Prospective nucleotide sequence analysis of methicillin resistant Staphylococcus aureus isolates from Sokoto state

Shuaibu Suleiman Adeiza, Josiah Ademola Onaolapo, Busayo Olalekan Olayinka

Department of Pharmaceutics and Pharmaceutical Microbiology, Faculty of Pharmaceutical Sciences, Ahmadu Bello University, Zaria, Kaduna, Nigeria

Summary

The epidemiology of MRSA has evolved considerably and a better understanding of the reservoir mix and phylogenetic resolutions of MRSA may inform strategies to prevent MRSA transmission and infection. Even though in Nigeria there are studies that have reported on the molecular characterization of MRSA, sadly, not much effort has been put into analysis and submission of sequenced data to Gene repositories. To the best of our knowledge, this study is the first of its kind in Sokoto to have successfully deposited nucleotide sequences of MRSA strains into Genbank database. The sequence reads of molecularly characterized strains should be deposited into Gene repositories by researchers to further populate the databank and improve the efficiency of local phylogenetic analysis to foster better bioinformatics as paucity of local data exists in these databases.

Introduction

Methicillin-resistant Staphylococcus aureus (MRSA) is a pathogen that is extremely impacting national and global health and has developed into a severe health burden (5). It is heralded by symptomless nasal carriage which may cause health difficulties in immunocompromised individuals (8). It has a remarkable ability to cause infections; from benign skin diseases to deep bone marrow infections (13). The epidemiology of MRSA has evolved considerably. In the past, MRSA strains could be traced back to a hospital or clinical source, but over the years there appears to be a decrease in hospital-acquired MRSA and an upsurge in the occurrence in community acquired strains (2). This may be due to improve hospital policies on MSA surveillance. However, the community-acquired strains have demonstrated more virulence than the hospital-acquired strains (1). Evolutionarily, MRSA arose as an effect of selective pressure engendered by over exposure to antibiotics (3). It has a remarkable ability to respond to environmental pressures. This facultative anaerobe can withstand strict environmental settings such as high temperatures, salt concentrations and dry environmental conditions (15).

Methods based on molecular techniques are gradually becoming less expensive (adjusted for inflation rate) and more prevalent in Nigeria with molecular companies now in a quest for cost-effective ways of collaborating with molecular laboratories and researchers. Even though in Nigeria there are studies that have described the molecular characterization of MRSA. Sadly, not much effort has been put into submission of sequenced data into Gene repositories. Sequence deposition into gene repositories fosters better bioinformatics like; phylogenetics and specific primer design. A better understanding of the reservoir mix and diversity of MRSA may inform strategies to prevent MRSA transmission and infection in Sokoto. Bioinformation is of particular importance in the management MRSA infection in patients under intensive care because routine biochemical tests are slow and often fail to characterize the pathogenic strain due to phenotypic divergence.

The goal of the study is to recognize the phylogenetic affiliation among MRSA strains recovered from Sokoto State-owned hospitals with a view to ascertain the genetic diversity of the hospital strains based on 16S rRNA and meca gene sequencing.

Materials and methods

Study center

The present research work was carried out on strains from Sokoto State, located at the extreme north-west of Nigeria, between...
The state has a population of nearly 5.4 million people (14), covers a terrestrial area of 32,000km² and shares a border with the Republic of Niger. The major ethnic groups in the state are the Hausa and Fulani groups. Over 80 percent of the people in the state practice agriculture as their major source of income while local crafts like black smithing, weaving, dyeing, carving and leather works form the remaining 20% (16).

**Bacterial strains**

In this study, 16 mecA positive MRSA clinical strains obtained from the nostrils of participants from Sokoto state-owned hospitals were selected for sequencing. All the enrolled strains were previously characterised using standard microbiological procedures (Gram stain, Catalase and Coagulase test), Microgen™ staph identification Kit, DNA extraction and PCR amplification. Additional characterization was carried out on these strains by DNA sequencing and phylogenetic analysis.

**Staphylococcal DNA extraction and PCR amplification**

Staphylococcal genomic DNA was harvested using Qiagen (Hilden, Germany) DNA extraction kit in adherence to the manufacturer’s protocols. For each strain, DNA fragment (about 956 bp) of the 16SrRNA and mecA (533 bp) genes were amplified according to methods previously described by Hassan (7).

