28

Detection of Essential Genes in *Streptococcus pneumoniae* Using Bioinformatics and Allelic Replacement Mutagenesis

Jae-Hoon Song and Kwan Soo Ko

1. Introduction

Although the emergence and spread of antimicrobial resistance in major bacterial pathogens for the past decades poses a growing challenge to public health, discovery of novel antimicrobial agents from natural products or modification of existing antibiotics cannot circumvent the problem of antimicrobial resistance. The recent development of bacterial genomics and the availability of genome sequences allow the identification of potentially novel antimicrobial agents. The cellular targets of new antimicrobial agents must be essential for the growth, replication, or survival of the bacterium. Conserved genes among different bacterial genomes often turn out to be essential ***(1, 2)***. Thus, the combination of comparative genomics and the gene knock-out procedure can provide effective ways to identify the essential genes of bacterial pathogens ***(3)***. Identification of essential genes in bacteria may be utilized for the development of new antimicrobial agents because common essential genes in diverse pathogens could constitute novel targets for broad-spectrum antimicrobial agents.

In this chapter, we introduce a rapid and efficient method for the identification of essential genes in *Streptococcus pneumoniae* that combines comparative genomics and allelic replacement mutagenesis.

2. Materials

2.1. Streptococcus pneumoniae

1. *S. pneumoniae* strain D39.
2. Todd-Hewitt broth or agar (Difco, Becton-Dickinson, Sparks, MD) supplemented with 0.5% yeast extract (Difco) (THYE).

From: *Methods in Molecular Biology, vol. 416: Microbial Gene Essentiality*
Edited by: A. L. Osterman and S. Y. Gerdes © Humana Press Inc., Totowa, NJ

3. Kanamycin (Sigma-Aldrich, St. Louis, MO).
4. Blood agar plate (BAP).
5. Resuspending solution: TE buffer (10 mM Tris-HCl, pH 8.0, and 1 mM EDTA), 0.005% sodium deoxycholate, and 0.01% SDS.
6. Proteinase K (Sigma-Aldrich).
7. Phenol/chloroform/isoamyl alcohol (25:24:1) (Invitrogen, Carlsbad, CA).
8. Oligonucleotide primers: Kan-F (5′-AAC AGT GAA TTG GAG TTC GTC TTG TTA TA-3′), Kan-R (5′-GCT TTT TAG ACA TCT AAA TCT AGG TA-3′), and others.
9. Agarose electrophoresis and polymerase chain reaction (PCR) equipment.
10. CoreOne PCR purification kit (CoreBioSystem, Seoul, Korea).
11. Competence medium: THYE, 0.2% bovine serum albumin, 0.01% $CaCl_2$, and 100 ng/mL peptide pheromone CSP (Takara Korea, Seoul, Korea): H-Glu-Met-Arg-Leu-Ser-Lys-Phe-Arg-Asp-Phe-Ile-Leu-Gln-Arg-Lys-Lys-Oh.

3. Methods

3.1. Selection of Target Genes by Bioinformatics

S. pneumoniae R6 genome sequence data are used for selection of target genes. Target genes are selected using the Microbial Concordance Tool ***(4, 5)*** as follows: the amino acid sequences of 2046 *S. pneumoniae* R6 open reading frames (ORFs) are compared with those of *Bacillus subtilis*, *Enterococcus faecalis*, *Escherichia coli*, and *Staphylococcus aureus*, and genes of more than 40% amino acid sequence identity to the corresponding genes in at least two of the other species are selected.

3.2. Preparation of Competent Cells

1. To prepare competent cells, *S. pneumoniae* is plated and cultured on a fresh blood agar (**Note 1**).
2. One colony is picked from a cultured plate and resuspended in 1.5 mL THYE. One hundred microliters of the resuspension is used to inoculate 50 mL of the same medium, which is grown at 37°C overnight.
3. Five milliliters of the culture is added to 45 mL fresh medium and is grown at 37°C to OD_{600} for 4 to 5 h.
4. Sterile glycerol is added to a final concentration of 10%, and cells are aliquoted in 1-mL samples, frozen in a dry ice–ethanol bath, and stored at −80°C.

3.3. Allelic Replacement Mutagenesis

3.3.1. Extraction of Genomic DNA

1. *S. pneumoniae* D39 is grown overnight on a blood agar plate at 37°C in 5% CO_2 for extraction of genomic DNA.
2. A single colony is removed with an inoculating loop and resuspended in 20 mL Todd-Hewitt agar supplemented with 0.5% yeast extract (THYE) with 400 μg/mL sterile sodium bicarbonate.

