ST534: The new sequence type of *Corynebacterium diphtheriae* causing diphteria in
Jakarta and surrounding areas, Indonesia

Abstract

**Background/aim:** The aim of this study was to find out characteristics and patterns of the spread of *C. diphtheriae* isolated from Jakarta and surrounding areas using whole genome sequencing (WGS) technique and multilocus sequence typing (MLST) approach.

**Materials and methods:** The study samples consisted of 86 *C. diphtheriae* isolates, which were isolated from diphtheria patients and close contacts of patients. The DNA sequencing was carried out using the whole genome sequencing (WGS) technique. Data conversion applied the U-gene software. Molecular typing was conducted through the multilocus sequence typing (MLST) approach, then followed by online data analysis.

**Results:** The results showed that as many as 43 (50%) of all samples examined were new types with the same allele profile, namely 9-1-13-4-3-3-4. New sequence type *C. diphtheriae* is registered in the MLST global database as ST534 based on the allele profile. Tox gene analysis in 43 isolates with ST 534 indicated that there were three mutation positions that all of which were silent mutations.

**Conclusion:** The main cause of diphtheria in Jakarta and surrounding areas is a new sequence type of *C. diphtheriae* registered as ST534.

**Key words:** *C. diphtheriae*, diphtheria, Jakarta, MLST

1. **Introduction**

Diphtheria is an acute infectious disease which generally attacks the upper respiratory tract with the typical symptoms of pseudomembranous formation in the focal area of infection followed by a systemic picture due to diphtheria toxin [1,2]. Diphtheria is still
a significant health problem in many parts of the world, including Indonesia [3,4]. Data
from the Ministry of Health and the World Health Organization (WHO) record that, in
the past few years, Indonesia ranks second to fourth country in the world with the most
cases of diphtheria [5,6]. The spread of the cases extends to almost all provinces. Even
cases of diphtheria have increased again at the end of 2017 until the beginning of 2018.
Most cases come from Jakarta and surrounding areas (Banten and West Java) and East
Java [7].

Various efforts were made to overcome diphtheria, including promotive, preventive, and
curative efforts. However, new cases continue to be found [8]. There are concerns about
the emergence of new variants of bacteria that are more virulent or resistant to vaccines
and antibiotics due to mutations like what occurs in other bacteria [9,10,11,12]. In this
case, it is necessary to carry out the molecular typing of diphtheria-causing bacteria in
Indonesia, especially in Jakarta and surrounding areas. Molecular typing is important to
find out the characteristics of bacteria and the spread pattern of disease as well as to
evaluate the success of the efforts that have been made [13]. In addition, this activity can
be implemented to identify the presence or absence of new circulating variants [14].

Diphtheria is caused by three bacterial species fused in the genus *Corynebacterium* that
are *Corynebacterium diphtheria*, *Corynebacterium ulcerans*, and *Corynebacterium
pseudotuberculosis* [15]. Molecular typing can be done using several types of methods
[16]. In this case, ribotyping is the gold standard for molecular typing bacteria which
cause diphtheria. However, ribotyping has limitations in terms of flexibility. Multilocus
sequence typing (MLST) is an alternative with differentiation capabilities equivalent to
ribotyping [17]. This study describes the results of molecular typing of diphtheria-causing
bacteria (*C. diphtheriae*) isolated from diphtheria in Jakarta and surrounding areas by the MLST approach.

2. **Materials and methods**

2.1 **Time, place and research sample**

The study was conducted at the Bacteriology Laboratory (Prof. Dr. Sri Oemijati Laboratory for Infectious Disease Research), Research Center for Biomedical and Basic Health Technology in 2018. Samples were in the forms of 86 stored isolates of *C. diphtheriae* that were isolated from diphtheria cases and close contact of patients in Jakarta and its surrounding areas (Banten and Jabar/West Java) in 2010-2017. The stored isolates were recultured on blood agar, incubated at 37°C for 24 hours. The colonies were harvested and put in a tube containing 0.5 ml of aquadest for DNA extraction. The sample proportion per province can be seen in figure below.

