



ELSEVIER

Contents lists available at ScienceDirect

Data in Brief

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

CrossMark

Data Article

Draft genome and sequence variant data of the oomycete *Pythium insidiosum* strain Pi45 from the phylogenetically-distinct Clade-III

Weerayuth Kittichotirat^a, Preecha Patumcharoenpol^{a,b},
Thidarat Rujirawat^c, Tassanee Lohnoo^c, Wanta Yingyong^c,
Theerapong Krajaejun^{d,*}

^a Systems Biology and Bioinformatics Research Group, Pilot Plant Development and Training Institute, King Mongkut's University of Technology Thonburi, Bangkhuntien, Bangkok, Thailand

^b Department of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, AR, USA

^c Research Center, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand

^d Department of Pathology, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand

ARTICLE INFO

Article history:

Received 5 September 2017

Received in revised form

4 October 2017

Accepted 18 October 2017

Available online 26 October 2017

Keywords:

Pythium insidiosum

Pythiosis

Draft genome

Sequence variant

ABSTRACT

Pythium insidiosum is a unique oomycete microorganism, capable of infecting humans and animals. The organism can be phylogenetically categorized into three distinct clades: Clade-I (strains from the Americas); Clade-II (strains from Asia and Australia), and Clade-III (strains from Thailand and the United States). Two draft genomes of the *P. insidiosum* Clade-I strain CDC-B5653 and Clade-II strain Pi-S are available in the public domain. The genome of *P. insidiosum* from the distinct Clade-III, which is distantly-related to the other two clades, is lacking. Here, we report the draft genome sequence of the *P. insidiosum* strain Pi45 (also known as MCC13; isolated from a Thai patient with pythiosis; accession numbers BCFM01000001-BCFM01017277) as a representative strain of the phylogenetically-distinct Clade-III. We also report a genome-scale data set of sequence variants (i.e., SNPs and INDELS) found in *P. insidiosum* (accessible online at the Mendeley database: <http://dx.doi.org/10.17632/r75799jy6c.1>).

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

* Corresponding author.

E-mail address: mr_en@hotmail.com (T. Krajaejun).

Specifications Table

Subject area	Biology
More specific subject area	Microbiology, Genomics
Type of data	Genome sequence, Sequence variants, Phylogenetic relationship
How data was acquired	IlluminaHiSeq 2500 Next Generation Sequencing Platform
Data format	Assembled genome sequence, Sequence variants [i.e., single-nucleotide polymorphisms (SNPs) and small insertions and deletions (INDELS)], Phylogenetic tree
Experimental factors	Genomic DNA was extracted from the <i>Pythium insidiosum</i> strain Pi45, which is categorized in the phylogenetically-distinct clade-III.
Experimental features	A rDNA-based phylogenetic tree of <i>P. insidiosum</i> was generated. Genome of the <i>P. insidiosum</i> strain Pi45 was sequenced and assembled. The reference genome sequence of the <i>P. insidiosum</i> strain Pi-S was mapped with sequence reads from the <i>P. insidiosum</i> strain Pi45 to identify SNPs and INDELS.
Data source location	The organism was isolated from a patient with pythiosis in Thailand.
Data accessibility	The draft genome sequence of the <i>P. insidiosum</i> strain Pi45 (also known as MCC13) has been deposited in the Data Bank of Japan (DDBJ) under the accession numbers: BCFM01000001-BCFM01017277. The sequence variant data (i.e., SNPs and INDELS) of the <i>P. insidiosum</i> strain Pi45 is accessible online at the Mendeley database (http://dx.doi.org/10.17632/r75799jy6c.1).

Value of the data

- The first draft genome sequence of a *P. insidiosum* strain from the rDNA-based phylogenetic-distinct clade-III is now available.
- Draft genome data of the *P. insidiosum* strain Pi45 will be valuable for comparative genomic studies of *Pythium* species and related oomycetes.
- Sequence variant data (i.e., SNPs and INDELS) will be applicable for identification of the organism, genetic polymorphism analyses, genotype-phenotype association studies, and epidemiological exploration.

