รายงานวิจัยฉบับสมบูรณ์

โครงการ: Identification, Functional Characterization, and Clinical Application of *Pythium insidiosum* Secreted and Surface-Associated Proteins as Diagnostic and Therapeutic Targets

โดย น.พ. ธีรพงษ์ กระแจะจันทร์ และคณะ

เดือน ปี ที่เสร็จโครงการ: พฤษภาคม 2562

Infection, Genetics and Evolution 54 (2017) 429–436

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Research paper

Single nucleotide polymorphism-based multiplex PCR for identification and genotyping of the oomycete *Pythium insidiosum* from humans, animals and the environment

Thidarat Rujirawat[a,b,c,1], Thanawat Sridapan[c,1], Tassanee Lohnoo[b], Wanta Yingyong[b], Yothin Kumsang[b], Pattarana Sae-Chew[b], Walaiporn Tonpitak[d], Theerapong Krajaejun[a,*]

[a] Department of Pathology, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand
[b] Research Center, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand
[c] Molecular Medicine Program, Multidisciplinary Unit, Faculty of Science, Mahidol University, Bangkok, Thailand
[d] Department of Microbiology, Faculty of Veterinary Medicine, Mahanakorn University of Technology, Bangkok, Thailand

ARTICLE INFO

Keywords:
Pythium insidiosum
Pythiosis
Multiplex PCR
rDNA
Genotyping
Detection

ABSTRACT

Pythium insidiosum causes a life-threatening infectious disease, called pythiosis, in humans and animals worldwide. Diagnosis of pythiosis is difficult and often delayed. Surgical removal of infected tissue is the main treatment option. Disabilities and death are common outcomes for pythiosis patients. Reports of *Py. insidiosum* infections are rising. While it would be useful for clinical, epidemiological, and microbiological studies, information on genetic variation in *Py. insidiosum* strains is limited. This limitation is, at least in part, due to the cost and time-requirements of DNA sequencing procedures. rDNA-sequence-based phylogenetic analyses categorize *Py. insidiosum* into three groups, in relation to geographic distribution: Clade-I (American strains), Clade-II (American, Asian, and Australian strains), and Clade-III (Thai and American strains). In rDNA sequence analyses, we observed single nucleotide polymorphisms (SNP) that were associated with the phylogenetic clades of *Py. insidiosum*. In this study, we aim to develop a multiplex PCR assay, targeting the identified SNPs, for rapid genotyping of *Py. insidiosum*. We also aim to assess diagnostic efficiency of the assay for identification of *Py. insidiosum*. Fifty-three isolates of *Py. insidiosum* from humans (n = 35), animals (n = 14), and the environment (n = 4), and 22 negative-control fungi were recruited for assay evaluation. Based on the pattern of amplicons, the multiplex PCR correctly assigned phylogenetic clades in 98% of the *Py. insidiosum* isolates tested. The assay exhibited 100% sensitivity and specificity for identification of *Py. insidiosum*. The assay successfully identified and genotyped the first proven isolate of *Py. insidiosum* from an animal with pythiosis in Thailand. In conclusion, the multiplex PCR provided accurate, sensitive and specific results for identifying and genotyping *Py. insidiosum*. Thus, this multiplex-PCR assay could be a simple, rapid, and cost-effective alternative to DNA sequencing for the identification and genotyping of *Py. insidiosum*.

1. Introduction

Pythium insidiosum is a member of the oomycetes, which is a unique group of morphologically fungus-like microorganisms that belong to the Kingdom Stramenopila (De Cock et al., 1987; Gaastra et al., 2010; Kamoun, 2003; Mendoza et al., 1996). While most pathogenic oomycetes are capable of infecting plants, *Py. insidiosum* can cause a life-threatening infectious condition, called pythiosis, in humans and animals (horses, dogs, cats, cattle) (Gaastra et al., 2010; Kamoun, 2003; Mendoza et al., 1996). Zoospores are the infective agents of *Py. insidiosum*, as they can adhere and germinate into host tissue (Mendoza et al., 1993). Target organs for infection include skin, intestines, blood vessels and eyes (Gaastra et al., 2010; Krajaejun et al., 2004, 2006; Mendoza et al., 1996). With the lack of other diagnostic tools (Chareonsirisuthigul et al., 2013; Inkomlue et al., 2016; Intaramat et al., 2016; Keeratijarut et al., 2009, 2015), diagnosis of pythiosis relies on culture identification, which is laborious and time-consuming. Conventional antimicrobial drugs are not generally effective against *Py. insidiosum*. Extensive surgical removal of the infected organ is the inescapable treatment option, and most often leads to disabilities in the

* Corresponding author.
E-mail address: mr_en@hotmail.com (T. Krajaejun).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.meegid.2017.08.004
Received 7 April 2017; Received in revised form 17 July 2017; Accepted 4 August 2017
Available online 05 August 2017
1567-1348/

affected patients. Post-surgery recurrence of the *Py. insidiosum* infection and death are the final outcomes in many patients with pythiosis (Krajaejun et al., 2004, 2006).

Pythiosis in animals has been increasingly diagnosed in tropical and subtropical regions, whereas the disease in humans has been mostly found in Thailand, where no animal case has been reported thus far (Gaastra et al., 2010; Krajaejun et al., 2006; Mendoza et al., 1996). The successful isolations of *Py. insidiosum* from swampy areas in Australia, Thailand, and the United States of America have been described (Miller, 1983; Presser and Goss, 2015; Supabandhu et al., 2008), suggesting that the pathogen is distributed worldwide. Several genes [i.e., ribosomal DNA (rDNA), *exo1*, and *cox II*] have been used to explore phylogenetic relationship of *Py. insidiosum* isolated from different hosts and geographic areas (Azevedo et al., 2012; Chaiprasert et al., 2009; Kammarnjesadakul et al., 2011; Krajaejun et al., 2010; Ribeiro et al., 2016; Schurko et al., 2003; Supabandhu et al., 2008). Among them, rDNA [which consists of 18S rRNA, internal transcribed spacer 1 (ITS1), 5.8S rRNA, ITS2, and 28S rRNA] has been a frequent target for phylogenetic studies of *Py. insidiosum* (Chaiprasert et al., 2009; Lerksuthirat et al., 2015; Schurko et al., 2003; Supabandhu et al., 2008; Vanittanakom et al., 2014). The rDNA-based phylogenetic analyses categorize *Py. insidiosum* into three groups, in association with geographic origins: Clade-I (containing American strains), Clade-II (American, Asian, and Australian strains), and Clade-III (mostly Thai strains) (Chaiprasert et al., 2009; Schurko et al., 2003).

According to the growing number of publications on pythiosis, awareness of the disease and its causative agent, *Py. insidiosum*, has increased among healthcare professionals and microbiologists. However, in order to support clinical, epidemiological and microbiological studies, a faster and simpler method is needed for identification and genotyping than that based on DNA sequence analysis. After analysis of the rDNA sequences previously-reported by our group (Lerksuthirat et al., 2015), we observed single nucleotide polymorphisms (SNP) within the rDNA region that show associations with the phylogenetic clades of *Py. insidiosum* (i.e., Clade-I, -II, and -III). SNPs are genetic markers widely used in biological and clinical studies, such as, genome-wide association, population genetics, pharmacogenomics, and disease susceptibility (Kim and Misra, 2007; Seeb et al., 2011; Vignal et al., 2002). To our knowledge, no SNP markers have been identified and utilized for molecular studies of *Py. insidiosum*. In the current study, we report on the development and evaluation of a multiplex PCR assay, targeting the SNPs identified in the rDNA region, for rapid and cost-effective genotyping of *Py. insidiosum*. We also test the diagnostic efficiency of the assay for identification of *Py. insidiosum* isolates. Finally, we demonstrate the efficacy of the multiplex PCR to identify and genotype the first proven isolate of *Py. insidiosum* from an infected animal in Thailand.

2. Materials and methods

2.1. Microorganisms and genomic DNA extraction

Fifty-three isolates of *Py. insidiosum* from human patients with pythiosis (n = 35), equine patients with pythiosis (n = 13), mosquito larva (n = 1), and environmental samples (n = 4) were used for this study (Table 1). Identities of all *Py. insidiosum* isolates were molecularly confirmed by rDNA sequence analysis with 99–100% homology to the rDNA sequences of *Py. insidiosum* obtained from GenBank (https://www.ncbi.nlm.nih.gov/genbank/) (Badenoch et al., 2001). Each isolate was incubated in a Petri dish with 10 ml of Sabouraud dextrose broth (pH 7.2), at room temperature for 10 days. The hyphae were removed from the broth, briefly washed with distilled water, and filtered through Whatman No.1 filter paper for genomic DNA (gDNA) extraction, using the salt-extract protocol described by Lohnoo and co-workers (Lohnoo et al., 2014).

To serve as negative controls, 22 clinically relevant fungi (4 *Candida* species, 2 *Mucor* species, 2 *Fusarium* species, 2 *Trichophyton* species, 2 *Trichosporon* species, and one isolate each of *Cryptococcus neoformans*, *Penicillium marneffei*, *Torulopsis glabrata*, *Scedeosporium apiospermum*, *Rhizopus* species, *Conidiobolus* species, *Aspergillus* species, *Acremorium* species, *Microsporum* species, and *Curvularia* species) were obtained from the Clinical Microbiology Laboratory, Department of Pathology, Ramathibodi Hospital, Bangkok, Thailand. The identity of each filamentous fungus was confirmed by colony and conidia morphology, while that of yeast was confirmed by carbohydrate and nitrogen assimilation, carbohydrate fermentation, and phenol oxidase/urease enzyme production. Fungal gDNAs were prepared using the gDNA extraction protocol described by Keeratijarut and co-workers (Keeratijarut et al., 2014). Quantity and purity of each gDNA sample was determined by a NanoDrop 2000 spectrophotometer (Thermo Scientific). All extracted gDNA were stored in Tris-EDTA buffer (pH 8.0) at − 20 °C.

2.2. Sequence alignment and primer design

The rDNA sequences of 18 *Py. insidiosum* isolates, classified in three different phylogenetic clades (Lerksuthirat et al., 2015), and six other oomycete microorganisms [i.e., *Pythium deliense* (accession number, AY151181.1), *Pythium aphanidermatum* (AY598622.2), *Phytophthora capsici* (AY726623.1), *Phytophthora infestans* (HQ643247.1), *Phytophthora parasitica* (KC768775.1), and *Hyaloperonospora arabidopsidis* (AY531434.1)] were retrieved from the NCBI database and aligned by using the BioEdit software (http://www.mbio.ncsu.edu/bioedit/bioedit.html). The Primer-BLAST program (Ye et al., 2012) was used to design the multiplex-PCR reverse primers R1 (5′-CCTCACATTCTGCCATCTCG-3′), R2 (5′-ATACCGCCAATAGAGGTCAT-3′), and R3 (5′-TTACCCGAAGGCGTCAAAGA-3′), based upon the clade-specific SNPs presented in the rDNA sequences of *Py. insidiosum* (Fig. 1). The fungal universal primer ITS1 (5′-TCCGTAGGTGAACCTGCGG-3′) was used as the shared multiplex-PCR forward primer, and also with the fungal universal ITS4 (5′-TCCTCCGCTAATTGATATGC-3′) for quality evaluation of the extracted gDNA samples (see below). The diagram, demonstrating all primer annealing locations, was prepared using the IBS program (Liu et al., 2015). All primer sequences were BLAST searched against the draft genome sequence of *Py. insidiosum* (Rujirawat et al., 2015) and the NCBI database for potential off-target priming.

2.3. Polymerase chain reaction

Multiplex-PCR amplification was carried out in a 25-μl reaction, which contained 100 ng of gDNA template, 0.65 U of Taq DNA polymerase (Fermentas, USA), 1 × Taq buffer with KCl, 2 mM of $MgCl_2$, 0.2 mM of dNTPs, 0.03 μM of the forward primer ITS1, and 0.07 μM each of the reverse primers R1, R2, and R3. The PCR reaction was performed in a Mastercycler Nexus thermal cycle machine (Eppendorf, Germany), using the following conditions: initial denaturation at 95 °C for 5 min, 20 cycles of denaturation at 95 °C for 30 s, annealing at 59 °C for 30 s, and extension at 72 °C for 45 s, and final extension at 72 °C for 10 min. To check for the presence of amplifiable DNA in the samples, conventional PCR amplification was conducted, using the above-mentioned PCR conditions, with two modifications: (i) using the primers ITS1 and ITS4 (0.1 μM each), and (ii) changing the annealing step to 25 cycles at 55 °C. The obtained PCR products were separated in a 1% agarose gel, stained with ethidium bromide, and visualized by a Gel Doc XR + Molecular Imager (Bio-Rad, USA).

2.4. DNA sequencing and nucleotide sequence accession numbers

The rDNA ~920-bp amplicons generated with primers ITS1 and ITS4 using genomic DNA from the 53 isolates of *Py. insidiosum* were sequenced using the primers ITS1 and ITS4, an ABI PRISM BigDyeTM terminator cycle sequencing ready reaction kit (Applied Biosystems, USA), and an ABI 3100 Genetic Analyzer (Applied Biosystems, USA).

Table 1
Fifty-three isolates of *Py. insidiosum* used for evaluation of the multiplex PCR assay. Information on strain ID, reference strain [indicate the strains that have been phylogenetically categorized by other investigators (Chaiprasert et al., 2009; Schurko et al., 2003)], sources of isolation (affected hosts and countries), phylogenetic clades, accession number for the rDNA sequences, and PCR products of each set of the PCR primers, are summarized in the table ('bp', base pair; '+', PCR product is detected; 'ND', PCR product is not detected; NA, data not available).