![Proportion of samples based on the area of origin of isolates.](image)

2.2 **DNA sequencing**

The DNA sequencing was carried out using the next generation sequencing (NGS) approach. The DNA extraction was conducted with a commercial QiaAmp DNA Minikit (Qiagen) kit following the manufacturer’s procedure. The quality and quantity of DNA
were measured before the sample preparation for the DNA sequencing stage using the Mi-seq (Illumina) machine. The sample preparation included genomic DNA tagging, amplify libraries, clean up libraries, normalization libraries, and pool libraries. The sequencing process was run by the Miseq (Illumina) machine by calculating the maximum number of isolates carried out in one process calculated according to the manufacturer’s formula.

2.3 Data analysis
Data in the format of "bam" were read, analyzed and converted into FASTA format using U-gene software. Molecular typing was done with the MLST approach. Profiling 7 loci were involved and sequence type determination was conducted by online based on the MLST global database. Further analysis to determine the suitability of bacterial strains with vaccines used in vaccination programs was carried out using BioEdit software. The DNA sequence of the tox gene encoding diphtheria toxin synthesis in the sample was compared with the reference strain of *C. diphtheriae* PW8, which was used as the vaccine seed.

3. Results

3.1 Allele profile and sequence type
A total of 43 (50%) of the 86 isolates examined showed the same allele profile, namely 9-1-13-4-3-3-4. Sequence type (ST) cannot be specified based on the existing database. This shows that the isolate is a new type that has never been reported before. Determination of sequence types is done by registering allele profiles to the MLST global database. The new sequence type was determined and registered with the code ST534.
The distribution pattern of isolates with ST534 by region and year of isolation can be seen in Table 1.

Table 1. Proportion of ST534 by province and year of isolation

<table>
<thead>
<tr>
<th>Year</th>
<th>Banten</th>
<th>Jakarta</th>
<th>West Java</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>2011</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>2012</td>
<td>2</td>
<td>1</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>2013</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>2014</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>2015</td>
<td>2</td>
<td>1</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>2016</td>
<td>-</td>
<td>3</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>2017</td>
<td>9</td>
<td>9</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>21 (49%)</td>
<td>14 (32%)</td>
<td>8 (19%)</td>
<td>43 (100%)</td>
</tr>
</tbody>
</table>

Table 1 shows that bacterial circulation tends to increase from year to year, although it was not found in 2013. The proportion of the spread in each province is proportional to the number of research samples in each province.

3.2 Analysis of tox genes

The results of the alignment of the tox gene (1683 bp) from 43 isolates with ST534 can be seen in Table 2.

Table 2. Variation and location of tox gene mutations in 43 samples with ST534

<table>
<thead>
<tr>
<th>Type of Bacteria</th>
<th>Mutation Location</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>84th Base</td>
</tr>
</tbody>
</table>
Table 2. shows that the mutation pattern of the tox gene can be grouped into 3 types and mutations occur at 3 positions from the 1683 nucleotide bases that set the tox gene.

Prediction of changes in amino acid caused by DNA mutations can be seen in Table 3.

Table 3. Prediction of codon translation at mutation location

<table>
<thead>
<tr>
<th>Mutation Location</th>
<th>Change in Base</th>
<th>Change in Amino Acid</th>
<th>Type of Mutation</th>
</tr>
</thead>
<tbody>
<tr>
<td>84th Base</td>
<td>GAT→GAC</td>
<td>Asp→Asp</td>
<td>silent mutation</td>
</tr>
<tr>
<td>415th Base</td>
<td>TTG→CTG</td>
<td>Leu→Leu</td>
<td>silent mutation</td>
</tr>
<tr>
<td>705th Base</td>
<td>AGG→AGA</td>
<td>Arg→Arg</td>
<td>silent mutation</td>
</tr>
</tbody>
</table>

Table 3. shows no amino changes due to tox gene mutations from all samples.

4. Discussion

Diphtheria is an infectious disease through direct contact with patients or carriers and indirect contact with the environment contaminated with the causative bacteria. The spread of disease is closely related to human mobility [18]. This study involved three provinces with very high population mobility, including Jakarta, Banten, and West Java. The most samples came from the Banten region; whereas the samples from West Java were the least (Figure) as they only came from buffer zones (Depok, Bogor and Bekasi),
excluding other regions. Therefore, the number of research samples did not represent the actual number of diphtheria cases in each province.

Molecular typing used in this study was the MLST method. This method has proven reliable for the needs of molecular typing of various types of bacteria, including diphtheria-causing bacteria [17,19]. In diphtheria-causing bacteria, MLST differentiation ability is equivalent to ribotyping, which is the gold standard for molecular typing [17].

