Comparative Genomics of Nonoutbreak *Pseudomonas aeruginosa* Strains Underlines Genome Plasticity and Geographic Relatedness of the Global Clone ST235

Holger Brüggemann¹, Leticia Busato Migliorini², Romario Oliveira de Sales², Paula Célia Mariko Koga³, Andrea Vieira de Souza², Anders Jensen¹, Anja Poehlein⁴, Elzbieta Brzuszkiewicz⁴, Andre Mario Doi³, Jacyr Pasternak³, Marines Dalla Valle Martin³, and Patricia Severino²,*

¹Department of Biomedicine, Aarhus University, Denmark
²Albert Einstein Research and Education Institute, Hospital Israelita Albert Einstein, Sao Paulo, Brazil
³Laboratorio Clínico, Hospital Israelita Albert Einstein, Sao Paulo, Brazil
⁴Department of Genomic and Applied Microbiology, Institute of Microbiology and Genetics, Georg-August University Göttingen, Germany

*Corresponding author: E-mail: patricia.severino@einstein.br.

Accepted: June 29, 2018

Data deposition: This project has been deposited at GenBank under the accession numbers QBET00000000, QBES00000000, QBER00000000, QBEOQ00000000, QBEP00000000, QBEO00000000, QBEN00000000, QBEM00000000, QBEL00000000, QBEM00000000, QBEK00000000, QBEL00000000, QBEO00000000, QBEB00000000.

Abstract

*Pseudomonas aeruginosa* is an important opportunistic pathogen in hospitals, responsible for various infections that are difficult to treat due to intrinsic and acquired antibiotic resistance. Here, 20 epidemiologically unrelated strains isolated from patients in a general hospital over a time period of two decades were analyzed using whole genome sequencing. The genomes were compared in order to assess the presence of a predominant clone or sequence type (ST). No clonal structure was identified, but core genome-based single nucleotide polymorphism (SNP) analysis distinguished two major, previously identified phylogenetic groups. Interestingly, most of the older strains isolated between 1994 and 1998 harbored *exoU*, encoding a cytotoxic phospholipase. In contrast, most strains isolated between 2011 and 2016 were *exoU*-negative and phylogenetically very distinct from the older strains, suggesting a population shift of nosocomial *P. aeruginosa* over time. Three out of 20 strains were ST235 strains, a global high-risk clonal lineage; these carried several additional resistance determinants including *aac(6')Ib-cr* encoding an aminoglycoside N-acetyltransferase that confers resistance to fluoroquinolones. Core genome comparison with ST235 strains from other parts of the world showed that the three strains clustered together with other Brazilian/Argentinean isolates. Despite this regional relatedness, the individuality of each of the three ST235 strains was revealed by core genome-based SNPs and the presence of genomic islands in the accessory genome. Similarly, strain-specific characteristics were detected for the remaining strains, indicative of individual evolutionary histories and elevated genome plasticity.

Key words: hospital infection, phylogenetics, antibiotic resistance, high-risk clone, molecular epidemiology.

Introduction

*Pseudomonas aeruginosa* is a ubiquitous Gram-negative bacterium and one of the most predominant organisms associated with hospital-acquired infections (Vincent et al. 2009). Treatment of infected patients is further complicated due to the intrinsic resistance of this bacterium to therapeutics and its elaborated capacity to develop resistance to antibiotics through the selection of mutations in chromosomal genes and through horizontal gene transfer (Kulasekara et al. 2006; Kitao et al. 2012; Hong et al. 2015). *P. aeruginosa* response to environmental stress as well as its extensive metabolic capacity have been associated with the ability of the
bacterium to persist in clinical settings (Frimmersdorf et al. 2010; Klockgether et al. 2011).