Strain ID	Reference strain	Source	Country (province, city or state) of origin	Phylogenetic clade	Accession number of rDNA	PCR product (bp)			
						Primers ITS1/ ITS4 (~920)	Primers ITS1/ R1 (~490)	Primers ITS1/ R2 (~660)	Primers ITS1/ R3 (~800)
Pi01	ATCC58639	Equine	Costa Rica (NA)	Clade-I	LC199875	+	+	+	ND
Pi02	ATCC58638	Equine	Costa Rica (NA)	Clade-I	AB971176	+	+	+	ND
Pi03	ATCC58640	Equine	Costa Rica (NA)	Clade-I	AB971177	+	+	+	ND
Pi04	ATCC58641	Equine	Costa Rica (NA)	Clade-I	AB898106	+	+	+	ND
Pi05	ATCC58642	Equine	Costa Rica (NA)	Clade-I	AB971178	+	+	+	ND
Pi06	ATCC58643	Equine	Costa Rica (NA)	Clade-I	AB971179	+	+	+	ND
Pi07	ATCC58644	Equine	Costa Rica (NA)	Clade-I	AY151158	+	+	+	ND
Pi08	ATCC58637	Equine	Costa Rica (NA)	Clade-I	AB898107	+	+	+	ND
Pi09	394	Equine	Brazil (NA)	Clade-I	AY151163	+	+	+	ND
Pi10	M6	Human	USA (Tennessee)	Clade-I	AB898108	+	+	+	ND
Pi11	–	Human	Thailand (NA)	Clade-II	AB898109	+	ND	+	ND
Pi12	Hu20	Human	Thailand (Lopburi)	Clade-II	LC199876	+	ND	+	ND
Pi13	–	Human	Thailand (NA)	Clade-II	AB898110	+	ND	+	ND
Pi14	Hu24	Human	Thailand (NA)	Clade-II	LC199877	+	ND	+	ND
Pi15	–	Human	Thailand (NA)	Clade-II	AB898111	+	ND	+	ND
Pi16	CBS119452	Human	Thailand (Nan)	Clade-II	AB971182	+	ND	+	ND
Pi17	–	Human	Thailand (NA)	Clade-II	AB898112	+	ND	+	ND
Pi18	CBS119453	Human	Thailand (Lumpang)	Clade-II	EF016853	+	ND	+	ND
Pi19	–	Human	Thailand (NA)	Clade-II	AB898113	+	ND	+	ND
Pi20	CBS119455	Human	Thailand (Chiang Mai)	Clade-II	EF016855	+	ND	+	ND
Pi21	–	Human	Thailand (NA)	Clade-II	AB898114	+	ND	+	ND
Pi22	Hu22	Human	Thailand (Yasothon)	Clade-II	GU137338	+	ND	+	ND
Pi23	–	Human	Thailand (Saraburi)	Clade-II	AB898115	+	ND	+	ND
Pi24	Hu10	Human	Thailand (Chonburi)	Clade-II	GU137327	+	ND	+	ND
Pi25	–	Human	Thailand (NA)	Clade-II	AB898116	+	ND	+	ND
Pi26		Human	Thailand (Patomthani)	Clade-II	AB898117	+	ND	+	ND
Pi27	Hu12	Human	Thailand (Korat)	Clade-II	GU137329	+	ND	+	ND
Pi28	–	Human	Thailand (NA)	Clade-II	AB898118	+	ND	+	ND
Pi29	Hu13	Human	Thailand (Suphanburi)	Clade-II	GU137330	+	ND	+	ND
Pi30	–	Human	Thailand (NA)	Clade-II	AB898119	+	ND	+	ND
Pi31	–	Human	Thailand (NA)	Clade-II	AB898120	+	ND	+	ND
Pi32	–	Human	Thailand (NA)	Clade-II	AB898121	+	ND	+	ND
Pi33	–	Human	Thailand (NA)	Clade-II	AB898122	+	ND	+	ND
Pi34	–	Human	Thailand (NA)	Clade-II	AB898123	+	ND	+	ND

(continued on next page)

Table 1 *(continued)*

Strain ID	Reference strain	Source	Country (province, city or state) of origin	Phylogenetic clade	Accession number of rDNA	PCR product (bp)			
						Primers ITS1/ ITS4 (~920)	Primers ITS1/ R1 (~490)	Primers ITS1/ R2 (~660)	Primers ITS1/ R3 (~800)
Pi35	–	Human	Thailand (NA)	Clade-II	AB898124	+	ND	+	ND
Pi36	M23	Equine	Australia (NA)	Clade-II	AY151174	+	ND	+	ND
Pi37	M15	Equine	Papua New Guinea (NA)	Clade-II	AY151171	+	ND	+	ND
Pi38	393	Human	India (NA)	Clade-II	AB898125	+	ND	+	ND
Pi39	ATCC46947	Equine	Japan (NA)	Clade-II	AY151170	+	ND	+	ND
Pi40	296	Mosquito larva	India (NA)	Clade-II	AY151169	+	ND	+	ND
Pi41	E55	Environment	Thailand (Chiang Mai)	Clade-II	EF016862	+	ND	+	ND
Pi42	E14	Environment	Thailand (Chaing Rai)	Clade-II	EF016903	+	+ (weak)	+	ND
Pi43	–	Environment	Thailand (NA)	Clade-II	AB898126	+	ND	+	ND
Pi44	CBS119454	Human	Thailand (Chiang Mai)	Clade-III	AB971185	+	ND	ND	+
Pi45	Hu08	Human	Thailand (Saraburi)	Clade-III	AB971186	+	ND	ND	+
Pi46	Hu26	Human	Thailand (Nakorn Srithamarat)	Clade-III	AB971187	+	ND	ND	+
Pi47	Hu29	Human	Thailand (NA)	Clade-III	AB971188	+	ND	ND	+
Pi48	Hu21	Human	Thailand (Rachaburi)	Clade-III	AB971189	+	ND	ND	+
Pi49	–	Human	Thailand (NA)	Clade-III	AB898127	+	ND	ND	+
Pi50	M19	Human	USA (TX)	Clade-III	AB971190	+	ND	ND	+
Pi51	–	Environment	Thailand (NA)	Clade-III	AB898128	+	ND	ND	+
Pi52	–	Human	Thailand (NA)	Clade-II	LC199888	+	ND	+	ND
Pi53	–	Equine	Thailand (Bangkok)	Clade-II	LC199889	+	ND	+	ND

All new rDNA sequences have been submitted to the DNA data bank of Japan (DDBJ), under the accession numbers shown in Table 1.

2.5. Phylogenetic analyses

The rDNA sequences from 53 isolates of *Py. insidiosum* (Table 1) were subjected to online phylogenetic tree construction software at http://www.phylogeny.fr (Dereeper et al., 2008). In brief, the sequences were aligned by the MUSCLE program (Edgar, 2004) and phylogenetically analyzed using the Maximum-Likelihood method in the PhyML 3.0 program with 500 bootstrap replicates (Guindon et al., 2010). A tree was constructed using the locally-installed software Dendroscope 3 (Huson and Scornavacca, 2012). The outgroups included rDNA sequences from *Pythium deliense* (accession number, AY151181) and *Pythium grandisporangium* (AY151182).

3. Results

3.1. DNA sequence analysis

Conventional PCR amplifications, using the fungal universal rDNA primers ITS1 and ITS4 (Fig. 1), were performed to provide template for sequence and phylogenetic analyses and to check for the presence of amplifiable gDNA in the samples extracted from *Py. insidiosum* and from the control samples (Fig. 2A). Only the PCR-positive samples (*Py. insidiosum*, $n = 53$; controls, $n = 22$) were used for the downstream multiplex PCR analyses (see below).

3.2. Generation of a phylogenetic tree

The rDNA sequences (ITS1/ITS4 amplicon, ~920 bp) from 53 strains of *Py. insidiosum*, isolated from different hosts and geographic locations (Table 1) were aligned, and analyzed for phylogenetic relationships. A resulting Maximum Likelihood-based phylogenetic tree categorized the *Py. insidiosum* strains into Clade-I ($n = 10$), Clade-II ($n = 35$), and Clade-III ($n = 8$), which are in agreement with the previous studies (Chaiprasert et al., 2009; Schurko et al., 2003), as shown in Fig. 3 and Table 1.

3.3. Setting up a multiplex PCR assay

To develop a simpler PCR-based assay for identification and genotyping (i.e., Clade identification) of *Py. insidiosum*, we used SNPs, identified in the sequence analyses, as the basis for primers for multiplex PCR. Three reverse primers (R1, R2, and R3) were designed to work in conjunction with the fungal universal forward primer ITS1 for the rDNA of *Py. insidiosum*, and to distinguish between the three different phylogenetic groups: Clade-I, -II, and -III (Fig. 1). The primer R1 should specifically prime the ITS-2 sequences of the Clade-I strains, the primer R2 should prime the ITS-2 sequences of both Clade-I and -II strains, and the primer R3 should only prime the ITS-2 sequences of the Clade-III strains (Fig. 1B). The three reverse primers did not match ITS-2 sequences of closely-related oomycete microorganisms, including *Py. deliense*, *Py. aphanidermatum*, *Ph. capsici*, *Ph. infestans*, *Ph. parasitica*, and *Hy. arabidopsidis* (Fig. 1B). To further check the annealing specificity, the primers were BLAST searched against the draft genome of *Py.*

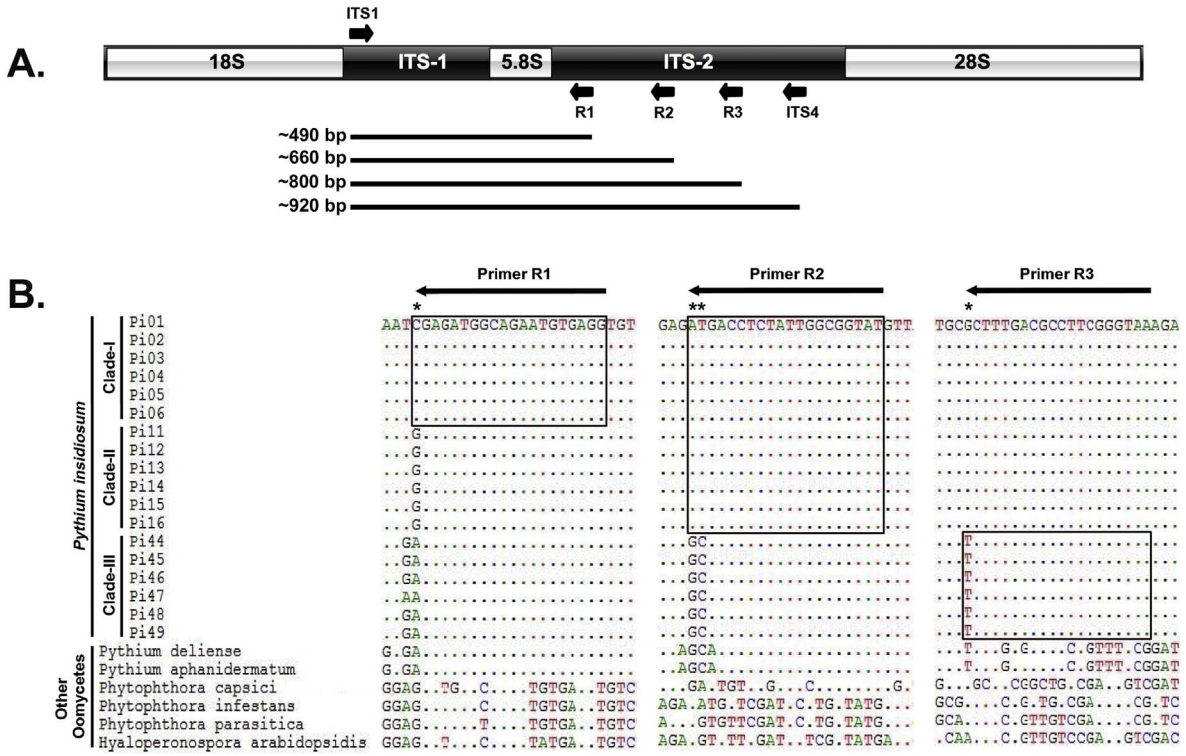


Fig. 1. The multiplex PCR primers used in this study: (A) The annealing locations of the primers ITS1, ITS2, R1, R2, and R3 in the rDNA region of *Py. insidiosum* (the horizontal bars indicate amplicon product sizes for each primer pair); (B) Alignment of the primer R1, R2, and R3-annealing sequences in the rDNA region of 18 *Py. insidiosum* isolates and 6 other oomycetes (the letters A, C, G, and T represent nucleotides; a dot represents the identical nucleotide; a box indicates the primer sequence).

insidiosum (Rujirawat et al., 2015), as well as, genes of all other organisms deposited in the NCBI database. No potential off-target annealing of these primers was identified. The reverse primers (R1, R2, and R3) and the shared forward primer ITS1 (Fig. 1A) were together used in the multiplex PCR assay. Among the strains tested, the amplicons obtained from each primer pair were similar in size with a few bases difference. The multiplex PCR assay resulted in the expected amplicon sizes of ~490 bp for the primer pair ITS-1/R1, ~660 bp for the primer pair ITS-1/R2, and ~800 bp for the primer pair ITS-1/R3 (Fig. 2B).

3.4. Multiplex PCR-based genotyping and detection

The multiplex PCR amplifications, using primers ITS1, R1, R2, and R3, provided the expected amplicons for gDNAs extracted from all *Py. insidiosum* isolates used in this study (Fig. 1). The amplicons had different sizes, so that the association with the phylogenetic groups of the organism, as previously defined by DNA sequence analyses, is dependent both on the presence of PCR products, as well as the sizes (Figs. 2B and 3). Two PCR products of ~490 and ~660 bp were amplified from all gDNAs of the Clade-I strains (n = 10). Only the PCR product of ~800 bp was observed in all Clade-III strains (n = 8). For the Clade-II strains (n = 35), 97% (n = 34) had one prominent (and expected) PCR product of ~660 bp. A single Cade-II strain (Pi42) showed two amplicons: a prominent band of ~660 bp and a weak band of ~490 bp. The multiplex PCR assay failed to amplify any product in all 22 negative-control gDNA samples extracted from various fungal species (The results of a few representative strains were shown in Fig. 2).

Because of the unexpected result of two amplicons with strain Pi42, we further analyzed the chromatogram of the rDNA sequence from strain Pi42 and found a unique SNP with a high intensity peak of guanosine triphosphate (G) and a relatively-low intensity peak of cytidine triphosphate (C), at the position corresponding to the annealing site of the 3′-end base of the primer R1 (Figs. 1B and 4D). No such finding was noted at the same SNP position in the rDNA sequences of the Clade-I strains (only C peak was observed; Fig. 4A), the Clade-III strains (only adenosine triphosphate (A) peak; Fig. 4C), and the other Clade-II strains (only G peak; Fig. 4B).

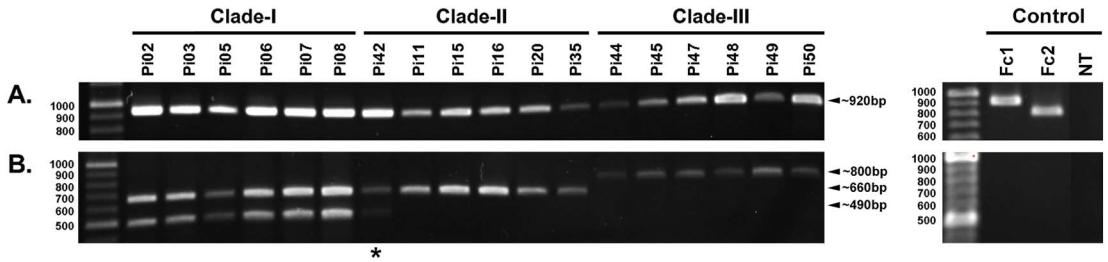


Fig. 2. PCR amplification of the rDNA sequences from representative strains of *Py. insidiosum* and control fungi: Agarose gel electrophoresis of PCR products amplified by (A) the primers ITS1/ITS4 (for checking the presence of gDNA; amplicon size, ~920 bp), and (B) the multiplex PCR primers ITS1, R1, R2, and R3 (the amplicon sizes for the primer pairs ITS1/R1, ITS1/R2, and ITS1/R3 are ~490 bp, ~660 bp, and ~800 bp, respectively). Fc1 and Fc2 represent the control fungi *Torulopsis glabrata* and *Conidiobolus* species, respectively. NT indicates no-template control. The 100-bp step-ladder markers are shown at the left side of the gel.

