Comparative whole-genome analysis of China and global epidemic Pseudomonas aeruginosa high-risk clones

Zhao, Yonggang; Xie, Lu; Wang, Chongzhi; Zhou, Qian; Jelsbak, Lars

Published in: Journal of Global Antimicrobial Resistance

Published: 01/12/2023

Document Version: Final Published version, also known as Publisher’s PDF, Publisher’s Final version or Version of Record

License: CC BY-NC-ND


Citing this paper Please note that where the full-text provided on CityU Scholars is the Post-print version (also known as Accepted Author Manuscript, Peer-reviewed or Author Final version), it may differ from the Final Published version. When citing, ensure that you check and use the publisher's definitive version for pagination and other details.

General rights Copyright for the publications made accessible via the CityU Scholars portal is retained by the author(s) and/or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights. Users may not further distribute the material or use it for any profit-making activity or commercial gain.

Publisher permission Permission for previously published items are in accordance with publisher's copyright policies sourced from the SHERPA RoMEO database. Links to full text versions (either Published or Post-print) are only available if corresponding publishers allow open access.

Take down policy Contact lbscholars@cityu.edu.hk if you believe that this document breaches copyright and provide us with details. We will remove access to the work immediately and investigate your claim.
Comparative whole-genome analysis of China and global epidemic Pseudomonas aeruginosa high-risk clones

Yonggang Zhao a,1, Lu Xie b,1, Chongzhi Wang a, Qian Zhou c,*, Lars Jelsbak a,2,***

a Department of Biotechnology and Biomedicine, Technical University of Denmark, Lyngby, Denmark
b Research Center for Micro-Ecological Agent Engineering and Technology of Guangdong Province, Guangzhou, Guangdong Province, China
c Department of Computer Science, City University of Hong Kong, Hong Kong, China

ARTICLE INFO
Article history:
Received 15 March 2023
Revised 15 August 2023
Accepted 31 August 2023
Available online 13 September 2023

Editor: Stefania Stefani
Keywords:
Pseudomonas aeruginosa
Core-Pan gene
High-risk clone
Phylogenetic analysis
Population structure
Antibiotic resistance

ABSTRACT

Objectives: The various sequence types (STs) of Pseudomonas aeruginosa (P. aeruginosa) high-risk clones (HiRiCs) have been sporadically reported in China, but the systematic analysis of genomes for these STs remains limited. This study aimed to address the evolutionary pathways underlying the emergence of HiRiCs and their routes of dissemination from Chinese and global perspectives.

Methods: The phylogenetic analysis was performed based on 416 newly sequenced clinical P. aeruginosa strains from Guangdong (GD), published genome sequences of 282 Chinese isolates, and 868 HiRiCs isolates from other countries. The genomic comparison study of global HiRIC ST244 was conducted to detect the model of global dissemination and local separation driven by association regional-specific antibiotic resistance genes. Furthermore, the evolutionary route of the emerging, China-specific HiRIC ST1971 was explored using Most Recent Common Ancestor (MRCA) analysis.

Results: Based on comparative genomics analysis, we found a clear geographical separation of ST244 isolates, yet with an association between ST244 isolates from GD and America. We identified a set of 38 AMR genes that contribute to the geographical separation in ST244, and we identified genetic determinants either positively (MexB) and negatively (opmD) associated with GD ST244. For the China-unique HiRIC ST1971, its evolutionary history across different continents before emerging as ST1971 in China was also deduced.

Conclusion: This study provides insight into the specific genetics underlying regional differences among globally disseminated P. aeruginosa HiRICs (ST244) as well as new understanding of the dissemination and evolution of a regional HiRIC (ST1971). Understanding the genetics of these other HiRICs may assist in controlling their emergence and further spread.

© 2023 The Author(s). Published by Elsevier Ltd on behalf of International Society for Antimicrobial Chemotherapy.

