The gene repertoire of Pythium porphyrae (Oomycota) suggests an adapted plant pathogen tackling red algae

Yacine Badis, Jong Won Han, Tatyana A Klochkova, Claire M M Gachon, Gwang Hoon Kim

The gene repertoire of *Pythium porphyrae* (Oomycota) suggests an adapted plant pathogen tackling red algae

Yacine Badis[1], Jong Won Han[2], Tatyana A. Klochkova[3], Claire M. M. Gachon[1,4] and Gwang Hoon Kim[5,*]

[1]*The Scottish Association for Marine Science, Scottish Marine Institute, Oban, Argyll PA37 1QA, UK*
[2]*National Marine Biodiversity Institute of Korea, Seocheon 33662, Korea*
[3]*Kamchatka State Technical University, Petropavlovsk-Kamchatsky, 683003, Russia*
[4]*Molécules de Communication et Adaptation des Micro-organismes, UMR 7245, Muséum National d'Histoire Naturelle, 75005 Paris, France*
[5]*Department of Biology, Kongju National University, Kongju 32588, Korea*

Tribute to This paper is dedicated to the memory of Professor Chul Hyun Sohn (1943~2019)

Pythium porphyrae is responsible for devastating outbreaks in seaweed farms of *Pyropia*, the most valuable cultivated seaweed worldwide. While the genus *Pythium* contains many well studied pathogens, the genome of *P. porphyrae* has yet to be sequenced. Here we report the first available gene repertoire of *P. porphyrae* and a preliminary analysis of pathogenicity-related genes. Using ab initio detection strategies, similarity based and manual annotation, we found that the *P. porphyrae* gene repertoire is similar to classical phytopathogenic *Pythium* species. This includes the absence of expanded RxLR effector family and the detection of classical pathogenicity-related genes like crinklers, glycoside hydrolases, cellulose-binding elicitor lectin-like proteins and elicitins. We additionally compared this dataset to the proteomes of 8 selected *Pythium* species. While 34% of the predicted proteome appeared specific to *P. porphyrae*, we could not attribute specific enzymes to the degradation of red algal biomass. Conversely, we detected several cellulases and a cutinase conserved with plant-pathogenic *Pythium* species. Together with the recent report of *P. porphyrae* triggering disease symptoms on several plant species in lab-controlled conditions, our findings add weight to the hypothesis that *P. porphyrae* is a reformed plant pathogen.

Key Words: aquaculture; disease; oomycete; *Pyropia*; *Pythium*; transcriptome

INTRODUCTION

Species of the red algal genus *Pyropia* (formerly *Porphyra*) are the most valuable seaweed crop worldwide. In 2017, global production of *Pyropia* spp. amounts to 2.5 Mt, for a value of USD 2.3 billion (http://www.fao.org/fishery/statistics/global-aquaculture-production).

The culture of *Pyropia* is seriously affected by the oomycete *Pythium porphyrae* which causes the red rot disease (Gachon et al. 2010, Kim et al. 2014). The genus *Pythium* includes more than 250 species (Van der Plaats-Niterink 1981), most of which are soil-borne saprobes or faculta-

tive pathogens on diverse hosts, including plants, invertebrates and mammals (Van der Plaats-Niterink 1981, Lévesque and de Cock 2004). The first extensive molecular phylogeny defined 11 clades (A to K) encompassing all *Pythium* species (Lévesque and de Cock 2004). Clade K was later reassigned to the genus *Phytopythium* (de Cock et al. 2015), and there is now strong evidence supporting the lack of monophyly of *Pythium* sensu lato (Ascunce et al. 2017). Clade A includes *P. porphyrae* and the closely related *Pythium chondricola* which was initially isolated from decaying *Chondrus*. Both are most likely conspecific as *P. chondricola* has been mostly observed parasitizing *Pyropia* sp. (Lee et al. 2017, Qiu et al. 2019) and most classical molecular markers used to reconstruct oomycete phylogenies are strictly identical to *P. porphyrae* (Lee et al. 2015). Additional members of the same clade A include the soil-borne plant pathogen *Pythium aphanidermatum* (Lévesque and de Cock 2004) and pathogens occasionally reported on algal hosts (e.g., *P. adhaerens*) (Sparrow 1931). Genome sequences are available for *Pythium* species of agronomical interest (notably *Pythium ultimum* var. *ultimum*) (Lévesque et al. 2010, Ah-Fong et al. 2017), the emerging human pathogen *Pythium insidiosum* (Krajaejun et al. 2011, 2018), and the mycoparasite *Pythium oligandrum* because of its potential as a biocontrol agent (Horner et al. 2012, Berger et al. 2016, Kushwaha et al. 2017). Additional draft genomes have been generated for the plant-pathogenic *Pythium ultimum* var. *sporangiferum*, *Pythium aphanidermatum*, *Pythium iwayamai*, *Pythium arrhenomanes*, *Pythium irregular*, and *Phytopythium vexans* (Adhikari et al. 2013). Thus, the diversity of hosts and lifestyle amongst *Pythium* species make them good models for comparative genomics to study the evolution of pathogenesis amongst oomycetes. A decade ago, genome sequencing combined with high-throughput functional assays dawned an era of systematic characterization of pathogen virulence factors (generally termed effectors) secreted into the host apoplast or even addressed to the host cytoplasm to facilitate infection. The best studied oomycete effectors are small proteins sharing a conserved N-terminal RxLR signature (Morgan and Kamoun 2007). These so-called RxLR effectors are delivered into the host cytoplasm and the precise implication of the conserved RxLR signature in this translocation mechanism is highly debated (Ellis and Dodds 2011, Sun et al. 2013, Wawra et al. 2017). While RxLR effectors are expanded amongst Peronosporales, they are much less represented in other oomycete orders, notably absent from *Pythium* species (Adhikari et al. 2013), and the detection of RxLR-like proteins in other