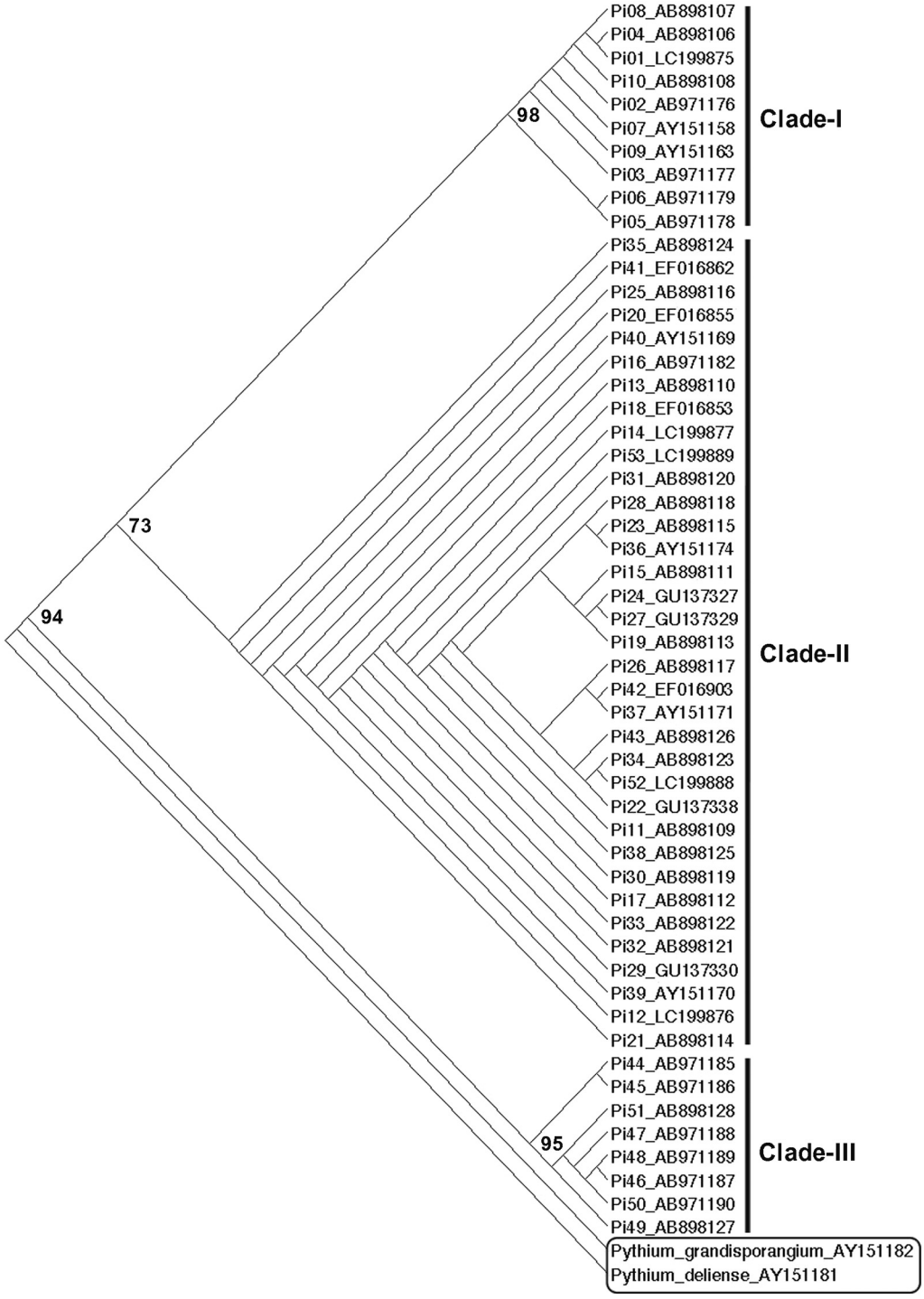


Fig. 3. Phylogenetic relationship of various *Py. insidiosum* isolates: rDNA sequences from 53 isolates of *Py. insidiosum* and 2 outgroup oomycetes (*Py. deliense* and *Py. grandisporangium*) are analyzed for phylogenetic relationship, using the Maximum Likelihood algorithm, with 500 bootstrap replicates. Only branch support values of 70% or more are shown.

4. Discussion

Currently, *Py. insidiosum* has been genotyped, using the costly, time-consuming, sequence-based phylogenetic analysis, which groups the organism into Clade-I (American strains), Clade-II (American, Asian, and Australian strains), and Clade-III (mostly Thai strains) (Chaiprasert et al., 2009; Schurko et al., 2003). Here, we have developed a multiplex PCR, targeting three SNPs identified in the rDNA region (Fig. 1), for simple and rapid genotyping of *Py. insidiosum*. Interpretation of the multiplex PCR-based genotyping relies on number and size of the resulting amplicons. The genotyping results from the multiplex PCR assay were generally consistent with that from the sequence-based phylogenetic analysis (Fig. 3 and Table 1). The Clade-I strains provided two amplicons (~490 and ~660 bp), whereas the Clades-II and -III strains showed only one amplicons (~660 and ~800 bp, respectively), with one exception for the strain Pi42 of Clade-II (Fig. 2). This finding indicates that the multiplex PCR assay provided ~98% overall accuracy for genotyping of *Py. insidiosum*.

The strain Pi42 was confirmed to be a Clade-II strain by the Maximum Likelihood-based phylogenetic analysis of the ITS rDNA region in this study (Fig. 3) and a previously reported study (Chaiprasert et al., 2009) (Table 1). Instead of having one PCR product (based on ITS1/R2 amplification), as usually observed in the other Clade-II strains, the strain Pi42 gave two PCR products that had compatible sizes

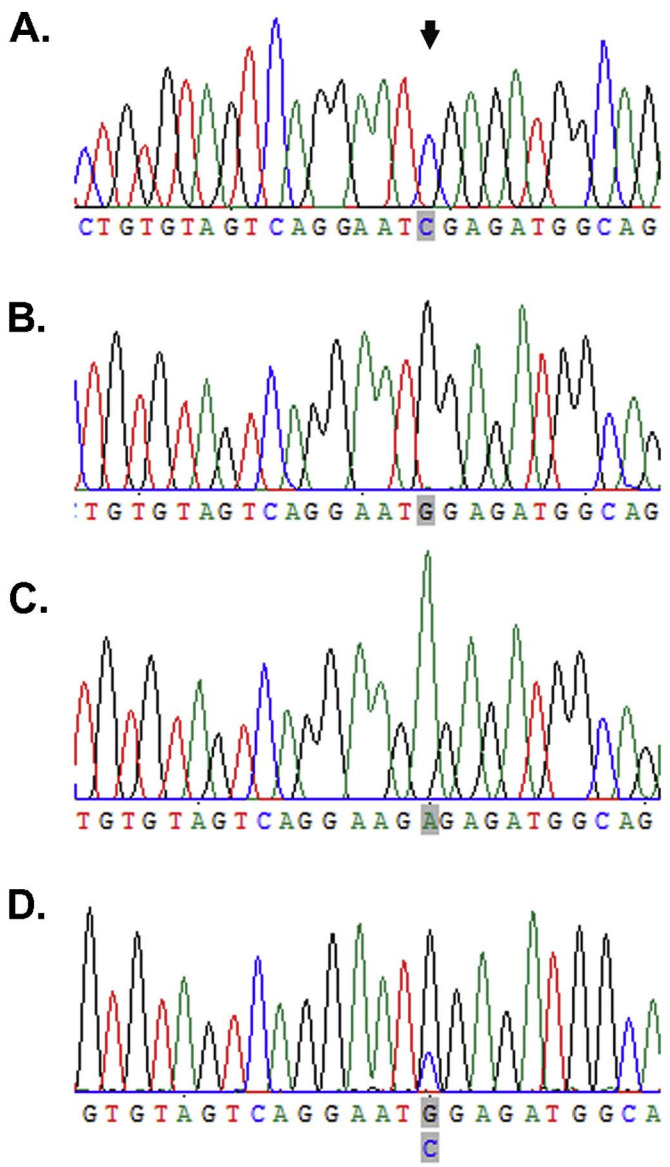


Fig. 4. Chromatograms of the rDNA sequences of *Py. insidiosum*: The arrow indicates the 3′-end nucleotide of the primer R1 annealing site in the rDNA region of the Clade-I (A), Clade-II (B), Clade-III (C), and Pi42 (D) strains [the letters A (green), C (blue), G (black), and T (red) represent nucleotides]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

with that of the Clade-I strains (with ITS1/R1 and ITS1/R2 products; Fig. 2). In addition, the chromatogram of the rDNA sequence of the strain Pi42, at the annealing site of the primer R1, shows two different peaks: a tall peak of nucleotide 'G' (the SNP found only in Clade-II strains) and an unexpected short peak of nucleotide 'C' (the SNP usually found in Clade-I strains) (Figs. 1B and 4D). Taken together with the fact that rDNA is a multi-copy locus (Grooters and Gee, 2002; Krajaejun et al., 2011), this information indicates that strain Pi42 has two different versions of the ITS region in its rDNA repeat sequence. In most of the genic units within the repeat, the sequence matches the expected Clade-II SNP, with a 'G', but in a minority of the genic units within the rDNA repeat, the 'G' is replaced with a 'C' (Clade-I SNP). This result with Pi42 suggests that in future use of this multiplex PCR for genotyping *Py. insidiosum*, if the gel analysis results are that Clade-I bands are not of the expected relative intensities, then further DNA analyses may be necessary to confirm the clade designation.

The multiplex PCR assay was also evaluated for its diagnostic efficiency for *Py. insidiosum*. The PCR results were read as positive when an expected PCR product is present, while read as negative if no PCR product is observed. All control samples from clinically relevant fungi (n = 22), including *Aspergillus* species, *Fusarium* species, and Zygomycetes (i.e., *Mucor*, *Rhizopus* and *Conidiobolus* species) that share microscopic features with *Py. insidiosum* were determined to be PCR negative, indicating that the assay has 100% detection specificity. On the other hand, all 53 gDNA samples extracted from *Py. insidiosum* were determined to be PCR positive, indicating that the assay has 100% detection sensitivity.

Among the *Py. insidiosum* samples, one gDNA, extracted from a strain of *Py. insidiosum* recently-isolated from an infected horse in Thailand (designated as the strain Pi53; Table 1), was also included in the multiplex PCR assay evaluation. There was only one PCR product at the size of ~660 bp, indicating that the organism was a Clade-II *Py. insidiosum* strain, as confirmed by the other methods, i.e., culture identification, sequence homology analysis, and rDNA-based phylogenetic study.

In conclusion, the SNP-based multiplex-PCR assay for identification and genotyping of *Py. insidiosum* was successfully developed. Based on the clade-specific patterns of PCR products, the assay provided 98% accuracy for genotyping, and 100% for both detection specificity and sensitivity. The newly-developed assay was also used to successfully identify, to our knowledge, the first proven isolate of *Py. insidiosum* from an animal in Thailand (clinical features of this animal pythiosis case will be described in details elsewhere). Taken together, the multiplex PCR is a simple and efficient assay for routine identification of *Py. insidiosum* isolated from clinical samples and the environment. The assay is also an alternative tool for convenient, rapid and cost-effective genotyping of the organism. Thus, correlation of the pathogen genotypes and clinical data is now more feasible, which can promote other areas of *Py. insidiosum* studies, i.e., molecular epidemiology, infection outbreaks, and host-pathogen associations.

Acknowledgements

This work is supported by the Faculty of Medicine, Ramathibodi Hospital, Mahidol University and the Thailand Research Fund (TK; Grant number, BRG5980009). We thank Thomas D. Sullivan for reviewing the manuscript and making valuable suggestions, and Angkana Chaiprasert, Nongnuch Vanittanakom, and Piriyaporn Chongtrakool for suggestions and material support.

References

Azevedo, M.I., Botton, S.A., Pereira, D.I.B., Robe, L.J., Jesus, F.P.K., Mahl, C.D., Costa, M.M., Alves, S.H., Santurio, J.M., 2012. Phylogenetic relationships of Brazilian isolates of *Pythium insidiosum* based on ITS rDNA and cytochrome oxidase II gene sequences. Vet. Microbiol. 159, 141–148. http://dx.doi.org/10.1016/j.vetmic.2012.03.030.

Badenoch, P.R., Coster, D.J., Wetherall, B.L., Brettig, H.T., Rozenbilds, M.A., Drenth, A., Wagels, G., 2001. *Pythium insidiosum* keratitis confirmed by DNA sequence analysis. Br. J. Ophthalmol. 85, 502–503.

Chaiprasert, A., Krajaejun, T., Pannanusorn, S., Prariyachatigul, C., Wanachiwanawin, W., Sathapatayavongs, B., Juthayothin, T., Smittipat, N., Vanittanakom, N., Chindamporn, A., 2009. *Pythium insidiosum* Thai isolates: molecular phylogenetic analysis. Asian Biomed. 3, 623–633.

Chareonsirisuthigul, T., Khositnithikul, R., Intaramat, A., Inkomlue, R., Sriwanichrak, K., Piromsontikorn, S., Kitiwanwanich, S., Lowhnoo, T., Yingyong, W., Chaiprasert, A., Banyong, R., Ratanabanangkoon, K., Brandhorst, T.T., Krajaejun, T., 2013. Performance comparison of immunodiffusion, enzyme-linked immunosorbent assay, immunochromatography and hemagglutination for serodiagnosis of human pythiosis. Diagn. Microbiol. Infect. Dis. 76, 42–45. http://dx.doi.org/10.1016/j.diagmicrobio.2013.02.025.

De Cock, A.W., Mendoza, L., Padhye, A.A., Ajello, L., Kaufman, L., 1987. *Pythium insidiosum* sp. nov., the etiologic agent of pythiosis. J. Clin. Microbiol. 25, 344–349.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.-F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.-M., Gascuel, O., 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 36, W465–469. http://dx.doi.org/10.1093/nar/gkn180.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. http://dx.doi.org/10.1093/nar/gkh340.

Gaastra, W., Lipman, L.J.A., De Cock, A.W.A.M., Exel, T.K., Pegge, R.B.G., Scheurwater, J., Vilela, R., Mendoza, L., 2010. *Pythium insidiosum*: an overview. Vet. Microbiol. 146, 1–16. http://dx.doi.org/10.1016/j.vetmic.2010.07.019.

Grooters, A.M., Gee, M.K., 2002. Development of a nested polymerase chain reaction assay for the detection and identification of *Pythium insidiosum*. J. Vet. Intern. Med. 16, 147–152.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321. http://dx.doi.org/10.1093/sysbio/syq010.

Huson, D.H., Scornavacca, C., 2012. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Syst. Biol. 61, 1061–1067. http://dx.doi.org/10.1093/sysbio/sys062.

Inkomlue, R., Larbcharoensub, N., Karnsombut, P., Lerksuthirat, T., Aroonroch, R., Lohnoo, T., Yingyong, W., Santanirand, P., Sansopha, L., Krajaejun, T., 2016. Development of an anti-elicitin antibody-based immunohistochemical assay for diagnosis of pythiosis. J. Clin. Microbiol. 54, 43–48. http://dx.doi.org/10.1128/JCM.02113-15.