3. The bacterial cells are grown at 37°C until an OD_{600nm} reaches 0.4 to 0.6 and are then chilled on ice and harvested by centrifugation at 5000 rpm for 15 min at 4°C.
4. The pellet is resuspended and washed once with 20 mL ice-cold TE buffer, centrifuged as above, and the resulting pellet quick frozen at –20°C.
5. The cells are thawed and resuspended in 5 mL TE buffer, and 0.005% sodium deoxycholate and 0.01% SDS added. Cells are lysed by incubation at 37°C for 10 min.
6. After cell lysis, 500 μg/mL proteinase K is added and additionally incubated for 10 min. The cell lysate is gently extracted with an equal volume of phenol/chloroform/isoamyl alcohol (25 : 24 : 1).
7. After centrifugation at 8000 rpm for 10 min, the upper layer is removed and extracted twice with an equal volume chloroform/isoamyl alcohol (24 : 1).
8. The final aqueous extract is brought up to 0.3 M sodium acetate and is overlaid with 2.2 volumes of ethanol. The DNA is spooled onto a glass rod and redissolved in 2 mL TE buffer overnight at 4°C.
9. This preparation is dialyzed against 400 volumes of TE buffer before storage at 4°C. DNA concentration is determined by absorbance at 260 nm and adjusted to 0.5 μg/μL.

3.3.2. Preparation of Kanamycin-Resistance Cassette

A kanamycin-resistance cassette (904 bp) containing KanR from *Staphylococcus aureus* ATCC43300 is amplified using 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′) ***(6, 7)***.

3.3.3. Two-Step PCR Protocol

For allelic replacement mutagenesis, two-step PCR is performed (**Fig. 1**). Two pairs of gene-specific primers, L-F/L-R and R-F/R-R, are 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 consist of 21 nucleotides (5′-GAC GAA CTC CAA TTC ACT GTT-3′ and 5′-AGA TTT AGA TGT CTA AAA AGC-3′, respectively), which are identical to the promoter region and the 3′-end of the Kan^R gene, plus 23 nucleotides of target gene–specific sequence.

PCR amplifications are run in 96-well format under 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 is purified using Core-One PCR purification kit (CoreBioSystem). A template mixture of the amplified Kan^R gene and two PCR products flanking the target gene are then subjected to a second PCR amplification to produce a linear fused product using primers L-F and R-R. The second PCR reaction mix contains in a total volume of 50 μL: 2 μL of each, 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 cycling condition are as follows: 30 cycles of 94°C for 40 s, 50°C for 40 s, and 72°C for 2 min 30 s, and the final extension of 72°C for 10 min (**Note 2**).

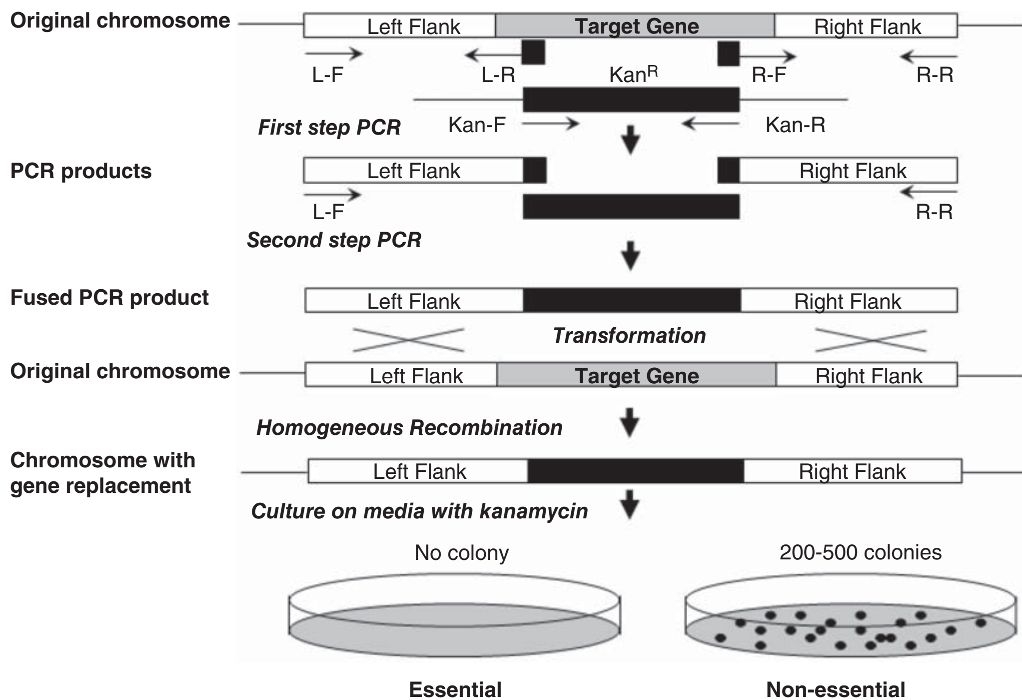

Fig. 1. Two-step PCR procedure to generate a fusion between kanamycin-resistance cassette (Kan^R) and the flanking regions of a target gene: introduction of a fused PCR product into *S. pneumoniae* chromosome via transformation and homologous recombination with the target gene.

