BMC Genomic Data

DATA NOTE

Open Access

Complete genome sequence of *Pseudomonas aeruginosa* YK01, a sequence type 16 isolated from a patient with keratitis

Shuo Jiang[1†], Mengmin Ye[2†], Ke Liu[3], Huiluo Cao[1*] and Xiaoshan Lin[4*]

Abstract

Objectives *Pseudomonas aeruginosa*, a Gram-negative opportunistic pathogen, is frequently associated with multidrug resistance and global epidemic outbreaks, contributing significantly to morbidity and mortality in hospitalized patients. However, *P. aeruginosa* belonging to the sequence type (ST) 16 was rarely reported. Here, this report presents the complete genome sequence of YK01, a ST16 *P. aeruginosa* isolate from a patient with keratitis. The complete reference genome of *P. aeruginosa* YK01 is expected to provide valuable data for investigating its genomic population, enhancing understanding of genetic basis of *P. aeruginosa* species complex.

Data description A complete genome of 6.3 Mb was obtained for *P. aeruginosa* YK01 by combining Illumina 150-bp short reads and Nanopore long reads. The assembly is fully complete with chromosomal genome size of 6,183,266 bp, presenting a GC content of 66.7%, and a plasmid with the size of 46,067 bp, presenting GC content of 59.0%. Predicted chromosomal genomic features include 5,709 CDS, 12 rRNAs, 63 tRNAs, 4 ncRNAs, and 5,788 genes. To our knowledge, this genome data represents the first complete genome of *P. aeruginosa* ST16, providing crucial information for further comparative genome analysis.

Keywords *Pseudomonas aeruginosa*, PA ST16, Complete genome

Objective

P. aeruginosa poses a significant threat to hospitalized patients, contributing to high morbidity and mortality rates [1, 2]. The emergence of multidrug-resistant and extensively drug-resistant *P. aeruginosa* strains, particularly carbapenem-resistant isolates, has become a critical concern, prompting the World Health Organization (WHO) to elevate *P. aeruginosa* to high priority on the 2024 Bacterial Priority Pathogens List [3]. *P. aeruginosa* is a ubiquitous opportunistic pathogen causing diverse infections, ranging from pulmonary infections in cystic fibrosis patients to cutaneous and ocular manifestations [4]. Its role as primary etiological agent in bacterial keratitis, especially among contact lens users [5], is well-established, with studies reporting a worldwide infections with incidence ranging 6.8% to 55% [6]. While genomic analyses have illuminated

†Shuo Jiang and Mengmin Ye contributed equally to this work.

*Correspondence:
Huiluo Cao
hcao@hku.hk
Xiaoshan Lin
lxs13751339320@sina.com
[1] Department of Microbiology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China
[2] Department of Laboratory Medicine, The First Affiliated Hospital, Sun Yat-Sen University, Guangzhou, China
[3] Shenzhen Eye Hospital, Jinan University, Shenzhen Eye Institute, Shenzhen, China
[4] The Fifth Affiliated Hospital, Southern Medical University, Guangzhou, China

the characteristics of certain *P. aeruginosa* lineages, including sequence type (ST) 308 in India [7], ST235 [8], ST1203 in USA [9], the genomic landscape of ST16 remains largely unexplored.

While the PubMLST *Pseudomonas aeruginosa* multilocus sequence typing (MLST) scheme (accessed 13 September 2024) encompasses 5,007 STs, only six records of ST16 are currently documented. To enhance our understanding of the conservation and evolutionary biology of this rare ST, we present the complete genome sequence of a ST16 *P. aeruginosa* strain YK01. This study presents the complete genome sequence of a clinically isolated *P. aeruginosa* strain belonging to ST16, which will contribute to the limited understanding of ST16, a rarely described lineage within the *P. aeruginosa* species. This study addresses this gap by providing the first comprehensive genomic characterization of a *P. aeruginosa* ST16 isolate.