Intaramat, A., Sornprachum, T., Chantrathonkul, B., Chaisuriya, P., Lohnoo, T., Yingyong, W., Jongruja, N., Kumsang, Y., Sandee, A., Chaiprasert, A., Banyong, R., Santurio, J.M., Grooters, A.M., Ratanabanangkoon, K., Krajaejun, T., 2016. Protein A/G-based immunochromatographic test for serodiagnosis of pythiosis in human and animal subjects from Asia and Americas. Med. Mycol. 54, 641–647. http://dx.doi.org/10.1093/mmy/myw018.

Kammarnjesadakul, P., Palaga, T., Sritunyalucksana, K., Mendoza, L., Krajaejun, T., Vanittanakom, N., Tongchusak, S., Denduangboripant, J., Chindamporn, A., 2011. Phylogenetic analysis of *Pythium insidiosum* Thai strains using cytochrome oxidase II (COX II) DNA coding sequences and internal transcribed spacer regions (ITS). Med. Mycol. 49, 289–295. http://dx.doi.org/10.3109/13693786.2010.511282.

Kamoun, S., 2003. Molecular genetics of pathogenic oomycetes. Eukaryot. Cell 2, 191–199.

Keeratijarut, A., Karnsombut, P., Aroonroch, R., Srimuang, S., Sangruchi, T., Sansopha, L., Mootsikapun, P., Larbcharoensub, N., Krajaejun, T., 2009. Evaluation of an in-house immunoperoxidase staining assay for histodiagnosis of human pythiosis. Southeast Asian J. Trop. Med. Public Health 40, 1298–1305.

Keeratijarut, A., Lohnoo, T., Yingyong, W., Nampoon, U., Lerksuthirat, T., Onpaew, P., Chongtrakool, P., Krajaejun, T., 2014. PCR amplification of a putative gene for exo-1, 3-beta-glucanase to identify the pathogenic oomycete *Pythium insidiosum*. Asian Biomed. 8, 637–644.

Keeratijarut, A., Lohnoo, T., Yingyong, W., Rujirawat, T., Srichunrusami, C., Onpeaw, P., Chongtrakool, P., Brandhorst, T.T., Krajaejun, T., 2015. Detection of the oomycete *Pythium insidiosum* by real-time PCR targeting the gene coding for exo-1,3-β-glucanase. J. Med. Microbiol. 64, 971–977. http://dx.doi.org/10.1099/jmm.0.000117.

Kim, S., Misra, A., 2007. SNP genotyping: technologies and biomedical applications. Annu. Rev. Biomed. Eng. 9, 289–320. http://dx.doi.org/10.1146/annurev.bioeng.9.060906.152037.

Krajaejun, T., Pracharktam, R., Wongwaisayawan, S., Rochanawutinon, M., Kunakorn, M., Kunavisarut, S., 2004. Ocular pythiosis: is it under-diagnosed? Am J. Ophthalmol. 137, 370–372. http://dx.doi.org/10.1016/S0002-9394(03)00908-5.

Krajaejun, T., Sathapatayavongs, B., Pracharktam, R., Nitiyanant, P., Leelachaikul, P., Wanachiwanawin, W., Chaiprasert, A., Assanasen, P., Saipetch, M., Mootsikapun, P., Chetchotisakd, P., Lekhakula, A., Mitarnun, W., Kalnauwakul, S., Supparatpinyo, K., Chaiwarith, R., Chiewchanvit, S., Tananuvat, N., Srisiri, S., Suankratay, C., Kulwichit, W., Wongsaisuwan, M., Somkaew, S., 2006. Clinical and epidemiological analyses of human pythiosis in Thailand. Clin. Infect. Dis. 43, 569–576. http://dx.doi.org/10.1086/506353.

Krajaejun, T., Keeratijarut, A., Sriwanichrak, K., Lowhnoo, T., Rujirawat, T., Petchthong, T., Yingyong, W., Kalambaheti, T., Smittipat, N., Juthayothin, T., Sullivan, T.D., 2010. The 74-kilodalton immunodominant antigen of the pathogenic oomycete *Pythium insidiosum* is a putative exo-1,3-beta-glucanase. Clin. Vaccine Immunol. 17, 1203–1210. http://dx.doi.org/10.1128/CVI.00515-09.

Krajaejun, T., Satapatayavong, B., Sullivan, T., 2011. Chapter 96: pythium. In: Molecular Detection of Human Fungal Pathogens. CRC Press, pp. 851–863.

Lerksuthirat, T., Lohnoo, T., Rujirawat, T., Yingyong, W., Jongruja, N., Krajaejun, T., 2015. Geographic variation in the elicitin-like glycoprotein, ELI025, of *Pythium insidiosum* isolated from human and animal subjects. Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis. 35, 127–133. http://dx.doi.org/10.1016/j.meegid.2015.08.010.

Liu, W., Xie, Y., Ma, J., Luo, X., Nie, P., Zuo, Z., Lahrmann, U., Zhao, Q., Zheng, Y., Zhao, Y., Xue, Y., Ren, J., 2015. IBS: an illustrator for the presentation and visualization of biological sequences: fig. 1. Bioinformatics 31, 3359–3361. http://dx.doi.org/10.1093/bioinformatics/btv362.

Lohnoo, T., Jongruja, N., Rujirawat, T., Yingyon, W., Lerksuthirat, T., Nampoon, U., Kumsang, Y., Onpaew, P., Chongtrakool, P., Keeratijarut, A., Brandhorst, T.T., Krajaejun, T., 2014. Efficiency comparison of three methods for extracting genomic DNA of the pathogenic oomycete *Pythium insidiosum*. J. Med. Assoc. Thail. 97, 342–348.

Mendoza, L., Hernandez, F., Ajello, L., 1993. Life cycle of the human and animal oomycete pathogen *Pythium insidiosum*. J. Clin. Microbiol. 31, 2967–2973.

Mendoza, L., Ajello, L., McGinnis, M.R., 1996. Infection caused by the Oomycetous pathogen *Pythium insidiosum*. J. Mycol. Med. 6, 151–164.

Miller, R.I., 1983. Investigations into the biology of three “phycomycotic” agents pathogenic for horses in Australia. Mycopathologia 81, 23–28.

Presser, J.W., Goss, E.M., 2015. Environmental sampling reveals that *Pythium insidiosum* is ubiquitous and genetically diverse in North Central Florida. Med. Mycol. 53, 674–683. http://dx.doi.org/10.1093/mmy/myv054.

Ribeiro, T.C., Weiblen, C., de Azevedo, M.I., de Avila Botton, S., Robe, L.J., Pereira, D.I.B., Monteiro, D.U., Lorensetti, D.M., Santurio, J.M., 2016. Microevolutionary analyses of *Pythium insidiosum* isolates of Brazil and Thailand based on exo-1,3-β-glucanase gene. Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis. 48, 58–63. http://dx.doi.org/10.1016/j.meegid.2016.11.020.

Rujirawat, T., Patumcharoenpol, P., Lohnoo, T., Yingyong, W., Lerksuthirat, T., Tangphatsornruang, S., Suriyaphol, P., Grenville-Briggs, L.J., Garg, G., Kittichotirat, W., Krajaejun, T., 2015. Draft genome sequence of the pathogenic oomycete *Pythium insidiosum* strain Pi-S, isolated from a patient with Pythiosis. Genome Announc. 3. http://dx.doi.org/10.1128/genomeA.00574-15.

Schurko, A.M., Mendoza, L., Lévesque, C.A., Désaulniers, N.L., de Cock, A.W.A.M., Klassen, G.R., 2003. A molecular phylogeny of *Pythium insidiosum*. Mycol. Res. 107, 537–544.

Seeb, J.E., Carvalho, G., Hauser, L., Naish, K., Roberts, S., Seeb, L.W., 2011. Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms: INTRODUCTION. Mol. Ecol. Resour. 11, 1–8. http://dx.doi.org/10.1111/j.1755-0998.2010.02979.x.

Supabandhu, J., Fisher, M.C., Mendoza, L., Vanittanakom, N., 2008. Isolation and identification of the human pathogen *Pythium insidiosum* from environmental samples collected in Thai agricultural areas. Med. Mycol. 46, 41–52. http://dx.doi.org/10.1080/13693780701513840.

Vanittanakom, N., Szekely, J., Khanthawong, S., Sawutdeechaikul, P., Vanittanakom, P., Fisher, M.C., 2014. Molecular detection of *Pythium insidiosum* from soil in Thai agricultural areas. Int. J. Med. Microbiol. 304, 321–326. http://dx.doi.org/10.1016/j.ijmm.2013.11.016.

Vignal, A., Milan, D., SanCristobal, M., Eggen, A., 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genet. Sel. Evol. 34, 275. http://dx.doi.org/10.1186/1297-9686-34-3-275.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., Madden, T.L., 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13, 134. http://dx.doi.org/10.1186/1471-2105-13-134.

Data in Brief 15 (2017) 896–900

Contents lists available at ScienceDirect

Data in Brief

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

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).

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

https://doi.org/10.1016/j.dib.2017.10.047

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 *Pythium insidiosum* strain Pi45, which is categorized in the phylogenetically-distinct clade-III.
Experimental features	A rDNA-based phylogenetic tree of *P. insidiosum* was generated. Genome of the *P. insidiosum* strain Pi45 was sequenced and assembled. The reference genome sequence of the *P. insidiosum* strain Pi-S was mapped with sequence reads from the *P. insidiosum* 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 *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 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 publically 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

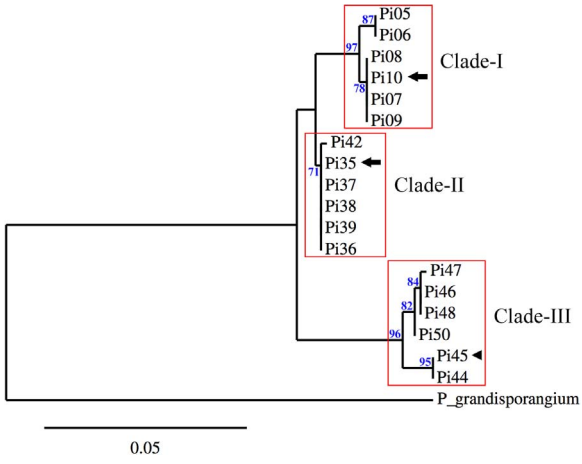


Fig. 1. Phylogenetic relationship of *Pythium insidiosum*: the rDNA-based maximum-likelihood phylogenetic tree categorizes 18 strains of *Pythium insidiosum* into three distinct clades: Clade-I, Clade-II, and Clade-III. Description of each strain of *P. insidiosum* can be found in Table 1. The arrows indicate the strains [i.e., CDC-B5653 (labeled as Pi10) and Pi-S (labeled as Pi35)] where genome sequences are publically available, while the arrow head indicates the strain Pi45 where genome data is reported here. The rDNA sequence from *Pythium granisporangium* (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; *N*50, 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. Pracharktam, P. Nitiyanant, P. Leelachaikul, W. Wanachiwanawin, A. Chaiprasert, 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. Pracharktam, S. Wongwaisayawan, M. Rochanawutinon, 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. Ascunce, 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. Yingyon, T. Lerksuthirat, U. Nampoon, Y. Kumsang, P. Onpaew, P. Chongtrakool, A. Keeratijarut, 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.

BMC Research Notes

DATA NOTE

Open Access

Data on whole genome sequencing of the oomycete *Pythium insidiosum* strain CBS 101555 from a horse with pythiosis in Brazil

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

Abstract

Objectives: The oomycete *Pythium insidiosum* infects humans and animals worldwide, and causes the life-threatening condition, called pythosis. Most patients lose infected organs or die from the disease. Comparative genomic analyses of different *P. insidiosum* strains could provide new insights into its pathobiology, and can lead to discovery of an effective treatment method. Several draft genomes of *P. insidiosum* are publicly available: three from Asia (Thailand), and one each from North (the United States) and Central (Costa Rica) Americas. We report another draft genome of *P. insidiosum* isolated from South America (Brazil), to serve as a resource for comprehensive genomic studies.

Data description: In this study, we report genome sequence of the *P. insidiosum* strain CBS 101555, isolated from a horse with pythiosis in Brazil. One paired-end (180-bp insert) library of processed genomic DNA was prepared for Illumina HiSeq 2500-based sequencing. Assembly of raw reads provided genome size of 48.9 Mb, comprising 60,602 contigs. A total of 23,254 genes were predicted and classified into 18,305 homologous gene clusters. Compared with the reference genome (the *P. insidiosum* strain Pi-S), 1,475,337 sequence variants (SNPs and INDELs) were identified in the organism. The genome sequence data has been deposited in DDBJ under the accession numbers BCFP01000001–BCFP01060602.

Keywords: *Pythium insidiosum*, Pythiosis, Oomycete, Genome, Gene cluster, Sequence variant

Objective

Pythium insidiosum is a fungus-like, aquatic, oomycetous microorganism that belongs to the kingdom Straminipila [1]. Microscopic features of *P. insidiosum* resemble that of filamentous fungi. The organism can be divided into three phylogenetical groups, in association with geographical origins: Clade-I strains (North, Central, and South Americas); Clade-II strains (Asia and Australia), and Clade-III strains (Thailand and the United States). In nature, *P. insidiosum* is observed in two forms: mycelium and zoospore (an infective unit) [2]. Several groups of investigators have successfully isolated *P. insidiosum* from swampy areas in Australia, Thailand, the United States, and Brazil [3–6]. While most pathogenic *Pythium* species infects plants, *P. insidiosum* infects humans and animals, and causes the life-threatening disease, called pythosis [7]. Case reports of the *P. insidiosum* infection in humans are almost exclusively from Asia, while that in animals are mainly from North, Central, and South Americas [1, 7]. Diagnosis of pythiosis is difficult. Treatment of this disease is challenging because effective drug and vaccine are lacking. Despite intensive cares are provided, most patients have their infected organs (i.e., eye, arm, leg) removed, and many patients die from the progressive infection [7].

Genome sequence can be used to explore pathobiology of an organism of interest. It is now feasible to sequence the genome of the non-model organism (i.e., *P.*

*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

insidiosum) using the next generation sequencing technologies. Comparative genomic analyses of different *P. insidiosum* strains could provide new insights into its biological processes and pathogenesis, which can lead to discovery of a novel method for pathogen control. Five draft genomes of *P. insidiosum* are deposited in the public repositories: three from Asia (Thailand; Clade-II and -III strains), and one each from North (the United States; Clade-I strain) and Central (Costa Rica; Clade-I strain) Americas [8–11]. Here, we report another draft genome data of *P. insidiosum* (Clade-I) isolated from South America (Brazil), as opposed to the other 5 strains (with published genome sequences) isolated from other regions of the world, to serve as a resource for comprehensive genomic studies in the future.