1. Data

Pythium insidiosum is a member of the oomycetes, a unique group of fungus-like microorganisms belonging to the Kingdom Stramenopiles [1]. *P. insidiosum* is distinguished from other oomycetes by its capacity to infect humans and animals [1–3]. The infectious condition called ‘pythiosis’ caused by this organism usually leads to life-long disability or death in affected individuals [2–5]. Genome sequence is a powerful resource that can be used to explore an organism of interest at the molecular level. Two draft genomes of the *P. insidiosum* strains CDC-B5653 [6] and Pi-S [7] are available in the public domain. *P. insidiosum* can be divided into three distinct clades: Clade-I (strains from Americas); Clade-II (strains from Asia and Australia); and Clade-III (strains from Thailand and the United States) (Table 1; Fig. 1). The strain CDC-B5653 (labeled as Pi10) is placed in the Clade-I, whereas the strain Pi-S (labeled as Pi35) is placed in the Clade-II (Fig. 1). The genome of *P. insidiosum* from the distinct Clade-III, which is distantly-related to the other two clades, is lacking. Here, we report genome data of the *P. insidiosum* strain Pi45, isolated from a Thai patient and categorized as Clade-III (Fig. 1). We also report a genome-scale data set of sequence variants (i.e., SNPs and INDELS) found in *P. insidiosum*.

2. Experimental design, materials and methods

2.1. rDNA-based phylogenetic tree

rDNA sequences from the strain Pi45 and 17 other strains of *P. insidiosum* were retrieved from the NCBI database (Table 1). The rDNA sequence from *Pythium grandisporangium* (accession number:

Table 1

Eighteen strains of *Pythium insidiosum* used for generation of the rDNA-based phylogenetic tree. Strain identification numbers, reference numbers, host sources, geographic origins, assigned phylogenetically-distinct clades, and accession numbers of the rDNA sequences of all strains are summarized in the table. ** indicates the strains, including Pi45, where genome sequences are publicly available.

ID	Reference ID	Source	Country of origin	Clade	Accession number
Pi05	CBS 575.85	Equine	Costa Rica	I	AB971178
Pi06	CBS 574.85	Equine	Costa Rica	I	AB971179
Pi07	CBS 573.85	Equine	Costa Rica	I	AB971180
Pi08	CBS 580.85	Equine	Costa Rica	I	AB898107
Pi09	CBS 101555	Equine	Brazil	I	AB971181
Pi10*	ATCC 200269	Human	USA	I	AB898108
Pi35*	Pi-S	Human	Thailand	II	AB898124
Pi36	ATCC 64221	Equine	Australia	II	LC199883
Pi37	ATCC 28251	Equine	Papua New Guinea	II	LC199884
Pi38	CBS 101039	Human	India	II	AB898125
Pi39	CBS 702.83	Equine	Japan	II	LC199885
Pi42	CR02	Environment	Thailand	II	AB971184
Pi44	MCC 17	Human	Thailand	III	AB971185
Pi45*	MCC 13	Human	Thailand	III	AB971186
Pi46	SIMI 3306–44	Human	Thailand	III	AB971187
Pi47	SIMI 2921–45	Human	Thailand	III	AB971188
Pi48	SIMI 4763	Human	Thailand	III	AB971189
Pi50	ATCC 90586	Human	USA	III	AB971190

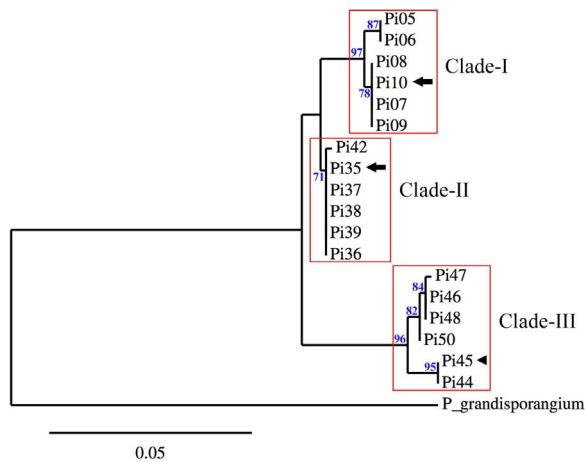


Fig. 1. Phylogenetic relationship of *Pythium insidiosum*: the rDNA-based maximum-likelihood phylogenetic tree categorizes 18 strains of *Pythium insidiosum* into three distinct clades: Clade-I, Clade-II, and Clade-III. Description of each strain of *P. insidiosum* can be found in Table 1. The arrows indicate the strains [i.e., CDC-B5653 (labeled as Pi10) and Pi-S (labeled as Pi35)] where genome sequences are publicly available, while the arrow head indicates the strain Pi45 where genome data is reported here. The rDNA sequence from *Pythium grandisporangium* (accession number: AY151182) is included as an outgroup. Branch support values of greater than 70% are demonstrated at the nodes. Nucleotide substitution per site is shown at the bottom.