Data description

YK01 was isolated from a 51-year-old patient with bacterial keratitis in Shenzhen, China (DataFile 1). Bacterial isolation was performed using sheep blood agar plate. YK01 was initially identified as *P. aeruginosa* by MALDI-TOF mass spectrometric analysis. To obtain genomic DNA of YK01 for sequencing, a single colony was inoculated in lysogeny broth (LB) at 37 °C with 220-rpm agitation for 20 h. 1 mL of bacterial culture was collected, and the genomic DNA was extracted using the EasyPure® Genomic DNA Kit (TransGen, China). The short-read library was prepared using Nextera XT DNA Library Preparation Kit (Illumina, USA) and sequenced by an Illumina NextSeq 500 platform with 150-bp paired-end reads (Novogene, China). Pair end raw reads ($n = 7{,}728{,}918$; SRA: SRR30916324; DataFile 2) were accessed and trimmed with FastQC v0.12.1 [10] and Trimmomatic v0.39 [11], respectively. The long-read library was prepared with Ligation Sequencing Kit V14 (SQK-LSK114) and sequenced using R10.4.1 flow cell and Nanopore MinION sequencer Mk1B (Oxford Nanopore Technologies, UK). Raw nanopore sequencing data was base-called to generate long reads (SRA: SRR30916323; DataFile 3). A hybrid assembly (DataFile 4 and 5) was performed using Unicycle v0.5.0 [12], followed by genomic functional annotation using the NCBI Prokaryotic Genome Annotation Pipeline [13]. The resulting complete genome of strain YK01 comprising chromosomal sequence and one plasmid sequence, exhibited an average coverage of 346-fold as determined by QUAST v5.2.0 [14]. The completeness of genome was verified using CheckM v1.2.1 [15] (DataFile 6). De novo assembly of strain YK01, analyzed using GTDB-Tk v2.3.2 [16], confirmed the species identification of strain YK01 as *P. aeruginosa*, closely related to strain NCTC10332 (RefSeq: GCF_001457615.1) with an average nucleotide identity of 99.36% (DataFile 7).

YK01 harbors a 6,183,266 bp chromosome (DataFile 8) and a 46,067 bp plasmid, designated pYK01 (DataFile 9). The genome encodes 239 virulence factors, 27 biocide and metal resistance genes, and five antibiotic resistance genes, all located on the chromosome (DataFile 10). Comparative analysis with PAO1 (GenBank: AE004091.2) [17] and PA14 (GenBank: NC_008463.1) [18] revealed the absence of *wzy* and *wzz* genes in YK01, along with partial deletions within the *pil* and *pvd* gene clusters (DataFile 11). The genes associated with biofilm formation in YK01 are same to those in PAO1 and PA14, primarily encompassing the alginate synthesis genes (*alg* gene cluster) and quorum sensing genes (*las* and *rhl* gene clusters) (DataFile 11). Alignment of pYK01 with the closely related plasmid pP9Me2 (GenBank: CP118640) indicated the absence of *acr* genes, which encode proteins that inhibit the Type I-F CRISPR-Cas system [19] (DataFile 12). CRISPRCas-Finder [20] identified five CRISPR arrays and a Type I-F CAS system within YK01 (DataFile 13). Furthermore, PHASTEST [21] identified a prophage closely related to *Pseudomonas* phage YMC11/02/R656 (RefSeq: NC_028657.1). AntiSMASH [22] analysis revealed 16 secondary metabolite biosynthetic gene clusters (BGCs) with potential links to virulence factor expression (DataFile 13 and 8).

To establish the global context of strain YK01, a core-genome single nucleotide polymorphism (SNP) analysis was conducted. This analysis utilized 42 *P. aeruginosa* public genomes with associated metadata, retrieved from the NCBI database (DataFile 14), and employed Snippy v4.6.0 with the reference genome of strain PAO1 as a reference (DataFile 15) [23]. A maximum-likelihood phylogenetic tree was constructed using IQ-TREE v2.2.0.3 (DataFile 16), employing the GTR+I+G model and 1,000 ultrafast bootstrap replicates [24]. This tree, along with metadata including source, location, and virulence factors, was rooted at sequence GCA 034812525.1 and visualized using iTOL v 6.9.1 (DataFile 17) [25]. All scripts in this study were written into DataFile 18 Table 1.

Limitations

While this study presents the first complete genome sequence of *P. aeruginosa* ST16, the analyses are limited by the current scale of available complete genomic data for this specific ST. To overcome this limitation and provide a broader understanding of this lineage, future

