Data in Brief 57 (2024) 110930

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Draft genome sequence data of *Pythium cedri* Chen 4, the causal pathogen of deodar cedar root rot

Jin-Feng Peng[a], Yusufjon Gafforov[b,c], Jian Yu[a], Hong-Jun Yang[a], Yuan-Yuan Chen[d], Yi-Fan Xing[d], Jia-Jia Chen[a,*]

[a] *College of Landscape Architecture, Jiangsu Vocational College of Agriculture and Forestry, Zhenjiang 212400, China*
[b] *Central Asian Center for Development Studies, New Uzbekistan University, Tashkent 100007, Uzbekistan*
[c] *Mycology Laboratory, Institute of Botany, Academy of Sciences of Republic of Uzbekistan, Tashkent 100125, Uzbekistan*
[d] *College of Forestry, Henan Agricultural University, Zhengzhou, Henan, China*

A R T I C L E I N F O

Article history:
Received 16 July 2024
Revised 30 August 2024
Accepted 4 September 2024
Available online 10 September 2024

Dataset link: Whole-Genome Sequence Resource of Pythium cedri Chen 4 (Original data)

Keywords:
Deodar cedar tree
Cedar root rot
Oomycete pathogens
Genome sequence

A B S T R A C T

Pythium species are distributed globally, with certain members playing significant roles as plant and animal pathogens. *Pythium cedri* Chen 4 has been identified as a pathogenic isolate responsible for causing root rot on *Cedrus deodara*. Here, a comprehensive genome-wide sequence of *P. cedri* strain Chen 4 utilizing the Illumina NovaSeq sequencing platform and a Pacific Biosciences Sequel sequencing platform is presented. The genome of *P. cedri* strain Chen 4 was assembled into 150 contigs containing a combined size of 41.25 Mb, N50 value of 1,717,859 bp and N90 value of 431,829 bp. Genome annotation revealed 14,077 protein-encoding genes and 364 of the 1016 predicted proteins were putative effectors. The present work enriches the genetic resources of *P. cedri* for studying its evolution and can contribute to a better understanding of *P. cedri*–host interaction.

* Corresponding author.
E-mail address: jiajiachen@jsafc.edu.cn (J.-J. Chen).

https://doi.org/10.1016/j.dib.2024.110930

Specifications Table

Subject	Omics
Specific subject area	Oomycetes Plant Pathogen Genomics
Data format	Genome Sequencing, Raw data, De novo Assembly, Filtered Reads, Analyzed sequences
Type of data	Table
Data collection	*Pythium cedri* is a newly defined member of the *Pythium* genus and was originally isolated from roots of *Cedrus deodara* growing in Nanjing, Jiangsu Province, China [1]. A draft genome was sequenced on Illumina NovaSeq sequencing platform and a Pacific Biosciences (PacBio) Sequel sequencing platform.
Data accessibility	Repository name: NCBI Data identification number: BioSample number SAMN37179841 and BioProject number PRJNA1010138. This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JAXISP000000000. Direct URL to data: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1010138/ The data has been deposited in Genbank SRA archive found at NCBI, meeting their requirements for submission.

1. Value of the Data

- This is the first report of a whole genome sequence of *Pythium cedri* from China.
- The genome provided in the present study is necessary for species identification and phylogenetic relationship study of *Pythium cedri*.
- The genome sequences can improve the genetic information of *Pythium* species and provided reference information for the whole-genome assembly of other *Pythium* species.
- The whole-genome sequences can provide reference information for future population genetics studies of *Pythium* species.

2. Background

Cedrus deodara (Roxb. ex D.Don) G.Don is a large evergreen, coniferous tree in the Pinaceae family with a cone-shaped crown and whorls of green needles. It is the most commercially important, widely distributed, and one of the world's most important timber woody species [2]. However, the frequent cedar root rot occurrence causes the death of young cedar seedlings in the nursery and many trees over ten years old, seriously affecting the green landscape and the forestry trade [3,4]. It has been reported that cedar root rot is often caused by *Phytophthora* spp. [4,5] and *Fusarium* spp. [6]. *Pythium cedri* strain Chen 4 was initially isolated from infected *C. deodara* rot [1]. The identification of *P. cedri* strain Chen 4 was accomplished through a combination of morphological and molecular analyses, with the construction of a phylogenetic tree based on ITS+*Cox1* sequences. Research on cedar root rot has focused on fungicide-based control in the fields [7], with only a few studies on the *P. cedri* infection process and pathogenicity mechanism in cedar [4]. This study presents a high quality genome sequence of *P. cedri* strain Chen 4 to address this knowledge gap.

3. Data Description

The *P. cedri* Chen 4 genome was sequenced, BioSample accession: SAMN37179841, BioProject ID PRJNA1010138. The data are at the URL: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1010138. Illumina NovaSeq sequencing platform and a Pacific Biosciences (PacBio) Sequel

Table 1
Summarized genome information of the *Pythium cedri* Chen-4.