AY151182) served as an outgroup. All rDNA sequences were subjected to phylogenetic analysis, using an array of online tools at www.phylogeny.fr [8–13].

2.2. Genome sequencing and assembly

Genomic DNA of the *P. insidiosum* strain Pi45 was extracted [14] and processed to prepare one paired-end library for genome sequencing, using the IlluminaHiSeq 2500 platform (Yourgene Bioscience, Taiwan). Raw reads underwent quality trimming (minimal read length, 35 bases) by CLC Genomics Workbench (<http://www.clcbio.com/>). Adaptor sequences were removed by Cutadapt 1.8.1 [15]. A total of 33,692,522 adaptor-removed, quality-validated reads, equivalent to 3,488,072,978 total bases, were subjected to genome assembly by Velvet 1.2.10 [16]. The assembled genome consisted of 65,230,783 bases ('N' composition, 0.6%) in 17,277 contigs (average length, 3776 bases; range, 300–209,930 bases; N50, 14,374 bases). Assessment of the resulting draft genome sequence by CEGMA [17,18] showed 78.6% genome completeness. A total of 26,058 open reading frames were predicted by MAKER2 [19].

2.3. Identification of sequence variants

A total of 7,843,910 adaptor-removed quality-validated reads (23.3% of all reads), derived from the *P. insidiosum* strain Pi45, can be aligned to the reference genome of the *P. insidiosum* strain Pi-S [7], using the Burrows-Wheeler Alignment tool [20]. A total of 865,332 variants (i.e., SNPs and INDELs) were identified by FreeBayes [21].

2.4. Data accessibility

The draft genome sequence of the *P. insidiosum* strain Pi45 (also known as MCC13) has been deposited in the Data Bank of Japan (DDBJ) under the accession numbers: BCFM01000001-BCFM01017277. The sequence variant data (i.e., SNPs and INDELs) of the *P. insidiosum* strain Pi45 can be accessible online at the Mendeley database (<http://dx.doi.org/10.17632/r75799jy6c.1>).

Acknowledgements

This work was supported by the Faculty of Medicine, Ramathibodi Hospital, Mahidol University; the Thailand Research Fund [Grant number, BRG5980009]; and the Mahidol University [Grant number, ngor-por 09/2560]. We thank Dr. T. Tristan Brandhorst for reviewing the manuscript.

Transparency document. Supplementary material

Transparency document associated with this article can be found in the online version at <https://doi.org/10.1016/j.dib.2017.10.047>.

References

- [1] S. Kamoun, Molecular genetics of pathogenic oomycetes, *Eukaryot. Cell.* 2 (2003) 191–199.
- [2] W. Gaastra, L.J.A. Lipman, A.W.A.M. De Cock, T.K. Exel, R.B.G. Pegge, J. Scheurwater, R. Vilela, L. Mendoza, *Pythium insidiosum*: an overview, *Vet. Microbiol.* 146 (2010) 1–16.
- [3] T. Krajaejun, B. Sathapatayavongs, R. Prachartam, P. Nitiyanant, P. Leelachaikul, W. Wanachiwanawin, A. Chairprasert, P. Assanasen, M. Saipetch, P. Mootsikapun, P. Chetchotisakd, A. Lekhakula, W. Mitarnun, S. Kalnauwakul, K. Supparatpinyo, R. Chaiwarith, S. Chiewchanvit, N. Tananuvat, S. Srisiri, C. Suankratay, W. Kulwichit, M. Wongsaisuwan, S. Somkaew, Clinical and epidemiological analyses of human pythiosis in Thailand, *Clin. Infect. Dis.* 43 (2006) 569–576.
- [4] T. Krajaejun, R. Prachartam, S. Wongwaisayawan, M. Rochanawutinin, M. Kunakorn, S. Kunavisarut, Ocular pythiosis: is it under-diagnosed? *Am. J. Ophthalmol.* 137 (2004) 370–372.

