Clonality Analysis of *Streptococcus pneumoniae* in Clinical Specimens

**Delly Chipta Lestari** 1,2,*, Pranee Somboonthum 2, Daisuke Motooka 3, Eiji Ishii 2, Shigeaki Matsuda 2, Anis Karuniawati 1 and Tetsuya Iida 2,3,4,*

1 Department of Clinical Microbiology, Faculty of Medicine, Universitas Indonesia/Dr. Cipto Mangunkusumo General Hospital, Jakarta 10430, Indonesia
2 Department of Bacterial Infections, Research Institute for Microbial Diseases, Osaka University, Osaka 565-0871, Japan
3 Department of Infection Metagenomics, Research Institute for Microbial Diseases, Osaka University, Osaka 565-0871, Japan
4 Center for Infectious Disease Education and Research (CiDER), Osaka University, Osaka 565-0871, Japan
* Correspondence: delly.lestari@ui.ac.id (D.C.L.); iida@biken.osaka-u.ac.jp (T.I.)

**Abstract:** Pneumococcal pneumonia is a significant cause of illness and death globally, particularly among young children and the elderly. The *cpsB* gene is involved in the biosynthesis of the capsule polysaccharide, and polymorphisms in the *cpsB* gene are the basis for sequetyping, a molecular biology-based approach to serotyping. In this study, we attempted the sequetyping of pneumococci directly from clinical sputum specimens collected from adult patients diagnosed with community-acquired pneumonia (CAP). We performed conventional PCR for the *cpsB* gene, followed by TA cloning and Sanger sequencing of the amplicon. The results showed the status of clonality of pneumococci in each specimen. We also performed real-time PCR targeting pneumococci for each specimen. It revealed a significant association between the Ct value of the real-time PCR and the clonality status of pneumococci among the specimens (p-value 0.0007 by Fisher’s exact test analysis). Specifically, when the Ct value was below 22, there was a high probability that pneumococcus existed as a single clone. Thus, this study demonstrates the possible correlation between pneumococcal clonality and bacterial load in clinical specimens, which might indicate the infection status.

**Keywords:** *Streptococcus pneumoniae*; pneumonia; clonality; Sanger sequencing

---

1. **Introduction**

*Streptococcus pneumoniae*, also known as pneumococcus, is a significant cause of illness and death globally, particularly among young children and the elderly [1]. Its ability to cause disease is directly linked to the production of a protective capsule, a polysaccharide structure outside the cell wall that enables the bacteria to evade the host’s immune system and resist phagocytosis [2]. Over 90 different serotypes of pneumococci have been identified based on variations in the capsular polysaccharides they possess [2]. The Quellung reaction is currently considered the standard method for serotyping pneumococcus based on the specific antigenic properties of different capsular polysaccharides. Although highly accurate, this technique requires costly and time-consuming antisera [3]. Therefore, many molecular-based alternative methods are proposed to substitute the traditional serotyping method.

The *cpsB* gene in pneumococcus is involved in the biosynthesis of the capsule polysaccharide, and the presence of serotype-associated polymorphisms in *cpsB* is the basis for a molecular biology-based approach to serotyping [4]. Leung et al. introduced a DNA sequencing serotyping method called sequetyping, which focuses on this gene. This approach was able to amplify and differentiate 84 and 46 serotypes, respectively [5].

Generally, both traditional and molecular-based serotyping are carried out on isolates that grow on solid media. An individual may carry more than one serotype of pneumococci
simultaneously; therefore, instead of performing sequetyping for isolated colonies, we
performed it directly on clinical sputum specimens. Conventional PCR for the cpsB gene
was conducted against sputum specimens collected from adult patients diagnosed with
community-acquired pneumonia (CAP). This was followed by TA cloning and Sanger
sequencing of the amplicon. The results showed the clonality status of pneumococci in
each specimen. The analysis revealed a significant association between the clonality status
and the cycle threshold (Ct) value of real-time PCR for pneumococci among specimens.

2. Materials and Methods

2.1. Samples

Eighty-three (83) sputum specimens stored at −80 °C at the Clinical Microbiology
Laboratory, Faculty of Medicine, Universitas Indonesia, were analyzed for this study. The
sputum was collected from adult patients diagnosed with CAP between September 2016
and August 2017 from hospitals in Jakarta.