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

1. Introduction

The emergence of multidrug-resistant (MDR) and extensively drug-resistant (XDR) P. aeruginosa leads a growing threat of worldwide nosocomial infections [1]. The global dissemination of high-risk clones (HiRiCs) is associated with complex interactions between different factors, including cumulative mutations, spread of multidrug-resistance genes, and expression of highly virulent exotoxins [2,3]. Several STs of P. aeruginosa, including ST235, ST111, ST233, ST244, ST357, ST308, ST175, ST277, ST654, and ST298, are of special concern because of their global dissemination, pathogenicity, and presence of carbapenemase and/or extended-spectrum β-lactamases (ESBLs) in their genomes [4]. These notorious STs are strongly associated with poor clinical outcomes and taking mass of horizontally acquired resistance determinants [2]. In recent years, HiRiCs emerged in many territories in China. In Zhejiang, ST463 posed a potential health threat due to its unique combination of pyocyanin production and virulence genes [5]. In Shenzhen, the detection of blaNDM-1 and colistin resistance mcr-1 MDR P. aeruginosa highlights the necessity for continuous surveillance in paediatric patients [6]. In Shanghai, a low prevalence of polymyxin-resistant P. aeruginosa, including the world-disseminated ST277, was discovered after polymyxin ad-

https://doi.org/10.1016/j.jgar.2023.08.020
2213-7165/© 2023 The Author(s). Published by Elsevier Ltd on behalf of International Society for Antimicrobial Chemotherapy. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
ministration in one tertiary teaching hospital [7]. Furthermore, the carbapenem-resistant *Pseudomonas aeruginosa* (CRPA) strains were reported in Shandong (ST244), Sichuan (ST277), and Hainan (ST357), respectively [8]. In a previous study, we conducted a genome-based epidemiology survey of the Guangdong *P. aeruginosa* clinical isolates and identified seven CRPA HiRiCs strains, including ST111, ST235, ST244, ST277, ST298, ST357, and ST1971 [9]. Among them, the isolates of ST1971 had significantly fewer SNPs (mean, 165 SNPs) than the isolates of the other HiRiCs. It suggested that the strains of ST1971 are closely related phylogenetically [10]. Interestingly, genetic structure analysis revealed that a ~45 kb (Kilobase) exogenous fragment was found to be inserted into the chromosome of ST1971 strains, which contained ~25 kb multiple antibiotic-resistant island [10].

To better understand the epidemiological characteristics of *P. aeruginosa* HiRiCs, many genotyping methods were applied, including O-serotype [11], pulsed-field gel electrophoresis (PFGE) [12], multiple locus variable number of tandem repeats analysis (MLVA) [13], and multilocus sequence typing (MLST) [14]. Because of the highly flexible genome of *P. aeruginosa*, the resolution of classical typing methods is limited [15]. Along the cost reduction of high-throughput DNA sequencing, the affordable whole genome sequencing (WGS) could provide unprecedented resolution for the population study of pathogens. The whole-genome SNP-based phylogenetic analysis can provide comprehensive molecular evidence, not only for sequence typing but also for MDR/XDR prediction, population structure determination, and phylogenetic inference [16].

Recently, the specific presence of genes and the significant relationship between HiRiCs and resistance mechanisms have been investigated [4]. However, the lack of genomic features of Chinese HiRiCs limits our ability to understand the epidemiology of these important clones and to improve infection control measures. In the present study, we aimed to uncover the phylogenetic distribution, population structure, and evolutionary origin of the Guangdong (GD) *P. aeruginosa* strains from national and international perspectives.

### 2. Materials and methods

#### 2.1. Genome sequences collection

We selected 416 newly sequenced clinical *P. aeruginosa* strains isolated from 10 tertiary hospitals in Guangdong Province, China [10]. All WGS data for 416 Guangdong *P. aeruginosa* isolates were able to be accessed through the NCBI SRA database under BioProject number PRJNAB23853 (Supplementary Table S1). Furthermore, data for 282 Chinese public *P. aeruginosa* strains (Supplementary Table S2) and 868 global HiRiCs strains were collected (Supplementary Table S3). The ST1971 was identified as unique to China HiRiCs [10]. To address the evolutionary routine of ST1971, we retrieved 244 highly similar strains within the same subclade from the NCBI database (Supplementary Table S4).

#### 2.2. Genotype profiling

The Comprehensive Antibiotic Resistance Database (CARD) (version 3.1.1, https://card.mcmaster.ca/home) was used for antimicrobial resistance (AMR) gene profiling [17]. The Virulence Factor Database (VFDB, http://www.mgc.ac.cn/VFS/) was used for virulence gene detection [18]. The serotype was identified by *Pseudomonas aeruginosa* serotype (Past) (v1.0), which enabled in silico serotyping of *P. aeruginosa* isolates using WGS data [19]. Sequence typing of all strains was performed via the MLST scheme (https://pubmlst.org/paeruginosa/; last accessed 11 August 2022) [20]. The data of genotypic profiling are listed in Supplementary Table S1.