Even though P. aeruginosa has a nonclonal population structure, multilocus sequence typing (MLST) has revealed that a few sequence types (STs) are widely distributed and have been called high-risk clones. ST235 is the most important clonal ST lineage associated with outbreaks and poor clinical outcomes (Hong et al. 2016; Miyoshi-Akiyama et al. 2017; Treepong et al. 2018). Some important features of ST235 are resistance to fluoroquinolones, aminoglycosides and beta-lactams and the presence of virulence factors such as a type III secretion system (Finck-Barbançon et al., 1997; Sato et al. 2003; Kulasekara et al. 2006). Additionally, through whole-genome sequencing (WGS) studies in recent years, detailed insights into the population structure of P. aeruginosa have been reported. It is now evident that the population consists of three phylogenetically distinct clades or groups (Stewart et al. 2014; Freschi et al. 2015). Most P. aeruginosa strains sequenced to date belong to group 1, including the widely used reference strain PAO1 (Stover et al. 2000). Group 2 strains include the ST235 clonal lineage, and are exoU-positive, in contrast to group 1 strains (Lee et al. 2006). Only few strains belong to group 3, including the reference strain PA7 (Roy et al. 2010).

The molecular epidemiology, resistance, and virulence properties of P. aeruginosa in nonoutbreak settings have received limited attention (Dettman et al. 2013; Hu et al. 2017). Here, we sequenced and analyzed epidemiologically unrelated P. aeruginosa strains, isolated from patients from a general hospital over a time period of two decades. We wanted to investigate to which degree these strains were related, if they belonged to a specific population group, clade or ST, and if detailed genome features of strains isolated at different times could indicate an environment pressure. Moreover, a comparison with global genomes was carried out in order to evaluate worldwide spread and molecular diversity of related clones.

Materials and Methods

The Clinical Laboratory from Hospital Israelita Albert Einstein, a Brazilian general hospital located in Sao Paulo, banks all isolates from normally sterile sites and all isolates obtained from documented nosocomial infections. Isolates are routinely identified to the species level using the Vitek 2 card system (Biomerieux, France). From this depository, 20 epidemiologically unrelated P. aeruginosa isolates collected over a period of 22 years were randomly selected for this study (table 1). Bacterial isolates were cultivated on Luria–Bertani broth for 18 h and genomic DNA was extracted using the DNeasy kit (Qiagen). Nextera XT shotgun libraries were prepared for the MiSeq (Illumina, San Diego, CA, USA) with a 300bp paired-end sequencing run. Libraries were prepared according to the manufacturer’s protocol at the Göttingen Genomics Laboratory, Germany. Raw reads were quality controlled with FastQC v0.11.2 (https://www.bioinformatics. babraham.ac.uk/projects/fastqc/; last accessed June 19, 2018), and subsequently trimmed using Trimmomatic 0.36 (http://www.usadellab.org/cms/? page=trimmomatic; last accessed June 19, 2018) to remove sequences with quality scores lower than 20 (Illumina 1.9 encoding) and remaining adaptor sequences. De novo assembly was performed using SPAdes v3.11 (Bankevich et al. 2012). Due to the Illumina sequencing technology used in this study, assembly problems related with repetitive sequences, including ribosomal RNA operons, transposases (often annotated as “mobile element protein”), and phage regions, are unavoidable. Low sequence read coverage at random positions is also often reported and may lead to assembly issues. We did not further investigate any particular gap or contig ends since our study focused on phylogenomic analyses. The ST of all strains analyzed in this work was determined from their whole genome sequence data using the multilocus sequencing typing (MLST) scheme (https://pubmlst.org/paeruginosa/; last accessed June 19, 2018) sited at the University of Oxford (Jolley and Maiden 2010).