**Sequencing**

The amplicons were purified using Monarch® bio-labs DNA clean-up kits according to the manufacturers’ recommendations. The purified PCR products was packaged and sent for Sanger sequencing at Inqaba Biotec South Africa.

**Bioinformatics and DNA Sequence Analysis**

The chromatograph sequence files were analyzed with the software package Mega™ 7 (version 7.0.26). After providing the input sequence (AB1 format) the software was used to mask out low quality sequence reads. The nucleotide sequences were converted into FASTA file format for further analysis.

**Gene annotation**

The FASTA formatted nucleotide sequences were marked for their Open reading frames (ORF), Coding sequences (CDS), Untranslated regions (UTR), START codons (ATG), STOP codons (TAA), Promoter region (-35, -10 and shine-Dergano sequences) and the terminator earpin using CLC genomic work bench software® (Version 3.6.5).

**Nucleotide blast**

The annotated sequences were entered into the Basic Local Alignment Search Tool (BLASTN) of National Center for Biotechnology Information (NCBI) database to assess the quality of the sequences. Sequences alignments from the 5’ to 3’ ends was also determined on the platform. The E-value, topology, percentage G and C component, pairwise identities and sequence length were also determined.

**Phylogenetic analysis**

All sequences were imported into Geneious® prime software (Bio-matters Ltd) program to further assess the query nucleotides in a customized environment and thus calculate the phylogenetic tree using the genetic distance model and a tree build method of neighbour joining neighbour. Pairwise comparison was used to rank the strains based on their phylogenetic affiliation; a comparison of row items (% identity) with column item (genetic distance). The smaller numbers in the comparison reveals sequences that are closer together or has a more recent common ancestor and vice-versa. The strains were assigned identification numbers 1 to 12 for convenience. The diagonal in the chart is an intercept of “like for like” where no comparisons were carried out. A combined Metatree (https://itol.embl.de) was further calculated in order to visualize the evolutionary relationships of the strains within a broader framework.

**Sequence submission to GenBank**

The source modifiers, the sequences of the primer region masked out, the sequencing methods (in this case Sanger) and other features of the nucleotide sequences were distinguished before they were submitted to GenBank for processing. The nucleotide sequences were submitted as follows; 16S ribosomal DNA partial CDS were submitted directly into GenBank repository and mecA gene partial CDS were submitted via Bankit. After full processing by gene bank Staffs and accepted for publication in their repository, accession numbers were issued as identifiers for each submitted sequence and a release date is pronounced depending on the queue. A FLATFILE was also supplied with a detailed preview of the GenBank format before publication.

**Results**

The submitted sequence data are accessible at GenBank under accession MK643256- MK643267 (16SrRNA) and MK659546- MK659557 (mecA). Figures 1-3 shows phylogenetic trees constructed after sequences were aligned with each other and with other strains from GenBank.

The cladogram (Figure 2) shows that 8 out of 12 (66.7%) of the MRSA strains are monophyletic and clustered together in a Single
clade. In the Figure, 25.0% (3 out of 12) of the sequences were polyphyletic and 8.3% of the strains are paraphyletic and appeared to have arisen from a distant ancestor relative to others.

The genetic relationships among the sequences of 12 MRSA strains with Sequences (in terms of their mecA gene) from GenBank were calculated in a cladogram (Figure 3). Based on the phylogeny, the strains were grouped into five clades. Two different clades contained 3 strains (25% each) and other two clades clustered 2 MRSA (16.7%) strains each. The cladogram spreads radially with the first clade (clockwise direction) occurring furthest away in the evolutionary tree.

Table 1 shows that the strain MK643264 was the most distant, with low similarity of 4.25% to 72% and an evolutionary distance of 0.35 to ∞, from all other strains. The next distant strain MK643261, exhibited 70.58%- 92.43% similarity and approximately 0.08-0.37 evolutionary distance from other strains. Sequences like MK643256, MK643258, MK643260, MK643262, MK643263, MK643265, MK643266 and MK643267 seemed to be grouped with higher similarities (red spectrum). The sequence groups shared 60.34-90.31% similarity with an evolutionary distance of 0.08-1.94.

The native strains of the major evolutionary cluster in Figure 2 appear to populate the red spectrum in the lower left side of Table 2 with high percentage similarity of 97.38% to 100% with little or no evolutionary distance from each other (0.00 to 0.3).