#### 2.3. SNP calling and whole-genome SNP based phylogenetic analysis

The NUCmer (NUcleotide MUMmer) function of MUMmer 3.23 was used to align genome sequences with the reference genome PAO1 [21]. Four other well-studied strains, PA7 [22], PA14 [23], PAK [24], and LESB58 [25], were also introduced as an outgroup. The alignment file was filtered by the delta-filter function, and the SNP locus was identified by the show-SNPs function using default parameters [26]. The SNPs, which are located in the repetitive regions or the distance between the two SNP loci is less than 100bp, will be discarded [26]. Then, we concatenated the rest of the SNPs as whole genome SNP sequences. The evolution analysis was conducted by PhyML with an HKY model and 1000 bootstrap. The online tool iTol was used to display the phylogenetic tree. The population structure was calculated by hierBAPS [27]. Clades and subclades were classified according to the phylogenetic tree, with a maximum number of populations (K) at 100, 200, and 300. The $\chi^2$ test was used for the analysis of the distribution of exou$^+$ and exos$^+$ within each clade.

#### 2.4. Core-Pan genes evolution

We first added the region tags after gene id, and the whole gene pool were used for Core-Pan gene analysis by the CD-HIT software, with a 50% pairwise identity and a 0.7 difference in length [28]. Region-specific and common genes were extracted. The UpSet diagram was used to illustrate the distribution and gene frequencies based on the presence/absence of genes across different regions. Similarities of pan-genes in different regions were shown as heatmaps. All antibiotic resistance genes of the HiRiCs ST244 strains were profiled and visualized using a heatmap.

#### 2.5. Single copy core gene-based phylogenetic analysis

Based on Core-Pan gene analysis, we selected the single-copy core genes for the evolutionary analysis. The protein sequences of those single-copy core genes were subjected to multiple sequence alignments by MUSCLE. The phylogenetic tree was inferred using the RAxML software (v8.2.12) [29]. Step 1: Use the parameters ‘-m PROTGAMMAAUTO -P 12345′ to build phylogenetic tree 1. Step 2: Use the parameter ‘-m PROTGAMMAAUTO -P 12345 -f a -x 12345 - # 100′ to build phylogenetic tree 2. Step 3: Use the parameters ‘-m PROTGAMMAAUTO -P 12345 -f b -t tree1 -z tree2′ to combine phylogenetic tree 1 and tree 2 into a final version of the phylogenetic tree.

#### 2.6. Correlation analysis between geographical regions and antibiotic resistance genes

Guangdong (GD) strains and public strains were organized into a two-dimensional matrix of geography and strains, and a two-dimensional matrix of strains and antibiotic resistance genes (numerical values represent the number of antibiotic resistance genes). The R package Himisc was used for calculating Spearman’s correlation, and significant relations were defined as adjusted P-value <0.05 and absolute value of correlation coefficient >0.5 [30].

#### 2.7. Most recent common ancestor (MRCA) analysis

Bayesian analysis of divergence times was performed using BEAST (v2.4.2) [29]. All the SNPs in which at least one isolate differed from the reference strain PAO1 were concatenated. BEAUTI
were used to estimate these SNPs and to calculate the phyloge- netic distances between isolates. Then, we used BEAST to infer the most recent common ancestor with the following user-determined settings: a lognormal relaxed molecular clock model and a general time-reversible substitution model with gamma correction [31]. Results were produced from one chain with 50 million steps, sampled every 1000 steps. The first 3 million steps were discarded as a burn-in. The maximum clade credibility tree was generated using the TreeAnnotator program from the BEAST package and displayed using FigTree (v1.4.2) [31]. Tree parameters were calculated by Tracer (v1.6) [32].

