *Streptococcus faecalis* plasmid gene encoding the 3′5′-aminoglycoside phosphotransferase type III. *Gene* **23**, 331–341.

Waller, A. S. and Clements, J. M. (2002) Novel approaches to antimicrobial therapy: peptide deformylase. *Curr. Opin. Drug Discov. Devel.* **5**, 785–792.

Zalacain, M., Biswas, S., Ingraham, K. A., Ambrad, J., Bryant, A., *et al.* (2004) A global approach to identify novel broad-spectrum antibacterial targets among proteins of unknown function. *J. Mol. Microbiol. Biotechnol.* **6**, 109–126.

Zhang, R., Ou, Z. Y., and Zhang, C. T. (2004) DEG: a database of essential genes. *Nucleic Acids Res.* **32**, D271–D272.