Figure 4 shows the difference amongst nucleotides sequences that were aligned. This difference was as low as one nucleotide per 100 residues and high as 100 nucleotides per wrap.

Discussion

Policies to control MRSA infections in hospitals and community environments require baseline knowledge on the genetic diversity of circulating strains. This study provides information on the analysis of local MRSA strains of unknown diversity collected from a hospital environment with the aim of uncovering otherwise undetected differences that may aid infection control strategies.

All 24 sequences (16srRNA(n=12) and mecA (n=12)) in this study were queried through a blast search and alignment for similarity among the local strains. Query sequences were evaluated and compared with the GenBank database sequences based on the following criteria: query coverage, e-value, and percent identity. Query coverage represents the percentage of query sequence length

Figure 2. Cladogram map based on genetic relatedness of 16srRNA gene of 12 MRSA strains with 60 query strains from GenBank and Genome.Net database. The positions of the twelve MRSA strains are indicated in blue within the circle. Circular representation of the phylogenetic tree obtained using Multiple Sequence Comparison by Log-Expectation (MUSCLE) algorithm.
that is included in the alignment. In this study the query coverages ranged from 40-100%, mostly due to the fact that most of the sequences obtained were partial and their coverage, when aligned with whole genome sequences was low.

The e-value is a parameter that represents the number of chance alignment and directly dependent on the size of the database, such that a smaller database has higher likelihood of containing the sequence of interest. Search queries were restricted to the database of the genes of interest. This filtering procedure leads to an e-value of zero (0) for all sequence alignment in this study.

The percent identity refers to the part of residues in the compared nucleotide sequences that match at the same nucleotide position. The percentage identical sites in this study varied from 17.8-100% for 16sRNA and 13.65%- 100% for meca. These findings agree with a previous work (11) that reported that organisms that have fewer than 97.0% homology of rRNA will not associate to more than 60% of genomic relatedness. The alignment was carried out to understand the sequence variations in order to estimate the relatedness for the strains. In this study we noticed nucleotide differences at various points in the alignment which may lead to a change in the amino acid sequences or phenotypic traits.

The evolutionary histories of 12 isolates based on 16SrRNA and meca sequences were inferred using the multiple sequence comparison by log expectation (MUSCLE) algorithm. The constructed phylogenetic tree clearly resolved the relationships among the sequences. The tree was calculated to scale, with branch lengths having identical measurements as those of the evolutionary distances used to generate the phylogenetic tree. In the case of the
The sequences were clustered into 4 clades and \textit{mecA} into five clades in accordance with their genetic similarities. The tree depicts the overall phylogeny of strain across the species. The number of nodes multiplies exponentially from the first node while the circumference increases linearly resulting in a tree with the external circumference packed with nodes (taxa or strains).

The tree data structure expands outwards and the length of each path increases with the radius (spiral outwards from the root of the tree). The number of nodes multiplies exponentially from the first node while the circumference increases linearly resulting in a tree with the external circumference packed with nodes (taxa or strains). The root of the tree indicates the ancestral state for the entire phylogeny. The tree data structure expands outwards and the length of each path increases with the radius (spiral outwards from the root of the tree). The number of nodes multiplies exponentially from the first node while the circumference increases linearly resulting in a tree with the external circumference packed with nodes (taxa or strains).

The data presented in the meta-tree shows that most (66.7\%) of the MRSA strains were monophyletic based on their \textit{16SrRNA} genes and supports the claim that the gene is highly conserved among bacteria strains (10). Conversely, the diversity of the outstanding 33.3\% suggests that conserved regions of the gene are probably not as conserved as anticipated, which is in agreement with a previous finding (9) where the conserved regions of the \textit{16SrRNA} gene revealed important variation between phylogenetic groups. However, diversities observed in this study may be due to differences in the DNA extraction and ribosequencing methods employed by the gene depositors. The mounting cases of horizontal transfer of \textit{16SrRNA} genes (Genetic promiscuity) among different bacterial strains reported by Miyazaki (12) also explicate our findings. Furthermore, mutation in the hypervariable region of 16S ribosome due to prolonged exposure to the antibiotics in the hospital environment, may also have contributed to the genetic diversity realized among the phylogenetic groups.

\textit{S. aureus} strains morphs upon acquisition (MRSA) and excision (MSSA) of staphylococcal cassette chromosome \textit{mec} or \textit{SCCmec} (4). These variations have occurred multiple times in the evolutionary clock of \textit{S. aureus} and may be responsible for the \textit{mecA} gene-based diversities observed in this study.