Data description

The *P. insidiosum* strain CBS 101555, isolated from a granulomatous lesion at the abdomen of a horse with pythiosis living in the southern region of Brazil, was cultured in Sabouraud dextrose broth at 37 °C for 1 week. Hyphal mat was harvested from the culture, and subjected to genomic deoxyribonucleic acid (DNA) extraction, using the conventional extraction method, optimized for *P. insidiosum* [12]. The identity of the strain was checked by single nucleotide polymorphism-based multiplex PCR and sequence homology analysis of the rDNA sequence (Accession number: AB971181) [13, 14]. The obtained genomic DNA was sequenced, using the Illumina next generation sequencing platform, as previously-described [8–10]. Briefly, the genomic DNA was processed to prepare a paired-end (180-bp insert) library for Illumina HiSeq 2500-based sequencing (Yourgene Bioscience, Taiwan). To guarantee read lengths of at least 35 bases, obtained raw reads underwent quality trims by CLC Genomics Workbench (Qiagen). The Cutadapt 1.8.1 [15] was used to remove the adaptor sequences. The resulting genome data contained 34,617,696 raw reads with an average length of 122 bases, providing 4,233,254,451 total bases. Genome assembly, performed by Velvet 1.2.10 [16], showed a total of 60,602 contigs, an average contig length of 806 bases (range 300–30,744), N_{50} of 953 bases, and 'N' composition of 0.9%. The draft assembled genome size of the organism was 48,855,945 bases. MAKER2 [17] predicted 23,254 genes in the draft genome. Basic Local Alignment Search Tool (BLAST) was used to annotate predicted genes by comparing to the NCBI non-redundant protein database using *E*-value cut off 10^{-6}. Product description of the best blast hit was used as the product description of the query gene. The genome sequence data has been deposited in the DNA Data Bank of Japan (DDBJ) under the Accession numbers BCFP01000001–BCFP01060602 (Data file 1; Table 1).

The 23,254 predicted genes can be classified into 18,305 homologous gene clusters (Data file 2; Table 1), using the method described by Kittichotirat et al. [18] and Rujirawat et al. [19], and the following setting: BLAST *E*-value of 10^{-6}, pairwise sequence identity of at least 30%, and pairwise alignment coverage for both query and subject sequences of at least 50%. Based on the BLAST search with *E*-value cut-off of 10^{-6} against the Clusters of Orthologous Groups of Proteins (COGs) database [20, 21], 3288 gene clusters (18%) were assigned to 24 COGs groups, while the rest (15,017 gene clusters [82%]; designated as uncharacterized cluster) did not match any COGs. Details on percentages and frequency of each assigned COGs group were shown in Data file 3 (Table 1).

The obtained draft genome was analysed for sequence variants, by using the Burrows–Wheeler Alignment tool [22]. Approximately, 44% of the processed reads (n = 15,084,792) of the *P. insidiosum* strain CBS 101555 can map the reference genome of the *P. insidiosum* strain Pi-S (the genome size of 53,239,050 bases, comprising 1192 contigs; Accession number BBXB00000000.1) [10]. FreeBayes [23] can identify 1,475,337 sequence variants, including single-nucleotide polymorphisms (SNPs) and insertion/deletion of bases (INDELs), in the genome of the organism (Data file 4; Table 1).

In conclusion, *P. insidiosum* is an understudied pathogen that causes the life-threatening condition, called

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	Whole genome sequence	FASTA	DDBJ (Accession numbers: BCFP01000001–BCFP01060602) (http://getentry.ddbj.nig.ac.jp/)
Data file 2	Gene clusters	MS Excel file (.xlsx)	Mendeley database (https://doi.org/10.17632/yjyzx5gk7s.1) (https://data.mendeley.com/datasets/yjyzx5gk7s/1)
Data file 3	Clusters of Orthologous Groups of Proteins (COGs)	MS Excel file (.xlsx)	Mendeley database (https://doi.org/10.17632/5rhfd4n37k.1) (https://data.mendeley.com/datasets/5rhfd4n37k/1)
Data file 4	Sequence variants	MS Excel file (.xlsx)	Mendeley database (https://doi.org/10.17632/4y8hdw7tb7.1) (https://data.mendeley.com/datasets/4y8hdw7tb7/1)

pythiosis, in humans and animals worldwide. We sequenced the draft genome of the *P. insidiosum* strain CBS 101555, isolated from a pythiosis horse living in the southern region of Brazil. The obtained genome will be a fundamental resource for exploring biology and pathogenesis of this invasive microorganism.

Limitations

The draft genome was obtained from short-read assembly of one Illumina-based paired-end (180-bp insert) library, without any mate pair library, resulting in as many as 60,602 contigs. The estimated genomic coverage is limited to ~ 87-fold. The mitochondrial genome sequences were not excluded from the nuclear genome assembly.

Abbreviations

DNA: deoxyribonucleic acid; DDBJ: DNA Data Bank of Japan; BLAST: Basic Local Alignment Search Tool; COGs: Clusters of Orthologous Groups of Proteins; SNP: single-nucleotide polymorphism; INDEL: insertion or deletion of bases.

Authors' contributions

WK and TK conceived the project. WK, PP, TR, TL, and WY performed the experiments. WK, PP, TR, and TK analyzed the data. WK and TK wrote the manuscript. WK and TK acquired the research funds. All authors read and approved the final manuscript.

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] Research Center, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand.

Acknowledgements

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Availability of data materials

The genome data described in this study can be accessed online at DDBJ (the draft genome sequence; Accession numbers BCFP01000001–BCFP01060602) and Mendeley database (i.e., gene clusters [https://doi.org/10.17632/yjyzx5gk7s.1], COGs [https://doi.org/10.17632/5rhfd4n37k.1], and sequence variants [https://doi.org/10.17632/4y8hdw7tb7.1]).

Consent for publication

Not applicable.

Ethics approval and consent to participate

This study is a part of the research project that has been approved by the Committee on Human Rights Related to Research Involving Human Subjects, at the Faculty of Medicine, Ramathibodi Hospital, Mahidol University (Protocol Number: ID 05-60-77).

Funding

This work was supported by the Faculty of Medicine, Ramathibodi Hospital, Mahidol University [Grant numbers CF_60001 and CF_61007] and the Thailand Research Fund [Grant Number BRG5980009]. The authors acknowledge the financial support provided by King Mongkut's University of Technology Thonburi through the "KMUTT 55th Anniversary Commemorative Fund". The funders had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Publisher's Note

Received: 8 October 2018 Accepted: 29 November 2018
Published online: 11 December 2018

References

1. Gaastra W, Lipman LJA, De Cock AWAM, Exel TK, Pegge RBG, Scheurwater J, et al. *Pythium insidiosum*: an overview. Vet Microbiol. 2010;146:1–16.
2. Mendoza L, Hernandez F, Ajello L. Life cycle of the human and animal oomycete pathogen *Pythium insidiosum*. J Clin Microbiol. 1993;31:2967–73.
3. Supabandhu J, Fisher MC, Mendoza L, Vanittanakom N. Isolation and identification of the human pathogen *Pythium insidiosum* from environmental samples collected in Thai agricultural areas. Med Mycol. 2008;46:41–52.
4. Presser JW, Goss EM. Environmental sampling reveals that *Pythium insidiosum* is ubiquitous and genetically diverse in North Central Florida. Med Mycol. 2015;53:674–83.
5. Miller RI. Investigations into the biology of three "phycomycotic" agents pathogenic for horses in Australia. Mycopathologia. 1983;81:23–8.
6. Zambrano CG, Fonseca AOS, Valente JSS, Braga CQ, Sallis ESV, Azevedo MI, et al. Isolation and characterization of Pythium species from swampy areas in the Rio Grande do Sul, Brazil, and evaluation of pathogenicity in an experimental model. Pesquisa Veterinária Brasileira. 2017;37:459–64.
7. 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.
8. 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 in Brief. 2017;15:896–900.
9. 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 in Brief. 2018;16:47–50.
10. 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.
11. 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. Genom Data. 2016;7:60–1.
12. 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.
13. 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.
14. Chaiprasert A, Krajaejun T, Pannanusorn S, Prariyachatigul C, Wanachiwanawin W, Sathapatayavongs B, et al. *Pythium insidiosum* Thai isolates: molecular phylogenetic analysis. Asian Biomed. 2009;3:623–33.
15. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17:10.
16. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de *Bruijn graphs*. Genome Res. 2008;18:821–9.
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. Kittichotirat W, Bumgarner RE, Asikainen S, Chen C. Identification of the pangenome and its components in 14 distinct *Aggregatibacter actinomycetemcomitans* strains by comparative genomic analysis. PLoS ONE. 2011;6:e22420.

19. 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.
20. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, et al. The COG database: an updated version includes eukaryotes. BMC Bioinform. 2003;4:41.
21. Galperin MY, Makarova KS, Wolf YI, Koonin EV. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res. 2015;43((Database issue)):D261–9.
22. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009;25:1754–60.
23. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv preprint. 2012. arXiv:1207.3907[q-bio.GN].

Data in Brief 16 (2018) 47–50

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Draft genome sequences of the oomycete *Pythium insidiosum* strain CBS 573.85 from a horse with pythiosis and strain CR02 from the environment

Preecha Patumcharoenpol [a,b], Thidarat Rujirawat [c], Tassanee Lohnoo [c], Wanta Yingyong [c], Nongnuch Vanittanakom [d], Weerayuth Kittichotirat [a,*], Theerapong Krajaejun [e,*]

[a] *Systems Biology and Bioinformatics Research Group, Pilot Plant Development and Training Institute, King Mongkut's University of Technology Thonburi, 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 Microbiology, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand*
[e] *Department of Pathology, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand*

ARTICLE INFO

Article history:
Received 4 October 2017
Received in revised form
5 October 2017
Accepted 2 November 2017
Available online 7 November 2017

Keywords:
Pythium insidiosum
Pythiosis
Draft genome sequence

ABSTRACT

Pythium insidiosum is an aquatic oomycete microorganism that causes the fatal infectious disease, pythiosis, in humans and animals. The organism has been successfully isolated from the environment worldwide. Diagnosis and treatment of pythiosis is difficult and challenging. Genome sequences of *P. insidiosum*, isolated from humans, are available and accessible in public databases. To further facilitate biology-, pathogenicity-, and evolution-related genomic and genetic studies of *P. insidiosum*, we report two additional draft genome sequences of the *P. insidiosum* strain CBS 573.85 (35.6 Mb in size; accession number, BCFO00000000.1) isolated from a horse with pythiosis, and strain CR02 (37.7 Mb in size; accession number, BCFR00000000.1) isolated from the environment.

* Corresponding authors.
E-mail addresses: weerayuth.kit@kmutt.ac.th (W. Kittichotirat), mr_en@hotmail.com (T. Krajaejun).

https://doi.org/10.1016/j.dib.2017.11.002

Specifications Table

Subject area	Biology
More specific subject area	Microbiology, Genomics
Type of data	Genome sequence data
How data was acquired	IlluminaHiSeq 2000 and IlluminaHiSeq 2500 Next Generation Sequencing Platforms
Data format	Assembled genome sequences
Experimental factors	Genomic DNA was extracted from the *Pythium insidiosum* strains CBS 573.85 (an animal isolate) and CR02 (an environmental isolate).
Experimental features	Genome of the *P. insidiosum* strains CBS 573.85 and CR02 were sequenced and assembled.
Data source location	*Pythium insidiosum* strain CBS 573.85 was isolated from a horse in Costa Rica, and strain CR02 was isolated from the environment in Thailand.
Data accessibility	The draft genome sequences of *P. insidiosum* have been deposited in the DNA Data Bank of Japan (DDBJ) under the accession numbers: BCFO00000000.1 (strain CBS573.85; https://www.ncbi.nlm.nih.gov/nuccore/BCFO00000000.1) and BCFR00000000.1 (strain CR02; https://www.ncbi.nlm.nih.gov/nuccore/BCFR00000000.1).

Value of the data

- Previously, only genome sequence data of *P. insidiosum* isolated from humans is available in the public databases.
- The first draft genome sequences of *P. insidiosum* isolated from a non-human animal with pythiosis and from the environment are now made available.
- The additional genome data will facilitate biology-, pathogenicity-, and evolution-related studies of *P. insidiosum*, through comparative genomic studies of *Pythium* species and related species.

1. Data

Pythium insidiosum is an aquatic oomycete microorganism that causes the lethal infectious condition, pythiosis, in humans and other animals [1,2]. The organism has been isolated from the environment in Australia, Thailand, Brazil and the United States [3–6]. Genome sequences of *P. insidiosum*, isolated from humans, are available and accessible in public databases [7,8]. We report two additional draft genome sequences of the organism isolated from a horse with pythiosis, as well as from the environment.

2. Experimental design, materials and methods

2.1. Genome sequencing and assembly

Genomic DNA samples were extracted from *P. insidiosum* strain CBS573.85 (from an infected horse in Costa Rica) and strain CR02 (from an agricultural area in Thailand), using the conventional extraction method described by Lohnoo and co-workers [9]. rDNA sequence analysis was performed to confirm the identity of the organism [10–12]. The extracted genomic DNA of each of these two strains was subjected to preparation of a paired-end library for genome sequencing, using the IlluminaHiSeq 2500 (strain CBS573.85) or IlluminaHiSeq 2000 (strain CR02) platform (Yourgene Bioscience, Taiwan). Quality trims of the raw reads were executed by CLC Genomics Workbench

(www.clcbio.com) to yield read lengths with at least 35 bases. The adaptor sequences were removed by Cutadapt 1.8.1 [13] to obtain a total of 34,651,034 reads with an average read length of 122 bases (4,238,414,330 total bases) for the strain CBS573.85, and a total of 27,436,541 reads with an average read length of 105 bases (2,888,290,738 total bases) for the strain CR02. All reads were assembled by Velvet 1.2.10 [14] to the total sequence length of 35,561,321 bases (number of contigs, 11,223; average contig length, 3169; N_{50}, 12,261; 'N' composition, 1.2%) for the strain CBS573.85, and 37,673,126 bases (number of contigs, 22,560; average contig length, 1670; N_{50}, 3553; 'N' composition, 2.7%) for the strain CR02. CEGMA analysis with 248 highly-conserved eukaryotic genes [15,16] reported 87% and 77% completeness of draft genomes of the strains CBS573.85 and CR02, respectively. MAKER2 [17] assigned 14,487 (strain CBS573.85) and 15,231 (strain CR02) open reading frames.

2.2. Data accessibility

The genome sequence data has been deposited in DDBJ under the accession numbers BCFO00000000.1 (strain CBS573.85) and BCFR00000000.1 (strain CR02).

Acknowledgements

This work was supported by the Faculty of Medicine, Ramathibodi Hospital, Mahidol University; the Thailand Research Fund (Grant no. BRG5980009); and the Mahidol University (Grant no. ngor-por 09/2560). We thank Thomas D. Sullivan for reviewing the manuscript.

Transparency document. Supplementary material

Transparency document associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2017.11.002.