The MLST approach to diphtheria-causing bacteria is based on analysis of DNA sequences at 7 loci (7 genes), including \textit{atpA}, \textit{dnaE}, \textit{dnaK}, \textit{fusA}, \textit{leuA}, \textit{odhA}, and \textit{rpoB}. Each allele profile at each locus is sorted to get a sequence type. In this study, sequence type (ST) 534 was determined based on the allele profile of the 7 loci. The description of the existence of ST534 that is almost every year, and even likely to increase (Table 1) indicates that disease prevention efforts have not been optimal. In addition, 43 ST534 isolates (50% of 86 isolates examined) illustrate that this strain is the main cause of diphtheria in Jakarta and surrounding areas. The proportion of isolates from each province explained that the cases of diphtheria in Jakarta and surrounding areas were related to one another. This strengthens the statement that population mobility greatly affects the spread of diphtheria.

ST534 is a new type sequence, but with allele profiles that already exist in the MLST global database. To find out whether the strain has a higher virulence and immune level, analysis of a tox gene encoding diphtheria toxin was performed, as the main virulence factor of bacteria (Tables 2 and 3). The level of immunity referred to here is immunity to the vaccine (antibodies formed by vaccination). This is predicted to occur if the DNA sequence of the tox gene from the isolate is different from the DNA sequence of the tox
gene from the vaccine seed (strain PW8). This illustration is seen in diphtheria cases in
developed countries with high vaccination coverage where diphtheria caused by
*Corynebacterium ulcerans* tends to increase [20,21], accompanied by the fact that the
sequence of the tox *C. ulcerans* gene differs significantly compared to *C. diphtheriae*
[22,23].

**Acknowledgement**

We would like to thank all of those who have played a role in the study and provided
input for this manuscript, especially Kambang Sariadji, Nelly Puspandari, Fauzul Muna,
Khariri, Bambang Heriyanto, Sarwo Handayani and Rita Marleta Dewi.

**References**

1. Meera M, Rajarao M. Diphtheria in Andhra Pradesh-a clinical-epidemiological
study. International Journal of Infectious Disease. 2014;19:74-78. doi:
10.1016/j.ijid.2013.10.017.

2. Jain A, Samdani S, Meena V, Sharma MP. Diphtheria: It is still prevalent!!!.

3. Blumberg LH, Prieto MA, Diaz JV, Blanco MJ, Valle B, Pla C, Durrheim DN. The
preventable tragedy of diphtheria in the 21st century. International Journal of


5. WHO. Diphtheria reported cases. [updated 2019 Jul 15; cited 2019 Aug 22].
Available from


Table 1. Proportion of ST534 by province and year of isolation

<table>
<thead>
<tr>
<th>Year</th>
<th>Banten</th>
<th>Jakarta</th>
<th>West Java</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>2011</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>2012</td>
<td>2</td>
<td>1</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>2013</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>2014</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>2015</td>
<td>2</td>
<td>1</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>2016</td>
<td>-</td>
<td>3</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>2017</td>
<td>9</td>
<td>9</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>21 (49%)</strong></td>
<td><strong>14 (32%)</strong></td>
<td><strong>8 (19%)</strong></td>
<td><strong>43 (100%)</strong></td>
</tr>
</tbody>
</table>

Table 2. Variation and location of tox gene mutations in 43 samples with ST534

<table>
<thead>
<tr>
<th>Type of Bacteria</th>
<th>Mutation Location</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>84th Base</td>
</tr>
</tbody>
</table>

Figure. Proportion of samples based on the area of origin of isolates.
Table 3. Prediction of codon translation at mutation location

<table>
<thead>
<tr>
<th>Mutation Location</th>
<th>Change in Base</th>
<th>Change in Amino Acid</th>
<th>Type of Mutation</th>
</tr>
</thead>
<tbody>
<tr>
<td>84th Base</td>
<td>GAT→GAC</td>
<td>Asp→Asp</td>
<td>silent mutation</td>
</tr>
<tr>
<td>415th Base</td>
<td>TTG→CTG</td>
<td>Leu→Leu</td>
<td>silent mutation</td>
</tr>
<tr>
<td>705th Base</td>
<td>AGG→AGA</td>
<td>Arg→Arg</td>
<td>silent mutation</td>
</tr>
</tbody>
</table>