Besides the 20 genomes sequenced in this study, we used all closed genomes of P. aeruginosa strains that were available from GenBank in December 2017 (in total, 124 available genomes), and 79 genomes of ST235 strains available in multiple contigs for phylogenomic analysis. Gene prediction and annotation of all genomes were done with RAST (Aziz et al. 2008). For the phylogenomic analysis, the core genome was aligned using Parsnp, a program that is part of the Harvest software package (Treangen et al. 2014). All sequences near the end of contigs (gaps) were excluded for this analyses in order to avoid using false SNPs for phylogenetic reconstruction (default settings in Parsnp). Phylogenetic trees were built in Mega v7 and visualized with iTOL (Kumar et al. 2016; Letunic and Bork 2016). To determine orthologous genes among the 20 sequenced P. aeruginosa strains we used Proteinortho (Lechner et al. 2011). As parameters the default settings of the program were used: BlastP e-value cutoff, 1e–05; minimum identity of best blast hits, 25%; minimum coverage of best blast hits, 50%. Genome comparison figures were generated using BLAST Ring Image Generator (BRIG) and based on BlastN analyses (Alikhan et al. 2011). Proteinortho and BRIG were used in combination to determine the accessory genome. For functional predictions and pathway analyses we used KEGG mapper (http://www.genome.jp/kegg/mapper.html; last accessed June 19, 2018).

Results and Discussion

WGS of 20 epidemiologically unrelated strains collected over a 20 year-period resulted in final genome assemblies with 23 to 124 contigs (average: 67) per strain; N50 and N90 values ranged from 212 to 1,331 kb (average: 366 kb) and 41 to 210 kb (average 93 kb), respectively (table 1). The GC-content of the strains varied between 65.7% and 66.5%,
and the genome size varied between 6,304 kbp and 6,970 kbp. The presence of different STs among the 20 strains illustrates the heterogeneity of the cohort.

The pan-proteome (all orthologous, core genome-based CDS and all CDS of the accessory genome) of the 20 strains compared with reference strains PA01 and UCBPP-PA14 comprises 7,874 CDS. The 22 strains share 4,844 orthologous CDS, representing the core functions as well as a high array of two-component systems (and proteins involved in quorum sensing, as predicted by KEGG pathway analyses (data not shown). In total, 117 CDS are strain-specific and 626 proteins are only found in maximal two genomes, representing the most diverse set of acquired resistance-associated genes, including the insertion sequences int1, tn7, and tn21 and also contained aac(6')Ib-cr encoding an aminoglycoside modifying N-acetyltransferase reported to confer resistance to fluoroquinolones (Jaillard et al. 2017). Taken together, these findings highlight how WGS of clinical isolates may contribute to the understanding of resistance markers spread, worldwide, through high-risk strains.

Noteworthy was also the widespread presence of the acquired blaOXA-51 and blaPAO genes, contributing to β-lactam resistance, and of the gene aph(3′)Iib, conferring resistance to aminoglycosides, in strains from both studied periods. The presence of genes encoding effectors secreted via the type III secretion system (exoS, exoT, exoU, and exoY) in clinical isolates of P. aeruginosa is variable. Here the three ST235 isolates were exoU-positive; three other STs, including a not previously reported one, also carried exoU. This gene encodes a major virulence factor, a phospholipase, which has been associated with poor clinical outcomes (Finck-Barbancon et al. 1997; Sato et al. 2003). Besides gene acquisition, the selection of chromosomal mutations is also a frequent occurrence associated with multidrug resistance in P. aeruginosa. Resistance to fluoroquinolones is mostly associated with the Quinolone-Resistance-Determining Region (QRDR) (Jaillard et al. 2017; del Barrio-Tofino et al. 2017). The three ST235 strains showed acquired mutations in QRDR, as also recently described in other Brazilian isolates (Araujo et al. 2016). The identification of antibiotic resistance markers in the genome correlates and often broadens observations obtained by traditional phenotypic assays (Jaillard et al. 2017). Taken
together, WGS helped in the identification of persistent resistance and virulence markers in *P. aeruginosa* strains within a clinical setting that may reflect antibiotic policies. As previously reported, *P. aeruginosa* population consists of three clades. Since most sequenced strains to date are isolates from human infections there could be a bias towards...
clinical isolates, but there is currently no convincing explanation for this specific population structure of *P. aeruginosa* (Hilker et al., 2015; Wiehlmann et al. 2015). It has been suggested that different environment habitats/niches and/or different host associations led to this evolutionary bifurcation (Freschi et al. 2015; Hilker et al. 2015; Wiehlmann et al. 2015). A phylogenetic analysis based on core genome-based SNPs separated the 20 *P. aeruginosa* strains sequenced in this study into these previously reported groups (fig. 1A). One strain, HIAE_PA16, is a hybrid strain with genomic regions from both clades; it belongs to the previously described group 3 that contains only few strains sequenced to date (Freschi et al. 2015). Additionally, this analysis showed that most of the older strains, isolated between 1994 and 1998, are phylogenetically distinct from strains isolated more recently (between 2011 and 2016). We were not able to find other reports that indicate this group shift, but we are also not aware of a study that has investigated nosocomial *P. aeruginosa* strains collected over a larger time period at the phylogenomic level.