Assembly statistics	Statistics
Number of Scaffolds	150
Assembly length (Mb)	41.25
Largest scaffold size (bp)	10,671
N_{50} length (bp)	1717,859
N_{90} length (bp)	431,829
GC (%)	51.94
Complete BUSCOs	86.2 %
Complete and single-copy	82.1 %
Complete and duplicated	4.1 %
Fragmented	1.7 %
Missing	12.1 %
Protein encoding genes	
Number of predicted genes	14,077
Number of predicted secreted proteins	1016
Number of predicted effector proteins	364
Number of Predicted cytoplasmic effectors	182
Number of Predicted apoplastic effectors	182

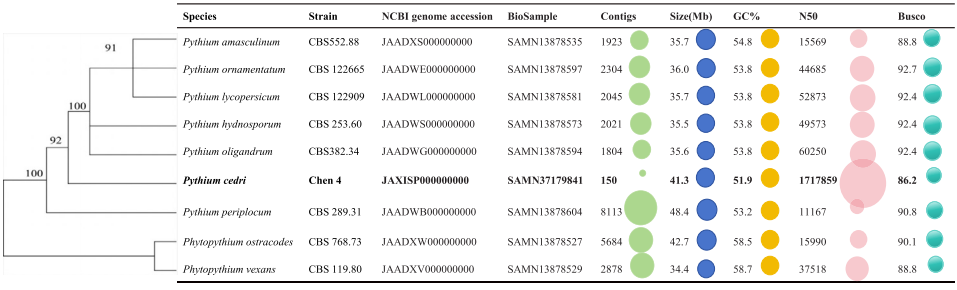

Fig. 1. Comparative analysis between the newly sequenced genome of *Pythium cedri* and a selection of closely related species publicly available. The *P. cedri* genome is highlighted in bold. On the left side, multilocus sequence typing (MLST) tree based on the concatenation of the partial sequences of following loci: ITS, LSU, *Cox1, Cox2* and *β-tubulin* (Table S1). Bubble plots report on assembly fragmentation, genome size, N50, GC content and completeness. Bubble sizes have been scaled to each panel and are not comparable across panels.

sequencing platform were used. Before the assembly, the *P. cedri* Chen 4 genome size was estimated to be 41.07 Mb (Table 1). The assembled genome was 41.25 Mb, with 150 contigs, the N50 value was 1717,859 bp, the N90 value was 431,829 bp, the GC content was 51.94 %, and the maximum scaffold size was 2917,360 bp (Table 1). Approximately 86.2 % of the orthologs were included in the assembled genome and a total of 14,077 protein-coding genes were generated. A total of 1016 were estimated to be secreted proteins in *P. cedri*, among those 364 have been generated to be candidate effectors (Table 1). A comparative analysis of the newly sequenced genome with those publicly available [9] revealed similar genomic features within closely related species (Fig. 1, Table S1).

4. Experimental Design, Materials and Methods

In our previous study, *P. cedri* strain Chen 4 was a newly defined member of the *Pythium* genus and originally isolated from roots of *Cedrus deodara* growing in Nanjing, Jiangsu Province, China [1]. Strain Chen 4 was cultured in corn meal agar (CMA) medium at 25 °C for 2–3 days. Then, mycelial growth was used for DNA extraction using a CTAB rapid plant genome extraction kit (Aidlab Biotechnologies Co., Ltd, Beijing) following the manufacturer's protocols. The quality

and quantity of the total DNA was evaluated by Nanodrop One, 0.38% w/v agarose gel electrophoresis and Qubit 3.0 Fluorometer (Thermo Fisher Scientific, USA), respectively. Then, the DNA was transferred to Nanjing Personalbio Co., Ltd. (Nanjing, China) to perform the genome sequencing. The sequencing errors [8–10] were corrected using the program Adapter Removal (version 2.0) and SOAPec (v 2.0), with a k-mers size of 17 as the default parameter. Subsequently, the Falcon tool assembled the *P. cedri* Chen 4 genome using data from the PacBio sequencing platform. The assembled contigs [11] were then polished using Illumina NovaSeq sequencing data in pilon v 1.18, and the gaps in the genome were filled using GapCloser. The genome assembly quality and the orthologs [12] was assessed by Benchmarking Universal Single-Copy Orthologs (v 3.0.2). Gene prediction was performed using Augustus (v 3.0.3), glimmerHMM (v 3.0.1), and GeneMark-ES (v 4.3.5). Subsequently, the protein sequences of closely related species were compared and homologous genes were predicted using Exonerate software (v 2.2.0). Protein-coding genes were generated using Evidence Modeler (v r2012-06-25) [13]. SignalP (v 5.0) and EffectorP (v 3.0) were used to predicted secreted proteins [14] and candidate effectors [15].