3. Results

3.1. Population structure of Chinese P. aeruginosa strains

The molecular epidemiology of 416 newly sequenced P. aerugi- nosa strains was characterized by high clonal diversity. Among them, the widely disseminated P. aeruginosa HiRiCs and the regional risk clone ST1971, which exhibited MDR/XDR phenotype, were identified (Table S1). In order to identify the population structure of the newly sequenced GD strains and public Chinese P. aeruginosa strains, the maximum likelihood phylogenetic tree was built based on 416 newly sequenced GD strains and 282 public Chi- nese strains. The result showed that the strains could be divided into two major clades and one small clade (Fig. 1). Among the two main clades, the public Chinese strains and the newly sequenced GD strains were distributed mainly in clade 1 (n = 287, 68.99% of GD strains and n = 217, 76.95% of public Chinese strains) and clade 2 (n = 128, 30.77% of GD strains and n = 63, 22.34% of public Chi- nese strains) (Fig. 2, Circle 1; Supplementary Table S1 and S2). It is worth noting that clade 3 only consists of three strains (Supple- mentary Tables S1 and S2). The gene profile of virulence factors showed that strains with exoS\textsuperscript{+} genes dominated clade 1 (471/504, 93.45%, \( \chi^2 = 475.03, P\)-value = 2.2e-16), whereas the strains with exoU\textsuperscript{+} genes had a low proportion and were located primarily in clade 2 (168/191, 87.96%, \( \chi^2 = 65.78, P\)-value = 5.2e-15). The P values of the \( \chi^2 \) test were all less than 0.05; there is a correlation between the distribution of exoS\textsuperscript{+} and exoU\textsuperscript{+} strains and the clades (Supplementary Table S1 and S2). In relation to serotype statistics, the three most common serotypes among the Chinese strains are O11 (n = 167, 23.93%), O6 (n = 145, 20.77%), and O2 (n = 111, 15.90%). The MLST distribution of P. aeruginosa isolates from GD and other regions in China demonstrated that the most widely dispersed and prominent ST was ST244 (n = 26, 3.72%) (Fig. 2, Circle 2). The ST463 strains also had a high proportion (n = 49, 7.02%), but the isolates of ST463 were mainly reported in Zhe- jiang. To specifically examine the GD strains, we did not identify the Chinese-reported HiRiCs ST233 and ST308, but the ST298 was exclusive to the GD strains. Additionally, the GD strains were found to be randomly distributed the entire phylogenetic tree, covering the majority of the subclades (Fig. 2, circle 4). The China-unique HiRiCs ST1971 and other STs, including ST316, ST357, ST773, ST1076, ST1284, ST1339, ST1621, ST207, ST2326, ST2850, ST2944, ST2957, ST235, ST352, ST313, and ST773, were located in the same branch clade (Fig. 2, circle 3). Interestingly, five GD ST1338 strains were located in the same subclade as ST244. Out of the seven house-keeping genes used for MLST identification, only one SNP differ- ence was observed in the nuoD gene between ST244 and ST1338. Based on above findings, it is speculated that ST1338 belongs to the same clonal complex (CC)244 (Fig. 2, circle 2).

3.2. Phylogenetic analysis of Guangdong P. aeruginosa HiRiCs in a global perspective

We placed the 416 newly sequenced GD P. aeruginosa genomes in the global context of 868 publicly available HiRiCs genomes (Table 1). This collection of public HiRiCs genomes represented 47 countries and 9 well-reported HiRiCs; 42.74% of the genomes were from North America (n = 371) and 30.53% from Europe (n = 265). Some types of HiRiCs have a certain regionality, such as ST277 strains that were mainly from South America (22/38, 57.89%), ST111 strains were mainly from Europe (133/179, 74.30%), and ST357 strains were mainly from Asia (including the strains from China, 17/30, 56.67%).

The phylogenetic tree of the global P. aeruginosa HiRiCs strains could be divided into 5 distinct clades and 16 subclades (Fig. 3). The different HiRiCs showed obvious clustering into distinct subclades as expected (cf. circles 1 and 4 in Fig. 3). However, in some cases these clusters were punctuated by other STs. For example, the subclade of ST244 contains additional STs constituting a clonal complex as described above (Fig. 3, circle 4). From a geographical perspective, the strains from America were classified into nine subclades and a large portion of them was in subclade 14 (47.70%, 207/434). The European strains were classified into 10 subclades with a large portion in subclade 3 (50.19%, 133/265), and the Asian public HiRiCs strains were mainly in sub- clade 13 (37.21%, 32/86). For some widely disseminated HiRiCs, such as ST111, ST235, and ST244, the same ST strains from different geographical locations were randomly distributed within the same subclade (Fig. 3), underscoring the conclusion that these clones are globally disseminated. Among GD strains, six globally dominant HiRiCs, including ST111, ST235, ST244, ST277, ST357, and ST298, were located in six different subclades, in which ST244 was in sub- clade 2 and ST1971 was in subclade 8.