In order to further explore this clade distribution of the here sequenced *P. aeruginosa* strains, we analyzed all available *P. aeruginosa* strains for which a closed genome sequence was available (fig. 1B). The total number of SNPs in the determined core genome (62% of the PAO1 reference genome) was 185,860. The analysis showed that almost all *P. aeruginosa* strains (98%) clustered into the above-mentioned group 1 and group 2, and three strains were considered possible hybrids of group 1 and group 2 genomes, belonging to the previously described group 3 (Mathee et al. 2008). There were few clade-specific genes present (data not shown). The presence of the genes *exoS* and *exoU* genes has been reported as mutually exclusive (Feltman et al. 2001). In fact, they were found only in group 1 and group 2 genomes, respectively. The gene *exoS* encodes the toxin exoenzyme S, which is an actin-ADP-ribosylating toxin, delivered by the type III secretion system (Engel and Balachandran. 2009).

We decided to have a closer look on group 2 since it contained strains belonging to ST235, a predominant global, high-risk clinical clonal lineage (Treepong et al. 2018). We wanted to know how closely related the three here sequenced ST235 strains are and how they relate to previously identified ST235 strains. In total, 79 genomes, for which metadata were available were used (Treepong et al. 2018). The respective strains originated from various geographic locations and were isolated between 1991 and 2014. The ST235 strains shared a core genome of 72% with a total number of 4,302 SNPs. The three Brazilian strains sequenced here clustered in close vicinity, revealing that they are closely related to each other and the closest relatives of these three strains were strains either isolated in Brazil or Argentina (fig. 1C). In order to visualize the core and the flexible genome of ST235 strains, a BRIG analysis was performed, using HIAE_PA17 as the reference genome (fig. 1D). Strain HIAE_PA17 contained an 85 kb genomic island (GI), inserted next to a tRNA (*trnA*-Gly) (fig. 1C). This GI has signatures of an integrating conjugative element (ICE), similar, but not identical, to other *P. aeruginosa* genomic islands (PAGIs), such as PAGI-2 and -3 (Kung et al. 2010); it encodes among many other functions a type IV secretion system (GI-type T4SS). There are two other GIs in this strain, carrying a GI-type T4SS, one of which is inserted next to a tRNA (*trnA*-Lys) and contains the *exoU* gene. The latter is also present in the other two ST235 genomes. Strain HIAE_PA20 possesses a strain-specific 36 kb prophage, and a second 37 kb prophage that is also absent in the other sequenced strains, except for HIAE_PA19.

Taken together, these results demonstrate the closer relationship between ST235 strains isolated in Brazil and Argentina when compared with ST235 strains isolated in other geographic locations, a noteworthy report since there is no epidemiological link between these isolates. However, the SNP differences in the core genome as well as the different genomic islands in the flexible genome underline the individuality of the three ST235 strains isolated in the same clinical setting and possibly emphasize the capacity of ST235 strains to quickly adapt and modify its genome according to its environment. In fact, WGS is becoming more accessible and the analysis of entire genomes may provide enough resolution for the discrimination of highly related lineages such as ST235 in routine screenings.