- [5] T. Chareonsirisuthigul, R. Khositnithikul, A. Intaramat, R. Inkomlue, K. Sriwanichrak, S. Piromsontikorn, S. Kitiwanwanich, T. Lowhnoo, W. Yingyong, A. Chaiprasert, R. Banyong, K. Ratanabanangkoon, T.T. Brandhorst, T. Krajaejun, Performance comparison of immunodiffusion, enzyme-linked immunosorbent assay, immunochromatography and hemagglutination for serodiagnosis of human pythiosis, *Diagn. Microbiol. Infect. Dis.* 76 (2013) 42–45.
- [6] M.S. Asuncion, J.C. Huguet-Tapia, E.L. Braun, A. Ortiz-Urquiza, N.O. Keyhani, E.M. Goss, Whole genome sequence of the emerging oomycete pathogen *Pythium insidiosum* strain CDC-B5653 isolated from an infected human in the USA, *Genom. Data* 7 (2016) 60–61.
- [7] T. Rujirawat, P. Patumcharoenpol, T. Lohnoo, W. Yingyong, T. Lerksuthirat, S. Tangphatsornruang, P. Suriyaphol, L. J. Grenville-Briggs, G. Garg, W. Kittichotirat, T. Krajaejun, Draft genome sequence of the pathogenic oomycete *Pythium insidiosum* strain Pi-S, isolated from a patient with pythiosis, *Genome Announc.* 3 (2015) (pii: e00574–15).
- [8] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res.* 32 (2004) 1792–1797.
- [9] J. Castresana, Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis, *Mol. Biol. Evol.* 17 (2000) 540–552.
- [10] S. Guindon, O. Gascuel, A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood, *Syst. Biol.* 52 (2003) 696–704.
- [11] M. Anisimova, O. Gascuel, Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative, *Syst. Biol.* 55 (2006) 539–552.
- [12] F. Chevenet, C. Brun, A.-L. Bañuls, B. Jacq, R. Christen, TreeDyn: towards dynamic graphics and annotations for analyses of trees, *BMC Bioinforma.* 7 (2006) 439.
- [13] A. Dereeper, V. Guignon, G. Blanc, S. Audic, S. Buffet, F. Chevenet, J.-F. Dufayard, S. Guindon, V. Lefort, M. Lescot, J.-M. Claverie, O. Gascuel, Phylogeny.fr: robust phylogenetic analysis for the non-specialist, *Nucleic Acids Res.* 36 (2008) W465–W469.
- [14] T. Lohnoo, N. Jongruja, T. Rujirawat, W. Yingyong, T. Lerksuthirat, U. Nampoorn, Y. Kumsang, P. Onpaew, P. Chongtrakool, A. Keeratjarut, T.T. Brandhorst, T. Krajaejun, Efficiency comparison of three methods for extracting genomic DNA of the pathogenic oomycete *Pythium insidiosum*, *J. Med. Assoc. Thail.* 97 (2014) 342–348.
- [15] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, *EMBnet. J.* 17 (2011) 10.
- [16] D.R. Zerbino, E. Birney, Velvet: algorithms for de novo short read assembly using de Bruijn graphs, *Genome Res.* 18 (2008) 821–829.
- [17] G. Parra, K. Bradnam, Z. Ning, T. Keane, I. Korf, Assessing the gene space in draft genomes, *Nucleic Acids Res.* 37 (2008) 289–297.
- [18] G. Parra, K. Bradnam, I. Korf, CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes, *Bioinformatics* 23 (2007) 1061–1067.
- [19] C. Holt, M. Yandell, MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects, *BMC Bioinforma.* 12 (2011) 491.
- [20] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform, *Bioinformatics* 25 (2009) 1754–1760.
- [21] E. Garrison, G. Marth, Haplotype-based variant detection from short-read sequencing, *ArXiv PreprArXiv:1207.3907 [q-GN]*, 2012.