3.3.4. S. pneumoniae Transformation

The linear fused product produced by the two-step PCR procedure is introduced into the chromosomal genome of *S. pneumoniae* D39 by transformation and homologous recombination (**Fig. 1**; *see* **Note 3**). Pneumococcal transformation is performed as follows ***(8)***:

1. 1 µg DNA and 200 µL *S. pneumoniae* D39 competent cells are diluted 1:10 in competence medium containing peptide pheromone CSP (Takara Korea; *see* **Note 4** and ***[8]***).
2. Cells are incubated at 37°C for 2.5 to 3 h without shaking and are plated on THYE with 400 µg/mL kanamycin (**Note 5**).
3. Plates are incubated at 37°C for 24 h in a CO_2 incubator.

As a result of introduction of the fused PCR product into the genome of *S. pneumoniae*, the Kan^R gene cassette replaces the chromosomal copy of the target gene, thereby creating a gene knockout. In all transformation experiments, THYE with 5% lysed sheep blood is used for growth of bacterial cells and preparation of competent cells.

3.3.5. Identification of Essential Genes

Typically, inactivation of a nonessential gene produces 300 to 500 Kan^R transformants. If no Kan^R colonies are obtained, the transformation is repeated at least two more

times. Genes are regarded as essential if no colonies are observed in all three transformations. If one or more Kan^R colonies are obtained, the target gene is considered nonessential.

3.4. Confirmation of Gene Replacement Events

Targeted gene replacement events are confirmed by PCR assay. Genomic DNA from mutant and wild-type strains are used as PCR templates along with primers L-F and R-R to verify the correct incorporation of a gene replacement construct into the genome. PCR reactions are carried out under the same conditions as described in **Section 3.2.3** (30 cycles of 95°C for 40 s, 50°C for 40 s, and 72°C for 2 min 30 s). The correct incorporation of a fused construct results in larger or smaller PCR product obtained for a mutant strain compared with that for the wild-type strain (**Fig. 2**).

3.5. Evaluation of Potential Polar Effects

The spr0004 and spr0005 genes (SP0004 and SP0005 in TIGR4) have been previously reported as essential ***(5)***. However, spr0004 was identified as nonessential by allelic replacement mutagenesis using Kan^R cassette without polarity (**Fig. 3**; *see* **Note 6**). Our data suggest that this allelic replacement method can effectively determine essentiality of monocistronic as well as polycistronic ORFs in *S. pneumoniae*.

3.6. Advantages of Allelic Replacement Mutagenesis Coupled with Comparative Genomics

Various methods have been proposed and performed to identify essential genes in several bacterial species ***(2, 5, 9–13)***. Compared with other techniques, our method has some advantages in identifying essential genes in pathogenic bacteria, including *S. pneumoniae* ***(14)***. First, stepwise filtering of ORFs through cross-genome comparison with other species based on simple criteria can effectively reduce the number of genes to be tested, as indicated in previous studies ***(4, 5)***. Second, *a priori* knowledge of target genes makes it unnecessary to sequence *a posteriori* for mutant identification. Third, allelic replacement mutagenesis by two-step PCR does not require cloning into a vector

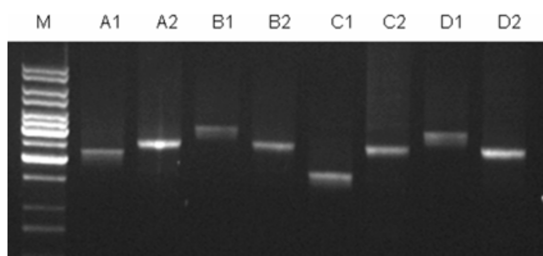

Fig. 2. Confirmation of gene replacement events. Lanes A1, B1, C1, and D1: PCR amplification of the wild-type genes spr0147, spr0232, spr0746, and spr1153 respectively. Lanes A2, B2, C2, and D2: PCR amplification of the corresponding mutant alleles. Lane M: 100-bp ladder marker. Note size differences between PCR products yielded by mutagenized and the corresponding wild-type genes.