3.3. Comparative genomic analysis of ST244

The Core-Pan gene analysis of ST244 strains (n = 106) consisted of 3672 core genes, 6633 flexible genes, and 3281 unique genes. The ST244 strains, isolated from Guangdong (GD) and America, shared the same 1635 genes, whereas 862 genes were unique to the GD ST244 strains (Fig. 4). Among the GD ST244-specific genes, 77 drug-resistant genes were identified by the CARD annotation.
Fig. 2. Single copy core gene based phylogenetic tree of 416 Guangdong P. aeruginosa strains and 282 Chinese public strains. The circles from the innermost to the outermost rings represent the Clade, MLST, the main clones with the same clade as ST1971, isolation area, and serotype of strains, respectively. The different colours represent the specific types or sources.

Table 1
The China and global P. aeruginosa high-risk clone strains collection.

<table>
<thead>
<tr>
<th>Sequence types</th>
<th>The number of Guangdong strains</th>
<th>Public data</th>
<th>China (Public data)</th>
<th>America</th>
<th>Europe</th>
<th>Oceania</th>
<th>Unknown</th>
<th>Total</th>
<th>Time period</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST11</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>33</td>
<td>133</td>
<td>8</td>
<td>183</td>
<td>1979–2016</td>
<td></td>
</tr>
<tr>
<td>ST235</td>
<td>5</td>
<td>4</td>
<td>30</td>
<td>2</td>
<td>82</td>
<td>73</td>
<td>14</td>
<td>217</td>
<td>1988–2016</td>
</tr>
<tr>
<td>ST244</td>
<td>17</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>34</td>
<td>21</td>
<td>17</td>
<td>4</td>
<td>104</td>
</tr>
<tr>
<td>ST274</td>
<td>13</td>
<td>1</td>
<td>2</td>
<td>212</td>
<td>16</td>
<td>1</td>
<td>4</td>
<td>249</td>
<td>1990–2021</td>
</tr>
<tr>
<td>ST277</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>29</td>
<td>1</td>
<td>4</td>
<td>43</td>
<td>1997–2022</td>
<td></td>
</tr>
<tr>
<td>ST298</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>39</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>48</td>
<td>1995–2017</td>
</tr>
<tr>
<td>ST313</td>
<td>11</td>
<td>1</td>
<td>6</td>
<td>3</td>
<td>12</td>
<td>1</td>
<td>3</td>
<td>36</td>
<td>1997–2018</td>
</tr>
<tr>
<td>ST316</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td>1960–2013</td>
</tr>
<tr>
<td>ST357</td>
<td>12</td>
<td>1</td>
<td>14</td>
<td>3</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>42</td>
<td>1983–2017</td>
</tr>
<tr>
<td>ST1971</td>
<td>12</td>
<td>1</td>
<td>8</td>
<td>31</td>
<td>434</td>
<td>265</td>
<td>27</td>
<td>45</td>
<td>1979–2016</td>
</tr>
<tr>
<td>Total</td>
<td>86</td>
<td>11</td>
<td>55</td>
<td>31</td>
<td>434</td>
<td>265</td>
<td>27</td>
<td>45</td>
<td>954</td>
</tr>
</tbody>
</table>

In total, 4953 P. aeruginosa genome sequences data were downloaded from the Pseudomonas Genome Database (https://www.pseudomonas.com/). Based on sampling information, 868 globally reported HiRiCs strains genome data were selected for the comparative genomic analysis.

(Supplementary Table S5), and 75 of them belonged to exogenous genes. These antibiotic resistance genes could be divided mainly into aminoglycosides-, carbapenems-, and macrolides-resistance genes.