Table 1 Overview of data files/data sets

Label	Name of Datafile/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
DataFile 1	Table S1, Health data of the patient	MS Excel file (.xlsx)	Figshare (https://doi.org/10.6084/m9.figshare.27907251) [26]
DataFile 2	Raw short Illumina sequencing reads 1 and 2	Sequence file (.fastq)	NCBI Sequence Read Archive (https://identifiers.org/ncbi/insdc.sra:SRR30916324) [27]
DataFile 3	Raw long nanopore sequencing reads	Sequence file (.fastq)	NCBI Sequence Read Archive (https://identifiers.org/ncbi/insdc.sra:SRR30916323) [28]
DataFile 4	Complete chromosomal genome and annotation of *P. aeruginosa* YK01	GenBank file (.gbk)	NCBI Sequence Read Archive (https://identifiers.org/ncbi/insdc:CP166913.1) [29]
DataFile 5	Complete plasmid genome and annotation of pYK01	GenBank file (.gbk)	NCBI Sequence Read Archive (https://identifiers.org/ncbi/insdc:CP166914.1) [30]
DataFile 6	Table S2, Quality control of assembly (QUAST and CheckM)	MS Excel file (.xlsx)	Figshare (https://doi.org/10.6084/m9.figshare.27168972.v3) [26]
DataFile 7	Table S3, Species identification of YK01 strain	MS Excel file (.xlsx)	Figshare (https://doi.org/10.6084/m9.figshare.27169149) [26]
DataFile 8	Figure S3, *P. aeruginosa* YK01 chromosomal genome plot	Portable Document Format file (.pdf)	Figshare (https://doi.org/10.6084/m9.figshare.27169251.v1) [26]
DataFile 9	Figure S4, *P. aeruginosa* YK01 plasmid genome plot	Portable Document Format file (.pdf)	Figshare (https://doi.org/10.6084/m9.figshare.27174903.v1) [26]
DataFile 10	Table S5, Summary of YK01 VFs, ARGs and antimetal resistance genes	MS Excel file (.xlsx)	Figshare (https://doi.org/10.6084/m9.figshare.27169194) [26]
DataFile 11	Figure S2, Comparative virulence of YK01, PAO1 and PA14	Portable Document Format file (.pdf)	Figshare (https://doi.org/10.6084/m9.figshare.27169245.v1) [26]
DataFile 12	Figure S5, *P. aeruginosa* YK01 plasmid alignments	Portable Document Format file (.pdf)	Figshare (https://doi.org/10.6084/m9.figshare.27174906.v1) [26]
DataFile 13	Table S6, Summary of YK01 CRISPR-cas system, prophage and biosynthetic gene clusters (BGCs)	MS Excel file (.xlsx)	Figshare (https://doi.org/10.6084/m9.figshare.27169200) [26]
DataFile 14	Table S4, *P. aeruginosa* ST16 metadata from NCBI database	MS Excel file (.xlsx)	Figshare (https://doi.org/10.6084/m9.figshare.27169191) [26]
DataFile 15	Table S7, Tab-separated columnar list of core SNP sites using PAO1 (AE004091.2) as a reference genome	Text file (.txt)	Figshare (https://doi.org/10.6084/m9.figshare.27898566) [26]
DataFile 16	MLTree S1, IQ-TREE – Maximum likelihood phylogenetic tree of *P. aeruginosa* ST16	Newick tree format (.txt)	Figshare (https://doi.org/10.6084/m9.figshare.27169227.v2) [26]
DataFile 17	Figure S1, Maximum likelihood phylogenetic tree of *P. aeruginosa* ST16	Portable Document Format file (.pdf)	Figshare (https://doi.org/10.6084/m9.figshare.27169233.v2) [26]
DataFile 18	Script S1, Summary of scripts of YK01	Text file (.txt)	Figshare (https://doi.org/10.6084/m9.figshare.27169269.v1) [26]

research should include a larger collection of complete ST16 genomes, potentially leveraging hybrid assembly approaches. Analysis of a broader dataset would facilitate a more robust investigation into the evolution of antibiotic resistance genes carried on plasmids and the comparative genomics of pathogenicity islands within the ST16 population.

Abbreviations

ST Sequence type
WHO World Health Organization
LB Lysogeny broth
MLST Multilocus sequence typing
BGC Biosynthetic gene cluster
SNP Single nucleotide polymorphism

Acknowledgements
The authors acknowledge Illumina and Oxford Nanopore Technologies for providing high-quality sequencing data.

Authors' contributions
SJ, HC and XL designed the study; SJ and MY performed experiments and genome analysis; KL provided research materials; HC acquired financial support; SJ and HC drafted the manuscript; HC and XL revised the manuscript.

Funding
This work was supported by the Health and Medical Research Fund (No. 19201901) from the Health Bureau of the Hong Kong Special Administrative Region, China, awarded to Huiluo Cao.

