BMC Research Notes

DATA NOTE

Open Access

Genome data of four *Pythium insidiosum* strains from the phylogenetically-distinct clades I, II, and III

Theerapong Krajaejun[1*] , Weerayuth Kittichotirat[2*], Preecha Patumcharoenpol[3], Thidarat Rujirawat[4], Tassanee Lohnoo[4] and Wanta Yingyong[4]

Abstract

Objectives: We employed the Illumina NGS platform to sequence genomes of 4 different strains of the pathogenic oomycete *Pythium insidiosum*, the causative agent of pythiosis. These strains were isolated from humans in Thailand (n = 3) and the United States (n = 1), and phylogenetically classified into clade-I, -II, and -III. Our study augmented the completeness of the *P. insidiosum* genome database for exploration of the biology, evolution, and pathogenesis of the pathogen.

Data description: One paired-end library (180-bp insert) was prepared from a gDNA sample of *P. insidiosum* strains ATCC200269 (clade-I), Pi19 (clade-II), MCC18 (clade-II), and SIMI4763 (clade-III) for whole-genome sequencing by Illumina HiSeq2000/HiSeq2500 NGS platform. A range of 28.4–59.4 million raw reads, accounted for 3.0–7.3 Gb, were obtained and assembled into the genome sizes of 47.1 Mb (15,153 contigs; 85% completeness; 19,329 open reading frames [ORFs]) for strain ATCC200269, 35.4 Mb (14,576 contigs; 83% completeness; 13,895 ORFs) for strain Pi19, 34.5 Mb (11,084 contigs; 84% completeness; 13,249 ORFs) for strain MCC18, and 47.1 Mb (15,162 contigs; 85% completeness; 19,340 ORFs) for strain SIMI4763. The genome data can be downloaded from the NCBI/DDBJ data-bases under the accessions BCFN00000000.1 (ATCC200269), BCFS00000000.1 (Pi19), BCFT00000000.1 (MCC18), and BCFU00000000.1 (SIMI4763).

Keywords: *Pythium insidiosum*, Pythiosis, Genome sequence, Next-generation sequencing

Objective

Next-generation sequencing (NGS) is a sophisticated technology that facilitates multiple genome sequencing of different strains of the same microbial species, in a short duration, and at a low cost [1]. Obtained data promise extensive comparative genomic analyses to better understand the biology, evolution, and pathogenesis

of a pathogen of interest. Besides, such data could serve as a comprehensive genetic resource for the identification of diagnostic and therapeutic microbial markers. Here, we employed the Illumina HiSeq2000/HiSeq2500 NGS platform to sequence the genomes of 4 different strains (i.e., ATCC200269, Pi19, MCC18, and SIMI4763) of *Pythium insidiosum*, a prominent pathogenic oomycete microorganism that infects humans and animals worldwide and causes an infectious condition with high mortality and morbidity, called pythiosis [2–4]. These strains were isolated from human patients with pythiosis from Thailand (n = 3) and the United States (n = 1), and have been phylogenetically classified into clade-I (n = 1), clade-II (n = 2), and clade-III (n = 1), based on

*Correspondence: mr_en@hotmail.com; weerayuth.kit@kmutt.ac.th
[1] Department of Pathology, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand
[2] Systems Biology and Bioinformatics Research Group, Pilot Plant Development and Training Institute, King Mongkut's University of Technology Thonburi, Bangkhuntien, Bangkok, Thailand
Full list of author information is available at the end of the article

Krajaejun *et al. BMC Res Notes*   (2021) 14:197

Page 2 of 4

the ribosomal deoxyribonucleic acid (rDNA) sequence analysis [5]. So far, the draft genome sequences from 7 strains of *P. insidiosum* (including the synonym species *Pythium destruens*), isolated from humans, horses, and the environment in various countries, are available in the public databases [6–12]. This study contributed additional genomic data to augment the completeness of the public *P. insidiosum* genome database. Researchers around the world can use this genome data as a basis to explore the biology, evolution, and pathogenesis of *P. insidiosum*, which could provide knowledge that can be adapted for the development of preventive measures, reliable diagnostic assay, and effective therapeutic modality for pythiosis.