References

[1] T. Krajaejun, B. Sathapatayavongs, R. Pracharktam, P. Nitiyanant, P. Leelachaikul, W. Wanachiwanawin, A. Chaiprasert, 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.
[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] R.I. Miller, Investigations into the biology of three "phycomycotic" agents pathogenic for horses in Australia, Mycopathologia. 81 (1983) 23–28.
[4] J. Supabandhu, M.C. Fisher, L. Mendoza, N. Vanittanakom, Isolation and identification of the human pathogen *Pythium insidiosum* from environmental samples collected in Thai agricultural areas, Med. Mycol. 46 (2008) 41–52.
[5] J.W. Presser, E.M. Goss, Environmental sampling reveals that *Pythium insidiosum* is ubiquitous and genetically diverse in North Central Florida, Med. Mycol. 53 (2015) 674–683.
[6] C.G. Zambrano, A.O.S. Fonseca, J.S.S. Valente, C.Q. Braga, E.S.V. Sallis, M.I. Azevedo, C. Weiblen, J.M. Santurio, S.A. Botton, D.I. B. Pereira, Isolation and characterization of Pythium species from swampy areas in the Rio Grande do Sul, Brazil, and evaluation of pathogenicity in an experimental model, Pesqui. Veter. Bras. 37 (2017) 459–464.
[7] M.S. Ascunce, 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, Genome Data 7 (2016) 60–61.
[8] 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).
[9] T. Lohnoo, N. Jongruja, T. Rujirawat, W. Yingyon, T. Lerksuthirat, U. Nampoon, Y. Kumsang, P. Onpaew, P. Chongtrakool, A. Keeratijarut, 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.
[10] T. Rujirawat, T. Sridapan, T. Lohnoo, W. Yingyong, Y. Kumsang, P. Sae-Chew, W. Tonpitak, T. Krajaejun, Single nucleotide polymorphism-based multiplex PCR for identification and genotyping of the oomycete *Pythium insidiosum* from humans, animals and the environment, Infect. Genet. Evol. 54 (2017) 429–436.

[11] P.R. Badenoch, D.J. Coster, B.L. Wetherall, H.T. Brettig, M.A. Rozenbilds, A. Drenth, G. Wagels, *Pythium insidiosum* keratitis confirmed by DNA sequence analysis, Br. J. Ophthalmol. 85 (2001) 502–503.
[12] A. Chaiprasert, T. Krajaejun, S. Pannanusorn, C. Prariyachatigul, W. Wanachiwanawin, B. Sathapatayavongs, T. Juthayothin, N. Smittipat, N. Vanittanakom, A. Chindamporn, *Pythium insidiosum* Thai isolates: molecular phylogenetic analysis, Asian Biomed. 3 (2009) 623–633.
[13] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, EMBnet J. 17 (2011) 10.
[14] D.R. Zerbino, E. Birney, Velvet: algorithms for de novo short read assembly using de Bruijn graphs, Genome Res. 18 (2008) 821–829.
[15] G. Parra, K. Bradnam, Z. Ning, T. Keane, I. Korf, Assessing the gene space in draft genomes, Nucleic Acids Res. 37 (2008) 289–297.
[16] G. Parra, K. Bradnam, I. Korf, CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes, Bioinformatics 23 (2007) 1061–1067. http://dx.doi.org/10.1093/bioinformatics/btm071.
[17] C. Holt, M. Yandell, MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects, BMC Bioinform. 12 (2011) 491.

DATABASE - Decision on Manuscript ID DATABASE-2018-0165.R1

DATABASE <onbehalfof@manuscriptcentral.com>
Thu 2019-05-30 08:09

To: mr_en@hotmail.com <mr_en@hotmail.com>
Cc: database.editorialoffice@oup.com <database.editorialoffice@oup.com>

29-May-2019

Dear Dr. Krajaejun,

It is a pleasure to accept your manuscript entitled "Oomycete Gene Table: An Online Database for Comparative Genomic Analyses of the Oomycete Microorganisms" in its current form for publication in DATABASE. The comments of the reviewer(s) who reviewed your manuscript are included at the foot of this letter.

In line with the expectations and standards of the community, authors are expected to ensure their databases and/or online resources remain available for at least 2 years following publication of their paper in DATABASE.

In order to publish your article, Oxford University Press requires that you complete a licence agreement online. A link to the online licensing system, and instructions on how to select and complete a licence, will be provided to you by the Production Editor at Oxford University Press in due course. You will receive your official acceptance date from Oxford University Press once you have signed your licence to publish. (N.B. If you are a UK-based author and are looking to comply with the HEFCE policy on open access in the Research Excellence Framework, you should use the official acceptance date when depositing in your repository).

All DATABASE articles are published under an open access model and are freely available to everybody online upon publication. Author publication charges make this possible. The open access publication charge per article is approximately 800 GBP/1600 USD/1200 EURO. When you sign the licence to publish online you will be prompted to proceed to the payment system, where you can pay online using a credit card, or request an invoice by email or post.

On behalf of the Editors of DATABASE, we look forward to your continued contributions to the Journal.

Yours sincerely,

David Landsman

==========
David Landsman, PhD
Editor-in-Chief
DATABASE - The Journal of Biological Databases and Curation
http://database.oxfordjournals.org/
==========

Reviewer(s)' Comments to Author:

Reviewer: 2

Comments to the Author
The authors adequately addressed all the comments and overall present a straight forward tutorial of an already existing tool. As noted, the Oomycete Gene Table complement other tools and provide additional options for oomycete genomics.

Reviewer: 1

Comments to the Author
Most of my concerns have been addressed.

Note the BLASTp link does not always work:
eg in p-cluster002929
See this error:
Software error:
---------- EXCEPTION: Bio::Root::Exception ----------
MSG: Could not open index file
../../../genomes/oomycetes//fungidb41/PultimumDAOMBR144/anndir/fasta/genes.faa.index: Permission denied
STACK: Error::throw

Title: Oomycete Gene Table: An Online Database for Comparative Genomic Analyses of the Oomycete Microorganisms

Authors: Thidarat Rujirawat[1,2,3], Preecha Patumcharoenpol[4], Weerayuth Kittichotirat[4,*], Theerapong Krajaejun[1,*]

Affiliations: [1]Department of Pathology, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; [2]Research Center, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; [3]Molecular Medicine Program, Multidisciplinary Unit, Faculty of Science, Mahidol University, Bangkok, Thailand; [4]Systems Biology and Bioinformatics Research Group, Pilot Plant Development and Training Institute, King Mongkut's University of Technology Thonburi, Bangkhuntien, Bangkok, Thailand;

***Corresponding authors:** Weerayuth Kittichotirat, weerayuth.kit@kmutt.ac.th; Theerapong Krajaejun, mr_en@hotmail.com

Keywords: Oomycete Gene Table, Oomycetes, Genome, Database, Comparative genomic analyses

Abstract

Oomycetes form a unique group of the fungal-like, aquatic, eukaryotic microorganisms. Lifestyle and pathogenicity of the oomycetes are diverse. Many pathogenic oomycetes affect a broad range of plants and cause enormous economic loss annually. Some pathogenic oomycetes cause destructive and deadly diseases in a variety of animals, including humans. No effective antimicrobial agent against the oomycetes is available. Genomic data of many oomycetes are currently available. Comparative analyses of the oomycete genomes must be performed to better understand the oomycete biology and virulence, as well as, to identify conserved and biologically-important proteins that are potential diagnostic and therapeutic targets of these organisms. However, a tool that facilitates comparative genomic studies of the oomycetes is lacking. Here, we described in detail the Oomycete Gene Table, which is an online user-friendly bioinformatic tool, designed to search, analyze, compare, and visualize gene contents of 20 oomycetes in a customizable table. Genomic contents of other oomycete species, when available, can be added to the existing database. Some of the applications of the Oomycete Gene Table include investigations of phylogenomic relationships, as well as identifications of biologically-important and pathogenesis-related genes of oomycetes. In summary, the Oomycete Gene Table is a simple and useful tool for comparative genomic analyses of oomycetes. The Oomycete Gene Table is available online at http://www.sbi.kmutt.ac.th/cgi-bin/gt/viewer?organism=oomycetes&build=190208.

Introduction

Oomycetes are fungal-like, aquatic, eukaryotic microorganisms that belong to the Stramenopiles-Alveolates-Rhizaria supergroup (1). Oomycetes include the organisms of the genera *Albugo*, *Aphanomyces*, *Bremia*, *Hyaloperonospora*, *Lagenidium*, *Peronospora*, *Phytophthora*, *Phytopythium, Plasmopara*, *Pythium* and *Saprolegnia* (2,3). Lifestyle and pathogenicity of the oomycetes are diverse. Many pathogenic oomycetes (e.g. *Phytophthora* and *Pythium* species) affect a broad range of plants and cause enormous economic loss each year (4). Some pathogenic oomycetes (e.g. *Pythium insidiosum*, *Pythium aphanidermatum*, *Aphanomyces invadans*, *Saprolegnia parasitica*, and *Lagenidium* species) infect a variety of animals, including humans, and can lead to a deadly disease (5–10). No effective antimicrobial agent against

oomycetes is available, making the control of infections caused by these organisms difficult and challenging.

Genomes of many oomycetes have been sequenced and deposited in public repositories, such as, the National Center for Biotechnology Information, the DNA Data Bank of Japan, and the European Bioinformatics Institute (6,11–22). Availability of the genome data provides opportunities to explore biology, pathogenicity, and evolution of oomycetes, which can lead to the discovery of a novel method for efficient control of the infections caused by these organisms. Comparative analyses of oomycete genomes must be performed to better understand the oomycete biology and virulence, as well as, to identify conserved and biologically-important proteins that are potential diagnostic and therapeutic targets of these organisms. An array of bioinformatic tools and well-trained personals are usually required for such complicated analyses. A few web-based tools, such as, FungiDB (23) and EuMicrobeDBLite (24), have been developed for genomic analysis of oomycetes. These online tools provide bioinformatic features, which include sequence retrieval, BLAST search engine, gene annotation, gene location, and gene and protein architecture. However, a bioinformatic tool that facilitates the comparative genomic studies of oomycetes is still lacking.

Recently, we have developed an online user-friendly bioinformatic tool, called 'Oomycete Gene Table', to facilitate the comparative genomic analysis of the human-pathogen *Pythium insidiosum* and 19 other oomycetes as well as 2 diatoms (3). The Oomycete Gene Table can be used to search, retrieve, sort, filter, compare, and display gene contents of the oomycete genome(s) of interest in a customizable and easy-to-understand table. The Oomycete Gene Table has been successfully used to investigate the phylogenomic relationships and identifying the conserve and species-specific genes of *Py. insidiosum* (3,25). As only a few bioinformatic tools for oomycetes are available, the Oomycete Gene Table is an additional, simple, and useful tool that can be applied for the genomic analyses of oomycetes. The key features of the Oomycete Gene Table are described in detail here to facilitate its utilization for genomic analyses of oomycetes. The Oomycete Gene Table is available online at http://www.sbi.kmutt.ac.th/cgi-bin/gt/viewer?organism=oomycetes&build=190208.

Contents of the Oomycete Gene Table

The Oomycete Gene Table was constructed based on sequence similarity analyses of gene contents from 20 oomycetes and 2 diatoms, using an algorithm described by Kittichotirat et al. (26). Briefly, genes were clustered based on four different BLAST strategies using *E*-value of $\leq 10^{-6}$, pairwise sequence identity of $\geq 30\%$, and pairwise sequence alignment coverage of $\geq 50\%$ (3). This clustering process was iterated until the gene members of all clusters are stable or no longer change. The Oomycete Gene Table contains a total of 382,450 protein-coding genes from these 22 microorganisms, which have been grouped into 105,541 unique gene clusters. In the Oomycete Gene Table, the rows show gene clusters while the columns show (i) Row numbers, (ii) Cluster identification numbers (ID), (iii) Annotations based on Clusters of orthologous groups of proteins (COG) (27,28), (iv) Functional descriptions, (v) Expected gene lengths, and (vi) the organism names included in the table (**Figure 1)**. Cluster IDs are unique identifiers that have been assigned to each of the 105,541 gene clusters (p-cluster000001-105541)

In the 'COG annotation' column, each code represents different COG-based protein description (e.g. COG0514 is for Superfamily II DNA helicase; COG1960 is for Acyl-CoA dehydrogenases; and COG0828 is for Ribosomal protein S21). Each COG annotation was assigned based on a BLASTP analysis of the longest protein from each gene cluster against the COG database (the BLASTP setting: sequence coverage $\geq 80\%$; protein identity $\geq 30\%$) (27,28). The letter in square brackets indicates a COG category (e.g. [A] is for RNA processing and modification, [I] is for Lipid transport and metabolism, and [K] is for Transcription). The 'Functional description' column displays protein functional information of the gene clusters.

Each gene cluster functional description is assigned based the most represented functional annotation among the gene members within the cluster (i.e. majority rule). Note that gene functional annotations from source databases (e.g. FungiDB, NCBI, and Joint Genome Institute) are directly used without any modification for consistency purpose. The 'Expected gene length' column indicates the length (in base pairs) of the longest gene in each gene cluster.

Each cell in the Oomycete Gene Table provides status and data of an individual gene, such as, gene ID, length, location, CG content, annotation, and sequence download links (both DNA and Protein). In each gene cluster, a light-gray cell indicates that one organism contains a sequence without a complete open reading frame that has significant sequence similarity to the genes of other organisms in the corresponding gene cluster. The lack of a complete open reading frame (or a gene) in one organism but not in the other organisms may due to the evolutionary loss of that particular gene from the genome or the incompleteness of the analyzed genome sequence. A black cell indicates that a gene or a similar sequence is absent from an organism.

User interface

The user interface of the Oomycete Gene Table comprises six drop-down windows (**Figure 2**), which can be used to search, analyze, and display the gene contents of the organisms included in the Oomycete Gene Table. The 'Show all options' and 'Hide all options' buttons can be used respectively to show and hide the 'query input' boxes and customizable options. The 'Refresh Gene Table' button can be used to update the data display of the Oomycete Gene Table according to the currently selected options. All display parameters can be managed and customized as described in detail below.

Keyword Search

Keywords can be used to search the gene contents stored in the Oomycete Gene Table (**Figure 2a**). Such keywords are cluster ID (e.g. p-cluster007458, p-cluster024940), COG IDs (e.g. COG0225, COG2183, COG5188), or protein/gene name, protein/gene family name or protein/gene identification number (e.g. chitinase, exo1, THAPS_005625). Multiple keywords can be used simultaneously to search the Oomycete Gene Table by separating the keywords with a comma, for example: 'p-cluster000001, THAPS_005625, adhesin' (**Figure 3**). Keyword search can be performed on the whole Oomycete Gene Table or focused on a specific genome by selecting from the drop-down box accordingly.

BLAST search

For the 'BLAST search' window (**Figure 2b**), either a DNA or a protein query sequence can be entered to search for its homologous sequences in the genomes that are available in the Oomycete Gene Table. Currently, only one sequence is allowed in each search. The BLAST search result shows matched organism(s), gene/protein names, bit score(s), and *E*-value(s). Each significant match is supplemented with a link to pair-wise sequence alignment of the query and subject sequences, as well as a link to view on the Oomycete Gene Table.

Row sorting and cell coloring

The number, order, and cell color of the displayed gene clusters can be managed and customized using the 'Row sorting and cell coloring' options (**Figure 2c**). By default, the Oomycete Gene Table shows 50 rows of gene clusters per page in ascending order of the Cluster IDs. The cells in the displayed table are color-coded to indicate the percentage of gene completeness as compared to the Expected gene length (which is the longest gene in each gene cluster). In addition to the Cluster IDs, the displayed gene clusters can be sorted according to COG annotations, functional descriptions, expected gene lengths, and gene positions of a selected organism (**Figure 4a**). In addition, each organism name shown in the column header can be

clicked to show sorting options based on gene positions or gene copy numbers in the corresponding genome (**Figure 4b**). The cells of the Oomycete Gene Table can be marked with different colors to demonstrate the selected property of the displayed genes (e.g. percentage of gene completeness, gene order in a genome, gene copy number, or CG content) (**Figure 4c**).