2.2. Sample Pretreatment and Nucleic Acid Extraction

Five hundred microliters of sterile distilled water (DW) containing 1% dithiothreitol
(DTT), a mucolytic agent, was added to 500 µL of sputum. The sputum was shaken at
1000 rpm at room temperature for 5–10 min or until homogenized. Next, we aliquoted
the treated sputum into two different tubes. Total nucleic acids were extracted from 600 µL
of the treated sputum using a QIAamp® MinElute® Virus Spin Kit (Qiagen, Hilden, Germany)
with 150 µL of the output eluate. A DNeasy PowerSoil® Kit (Qiagen) was used to extract
dNA from 400 µL of treated sputum with 100 µL of output eluate. All nucleic acids were
kept at −80 °C until further investigation.

2.3. Real-Time PCR

An 80 µL total nucleic acid extract was screened for respiratory viruses and bacteria using
FTD® Respiratory Pathogens 33 (Fast Track Diagnostics Ltd., Luxembourg). The real-time
PCR machine was a LightCycler® 96 Instrument (Roche Diagnostics International, Basel,
Switzerland). All procedures were conducted according to the manufacturer’s instructions.

2.4. Clonality Analysis by Sanger Sequencing

2.4.1. Control Strains

Before examining our clinical samples, we optimized the PCR and nucleotide sequencing
methods targeting the cpsB gene using control strains (Figure 1). Streptococcus pneumoniae
RIMD3122004 (ATCC 33400), RIMD3122033 (serotype 14), RIMD3122093 (serotype 7F), and
RIMD3122027 (serotype 6A) provided by the Pathogenic Microbes Repository Unit of the
Research Institute for Microbial Diseases, Osaka University, through the National BioResource
Project (Pathogenic Bacteria) of MEXT/AMED, Japan, were used as positive controls.

![Figure 1](image-url)
2.4.2. PCR of cpsB Gene

The primers were forward (5′-CGA ACC ATT GTC TCT ACC TCT CA-3′) and reverse (5′-CAA TKA iniCTG GMG TRA TTC CC-3′). The primer coverage starts at position 103 and ends at position 403 in the 732-bp region of cpsB. The reaction mixture contained 1 µL of DNA, 0.2 µL of TaKaRa Ex Taq® DNA Polymerase (initial concentration 5 U/µL), 2 µL of 10× PCR buffer, 1 µL of MgCl₂ (25 mM), 1.6 µL of the deoxynucleoside triphosphate mixture (2.5 mM), 2 µL of 10% Tween 20, 0.2 µL of bovine serum albumin (0.2 µg/µL), and 0.4 µL of each primer (10 µM), which made up a final volume of 20 µL with DNase/RNase-free distilled water. The reaction cycle consisted of an initial denaturation step at 94 °C for 4 min, followed by 35 amplification cycles of 94 °C for 45 s, 55 °C for 45 s, and 72 °C for 1 min 30 s, with an extension step at 72 °C for 5 min. Amplicons from the PCR were analyzed with 2% agarose gel electrophoresis.

2.4.3. Nucleotide Sequencing

Amplicons with the expected cpsB band size (~301 bp) were purified using a QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer’s instructions. DNA cloning was performed using a TOPO-TA cloning kit (Thermo Fisher Scientific, Waltham, MA, USA), transformed into DH5α competent cells, and incubated overnight at 37 °C. Colony PCR was performed with an initial denaturation step at 94 °C for 3 min, followed by 30 amplification cycles of 94 °C for 30 s, 55 °C for 30 s, and 72 °C for 1 min 30 s, with an extension step at 72 °C for 7 min. The T7 and M13 primers were used for the forward and reverse primers, respectively. Amplicons from the PCR were analyzed with 2% agarose gel electrophoresis. Amplicons purified from the gel with the expected cpsB band size (~500 bp) were diluted ten times. The cycle was sequenced with the Applied Biosystems™ BigDye™ Sequence Terminator v.3.1 Cycle Sequencing Kit (Thermo Fisher Scientific, USA) according to the manufacturer’s protocol. Sequences were analyzed using Applied Biosystems™ Genetic Analyzer 3130 (Thermo Fisher Scientific, USA) and viewed with Applied Biosystems™ Sequencing Analysis Software v5.2 Patch 2 (Thermo Fisher Scientific, USA).

2.4.4. Clonality Identification

For clonality identification, at least ten sequences were analyzed for one specimen. Sequences were classified into a single clone lineage when the SNPs were present in at least two sequences at the same position. Multiple alignments were displayed using Genetyx 8.

2.5. Statistical Analysis

The correlation between the cycle threshold (Ct) value and pneumococcal clonality was calculated using Fisher’s exact test using QI macros Excel add-in features. Differences were considered statistically significant at p-values of <0.05.