Data description

The *P. insidiosum* strain ATCC200269 (phylogenetic clade-I) was isolated from a human patient in the United States, while the strains Pi19 (clade-II), MCC18 (clade-II), and SIMI4763 (clade-III) were isolated from human patients in Thailand. The identity (i.e., species) and genotype (i.e., clade) of each strain were confirmed by the rDNA sequence analysis [accession numbers: AB898108 (for strain ATCC200269), AB898113 (Pi19), AB971183 (MCC18), and AB971189 (SIMI4763)] [5]. These organisms were cultured in Sabouraud dextrose broth with shaking (50–150 rounds per min) for one week at 37 °C. The resulting hyphal material of each strain was harvested and subjected to genomic deoxyribonucleic acid (gDNA) extraction, using an established method [13]. The identity of each strain was re-assessed by the rDNA sequence analysis, using the obtained gDNA [5]. One paired-end library with a 180-bp gap was prepared for each gDNA sample before proceeding to whole-genome sequencing by the Illumina HiSeq2000 (for strains Pi19 and MCC18) and HiSeq2500 (for strains ATCC200269 and SIMI4763) NGS platforms (Yourgene Bioscience, Taiwan), as previously described [6, 7, 10, 12]. In brief, the Qiagen CLC Genomics Workbench software trimmed raw reads to ensure a read length of at least 35 bases. Cutadapt 1.8.1 [14] removed the adaptor sequences from all reads. A total of 59,442,302 raw reads (average length: 122.2 bases) from the strain ATCC200269; 30,517,195 raw reads (average length: 92.5 bases) from the strain Pi19; 28,443,839 raw reads (average length: 94.7 bases) from the strain MCC18; and 28,531,434 raw reads (average length: 122.3 bases) from the strain SIMI4763 were obtained. Velvet 1.2.10 [15] assembled the raw reads of the strain ATCC200269 into 15,153 contigs [average length: 3111.1 (range: 300–182,581); N50: 11,266; total bases: 47,142,494; %N: 0.7%; genome coverage: 154×]; the strain Pi19 into 14,576 contigs [average length: 2426.8 (range: 300–111,336); N50: 6208; total bases: 35,372,432; %N: 2.4%; genome coverage: 91×]; the strain MCC18 into 11,084 contigs [average length: 3116.3 (range: 300–150,908); N50: 8946; total bases: 34,541,218; %N: 2.3%; genome coverage: 87×]; and the strain SIMI4763 into 15,162 contigs [average length: 3109.2 (range: 300–182,337); N50: 11,187; total bases: 47,141,692; %N: 0.7%; genome coverage: 74×]. BLAST search analyses of the assembled sequences of the strains ATCC200269, Pi19, MCC18 and SIMI4763, using the "Core Eukaryotic Genes Mapping Approach (CEGMA)" panel (containing 248 highly-conserved eukaryotic genes) [16] demonstrated 85%, 83%, 84%, and 85% genome completeness, respectively. MAKER2 pipeline [17] assigned 19,329; 13,895; 13,249 and 19,340 open reading frames (ORFs) in the genomes of the strains ATCC200269, Pi19, MCC18 and SIMI4763, respectively. All contig sequences have been deposited in the National Center for Biotechnology Information (NCBI) and DNA Data Bank of Japan (DDBJ) databases under the accessions BCFN00000000.1 (for strain ATCC200269), BCFS00000000.1 (Pi19), BCFT00000000.1 (MCC18), and BCFU00000000.1 (SIMI4763) (Table 1).

Table 1 Overview of data files/data sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	*Pythium insidiosum* strain ATCC200269, whole genome shotgun sequencing project	FASTA	GenBank (https://identifiers.org/ncbi/insdc:BCFN00000000.1) [18]
Data file 2	*Pythium insidiosum* strain Pi19, whole genome shotgun sequencing project	FASTA	GenBank (https://identifiers.org/ncbi/insdc:BCFS00000000.1) [19]
Data file 3	*Pythium insidiosum* strain MCC18, whole genome shotgun sequencing project	FASTA	GenBank (https://identifiers.org/ncbi/insdc:BCFT00000000.1) [20]
Data file 4	*Pythium insidiosum* strain SIMI4763, whole genome shotgun sequencing project	FASTA	GenBank (https://identifiers.org/ncbi/insdc:BCFU00000000.1) [21]

Krajaejun *et al. BMC Res Notes* (2021) 14:197

Page 3 of 4

In summary, the draft genomes of *P. insidiosum* strains ATCC200269 (genome size: 47.1 Mb), Pi19 (35.4 Mb), MCC18 (34.5 Mb), and SIMI4763 (47.1 Mb) isolated from human patients with pythiosis living in Thailand and the United States, have been generated and publicly available. The obtained genome data could be a useful dataset to enhance the exploration of the biology, evolution, and pathogenesis of *P. insidiosum*, which can lead to clinical applications for better management of patients with pythiosis.