Row filtering

The 'Row filtering' options can be used to filter the gene contents and list only the gene clusters with a desired COG category (**Figure 2d**). The drop-down window shows COG functional categories (27,28) together with numbers and percentages of gene clusters that have been assigned to them **(Figure 5**). A minimal length of the displayed genes can be specified by putting in a desired gene length (in base pairs) in the provided space. There are also options for displaying clusters that contain genes that are located within 300 bp from the ends of the genomic contigs (an asterisk is marked in the cell) as well as genes with possible frameshift mutation (the letter 'F' is marked in the cell).

'Column(s) to display' and 'Show genes only present in' options

The Oomycete Gene Table is set by default to display the gene contents of all organisms (see the organism columns in **Figure 1**). However, users can select and show only gene contents of the organism(s) of interest by checking the desired organism(s) shown in the 'Column(s) to display' box (**Figure 2e**). For example, when '*Phytophthora infestans*' and '*Pythium ultimum*' are checked, only gene contents of these two organisms are displayed upon refreshing the Oomycete Gene Table. Links to the source databases of the available genomes are also provided in this box.

The 'Show genes only present in' window (**Figure 2f**) can be used to display genes that are present only in the selected organism(s). For example, when '*Pythium insidiosum*' is checked, a total of 998 *Py. insidiosum*-specific gene clusters are shown in the Oomycete Gene Table. If multiple organisms are selected, only the gene clusters that are present in the selected organisms and absent from the other unselected organisms are displayed. For example, when '*Pythium. insidiosum*' and '*Pythium aphanidermatum*' are checked, a total of 91 gene clusters that are specific to just these two oomycetes are shown. In addition, if the 'All genomes (Core genes)' box is checked, only the set of gene clusters that are present in all organisms are visualized. In contrast, if the 'Some genomes (Variable genes)' box is checked, only gene clusters that are present in some but not all of the organisms are displayed.

Retrieval of genomic data

Various genomic data can be retrieved from the Oomycete Gene Table. By clicking a cluster ID of interest (e.g. p-cluster097753; **Figure 6a**), a pop-up window that contains links for obtaining an overview as well as displaying sequences (in FASTA format) of all genes/proteins belonging to the corresponding gene cluster is shown. The overview page contains detail of gene members such as genome IDs, gene/contig IDs, gene locations, gene lengths, functional descriptions, links to gene/protein sequences, links to source databases, and links to NCBI Protein-BLAST tool (**Figure 6b**). These information can also be retrieved by clicking a cell on the displayed Gene Table. For example, **Figure 6c** shows the retrievable gene data of '*Saprolegnia declina*' that is found in the gene cluster 'p-cluster097753'.

Discussion

Gene gain, loss, and modification can accumulatively occur in different genomes, and contribute to variations in lifestyles, pathogenicity, and host specificities of the saprophytic and pathogenic oomycetes. Genome sequences of many oomycetes (6,11–22) are now available and serve as an important resource for comparative genomic studies. Among the microbiologists and

researchers in the field of oomycete molecular genetics, the lack of bioinformatic experiences and technical skills limits their capabilities to thoroughly use the available genome data to explore biology, pathogenicity, and evolution of oomycetes. We have developed an online user-friendly bioinformatic tool called 'Oomycete Gene Table' to facilitate comparative analysis and data retrieval of oomycete genomes (3). The interface of the Oomycete Gene Table has been designed for users who are unfamiliar with or have limited experiences in big data mining and analysis of multiple genomes.

A large repertoire of ~380,000 genes from 20 oomycetes and 2 diatom genomes that have been grouped into ~100,000 homologous gene clusters can be found and analyzed in the Oomycete Gene Table. A number of useful features have been implemented, which allow users to easily search, retrieve, and compare gene(s) of interest across the available genomes. Core genes of selected oomycetes can be extracted from the Oomycete Gene Table for phylogenomic analyses. Some core genes of oomycetes can be potential targets for drug design and development. Species-specific genes are important for elucidating the lifestyle and pathogenicity of a particular oomycete. For example, unlike most other pathogenic oomycetes that infect plants, *Py. insidiosum* and *Py. aphanidermatum* are capable of infecting humans. The Oomycete Gene Table can identify ~90 gene clusters that are only present in these two oomycetes, which may possibly contribute to their specific virulence in the human host.

Other applications of the Oomycete Gene Table include the identification of elicitins (3) and ureases (25) among oomycetes. Elicitins form a protein family that is originally identified in *Phytophthora* and *Pythium* species but not in other oomycetes or other organisms (29–31). The Oomycete Gene Table has been used to explore the presence of elicitins in various oomycetes (3). By simply using the keyword 'elicitin' in a search (as demonstrated in **Figure 3**), the result shows that *Phytophthora* and *Pythium* species contain a more extensive repertoire of the elicitin genes than the other oomycetes. Regarding the ureases, some pathogens (e.g. the bacterium *Helicobacter pylori* and the fungus *Cryptococcus neoformans*) produce urease as a virulence factor to facilitate their pathogenesis (32–34). Since the role of urease in biology and pathogenesis of oomycetes has not been characterized, the Oomycete Gene Table was searched by using the keyword 'urease' to identify urease-encoding genes in oomycetes (25). The proper function of urease requires the presence of several urease accessory proteins (e.g. urease accessory protein D, F, and G) (35). The Oomycete Gene Table shows that, although the ureases are conserved in oomycetes, the presence of urease accessory protein-encoding genes are markedly diverse in these organisms (25).

As shown above, the Oomycete Gene Table is a useful bioinformatic tool. However, it is important to note that the quality of the genome sequences included in the database, all of which are still incomplete, may compromise the reliability of analysis results. For example, an absence of a gene in one organism may actually be due to missing or unknown genome sequence. It is also important to note that since genome data in the Oomycete Gene Table are obtained from various public resources, they may not be annotated with the same annotation pipeline. Users should always consider these limitations when interpreting the comparative analysis results. As genome sequencing technologies continue to improve in term of cost and efficiency, better quality genome sequences of oomycetes will be available and can be added into the Oomycete Gene Table to improve the database quality.

In conclusion, we described here the Oomycete Gene Table, which is an online bioinformatic tool for comparative analysis and data retrieval of oomycete genomes. The Oomycete Gene Table is accessible via a web browser and its functionalities can be operated via an easy-to-use interface. The inexperienced users (e.g. microbiologists, researchers, and clinicians) in the field can seamlessly analyze and compare the big data of multiple oomycete genomes. The genomic data in the Oomycete Gene Table will be updated regularly to increase

their completeness and accuracy, which will allow the Oomycete Gene Table to serve as a useful bioinformatic tool for exploring the biology, pathogenicity, and evolution of oomycetes.

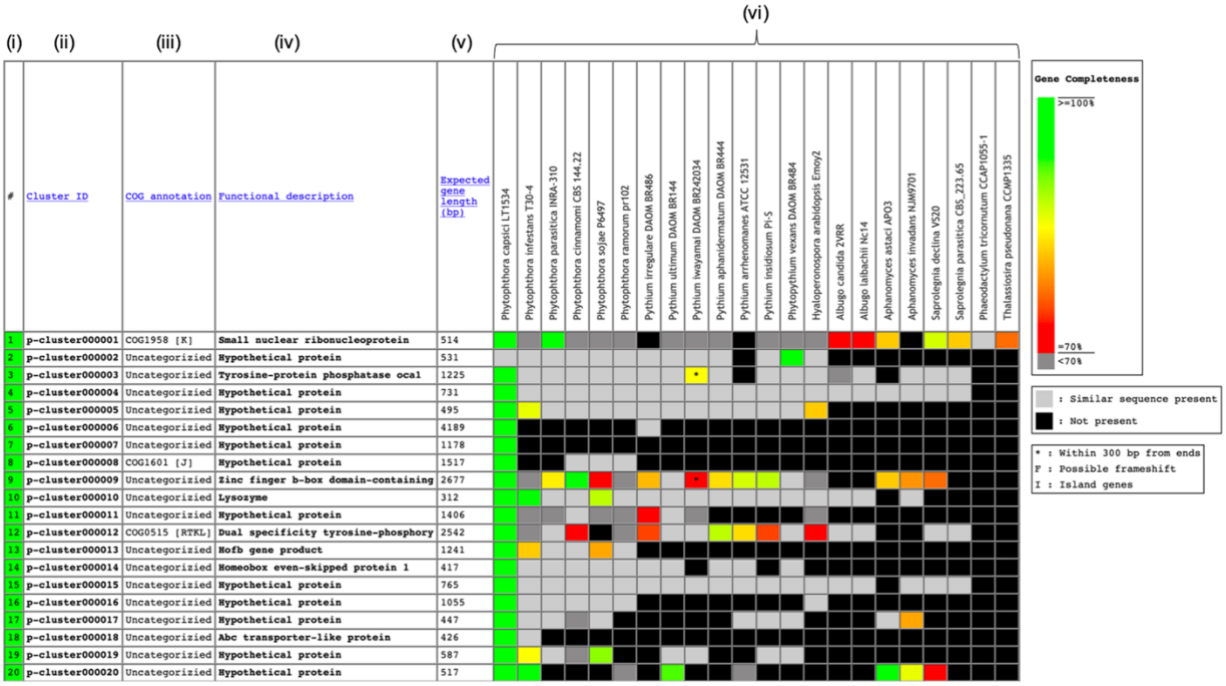
Funding

This work is supported by the Faculty of Medicine, Ramathibodi Hospital, Mahidol University [Grant number CF_60001] and the Thailand Research Fund [Grant number BRG5980009]. The authors acknowledge the financial support provided by King Mongkut's University of Technology Thonburi through the "KMUTT 55th Anniversary Commemorative Fund". The funding agencies have no role in the design of the study, the collection, analysis and interpretation of data, and the writing of the manuscript.

Acknowledgement

Conflict of interest. None declared.

Figures



#	Cluster ID	COG annotation	Functional description	Expected gene length (bp)
1	p-cluster000001	COG1958 [K]	Small nuclear ribonucleoprotein	514
2	p-cluster000002	Uncategorized	Hypothetical protein	531
3	p-cluster000003	Uncategorized	Tyrosine-protein phosphatase oca1	1225
4	p-cluster000004	Uncategorized	Hypothetical protein	731
5	p-cluster000005	Uncategorized	Hypothetical protein	495
6	p-cluster000006	Uncategorized	Hypothetical protein	4189
7	p-cluster000007	Uncategorized	Hypothetical protein	1178
8	p-cluster000008	COG1601 [J]	Hypothetical protein	1517
9	p-cluster000009	Uncategorized	Zinc finger b-box domain-containing	2677
10	p-cluster000010	Uncategorized	Lysozyme	312
11	p-cluster000011	Uncategorized	Hypothetical protein	1406
12	p-cluster000012	COG0515 [RTKL]	Dual specificity tyrosine-phosphory	2542
13	p-cluster000013	Uncategorized	Hofb gene product	1241
14	p-cluster000014	Uncategorized	Homeobox even-skipped protein 1	417
15	p-cluster000015	Uncategorized	Hypothetical protein	765
16	p-cluster000016	Uncategorized	Hypothetical protein	1055
17	p-cluster000017	Uncategorized	Hypothetical protein	447
18	p-cluster000018	Uncategorized	Abc transporter-like protein	426
19	p-cluster000019	Uncategorized	Hypothetical protein	587
20	p-cluster000020	Uncategorized	Hypothetical protein	517

Figure 1. Components of the Oomycete Gene Table. Each row represents a gene cluster. Several columns are arranged to present: **(i)** Row numbers, **(ii)** Cluster identification numbers, **(iii)** Clusters of orthologous groups of proteins (COG), **(iv)** Functional descriptions, **(v)** Expected gene lengths, and **(vi)** organism names (e.g. 20 oomycetes and 2 diatoms). Percent completeness of a gene in each cell is indicated by colour shading. A light-gray cell shows that a sequence with significant similarity (but without a complete open reading frame) is identified in a particular organism. A black cell indicates that no gene is identified in the corresponding genome.

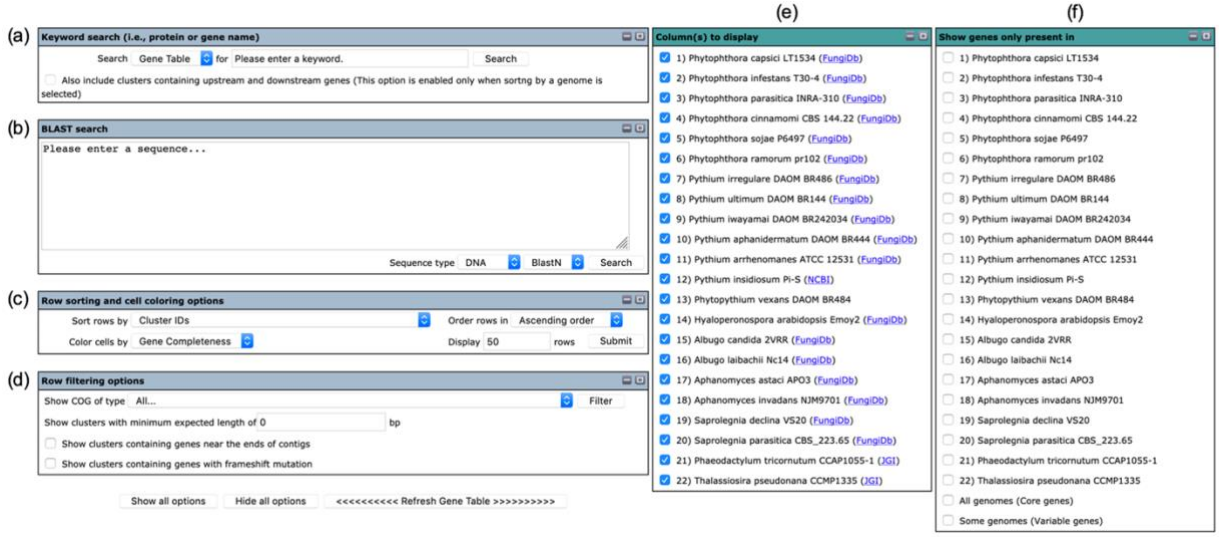


Figure 2. User interface of the Oomycete Gene Table comprises of six drop-down windows: **(a)** Keyword search, **(b)** BLAST search, **(c)** Row sorting and cell coloring options, **(d)** Row filtering options, **(e)** Column (s) to display, and **(f)** Show genes only present in.

Figure 3. The 'Keyword search' window of the Oomycete Gene Table: **(a)** a single or combined keyword(s) (e.g. cluster ID, COG number, and protein/gene name) can be entered in the 'input' box for searching the genomic data of all 22 organisms and **(b)** the Oomycete Gene Table shows a search result of the matched gene contents of all or selected genomes.