3. Results

3.1. Detection of Pneumococcus from Clinical Specimens Using Real-Time PCR

Real-time PCR FTD® Respiratory Pathogens 33 was used to detect pneumococcus from stored sputum specimens collected from hospitalized adults in Jakarta, Indonesia, who were clinically diagnosed with CAP. Pneumococcus was detected in 39.8% of cases (33 out of 83 samples). The Ct values varied from 11.9 to 34.7, with a median of 26.79 (Table 1).
Table 1. Pneumococcal detection by real-time PCR and clonality status.

<table>
<thead>
<tr>
<th>No</th>
<th>Sample Code</th>
<th>Real-Time PCR (Ct)</th>
<th>Sanger Sequencing Results *</th>
<th>Clone(s) Number</th>
<th>Interpretation **</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C5136</td>
<td>11.9</td>
<td></td>
<td>1</td>
<td>Single</td>
</tr>
<tr>
<td>2</td>
<td>A5001</td>
<td>12.08</td>
<td></td>
<td>5</td>
<td>Multiple</td>
</tr>
<tr>
<td>3</td>
<td>C5119</td>
<td>13.05</td>
<td></td>
<td>1</td>
<td>Single</td>
</tr>
<tr>
<td>4</td>
<td>C5004</td>
<td>14.31</td>
<td></td>
<td>1</td>
<td>Single</td>
</tr>
<tr>
<td>5</td>
<td>C5071</td>
<td>15.94</td>
<td></td>
<td>1</td>
<td>Single</td>
</tr>
<tr>
<td>6</td>
<td>C5003</td>
<td>15.98</td>
<td></td>
<td>4</td>
<td>Multiple</td>
</tr>
<tr>
<td>7</td>
<td>A5019</td>
<td>16.74</td>
<td></td>
<td>1</td>
<td>Single</td>
</tr>
<tr>
<td>8</td>
<td>C5134</td>
<td>17.83</td>
<td></td>
<td>1</td>
<td>Single</td>
</tr>
<tr>
<td>9</td>
<td>C5124</td>
<td>18.09</td>
<td></td>
<td>5</td>
<td>Multiple</td>
</tr>
<tr>
<td>10</td>
<td>C5014</td>
<td>18.56</td>
<td></td>
<td>4</td>
<td>Multiple</td>
</tr>
<tr>
<td>11</td>
<td>C5078</td>
<td>18.8</td>
<td></td>
<td>1</td>
<td>Single</td>
</tr>
<tr>
<td>12</td>
<td>D5017</td>
<td>19.47</td>
<td></td>
<td>1</td>
<td>Single</td>
</tr>
<tr>
<td>13</td>
<td>D5008</td>
<td>21.8</td>
<td></td>
<td>1</td>
<td>Single</td>
</tr>
<tr>
<td>14</td>
<td>A5004</td>
<td>23.53</td>
<td></td>
<td>3</td>
<td>Multiple</td>
</tr>
<tr>
<td>15</td>
<td>C5133</td>
<td>24.44</td>
<td></td>
<td>4</td>
<td>Multiple</td>
</tr>
<tr>
<td>16</td>
<td>C5108</td>
<td>24.67</td>
<td></td>
<td>4</td>
<td>Multiple</td>
</tr>
<tr>
<td>17</td>
<td>C5069</td>
<td>26.79</td>
<td></td>
<td>4</td>
<td>Multiple</td>
</tr>
<tr>
<td>18</td>
<td>C5109</td>
<td>28.8</td>
<td></td>
<td>4</td>
<td>Multiple</td>
</tr>
<tr>
<td>19</td>
<td>D5007</td>
<td>29.8</td>
<td></td>
<td>6</td>
<td>Multiple</td>
</tr>
<tr>
<td>20</td>
<td>A5014</td>
<td>30.17</td>
<td></td>
<td>4</td>
<td>Multiple</td>
</tr>
<tr>
<td>21</td>
<td>C5019</td>
<td>30.38</td>
<td></td>
<td>4</td>
<td>Multiple</td>
</tr>
<tr>
<td>22</td>
<td>C5075</td>
<td>30.38</td>
<td></td>
<td>3</td>
<td>Multiple</td>
</tr>
<tr>
<td>23</td>
<td>C5095</td>
<td>31.16</td>
<td></td>
<td>3</td>
<td>Multiple</td>
</tr>
<tr>
<td>24</td>
<td>C5005</td>
<td>31.65</td>
<td></td>
<td>4</td>
<td>Multiple</td>
</tr>
<tr>
<td>25</td>
<td>A5018</td>
<td>32.0</td>
<td></td>
<td>1</td>
<td>Single</td>
</tr>
<tr>
<td>26</td>
<td>C5017</td>
<td>32.17</td>
<td></td>
<td>3</td>
<td>Multiple</td>
</tr>
<tr>
<td>27</td>
<td>C5110</td>
<td>32.22</td>
<td></td>
<td>4</td>
<td>Multiple</td>
</tr>
<tr>
<td>28</td>
<td>C5107</td>
<td>32.71</td>
<td></td>
<td>1</td>
<td>Single</td>
</tr>
<tr>
<td>29</td>
<td>C5074</td>
<td>33.01</td>
<td></td>
<td>4</td>
<td>Multiple</td>
</tr>
<tr>
<td>30</td>
<td>C5024</td>
<td>33.21</td>
<td></td>
<td>4</td>
<td>Multiple</td>
</tr>
<tr>
<td>31</td>
<td>A5016</td>
<td>33.92</td>
<td></td>
<td>3</td>
<td>Multiple</td>
</tr>
<tr>
<td>32</td>
<td>C5104</td>
<td>34.17</td>
<td></td>
<td>4</td>
<td>Multiple</td>
</tr>
<tr>
<td>33</td>
<td>C5137</td>
<td>34.7</td>
<td></td>
<td>5</td>
<td>Multiple</td>
</tr>
</tbody>
</table>