Limitations

We used the Illumina HiSeq2000/HiSeq2500 short-read NGS platform to sequence 4 genomes of *P. insidiosum* (strains ATCC200269, Pi19, MCC18, and SIMI4763). Users of the genome data should be aware that the sequencing-by-synthesis technique in the Illumina platforms constructs a library base on DNA amplification, which could result in sequence coverage biases and substitution errors. As seen in the genome data of these *P. insidiosum* strains, the total bases ranged from 3.0 to 7.3 Gb, and the genome sequence coverages ranged from 74× to 154×. Another limitation of the study is the number and type of the DNA library. The genome sequences of each *P. insidiosum* strain were obtained from only one paired-end library. As expected, all strains showed a less complete genome (83–85% CEGMA-based genome completeness), a higher number of contigs (11,084–15,162 contigs), and a smaller genome size (34.5–47.1 Mb), when compared with the *P. insidiosum's* reference genome (92% completeness; 1192 contigs; 53.2-Mb size) generated from one paired-end and three mate-pair libraries [8].

Abbreviations
CEGMA: Core Eukaryotic Genes Mapping Approach; DDBJ: DNA Data Bank of Japan; gDNA: Genomic deoxyribonucleic acid; NCBI: National Center for Biotechnology Information; NGS: Next-generation sequencing; ORF: Open reading frame; rDNA: Ribosomal deoxyribonucleic acid.

Authors' contributions
W.K. and T.K. conceived the project. W.K., P.P., T.R., T.L., and W.Y. performed the experiments. W.K., P.P., T.R., and T.K. analyzed the data. W.K. and T.K. wrote the manuscript. All authors reviewed the manuscript. W.K. and T.K. acquired the research funds.

Availability of data and materials
Please see Table 1 and references [18–21] for details and links to the data. The draft genome sequence of the *P. insidiosum* strain ATCC200269 comprising 15,153 contigs (accession numbers BCFN01000001-BCFN01015153), is available in GenBank here: https://identifiers.org/ncbi/insdc:BCFN00000000.1 [18].
The draft genome sequence of the *P. insidiosum* strain Pi19 comprising 14,576 contigs (accession numbers BCFS01000001-BCFS01014576), is available in GenBank here: https://identifiers.org/ncbi/insdc:BCFS00000000.1 [19].
The draft genome sequence of the *P. insidiosum* strain MCC18 comprising 11,084 contigs (accession numbers BCFT01000001-BCFT01011084), is available in GenBank here: https://identifiers.org/ncbi/insdc:BCFT00000000.1 [20].
The draft genome sequence of the *P. insidiosum* strain SIMI4763 comprising 15,162 contigs (accession numbers BCFU01000001-BCFU01015162), is available in GenBank here: https://identifiers.org/ncbi/insdc:BCFU00000000.1 [21].

Declarations

Ethics approval and consent to participate
This study was approved by the Human Research Ethics Committee, Faculty of Medicine, Ramathibodi Hospital, Mahidol University (approval numbers: MURA2020/966).

Consent for publication
Not applicable.

Competing interests
None.

Author details
[1]Department of Pathology, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand. [2]Systems Biology and Bioinformatics Research Group, Pilot Plant Development and Training Institute, King Mongkut's University of Technology Thonburi, Bangkhuntien, Bangkok, Thailand. [3]Interdisciplinary Graduate Program in Bioscience, Faculty of Science, Kasetsart University, Bangkok, Thailand. [4]Research Center, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand.