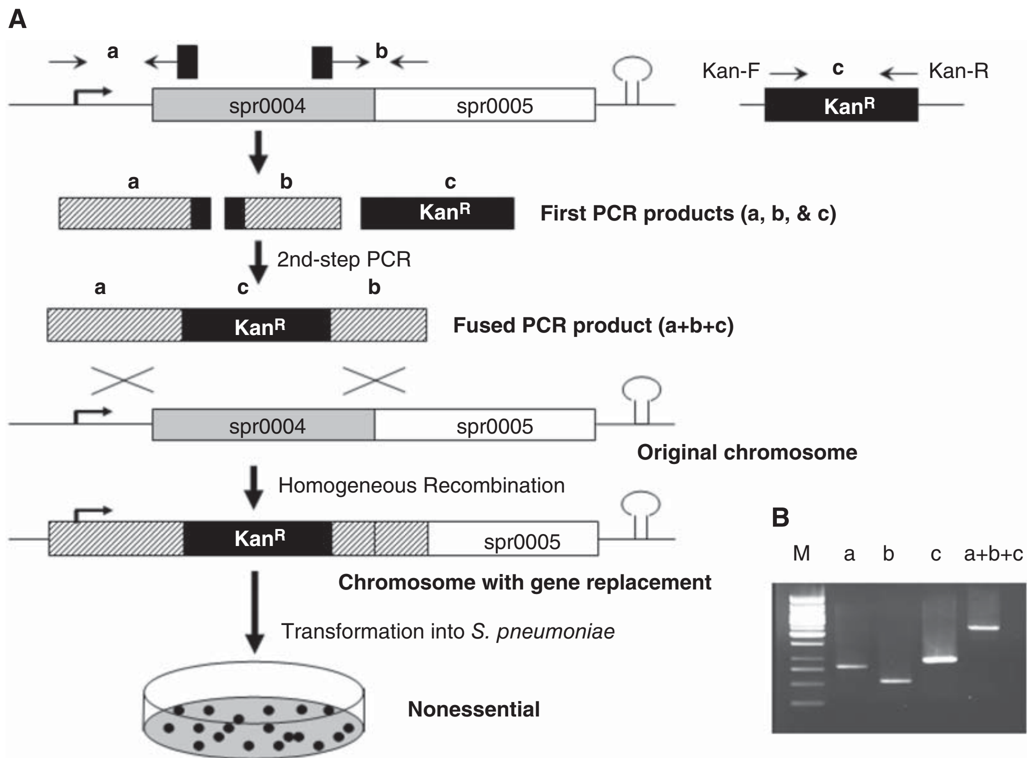

Fig. 3. Gene knock-out of spr0004 to confirm the removal of a potential polar effect. In the first PCR reaction, the KanR gene with transcriptional termination signal removed and up- and downstream regions of spr0004 were amplified. As a result, the upstream (**a**, 813 bp) and downstream (**b**, 558 bp) regions and the KanR gene (**c**, 904 bp) were obtained. The ~2.3-kb-long fused PCR product (**a** + **b** + **c**, 2274 bp) was consequently produced by the second PCR. Upon transformation, hundreds of colonies were obtained on THYE agar plates containing kanamycin, indicating that spr0004 is nonessential.

for recombination. Fourth, this method can minimize potential polar effect and is applicable for both monocistronic and polycistronic genes.

3.7. Bacterial Essential Genes as Potential Targets for Novel Antimicrobial Agents

Identification of essential genes in bacterial pathogens can be applied to the development of new antimicrobial agents because common essential genes in diverse bacterial species could constitute novel targets for broad-spectrum antimicrobial agents ***(15)***. Because of the explosion in the number of available complete bacterial genome sequences, microbial genomics can be applied to evaluate the suitability of potential targets for new antimicrobial drugs, based on the criteria of "essentiality" or "selectivity" ***(16)***. Several studies based on genomics-driven, target-focused approaches have provided a valuable inventory of essential genes that can be used to select and validate antimicrobial agents ***(3, 5, 9–11, 14, 17)***. For example, peptide deformylase (PDF)

inhibitors are the products of a genomics-driven approach to discovery of novel antimicrobial agents. Although the identification of new antimicrobial drug targets does not guarantee the development of new chemical compounds, it is an important first step.