Data availability
The genomic data described in this Data note can be freely and openly accessed on GenBank of NCBI under Bioproject: PRJNA1145087. The results of genome analysis have been uploaded to the Figshare. Please see Table 1 for details and links to data.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 10 October 2024 Accepted: 10 January 2025
Published online: 22 January 2025

References

1. Raman G, Avendano EE, Chan J, Merchant S, Puzniak L. Risk factors for hospitalized patients with resistant or multidrug-resistant Pseudomonas aeruginosa infections: a systematic review and meta-analysis. Antimicrob Resist Infect Control. 2018;7:1–14.
2. Elfadadny A, Ragab RF, AlHarbi M, Badshah F, Ibáñez-Arancibia E, Farag A, et al. Antimicrobial resistance of Pseudomonas aeruginosa: navigating clinical impacts, current resistance trends, and innovations in breaking therapies. Front Microbiol. 2024;15:1374466.
3. Jesudason T. WHO publishes updated list of bacterial priority pathogens. Lancet Microbe. 2024;5(9):100940. https://www.thelancet.com/journals/lanmic/article/PIIS2666-5247(24)00193-9/fulltext
4. Malhotra S, Hayes Jr D, Wozniak DJ. Cystic fibrosis and Pseudomonas aeruginosa: the host-microbe interface. Clin Microbiol Rev. 2019;32(3). https://doi.org/10.1128/CMR.00138-18.
5. Gjerde H, Mishra A. Contact lens–related Pseudomonas aeruginosa keratitis in a 49-year-old woman. CMAJ. 2018;190(2):E54.
6. Teweldemedhin M, Gebreyesus H, Atsbaha AH, Asgedom SW, Saravanan M. Bacterial profile of ocular infections: a systematic review. BMC Ophthalmol. 2017;17:1–9.
7. Khan M, Willcox MD, Rice SA, Sharma S, Stapleton F. Development of antibiotic resistance in the ocular Pseudomonas aeruginosa clone ST308 over twenty years. Exp Eye Res. 2021;205:108504.
8. Fischer S, Dethlefsen S, Klockgether J, Tümmler B. Phenotypic and genomic comparison of the two most common ExoU-positive Pseudomonas aeruginosa clones, PA14 and ST235. Msystems. 2020;5(6). https://doi.org/10.1128/mSystems.01007-20.
9. CDC. Outbreak of extensively drug-resistant Pseudomonas aeruginosa associated with artificial tears. CDC; 2023. https://archive.cdc.gov/www_cdc_gov/hai/outbreaks/crpa-artificial-tears.html.
10. Simons A. A quality control tool for high throughput sequence data. A quality control tool for high throughput sequence data. 2010;10:f1000research.
11. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
12. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13(6):e1005595.
13. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, et al. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 2016;44(14):6614–24.
14. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29(8):1072–5.
15. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25(7):1043–55.
16. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics. 2020;36(6):1925–192. https://doi.org/10.1093/bioinformatics/btz848.
17. Stover C, Pham X, Erwin A, Mizoguchi S, Warrener P, Hickey M, et al. Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen. Nature. 2000;406(6799):959–64.
18. Mikkelsen H, McMullan R, Filloux A. The Pseudomonas aeruginosa reference strain PA14 displays increased virulence due to a mutation in ladS. PLoS One. 2011;6(12):e29113.
19. Davidson AR, Lu W-T, Stanley SY, Wang J, Mejdani M, Trost CN, et al. Anti-CRISPRs: protein inhibitors of CRISPR-Cas systems. Annu Rev Biochem. 2020;89(1):309–32.
20. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res. 2018;46(W1):W246–51.
21. Wishart DS, Han S, Saha S, Oler E, Peters H, Grant JR, et al. PHAST-EST: faster than PHASTER, better than PHAST. Nucleic Acids Res. 2023;51(W1):W443–50.
22. Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M, et al. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. Nucleic Acids Res. 2023;51(W1):W46–50.
23. Seemann T. Snippy: fast bacterial variant calling from NGS reads. 2015.
24. Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268–74.
25. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49(W1):W293–6.
26. Complete Genome of Pseudomonas aeruginosa YK01. Available from: https://figshare.com/projects/Pseudomonas_aeruginosa_YK1_ST16/222975.
27. Short reads of Pseudomonas aeruginosa YK01. NCBI Sequence Read Archive. Available from: https://identifiers.org/ncbi/insdc.sra:SRR30916324.
28. Long reads of Pseudomonas aeruginosa YK01. NCBI Sequence Read Archive. Available from: https://identifiers.org/ncbi/insdc.sra:SRR30916323.
29. Chromosomal genome and annotation of Pseudomonas aeruginosa YK01. NCBI Sequence Read Archive. Available from: https://identifiers.org/ncbi/insdc:CP166913.1.
30. Plasmid genome and annotation of Pseudomonas aeruginosa YK01. NCBI Sequence Read Archive. Available from: https://identifiers.org/ncbi/insdc:CP166914.1.

Publisher's Note