References
1. Kittichotirat W, Krajaejun T. Application of genome sequencing to study infectious diseases. J Infect Dis Antimicrob Agents. 2019;36:47–58.
2. Gaastra W, Lipman LJ, De Cock AW, Exel TK, Pegge RB, Scheurwater J, et al. *Pythium insidiosum*: an overview. Vet Microbiol. 2010;146:1–16.
3. Krajaejun T, Sathapatayavongs B, Pracharktam R, Nitiyanant P, Leelachaikul P, Wanachiwanawin W, et al. Clinical and epidemiological analyses of human pythiosis in Thailand. Clin Infect Dis. 2006;43:569–76.
4. Chitasombat MN, Jongkhajornpong P, Lekhanont K, Krajaejun T. Recent update in diagnosis and treatment of human pythiosis. PeerJ. 2020;8:e8555.
5. Rujirawat T, Sridapan T, Lohnoo T, Yingyong W, Kumsang Y, Sae-Chew P, et al. Single nucleotide polymorphism-based multiplex PCR for identification and genotyping of the oomycete *Pythium insidiosum* from humans, animals and the environment. Infect Genet Evol. 2017;54:429–36.
6. Kittichotirat W, Patumcharoenpol P, Rujirawat T, Lohnoo T, Yingyong W, Krajaejun T. Draft genome and sequence variant data of the oomycete *Pythium insidiosum* strain Pi45 from the phylogenetically-distinct Clade-III. Data Brief. 2017;15:896–900.

Krajaejun *et al. BMC Res Notes*   (2021) 14:197

Page 4 of 4

7. Patumcharoenpol P, Rujirawat T, Lohnoo T, Yingyong W, Vanittanakom N, Kittichotirat W, et al. Draft genome sequences of the oomycete Pythium insidiosum strain CBS 573.85 from a horse with pythiosis and strain CR02 from the environment. Data Brief. 2018;16:47–50.

8. Rujirawat T, Patumcharoenpol P, Lohnoo T, Yingyong W, Lerksuthirat T, Tangphatsornruang S, et al. Draft genome sequence of the pathogenic oomycete *Pythium insidiosum* Strain Pi-S, isolated from a patient with pythiosis. Genome Announc. 2015;3:e00574-e615.

9. Ascunce MS, Huguet-Tapia JC, Braun EL, Ortiz-Urquiza A, Keyhani NO, Goss EM. Whole genome sequence of the emerging oomycete pathogen *Pythium insidiosum* strain CDC-B5653 isolated from an infected human in the USA. Genomics Data. 2016;7:60–1.

10. Krajaejun T, Kittichotirat W, Patumcharoenpol P, Rujirawat T, Lohnoo T, Yingyong W. Data on whole genome sequencing of the oomycete *Pythium insidiosum* strain CBS 101555 from a horse with pythiosis in Brazil. BMC Res Notes. 2018;11:880.

11. Rujirawat T, Patumcharoenpol P, Lohnoo T, Yingyong W, Kumsang Y, Payattikul P, et al. Probing the phylogenomics and putative pathogenicity genes of *Pythium insidiosum* by oomycete genome analyses. Sci Rep. 2018;8:4135.

12. Krajaejun T, Kittichotirat W, Patumcharoenpol P, Rujirawat T, Lohnoo T, Yingyong W. Draft genome sequence of the oomycete *Pythium destruens* strain ATCC 64221 from a horse with pythiosis in Australia. BMC Res Notes. 2020;13:329.

13. Lohnoo T, Jongruja N, Rujirawat T, Yingyon W, Lerksuthirat T, Nampoon U, et al. Efficiency comparison of three methods for extracting genomic DNA of the pathogenic oomycete *Pythium insidiosum*. J Med Assoc Thai. 2014;97:342–8.

14. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17:10.

15. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008;18:821–9.

16. Parra G, Bradnam K, Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics. 2007;23:1061–7.

17. Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinform. 2011;12:491.

18. Rujirawat T, Patumcharoenpol P, Kittichotirat W, Krajaejun T. *Pythium insidiosum* strain ATCC200269, whole genome shotgun sequencing project. GenBank. 2019. https://www.ncbi.nlm.nih.gov/nuccore/BCFN00000000.1.

19. Rujirawat T, Patumcharoenpol P, Kittichotirat W, Krajaejun T. *Pythium insidiosum* strain Pi19, whole genome shotgun sequencing project. GenBank. 2019. https://www.ncbi.nlm.nih.gov/nuccore/BCFS00000000.1.

20. Rujirawat T, Patumcharoenpol P, Kittichotirat W, Krajaejun T. *Pythium insidiosum* strain MCC18, whole genome shotgun sequencing project. GenBank. 2019. https://www.ncbi.nlm.nih.gov/nuccore/BCFT00000000.1.

21. Rujirawat T, Patumcharoenpol P, Kittichotirat W, Krajaejun T. *Pythium insidiosum* strain SIMI4763, whole genome shotgun sequencing project. GenBank. 2019. https://www.ncbi.nlm.nih.gov/nuccore/BCFU00000000.1.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.