Table 1. Pairwise comparison of Percentage identity (lower left) and genetic distance (Upper right) of twelve \textit{16SrRNA} sequences of MRSA.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>MK643260</td>
<td>1</td>
<td>0.29</td>
<td>0.10</td>
<td>0.16</td>
<td>0.19</td>
<td>0.11</td>
<td>0.22</td>
<td>0.35</td>
<td>0.23</td>
<td>1.02</td>
<td>1.92</td>
<td></td>
</tr>
<tr>
<td>MK643267</td>
<td>2</td>
<td>76.23</td>
<td>0.36</td>
<td>0.22</td>
<td>0.49</td>
<td>0.37</td>
<td>0.32</td>
<td>0.72</td>
<td>0.54</td>
<td>0.48</td>
<td>0.95</td>
<td>1.88</td>
</tr>
<tr>
<td>MK643260</td>
<td>3</td>
<td>90.31</td>
<td>71.49</td>
<td>0.22</td>
<td>0.17</td>
<td>0.08</td>
<td>0.22</td>
<td>0.34</td>
<td>0.20</td>
<td>1.21</td>
<td>2.60</td>
<td></td>
</tr>
<tr>
<td>MK643256</td>
<td>4</td>
<td>85.30</td>
<td>80.69</td>
<td>80.60</td>
<td>0.35</td>
<td>0.24</td>
<td>0.18</td>
<td>0.55</td>
<td>0.39</td>
<td>0.91</td>
<td>1.73</td>
<td></td>
</tr>
<tr>
<td>MK643265</td>
<td>5</td>
<td>83.57</td>
<td>64.15</td>
<td>84.78</td>
<td>71.84</td>
<td>0.17</td>
<td>0.34</td>
<td>0.20</td>
<td>0.18</td>
<td>1.55</td>
<td>1.55</td>
<td></td>
</tr>
<tr>
<td>MK643261</td>
<td>6</td>
<td>90.20</td>
<td>70.58</td>
<td>92.43</td>
<td>79.78</td>
<td>85.06</td>
<td>0.23</td>
<td>0.33</td>
<td>0.19</td>
<td>1.26</td>
<td>2.83</td>
<td></td>
</tr>
<tr>
<td>MK643263</td>
<td>7</td>
<td>81.18</td>
<td>74.15</td>
<td>81.02</td>
<td>83.87</td>
<td>72.38</td>
<td>80.13</td>
<td>0.56</td>
<td>0.39</td>
<td>0.96</td>
<td>1.85</td>
<td></td>
</tr>
<tr>
<td>MK643258</td>
<td>8</td>
<td>71.73</td>
<td>53.83</td>
<td>72.48</td>
<td>69.90</td>
<td>82.30</td>
<td>73.40</td>
<td>60.34</td>
<td>0.20</td>
<td>0.99</td>
<td>1.85</td>
<td></td>
</tr>
<tr>
<td>MK643262</td>
<td>9</td>
<td>80.00</td>
<td>61.34</td>
<td>82.21</td>
<td>69.52</td>
<td>84.32</td>
<td>83.11</td>
<td>69.41</td>
<td>82.48</td>
<td>1.94</td>
<td>1.94</td>
<td></td>
</tr>
<tr>
<td>MK643257</td>
<td>10</td>
<td>44.15</td>
<td>64.39</td>
<td>39.93</td>
<td>47.16</td>
<td>34.54</td>
<td>39.07</td>
<td>45.82</td>
<td>23.87</td>
<td>30.68</td>
<td>0.53</td>
<td>1.16</td>
</tr>
<tr>
<td>MK643259</td>
<td>11</td>
<td>30.76</td>
<td>46.22</td>
<td>27.36</td>
<td>32.43</td>
<td>23.58</td>
<td>26.72</td>
<td>31.40</td>
<td>14.70</td>
<td>20.04</td>
<td>62.00</td>
<td>0.35</td>
</tr>
<tr>
<td>MK643264</td>
<td>12</td>
<td>17.80</td>
<td>31.09</td>
<td>14.53</td>
<td>18.49</td>
<td>11.55</td>
<td>14.14</td>
<td>17.56</td>
<td>4.25</td>
<td>8.31</td>
<td>40.95</td>
<td>72.00</td>
</tr>
</tbody>
</table>