The genetic similarity matrix generated by the accessory genes (unique and flexible genes) of all ST244 strains was displayed using a heatmap (Fig. 5A). Compared with ST244 strains collected from other regions, the accessory genes of GD ST244 strains and American ST244 strains were most similar and grouped into the same cluster (Fig. 5A). Similarly, a focus specifically on AMR genes also revealed a close relationship between GD ST244 and American ST244 (Fig. 5B). Overall, these results either suggest that the ST244 strains found in America and China share a common ancestor or that strains are shared more frequently between these regions compared with others. To further explore the genetics underlying the geographical separation of ST244 isolates (as shown in Fig. 5), we next investigated the association between AMR genes and the geographical locations of the 106 ST244 strains. Here, 38 AMR genes were found to be highly associated (either positive or negative correlations) with geographical location (P-value <0.05, and absolute value of correlation coefficient >0.5). The strains isolated from Oceania showed the most significant regional associations compared with the other regions, with 22 genes containing nine positively correlated and 13 negatively correlated genes. Although the colour of the AMR genes pattern in Europe and America is moderately lighter in contrast with other regions, some geographically associated genes were found in Europe and America, and the correlation value of these genes is higher than that in other regions. It may due to the varying strategies against AMR and the differences in antibiotic prescription among different states.
in America and countries in Europe. In addition, the high levels of migration between these regions could result in more frequent interactions between strains, potentially influencing their evolution. The finding that highly homologous genes are present in strains from both American and European regions further supports this hypothesis. Overall, this analysis show that ST224 isolates from different geographical regions are associated with different subsets of AMR genes (Fig. 6). Focusing on the GD strains, three antibiotic resistance genes/alleles, including OpmD, MexB, and EcoEFTu_PLV, showed significant correlations (Fig. 6). Both MexB and EcoEFTu_PLV resistance alleles were positively correlated with GD strains, but OpmD showed a negative correlation (Fig. 6). Furthermore, the MexB gene is one of the validated unique genes of the GD ST244 strains, suggesting that, compared with...
the ST244 strains in other regions, the MexB gene is highly associated with the GD ST244 strains. Furthermore, the phylogenetic tree combined with the heatmap of the 38 AMR genes showed abundant gene copy numbers of TtxR and OpmD in ST244 strains (Fig. 7).

3.4. The phylogeny of China’s unique HiRiC ST1971

This clone was surrounded by various other ST strains (Fig. 2), which suggested that ST1971 has a close evolutionary relationship with these STs. The MRCA analysis of ST1971 and closely related STs (ST316, ST357, ST773, and ST1076) indicated that these strains evolved from the same ancestral strain in South America. ST1284, ST773, ST532, ST1971, and ST1076, belonged to the same group, and their common ancestor strain spread from South America to Europe and North America and experienced a certain mutual transmission between these two continents. The common ancestor strain of ST1284, ST1971, and ST1076 spread from South America to Asia. Finally, ST1971 was introduced into China from Asia (Fig. 8). Our findings suggest that ST1076 and ST1971 are closely related and belong to the same subclade. Despite PA_D1 and PA_D9 of ST1971 being first detected in Guangxi Province in 2013–2014, we believe that the GD ST1971 strain appeared earlier than these two strains and other ST1971 strains in China (Fig. 8).

4. Discussion

Past research has focused mainly on the P. aeruginosa HiRiCs isolates from Europe and North America [15,33–35], with only sparse cases reported in China. It is widely observed that isolates collected from any region may contain a specific set of spatiotemporal characteristics [36]. Nosocomial infections by P. aeruginosa present a growing health threat in China [37], suggesting the necessity of analysing the phylogenetic and geographical relationship of P. aeruginosa HiRiCs. By combining newly sequenced GD strains and public strains from China and other countries, we depicted a complete picture of the phylogenetic relationship of P. aeruginosa from both a Chinese and global perspective.

Analysis of the population structure demonstrated that the P. aeruginosa strains in China could be divided into two major clades and one small clade, which is consistent with previous studies [38]. The representative strains PA14 and PAO1 were located in the two major clades respectively, while PA7 was observed in the small clade (Fig. 1, clade 3). It has been reported that by analysing strains from different environments, the clade in which PAO1 is located is naturally more abundant than that in which PA14 is located. The strains in the present datasets are all clinical isolates still support this conclusion [15].

P. aeruginosa contains a large number of virulence factors [39]. One of the important and widely studied virulence determinants is the type III secretion system (TTSS) [40]. Cytotoxins including ExoS, ExoT, ExoU and/or ExoI, and ExoU are delivered to the host by TTSS and show strong association with poor prognosis [41]. There is an association between TTSS genotypes and patterns of antibiotic resistance; for example, the exoU+ genotype has been reported to be less common in MDR strains [5]. However, from the perspective of GD strains, many exoU+ strains also exhibited MDR characteristics (Supplementary Table S1).