* All 33 specimens in which pneumococci were detected by real-time PCR were also positive for cpsB. ** The presence of single-nucleotide polymorphisms (SNPs) in at least two out of ten sequences is classified as a single clone lineage. Single = single clone; Multiple = multiple clones.

3.2. Clonality Analysis Using Sanger Sequencing

Sanger sequencing on the PCR products targeting the cpsB gene was conducted to obtain information about the sequence type of pneumococcus in clinical specimens. We performed conventional PCR targeting cpsB, with expected PCR products of 301 base pairs (bp) in size. The primers used in this study were similar to the previous study [6], with slight modifications considering our consensus sequence from the multiple sequence alignment (MSA) of 461 pneumococcal cpsB sequences (not from other streptococcal species) downloaded from GenBank. We selected a primer set that starts at position 103 and ends at position 403 in the 732-bp region of cpsB. Previous studies had indicated that pneumococcal serotype determination would be optimal when the central 732-bp region of the cpsB amplicon was used [7,8]. We found PCR amplicons of cpsB in all 33 pneumococcal real-time PCR-positive samples. Subsequently, we performed TA cloning followed by Sanger sequencing on the cpsB amplicons. We found that in 11 out of 33 (33%) specimens, pneumococci were present as a single clone (Table 1, Figure 2). By Fisher’s exact test analysis, we found that the clonality status (either single or multiple) of pneumococcus correlates with the Ct value of real-time PCR for pneumococci when a Ct value cut-off of 22 [9] was used (p-value 0.0007). Specifically, when the Ct value was below 22, there was a high probability that pneumococcus existed as a single clone.
Figure 2. Examples of clonality analysis using Sanger sequencing. Single clone (A) C5078; (B) D5017; (C) D5008; Multiple clones (D) C5014; (E) C5109; (F) C5137. Multiple alignments were displayed using Genetyx 8. The black box implies different clones.
3.3. Clonality Tendencies by Underlying Factors

We obtained clinical information related to the underlying factors and comorbidities of each patient. The fact that the pool sample size is small and with uneven distribution, analysis using inferential statistics could be inaccurately generalizable to the broader population. Thus, we display the frequency of the factors in percentages and make a comparison by clonality to assess the tendency (Figure 3). Among all variables or factors analyzed, we found that at least five factors showed an obvious tendency to become a single clone with a difference of 20% or more. The factors are gender, COPD (chronic obstructive pulmonary disease), hyponatremia, right-lobe infiltrate, and virus coinfections.

Figure 3. Clonality tendency by underlying factors. The axis is the percentage of each factor. PSI: pneumonia severity index; COPD: chronic obstructive pulmonary disease.

4. Discussion

Real-time PCR identified pneumococci in approximately 40% of adult CAP cases from hospitals in Jakarta between September 2016 and August 2017, exhibiting a broad range of Ct values. All positive specimens for pneumococcus were subjected to the conventional PCR for cpsB, which yielded positive results. We employed TA cloning and Sanger sequencing to obtain additional information about the clonality of pneumococci from sputum samples, focusing on the cpsB gene.