Notes

1. One should use fresh competence medium for making *S. pneumoniae* competent cells.
2. Second PCR reaction may produce several bands such as linear fused PCR product, products of left and right blanks, and others. Thus, elution of the fused PCR product from an agarose gel may be necessary (identified by expected DNA fragment size).
3. Large amounts of linear fused PCR product might be necessary to achieve sufficient transformation rates. Because the optimal PCR conditions for the production of a gene replacement construct may differ from gene to gene, several sets of conditions might need to be tested.
4. The peptide pheromone CSP (Takara) is commonly used to increase pneumococcal transformation efficacy.
5. This kanamycin concentration has been empirically determined in a preliminary study not to give rise to background kanamycin resistance. If concentration of kanamycin in the medium is higher than 400 μg/mL, pneumococcal cells may acquire resistance to it due to reason(s) other than transformation with gene replacement constructs.
6. In order to minimize potential polar effect of mutagenesis, primers are designed so that flanking genes and intergenic regions, including potential promoters, would remain intact in the mutants. In addition, transcriptional termination signals are removed from kanamycin-resistance gene marker (Kan^R), and the cassettes are designed to integrate in the same orientation as the target genes to ensure transcription of the downstream ORFs.

Acknowledgments

This work was supported partly by Samsung Biomedical Research Institute (SBRI) and partly by the IMT-2000 project of the Ministry of Commerce, Industry, and Energy, Republic of Korea.

References

1. Jordan, I. K., Rogozin, I. B., Wolf, Y. I., and Koonin, E. V. (2002) Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. *Genome Res.* **12**, 962–968.
2. Kobayashi, K., Ehrlich, S. D., Albertini, A., Amati, G., Andersen, K. K., Arnaud, M., et al. (2003) Essential *Bacillus subtilis* genes. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 4678–4683.
3. Arigoni, F., Talabot, F., Peitsch, M., Degerton, M. D., Meldrum, E., Allet, E., et al. (1998) A genome-based approach for the identification of essential bacterial genes. *Nat. Biotechnol.* **16**, 851–858.
4. Bruccoleri, R. E., Dougherty, T. J., and Davison, D. B. (1998) Concordance analysis of microbial genomes. *Nucleic Acids Res.* **16**, 4482–4486.
5. 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.
6. Pierce, B. J., Ianelli, F., and Pozzi, F. (2002) Construction of new unencapsulated (rough) strains of *Streptococcus pneumoniae. Res. Microbiol.* **153**, 243–247.

7. 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.
8. 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. U.S.A.* **92**, 11140–11144.
9. Akerley, B. J., Rubin, E. J., Novick, V. L., Amaya, K., Judson, N., and Mekalanos, J. J. (2002) A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenzae*. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 966–971.
10. Hutchison, C. A., Pterson, S. N., Gill, S. R., Cline, R. T., White, O. Fraser, C. M., et al. (1999) Global transposon mutagenesis and a minimal *Mycoplasma* genome. *Science* **286**, 2165–2169.
11. Forsyth, R. A., Haselbeck, R. J., Ohlsen, K. L., Yamamoto, R. T., Xu, H., Trawick, J. D., et al. (2002) A genome-wide strategy for the identification of essential genes in *Staphylococcus aureus*. *Mol. Microbiol.* **43**, 1387–1400.
12. Ji, Y., Zhang, B., Van Horn, S. F., Warren, P., Woodnutt, G., Burnham, M. K. R., and Rosenberg, M. (2001) Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. *Science* **293**, 2266–2269.
13. Sassetti, C. M., Boyd, D. H., and Rubin, E. J. (2001) Comprehensive identification of conditionally essential genes in mycobacteria. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12712–12717.
14. Song, J. H., Ko, K. S., Lee, J. Y., Baek, J. Y., Oh, W. S., Yoon, H. S., et al. (2005) Identification of essential genes in *Streptococcus pneumoniae* by allelic replacement mutagenesis. *Mol. Cells* **19**, 365–374.
15. Zhang, R., Ou, Z. Y., and Zhang, C. T. (2004) DEG: a database of essential genes. *Nucleic Acids Res.* **32**, D271–D272.
16. Sakharkar, K. R., Sakharkar, M. K., and Chow, V. T. K. (2004) A novel genomics approach for the identification of drug targets in pathogens, with special reference to *Pseudomonas aeruginosa*. *In Silico Biol.* **4**, 355–360.
17. Ko, K. S., Lee, J. Y., Song, J. H., Baek, J. Y., Oh, W. S., Chun, J., and Yoon, H. S. (2006) Screening of essential gene in *Staphylococcus aureus* N315 using comparative genomics and allelic replacement mutagenesis. *J. Microbiol. Biotechnol.* **16**, 623–632.