The distribution of the newly sequenced GD strains was found throughout the entire phylogenetic tree, indicating that the P. aeruginosa population of GD strains largely represents the Chinese diversity of this pathogen. This corresponds well with the large population inflow in the Guangdong Province, China [42]. When studying the evolutionary relationship between GD and global P. aeruginosa HiRiCs genomes, we found that HiRiCs of the same ST from different regions showed a mixed distribution within the same subclade, which may be easily explained by cross-border trade and travel. Global communication is an important factor that increase the risk of spreading P. aeruginosa HiRiCs.

From a global perspective, most of HiRiCs types correspond to one serotype, even if the same ST comes from different regions. The notable exceptions were HiRIC ST111 and ST244. The
Fig. 6. Heatmap showing association frequency of antimicrobial resistance genes and geographical locations of 106 global high-risk clone ST244 strains; 38 antimicrobial resistance genes among 106 ST244 strains were found to be highly associated with geographical locations. The strains isolated from Oceania showed the most significant regional associations compared with the other regions, with 22 genes containing nine positively correlated and 13 negatively correlated genes. The antimicrobial resistance genes of OpmD, MexB, and Ecol_EFTu_PLV had the highest association with Guangdong (GD) ST244 strains. The grading numbers in the colour strip depict the correlation value and the asterisk number represents the P value. *P < 0.05; **P = 0.01; ***P = 0.0001.

ST111 is mainly distributed throughout the world with serotype O12 [1]. However, strains of ST111 in China are mainly classified as serotype O4 (Fig. 3, Supplementary Tables S1 and S3). We have previously shown that ST111 expressing serotype O12 has evolved via serotype switching, where the native serotype O4 biosynthesis gene cluster was replaced by the O12 cluster by horizontal gene transfer and recombination [43,44]. Furthermore, the process of serotype switching (from O4 to O12) has occurred repeatedly and independently in ST111 across different continents and countries [43,44]. The observation of ST111 O4 and O12 isolates in China may suggest similar serotype switching processes. The ST244 is the founder of the CC244 clonal complex and is commonly associated with the serotype O2 [4], but the GD ST244 strains in China could be assigned as O2 (major serotype) and two minor serotypes, O12 and O5 (Fig. 3). Whether these different serotype expressing ST244 are examples of serotype switching processes similar to the one observed in ST111 in currently not known.

Recently, more and more STs were categorized as HiRIcs because of clinical characteristic alterations following extrinsic stress and/or internal genomic variation. ST235, one of the most prominent HiRIcs, may have a relationship with the previous use of antipseudomonal fluoroquinolones [36]. ST111 was identified in UK and associated with horizontal gene transfer and mostly carried VIM-type MBLs [1]. CC446 (STs 298 and 446) with MDR/XDR phenotypes was reported harbouring a novel antibiotic resistance integron [45]. ST244 is a ubiquitous HiRIC, causing multidrug-resistant nosocomial outbreaks [1]. In the present study, through the comparative genomic analysis, the GD ST244 strains were most similar to the ST244 strains isolated in America. Thirty-eight AMR genes with a high geographical association frequency of HiRIC ST244 were identified, and the phylogenetic tree, combined with the heatmap of these AMR genes, showed abundant gene copy numbers of TnXr and OpmD in ST244 strains. TnXr plays a critical role in bacterial development of tetracycline resistance [46], and OpmD is the outer membrane channel protein of the efflux complex MexGH–OpmD [47]. Among the GD ST244 strains, OpmD, MexB, and Ecol_EFTu_PLV showed significant geographical correlations, and, interestingly, antimicrobial resistance genes of
Opmd and MexB represented a mutually exclusive status. MexB is the inner membrane multidrug exporter of the efflux complex MexAB-OprM. The balance between efflux complex MexAB-OprM and MexGHI-OpmD in ST244 may be explained by different routes of adaptation to clinical environments in a most economical metabolic form [48]. In a previous study, for the MexAB-OprM efflux pump system of GD strains, the mutation rate of the mexB gene reached 21.9% [10]; such a high mutation frequency ultimately leads to phenotypic changes that require further study. It was reported that the sequence variants of Ecol_EFThi_PLV confer resistance to Pulvomycin [49]. Because it is not routine drug for the treatment of P. aeruginosa infection, drug susceptibility testing for Pulvomycin was not performed in this project.

Furthermore, through phylogenetic comparative methods, we validated that ST1338 belongs to the same clonal complex as ST244. This observation point towards a value in paying attention to the emergence of new STs of the global HiRICs.