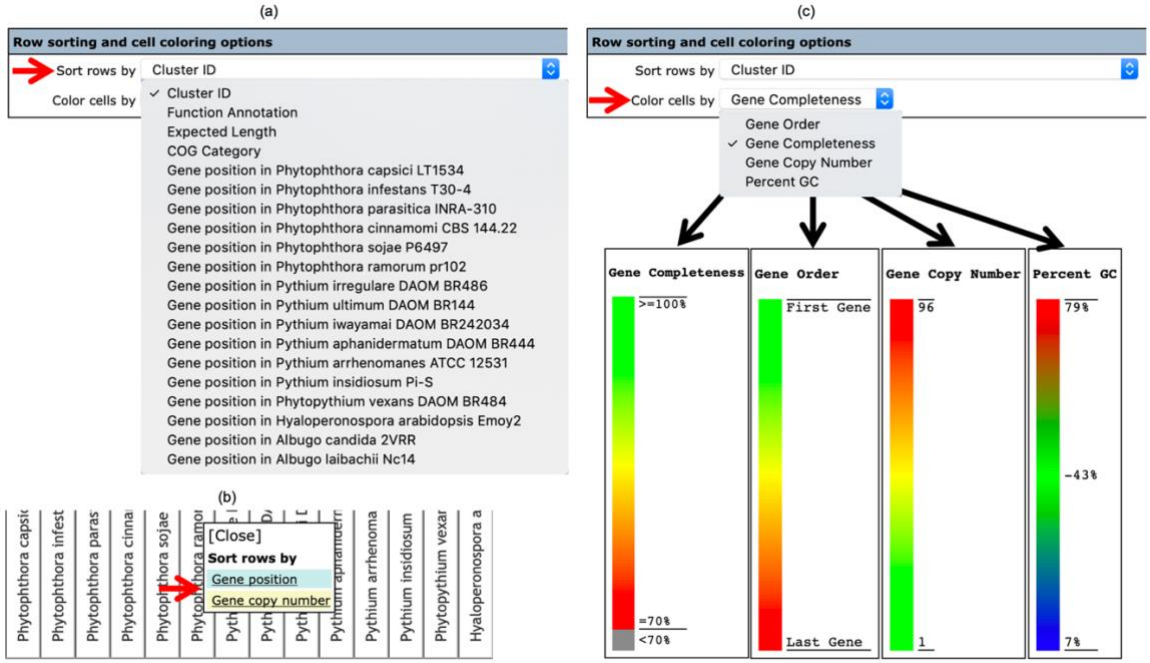


Figure 4 The 'Row sorting and cell coloring' options of the Oomycete Gene Table. Display of the gene clusters and their associated data can be customized, arranged, and sorted based on: **(a)** cluster IDs, functional annotations, expected gene lengths, and COG categories; **(b)** position or copy number of gene(s) in the genome; and **(c)** a selected property, such as gene completeness, gene order in a genome, gene copy number, or CG content.

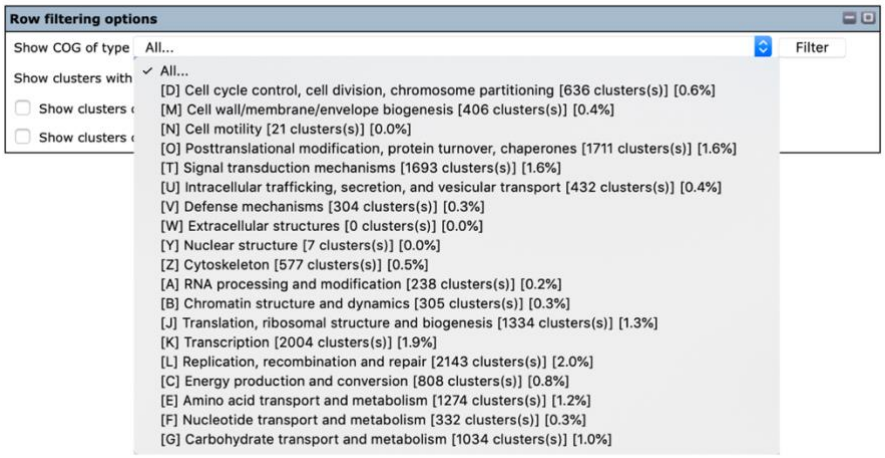


Figure 5. The 'Row filtering' options of the Oomycete Gene Table. The drop-down window is used to customize and display only gene clusters with a desired COG category. The COG codes (e.g. [D], [M], [N]), together with the number and percentage of each gene cluster identified in the organism(s) are listed.

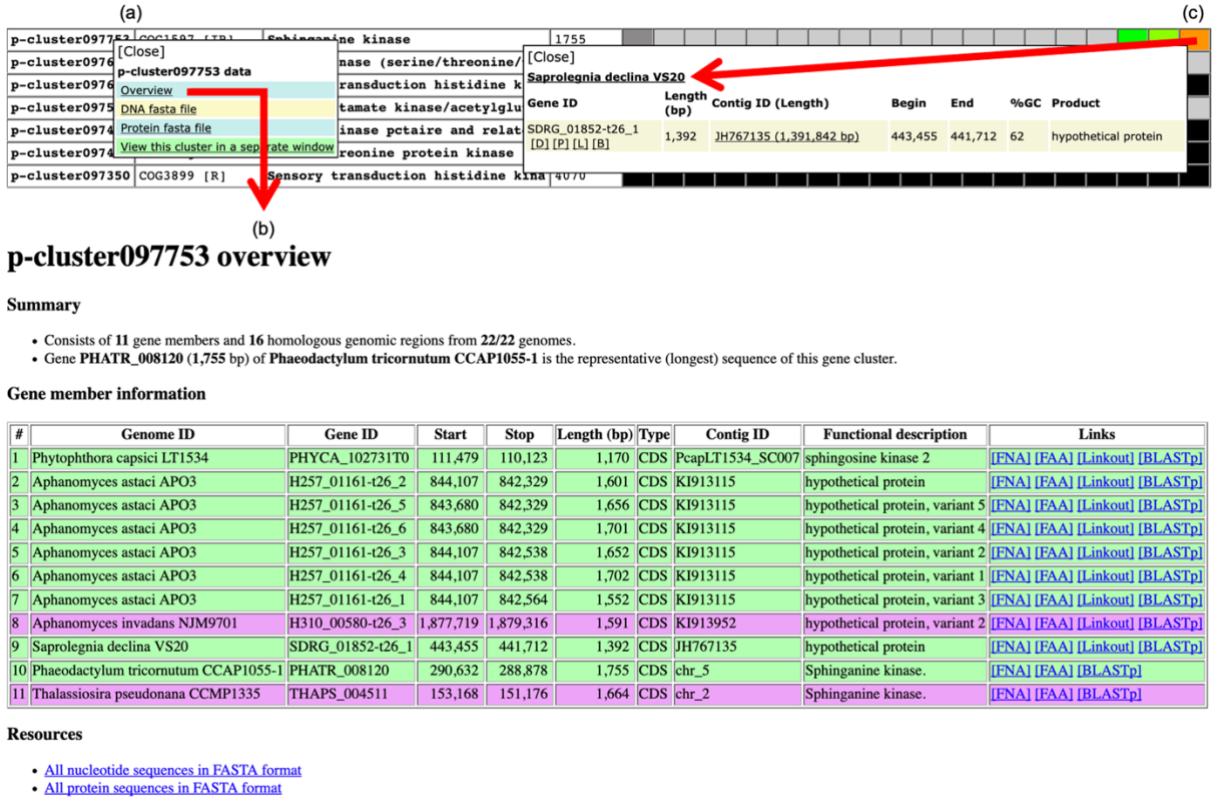


Figure 6. Genomic data retrieval features of the Oomycete Gene Table: **(a)** when a gene cluster ID is selected, several links are shown as gateways to access the genomic data of all organisms contained in that gene cluster; **(b)** overview genomic data includes genome IDs, contig/gene IDs, genomic location and length of the genes, functional annotations, links to gene/protein sequences, links to source databases, and links to NCBI Protein-BLAST tool; and **(c)** when an individual cell is selected, the genomic data of an individual organism can be obtained.

REFERENCES

1. Burki, F. (2014) The eukaryotic tree of life from a global phylogenomic perspective, *Cold Spring Harb Perspect Biol*, **6**, a016147.
2. Kamoun, S. (2003) Molecular genetics of pathogenic oomycetes, *Eukaryotic Cell*, **2**, 191–199.
3. Rujirawat, T., Patumcharoenpol, P., Lohnoo, T., et al. (2018) Probing the Phylogenomics and Putative Pathogenicity Genes of Pythium insidiosum by Oomycete Genome Analyses, *Sci Rep*, **8**, 4135.
4. Kamoun, S., Furzer, O., Jones, J. D. G., et al. (2015) The Top 10 oomycete pathogens in molecular plant pathology, *Mol. Plant Pathol.*, **16**, 413–434.
5. Beakes, G. W., Glockling, S. L. and Sekimoto, S. (2012) The evolutionary phylogeny of the oomycete "fungi," *Protoplasma*, **249**, 3–19.
6. Jiang, R. H. Y., de Bruijn, I., Haas, B. J., et al. (2013) Distinctive Expansion of Potential Virulence Genes in the Genome of the Oomycete Fish Pathogen Saprolegnia parasitica, *PLoS Genetics*, **9**, e1003272.
7. Krajaejun, T., Sathapatayavongs, B., Pracharktam, R., et al. (2006) Clinical and epidemiological analyses of human pythiosis in Thailand, *Clin. Infect. Dis.*, **43**, 569–576.
8. Calvano, T. P., Blatz, P. J., Vento, T. J., et al. (2011) Pythium aphanidermatum infection following combat trauma, *J. Clin. Microbiol.*, **49**, 3710–3713.
9. Reinprayoon, U., Permpalung, N., Kasetsuwan, N., et al. (2013) Lagenidium sp. ocular infection mimicking ocular pythiosis, *J. Clin. Microbiol.*, **51**, 2778–2780.
10. Mendoza, L., Ajello, L. and McGinnis, M. R. (1996) Infection caused by the Oomycetous pathogen Pythium insidiosum, *J. Mycol. Med.*, **6**, 151–164.
11. Ascunce, M. S., Huguet-Tapia, J. C., Braun, E. L., et al. (2016) Whole genome sequence of the emerging oomycete pathogen Pythium insidiosum strain CDC-B5653 isolated from an infected human in the USA, *Genom Data*, **7**, 60–61.
12. McCarthy, C. G. P. and Fitzpatrick, D. A. (2017) Phylogenomic Reconstruction of the Oomycete Phylogeny Derived from 37 Genomes, *mSphere*, **2**.
13. Baxter, L., Tripathy, S., Ishaque, N., et al. (2010) Signatures of adaptation to obligate biotrophy in the Hyaloperonospora arabidopsidis genome, *Science*, **330**, 1549–1551.
14. Haas, B. J., Kamoun, S., Zody, M. C., et al. (2009) Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans, *Nature*, **461**, 393–398.
15. Lamour, K. H., Mudge, J., Gobena, D., et al. (2012) Genome Sequencing and Mapping Reveal Loss of Heterozygosity as a Mechanism for Rapid Adaptation in the Vegetable Pathogen Phytophthora capsici, *Mol. Plant Microbe Interact.*, **25**, 1350–1360.
16. Waugh, M., Hraber, P., Weller, J., et al. (2000) The phytophthora genome initiative database: informatics and analysis for distributed pathogenomic research, *Nucleic Acids Res*, **28**, 87–90.
17. Tyler, B. M., Tripathy, S., Zhang, X., et al. (2006) Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis, *Science*, **313**, 1261–1266.
18. Lévesque, C. A., Brouwer, H., Cano, L., et al. (2010) Genome sequence of the necrotrophic plant pathogen Pythium ultimum reveals original pathogenicity mechanisms and effector repertoire, *Genome Biol.*, **11**, R73.
19. Rujirawat, T., Patumcharoenpol, P., Lohnoo, T., et al. (2015) Draft Genome Sequence of the Pathogenic Oomycete Pythium insidiosum Strain Pi-S, Isolated from a Patient with Pythiosis, *Genome Announc*, **3**, e00574-15.
20. Adhikari, B. N., Hamilton, J. P., Zerillo, M. M., et al. (2013) Comparative Genomics Reveals Insight into Virulence Strategies of Plant Pathogenic Oomycetes, *PLoS ONE*, **8**, e75072.

21. Kittichotirat, W., Patumcharoenpol, P., Rujirawat, T., et al. (2017) Draft genome and sequence variant data of the oomycete Pythium insidiosum strain Pi45 from the phylogenetically-distinct Clade-III, *Data in Brief*, **15**, 896–900.
22. Patumcharoenpol, P., Rujirawat, T., Lohnoo, T., et al. (2018) Draft genome sequences of the oomycete Pythium insidiosum strain CBS 573.85 from a horse with pythiosis and strain CR02 from the environment, *Data in Brief*, **16**, 47–50.
23. Stajich, J. E., Harris, T., Brunk, B. P., et al. (2012) FungiDB: an integrated functional genomics database for fungi, *Nucleic Acids Research*, **40**, D675–D681.
24. Panda, A., Sen, D., Ghosh, A., et al. (2016) EuMicrobedbLite: A lightweight genomic resource and analytic platform for draft oomycete genomes.
25. Krajaejun, T., Rujirawat, T., Kanpanleuk, T., et al. (2018) Biochemical and genetic analyses of the oomycete Pythium insidiosum provide new insights into clinical identification and urease-based evolution of metabolism-related traits, *PeerJ*, **6**, e4821.
26. Kittichotirat, W., Bumgarner, R. E., Asikainen, S., et al. (2011) Identification of the pangenome and its components in 14 distinct Aggregatibacter actinomycetemcomitans strains by comparative genomic analysis, *PLoS ONE*, **6**, e22420.
27. Tatusov, R. L., Fedorova, N. D., Jackson, J. D., et al. (2003) The COG database: an updated version includes eukaryotes, *BMC Bioinformatics*, **4**, 41.
28. Galperin, M. Y., Makarova, K. S., Wolf, Y. I., et al. (2015) Expanded microbial genome coverage and improved protein family annotation in the COG database, *Nucleic Acids Res.*, **43**, D261-269.
29. Jiang, R. H. Y., Tyler, B. M., Whisson, S. C., et al. (2006) Ancient origin of elicitin gene clusters in Phytophthora genomes, *Mol. Biol. Evol*, **23**, 338–351.
30. Jiang, R. H. Y., Dawe, A. L., Weide, R., et al. (2005) Elicitin genes in Phytophthora infestans are clustered and interspersed with various transposon-like elements, *Mol Genet Genomics*, **273**, 20–32.
31. PanabiÈRes, F., Ponchet, M., Allasia, V., et al. (1997) Characterization of border species among Pythiaceae: several Pythium isolates produce elicitins, typical proteins from Phytophthora spp., *Mycological Research*, **101**, 1459–1468.
32. Cox, G. M., Mukherjee, J., Cole, G. T., et al. (2000) Urease as a virulence factor in experimental cryptococcosis, *Infect. Immun.*, **68**, 443–448.
33. Rutherford, J. C. (2014) The emerging role of urease as a general microbial virulence factor, *PLoS Pathog.*, **10**, e1004062.
34. Mora, D. and Arioli, S. (2014) Microbial urease in health and disease, *PLoS Pathog.*, **10**, e1004472.
35. Witte, C.-P., Rosso, M. G. and Romeis, T. (2005) Identification of three urease accessory proteins that are required for urease activation in Arabidopsis, *Plant Physiol.*, **139**, 1155–1162.