Limitations

Not applicable.

Ethics Statement

The current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

CRediT Author Statement

CJJ: conceived the project. **PJF, YG, YJ and CJJ:** performed the experiments. **PJF, YG,YJ, YHJ, CYY and CJJ:** analyzed the data. **PJF, XYF and CJJ:** wrote the manuscript. All authors read and approved the final manuscript.

Data Availability

The details of the Chen-4 isolate genome are available under BioSample number SAMN37179841 and BioProject number PRJNA1010138.

Whole-Genome Sequence Resource of Pythium cedri Chen 4 (Original data) (NCBI).

Acknowledgments

The research was supported by the National Natural Science Foundation of China (32070022), the Science Fund of the Jiangsu Vocational College of Agriculture and Forestry (2023kj16 & 2021kj91), Natural Science Foundation of the Jiangsu Higher Education Institutions of China (24KJA180003) and State Scientific and Technical Program of Institute of Botany of Uzbekistan Academy of Sciences (2021–2025).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary Materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2024.110930.

References

[1] J.J. Chen, L. Li, W.W. Ye, Y.C. Wang, X.B. Zheng, *Pythium cedri* sp. nov. (pythiaceae, pythiales) from southern china based on morphological and molecular characters, Phytotaxa 309 (2017) 135–142.
[2] D. Sinha, A review on phytochemical, ethnobotanical, pharmacological, and antimicrobial importance of *Cedrus deodara* (Roxb. Ex D. Don) G. Don, Int. J. Green Pharm. 13 (2019) 13.
[3] T. Doğmuş-Lehtijärvi, A.G.A. Kaya, A. Lehtijärvi, T. Jung, First report of *Phytophthora syringae* on *Cedrus libani* in Turkey, Plant Dis. 98 (2014) 846.
[4] C.G. McCarthy, D.A. Fitzpatrick, Systematic search for evidence of interdomain horizontal gene transfer from prokaryotes to Oomycete lineages, mSphere 1 (2016) e00195-16.
[5] A. Szigethy, Z.Á. Nagy, A.M. Vettraino, A. Józsa, S.O. Cacciola, R. Faedda, J. Bakonyi, First report of *Phytophthora* × *pelgrandis* causing root rot and lower stem necrosis of common box, lavender and port-orford-cedar in Hungary, Plant Dis. 97 (2013) 152.
[6] J. Xu, X. Yang, C. Wu, Z. Chen, T. Dai, Recombinase polymerase amplification-lateral flow dipstick assay for rapid detection of *Fusarium circinatum* based on a newly identified unique target gene, Plant Dis. 107 (2023) 1067–1074.
[7] D.H. Phillips, Fungicides in forestry in great Britain, J. Sci. Food Agr. 20 (2010) 503–504.
[8] M. Schubert, S. Lindgreen, L. Orlando, AdapterRemoval v2: rapid adapter trimming, identification, and read merging, BMC Res. Notes 9 (2016) 88.
[9] H.D.T. Nguyen, A. Dodge, K. Dadej, T.L. Rintoul, E. Ponomareva, F.N. Martin, A.W.A.M de Cock, C.A. Lévesque, S.A. Redhead, C.F.J. Spies, Whole genome sequencing and phylogenomic analysis show support for the splitting of genus *Pythium*, Mycologia 114 (2022) 501–515.
[10] B. Liu, C.M. Liu, D. Li, Y. Li, H.F. Ting, S.M. Yiu, R. Luo, T.W. Lam, BASE: a practical de novo assembler for large genomes using long NGS reads, BMC Genomics 17 (2016) 499.
[11] B.J. Walker, T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C.A. Cuomo, Q. Zeng, J. Wortman, S.K. Young, A.M. Earl, Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement, PLoS ONE 9 (2014) e112963.
[12] F.A. Simão, R.M. Waterhouse, P. Ioannidis, E.V. Kriventseva, E.M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs, Bioinformatics 31 (2015) 3210–3212.
[13] B.J. Haas, S.L. Salzberg, W. Zhu, M. Pertea, J.E. Allen, J. Orvis, O. White, C.R. Buell, J.R. Wortman, Automated eukaryotic gene structure annotation using evidencemodeler and the program to assemble spliced alignments, Genome Biol. 9 (2008) R7.
[14] J.J. Almagro Armenteros, K.D. Tsirigos, C.K. Sønderby, T.N. Petersen, O. Winther, S. Brunak, G. von Heijne, H. Nielsen, SignalP 5.0 improves signal peptide predictions using deep neural networks, Nat. Biotechnol. 37 (2019) 420–423.
[15] J. Sperschneider, P.N. Dodds, EffectorP 3.0: prediction of apoplastic and cytoplasmic effectors in fungi and oomycetes, Mol. Plant Microbe Interact. 35 (2022) 146–156.