In conclusion, analysis of our strain cohort corroborates data of recent studies that have presented two major clades for the classification of *P. aeruginosa* population. Reasons for this distribution remain to be found. Each strain showed a distinct evolutionary history as judged from the many strain-specific core genome-based SNPs and the presence of strain-specific GIs. This illustrates the high diversification of the species *P. aeruginosa* in this clinical setting, including the three ST235 isolates. The three ST235 strains were highly similar to other ST235 isolates from Brazil and Argentina, but not with isolates from other parts of the world, a new insight into epidemiologically unrelated ST235 strains. WGS also allowed for the identification of resistance markers that have persisted in *P. aeruginosa* genomes during the past two decades, despite dissimilar antibiotic policies during this period.

**Supplementary Material**

Supplementary data are available at Genome Biology and Evolution online.

**Acknowledgments**

The authors thank Lise Hald Schultz for excellent technical assistance. This work was supported by funds of the Aarhus University to H.B. (No. 11544). L.B.M. and R.O.S. are supported by a Hospital Israelita Albert Einstein fellowship.
Literature Cited

Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics 12:402.

Araujo BF, et al. 2016. Clinical and molecular epidemiology of multidrug-resistant P. aeruginosa carrying aac(6’)-Ib-cr, qnrS1 and blsA5PM genes in Brazil. PLoS One 11(5):e0155914.

Aziz RK, et al. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9:75.

Bankewich A, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19(5):455–477.

Dettman JR, Rodrigue N, Aaron SD, Kassen R. 2013. Evolutionary genomics of epidemic and nonepidemic strains of Pseudomonas aeruginosa. Proc Natl Acad Sci U S A. 110(32):21065–21070.

Engel J, Balachandran P. 2009. Role of Pseudomonas aeruginosa type III effectors in disease. Curr Opin Microbiol. 12: 61–66.

Feltman H, et al. 2001. Prevalence of type III secretion genes in clinical and environmental isolates of Pseudomonas aeruginosa. Microbiology 147(10):2659–2669.

Finck-Barbançon V, et al. 1997. ExoU expression by Pseudomonas aeruginosa correlates with acute cytotoxicity and epithelial injury. Mol Microbiol. 25(3):547–557.

Freschi L, et al. 2015. Clinical utilization of genomics data produced by the international Pseudomonas aeruginosa consortium. Front Microbiol. 6:1036.

Frimmersdorf E, Horatzek S, Pelinkevich A, Weihlmann L, Schomburg D. 2010. How Pseudomonas aeruginosa adapts to various environments: a metabolomic approach. Environ Microbiol. 12(6):1734–1747.

Gerich D, Naas T, Nordmann P. 2004. Biochemical characterization of the naturally occurring oxacillinase OXA-50 of Pseudomonas aeruginosa. Antimicrob Agents Chemother. 48(6):2043–2048.

Hilker R, et al. 2015. Intercellular gradient of virulence in the Pseudomonas aeruginosa pangenome from disease and environment. Environ Microbiol. 17(1):29–46.

Hong DJ, et al. 2015. Epidemiology and characteristics of metallo-β-lactamase-producing Pseudomonas aeruginosa. Infect Chemother. 47(2):81–97.

Hong JS, Yoon EJ, Lee H, Jeong SH, Lee K. 2016. Clonal dissemination of Pseudomonas aeruginosa sequence type 235 carrying blaSPm-6 and emergence of blaQms-2 and blaQms-10 on novel genomic islands PAGI-15 And -16 in Korea. Antimicrob Agents Chemother. 60:7216–7223.

Hu P, et al. 2017. Molecular epidemiology, resistance, and virulence properties of Pseudomonas aeruginosa cross-colonization clonal isolates in the non-outbreak setting. Infect Genet Evol. 55:288–296.

Jalilard M, et al. 2017. Correlation between phenotypic antibiotic susceptibility and the resistome in Pseudomonas aeruginosa. Int J Antimicrob Agents 50(2):210–218.

Jolley KA, Maiden MC. 2010. BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 11:595.

Kita T, et al. 2012. Emergence of a novel multidrug-resistant Pseudomonas aeruginosa strain producing IMP-type metallo-β-lactamases and AAC(6’)-Ia in Japan. Int J Antimicrob Agents 39(6):518–521.