We found a significant association between the Ct value and clonality status, as indicated by Fisher’s exact test analysis (p-value 0.0007). When the Ct value was below 22, there was a high likelihood that pneumococcus was present as a single clone (Table 1). The Ct value represents the amount of target DNA/RNA present in the sample. The low Ct value, in this case, indicates the abundance of pneumococci in the sputum sample. This suggests that when the pneumococcal load is high in clinical specimens, there is a tendency for pneumococci to exist as a single clone. The current results may suggest that when pneumococci are present as commensal (generally low abundance), they tend to be multiple clones in clinical specimens, while when causing infection (resulting in high
abundance), they can be a single clone because a particular virulent clone expands and dominates the others in the human body. If this is the case, the clonality status could be used to distinguish whether the pneumococci are a causative agent or just commensal. Meanwhile, we also noticed that 30% of specimens with a Ct value below 22 were multiple clones. Former studies reported that dual serotypes were found in patients with invasive pneumococcal disease (IPD) [10] and pneumococcal pneumonia [11], both using colonies-based serotype identification.

This study is retrospective and adding the number of samples is not possible; thus, we analyzed the underlying factors descriptively. We found that pneumococcus tends to be a single clone in male CAP patients and patients with COPD, hyponatremia, right-lobe infiltrate, and coinfection with viruses. Our study concurs with earlier reports that the prevalence of CAP in males is higher than in females [12–15], and COPD is one of the risk factors for CAP infection in adults [13,14]. Hyponatremia is commonly found in patients with respiratory infections, including pneumonia, with worsened outcomes [16–18]. We observe a higher proportion of single-clone pneumococcus in patients with infiltration in the lung’s right lobe. It is unclear whether the infiltrate is due to pneumococcal infection or because of the anatomical position that it has a larger caliber and more vertical orientation of the right mainstem bronchus, thereby facilitating the accumulation of the pathogens. It was also already known that pneumococcus is the most frequently isolated influenza-associated pathogen [19,20]. To the best of our knowledge, there were no previous studies that investigated the relationship between clinical manifestations and clonality of pneumococcus. Our findings could pave the way to establish the connection with higher confidence.

We acknowledge that the main limitation of this study is the small number of samples analyzed. As a consequence, our findings should be considered cautiously because the small sample size hindered our efforts to control heterogeneity among patients, thus the generalizability is also limited. Hence, we recommend future studies to be conducted using prospective design instead of retrospective, so that more robust statistical analysis can be performed whenever sufficient samples are obtained. In addition, the study was focused on exploring the role of the cpsB gene for sequentyping. Therefore, the genetic diversity of pneumococcal strains might not be fully captured. However, future studies can be expanded to include other relevant genes to further explore the polymorphism among pneumococcal serotypes.

In conclusion, this study demonstrated the possible correlation between pneumococcal clonality and bacterial load in clinical specimens. It should be noted that our study focused specifically on pneumococcus as an example. The clonality analysis approach we employed has the potential to be applied to other bacteria and various types of infections. For instance, it could be valuable in analyzing Escherichia coli in diarrhea cases. It is worthwhile to further explore the possibility that the clonality analysis can be applied to examine the infection status of bacteria in patients. Furthermore, with technological advancements, particularly next-generation sequencing (NGS), a metagenomic approach can be employed for clonality analysis. While NGS has been widely utilized for clonality analysis in neoplasms or tumors [21,22], its application in infectious diseases has yet to be extensively explored. Understanding bacterial clonal dynamics could be useful in developing targeted interventions and treatment strategies for infectious diseases.

Author Contributions: Conceptualization, T.I. and D.C.L.; methodology, T.I., D.C.L. and P.S.; software, D.C.L. and P.S.; validation, T.I. and D.C.L.; formal analysis, D.C.L.; investigation, D.C.L. and P.S.; resources, T.I., A.K., S.M., E.I. and D.M.; data curation, D.C.L.; writing—original draft preparation, D.C.L.; writing—review and editing, all; visualization, D.C.L.; supervision, T.I.; project administration, P.S. and D.C.L.; funding acquisition, T.I. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a Grant-in-Aid from the United States-Japan Cooperative Medical Science Program (USJCMSF), AMED, Japan. Additionally, this work was supported by the Center for Infectious Disease Education and Research (CiDER), Osaka University.
**Institutional Review Board Statement:** This project was approved by the Ethics Committee of the Faculty of Medicine, Universitas Indonesia (no. 0052/UN2.F1/ETIK/2019).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding authors.

**Acknowledgments:** We thank the Pathogenic Microbes Repository Unit of the Research Institute for Microbial Diseases, Osaka University, for providing the *S. pneumoniae* control strains. D.C.L. is a scholarship awardee of the Hitachi Global Foundation and Ichiro Kanehara Foundation, Japan.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest.

**References**


**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.