Previously, SNP-based phylogenetic analysis reveals that there was a close evolutionary relationship between China-unique HiRIC ST1971 strains, and the chromosomal gene island insertion fragment endows the genetic structure basis of multidrug resistance of the strains. All ST1971 strains are with exoU genotype, and virulence experiments proved that they had high virulence [10]. Therefore, the multidrug-resistant phenotype, high potential of transmissibility, and virulence of these strains infer that they pose a serious threat in hospitals. Information on its evolutionary background is well worth further study. Our phylogenetic and MLST analysis indicated that ST1971 may be a newly emerged HiRIC emerging in China. From a geographical perspective, ST1971 strains PA_D1 and PA_D9 were isolated in the city of Nanning in the Guangxi region during 2013–2014, and the ST1076 strain AZPAE14988 was isolated in Shatin, Hong Kong in 2010. Guangdong, Guangxi, and Hong Kong are relatively close geographically. As the window areas of China’s foreign trade, these areas have experienced many outbreaks of infectious diseases [50,51]. International trade and travel facilitated the spread of this pathogen. More interestingly, an outbreak of ST1076 was previously reported in the burn ward of the University Hospital of Lausanne from 2010 to 2012 [52]. One of the GD ST1971 strains was also isolated from the burn and plastic surgery ward. The previous study identified the high virulence and multidrug resistance characteristics of GD ST1971 strains [53]. These results are a reminder to keep focus on the surveillance on the GD ST1971 clone and prevent new P. aeruginosa HiRICs outbreaks.

There are some limitations that need to be addressed. First, the data of the HiRICs strains have regional biases. Although we collected genome data as widely as possible, including 888 strains...
from seven continents, European and North American strains account for a higher proportion (73.27%, 636/868). This is related to the outbreak status and the antibiotic policy of dealing with P. aeruginosa infections in different regions. Second, the source diversity of the Chinese strains was collected mainly from clinical isolates, although we added five well-studied control strains for reference. Finally, through the MRCA analysis, we can speculate that the earliest known occurrence of ST1971 in China was in Guangdong, but we had no earlier archived GD ST1971 strains to use as direct evidence. Considering the above-mentioned limitations, our future aim is to collect more environmental isolates and broaden the time span of our strain collection to enable more detailed phylogenetic analysis to better understand the evolution of P. aeruginosa and their global and regional dissemination.

In conclusion, this work is the first large-scale population genomic study of the Chinese P. aeruginosa HiRIC strains. The present results reveal the diversity of Chinese representative STs and demonstrate the global phylogeny and evolutionary relationship represented by the global HiRICs ST244 and the Chinese unique endemic HiRIC ST1971. Overall, our study (of ST244) supports a model of global dissemination and local separation driven by association to region-specific antibiotic resistance genes. In addition, our study of the emerging HiRIC (ST1971) underlines the importance of including global sequence data to determine the evolutionary route towards HiRIC. The study will broaden the theoretical basis on the evolution and spread of P. aeruginosa HiRICs and will provide new insights into local control of P. aeruginosa.

Appendix A. Supplementary data

Supplementary material related to this article can be found in the online version.

Funding: This work was supported by a research grant to LJ from the Independent Research Fund Denmark (9039-00350A).

Competing interests: The authors declare that we have no conflicts of interest to this work.

Ethical approval: The study was approved by the Ethics Committee of the Zhubiang Hospital of Southern Medical University (reference number 2021-KY-046-01). The samples were obtained with written informed consent and reviewed by the ethical board, in accordance with the tenets of the Declaration of Helsinki.

Data statement: All the genome sequence data used in this project can be publicly accessed from NCBI. The accession numbers are available in Supplementary Materials. The MLST analysis was performed for 416 recently sequenced strains, and the nucleic acid sequences of the seven housekeeping genes for each strain can be accessed at https://doi.org/10.11583/DTU.22100552.v1. Detailed information on software and databases covered in the article can be accessed at https://doi.org/10.11583/DTU.22100573.v1.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jgar.2023.08.020.

References


[32] BioWeb -tracer v.1.6 n.d. https://bioweb.pasteur.fr/packages/pack@Tracer@v1.6[accessed 29 October 2022].


[59] Risom D, Behlin J, Nguyen TT, Jessuruh A, Schurr TG. Evolution and differential expression of mitochondrial DNA haplogroup US in Northern Europe: insights...