Klockgether J, Cramer N, Weihlmann L, Davenport C, Tümmler B. 2011. Pseudomonas aeruginosa genomic structure and diversity. Front Microbiol. 2:150.

Klockgether J, Tümmler B. 2017. Recent advances in understanding Pseudomonas aeruginosa as a pathogen. F1000Res 6:1261.

Köser CU, et al. 2012. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. PLoS Pathog. 8(8):e1002824.

Kulasekara BR, et al. 2006. Acquisition and evolution of the exolocus in Pseudomonas aeruginosa. J Bacteriol. 188(11):4037–4050.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 33(7):1870–1874.

Kung VL, Ozer EA, Hauser AR. 2010. The accessory genome of Pseudomonas aeruginosa. Microbiol Mol Biol Rev. 74(4):621–641.

Lechner M, et al. 2011. Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinformatics 12:124.

Lee DG, et al. 2006. Genomic analysis reveals that Pseudomonas aeruginosa virulence is combinatorial. Genome Biol. 7(10):R90.

Letunic I, Bork P. 2016. Interactive Tree Of Life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 44(W1):W242–W245.

Libisch B, et al. 2008. Identification of PER-1 extended-spectrum β-lactamase producing Pseudomonas aeruginosa clinical isolates of the international clonal complex CC11 from Hungary and Serbia. FEMS Immunol Med Microbiol., 54(3): 330–338.

Mathee K, et al. 2008. Dynamics of Pseudomonas aeruginosa genome evolution. Proc Natl Acad Sci U S A. 105(8):3100–3105.

Miyoishi-Akiyama T, et al. 2017. Emergence and spread of epidemic multidrug-resistant Pseudomonas aeruginosa. Genome Biol Evol. 9(12):3238–3245.

Rahme LG, et al. 1995. Common virulence factors for bacterial pathogenicity in plants and animals. Science 268(5219):1899–1902.

Ramirez MS, Nikolaidis N, Tolmasky ME. 2013. Rise and dissemination of aminoglycoside resistance: the aac(6’)-Ib paradigm. Front Microbiol. 4:121.

Roy RH, et al. 2010. Complete genome sequence of the multiresistant taxonomic outlier Pseudomonas aeruginosa PA7. PLoS One 5(1):e88842.

Sato H, et al. 2003. The mechanism of action of the Pseudomonas aeruginosa-encoded type III cytotoxin, ExoU. EMBO J. 22(12):2959–2969.

Stewart L, et al. 2014. Draft genomes of 12 host-adapted and environmental isolates of Pseudomonas aeruginosa and their positions in the core genome phylogeny. Pathog Dis. 71(1):20–55.

Stover CK, et al. 2000. Complete genome sequence of the multiresistant international clonal complex CC11 from Hungary and Serbia. FEMS Immunol Med Microbiol., 54(3): 330–338.

Treangen TJ, Ondov BD, Koren S, Phillippy AM. 2014. The Harvest suite for rapid core-genome alignment and visualization of thousands of intra-specific microbial genomes. Genome Biol. 15(11):524.

Treonpo P, et al. 2018. Global emergence of the widespread Pseudomonas aeruginosa ST235 clone. Clin Microbiol Infect. 24:258–266.

Vincent JL, et al. 2009. International study of the prevalence and outcomes of infection in intensive care units. JAMA 302(21):2323–2329.

Weihlmann L, et al. 2007. Population structure of Pseudomonas aeruginosa. Proc Natl Acad Sci U S A. 104(19):8101–8106.

Weihlmann L, Cramer N, Tümmler B. 2015. Habitat-associated skew of the Pseudomonas aeruginosa population. Environ Microbiol Rep. 7(6):955–960.

Zeng L, Jin S. 2003. aph(3')-IIb, a gene encoding an aminoglycoside-modifying enzyme, is under the positive control of surrogate regulator HpaA. Antimicrob Agents Chemother. 47(12):3867–3876.

Associate editor: Howard Ochman