Key: Comparison gradient runs from Purple (min) to Red (max)

Table 2. Pairwise comparison of Percentage identity (lower left) and genetic distance (Upper right) of twelve \textit{mecA} genes sequences of MRSA.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>MK669547</td>
<td>mecA</td>
<td>1</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>MK669553</td>
<td>mecA</td>
<td>2</td>
<td>99.44</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>MK669554</td>
<td>mecA</td>
<td>3</td>
<td>99.44</td>
<td>100.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>MK669556</td>
<td>mecA</td>
<td>4</td>
<td>99.44</td>
<td>100.00</td>
<td>100.00</td>
<td>0.01</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>MK669548</td>
<td>mecA</td>
<td>5</td>
<td>99.81</td>
<td>99.25</td>
<td>99.25</td>
<td>99.25</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>MK669546</td>
<td>mecA</td>
<td>6</td>
<td>97.76</td>
<td>97.38</td>
<td>97.38</td>
<td>97.38</td>
<td>97.38</td>
<td>97.57</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>MK669567</td>
<td>mecA</td>
<td>7</td>
<td>97.76</td>
<td>97.38</td>
<td>97.38</td>
<td>97.38</td>
<td>97.38</td>
<td>97.57</td>
<td>100.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>MK669550</td>
<td>mecA</td>
<td>8</td>
<td>22.40</td>
<td>22.14</td>
<td>22.14</td>
<td>22.14</td>
<td>22.54</td>
<td>21.86</td>
<td>21.86</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>MK669551</td>
<td>mecA</td>
<td>9</td>
<td>22.40</td>
<td>22.14</td>
<td>22.14</td>
<td>22.14</td>
<td>22.54</td>
<td>21.86</td>
<td>21.86</td>
<td>100.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>MK669552</td>
<td>mecA</td>
<td>10</td>
<td>22.40</td>
<td>22.14</td>
<td>22.14</td>
<td>22.14</td>
<td>22.54</td>
<td>21.86</td>
<td>21.86</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>MK669555</td>
<td>mecA</td>
<td>12</td>
<td>20.13</td>
<td>19.87</td>
<td>19.87</td>
<td>19.87</td>
<td>20.13</td>
<td>20.40</td>
<td>20.40</td>
<td>13.65</td>
<td>13.65</td>
<td>13.65</td>
</tr>
</tbody>
</table>
Figure 4. Multiple sequence alignment of twelve (12) nucleotide sequences of the 16S rRNA gene of MRSA isolates from Sokoto. The coloured bases indicate regions of nucleotides diversity among isolates. The identical residues are represented with dots. The sequence layout is set to wrap at every 100 residues.
The distance between the strains is a measure of the genetic distances between the strains. It is also a measure of times since the isolates diverge in evolution (17). Consequently, the strains with GenBank accession Identities MK643264 (16S rRNA), MK659555 and MK659549 are more distantly related to all other strains in the tree. The high nucleotides similarity between 16S rRNA genes of strains that originated from racing horse (SA17S6, Korea) and human (Blood and Pus form RSAU000415, Japan, SA957, Australia and SABB06263, Brazil) gives credence to the probability of human-to-animal host jump events (6).

It can be argued from the tree that a number of diversities have taken place among the Sokoto-strains. The tree shows that the capacity to mutate has evolved on several occasions during the evolutionary clock of the strains. This may be due to mutations owing to selective pressures (Antimicrobial resistance) or human-to-animal host jump events. These jumps are not startling because S. aureus has a clonal population structure, it grows and diversifies by mutation more often than by recombination (horizontal acquisition of short stretches of DNA and between different strains which then recombine and replace existing stretches of DNA).

Conclusions

To the best of our knowledge, this study is the first of its kind in Sokoto to have successfully deposited nucleotide sequences of studied MRSA strains into Genbank database. It has also revealed the phylogenetic diversity of the strains. Based on sequence analysis it was found that a great deal of diversity has taken place among Sokoto-strains identified in this study. The sequence reads of molecularly studied strains should be deposited into GenBank database by researchers to further populate the database and improve the efficiency of local phylogenetic analysis to foster better bioinformatics as paucity of local data exists in these databases.

References

11. Lan Y, Rosen G, Hershberg R. Marker genes that are less conserved in their sequences are useful for predicting genome-wide similarity levels between closely related prokaryotic strains. Microbiome 2016;4:18.