orders typically yields high numbers of false-positives (McGowan and Fitzpatrick 2017). Within the oomycete secretome, additional families of typical effectors include crinkling and necrosis factors (CRNs), necrosis inducing-proteins, elicitins, protease inhibitors and an array of peptidases and glycoside hydrolases (Jiang and Tyler 2012, McGowan and Fitzpatrick 2017).

As no genome or transcriptome data is yet available on *P. porphyrae*, we report here an expressed sequenced tag (EST) dataset generated using 454 sequencing. We provide a proteome and secretome prediction as well as a preliminary account of the genes potentially involved in *P. porphyrae* pathogenicity. We also performed gene clustering and *de-novo* approaches to assess if new pathogenicity-related proteins or host translocation mechanisms could be identified with this EST dataset. Finally, we also compared the *P. porphyrae* gene complement with that of known mammal and plant *Pythium* pathogens to get first molecular insights into the specialization towards a red algal host within the *Pythium* genus.

MATERIALS AND METHODS

Materials

A clonal strain of *P. porphyrae* was isolated from *Porpyra* sp. blades collected in Seochon (36°12′ N, 126°50′ E, Western Sea) and maintained at 6°C, under 15 μmol photons m^{-2} s^{-1} (10 h light, 14 h dark photoperiod) on corn meal agar plates. To generate free living hyphae of *P. porphyrae*, colonized agar plugs were transferred into liquid Arasaki B medium (Arasaki et al. 1968) and maintained for 7-10 days at 20°C in this suspension. Blades of *Pyropia* sp. host were grown in Modified Grund Medium (Klochkova et al. 2012) with constant aeration at 10°C under 30 μmol photons m^{-2} s^{-1} (16 h light, 8 h dark photoperiod). For the generation of EST libraries, *P. porphyrae* free living hyphae were inoculated into *Pyropia* sp. MGM cultures and harvested (three biological replicates) at 0, 4, and 30 h after inoculation. Mycelium was collected and thoroughly washed 4 times with sterile seawater to remove host debris and immediately flash frozen in liquid nitrogen.

RNA extraction and EST library assembly

Total RNA from *P. porphyrae* was isolated using TRI reagent (MRC, Los Angeles, CA, USA) according to the manufacturer's protocol, treated enzymatically with

DNase I (TaKaRa, Tokyo, Japan) to remove contaminant DNA, and purified with the Oligotex purification kit (Qiagen, Los Angeles, CA, USA). RNA quality was assessed both visually (1.2% formaldehyde gel electrophoresis) and using an Agilent 2100 Bioanalyzer (Agilent, Hambruk, Germany). 454 pyrosequencing library preparation and sequencing were carried out by GNCBio (Daejeon, Korea) using a GS Titanium Library Preparation kit (454 Life Sciences, Branford, CT, USA). cDNA libraries were validated using a High Sensitivity Chip on the Agilent 2100 Bioanalyzer, and emulsion PCR reactions were performed to enrich cDNA templates. Tagged libraries were combined in a picotiter plate and sequenced using a 454 Genome Sequencer FLX System (Roche, Los Angeles, CA, USA). The resulting Standard Flowgram Format file containing raw read data was used to filter low-quality reads (cut-off value of 40), primer and adapter sequences. Reads shorter than 50 bp were removed, and after pooling all conditions, the remaining dataset was *de novo* assembled using the GS De Novo Assembler software (Newbler v. 2.7; 454 Life Sciences). Datasets are available on the Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) under the run accessions SRR11079882, SRR11079883, and SRR11079884.

Proteome prediction, clustering, and annotation of multigenic families

For the definition of open reading frames (ORFs), the EST dataset was locally blasted against the *Pythium ultimum* BR144 reference genome assembly (E-value cut-off, 1.0 E-5) and the resulting BlastX result file was fed into the OrfPredictor server (Min et al. 2005). The predicted proteome was further quality filtered with the CAST algorithm (Promponas et al. 2000), with a threshold set manually at 90, an all against all BlastP (E-value cut-off, 1.0 E-5) and a minimum similarity score of 20. The predicted ORF proteome was subjected to local all against all BlastP analysis (E-value threshold of 1.0 E-5) for the definition of gene clusters with TRIBE-MCL (Enright et al. 2002), using an inflation value of 2 (similar results were obtained with inflation values up to 7). A preliminary automated annotation of the proteome was performed using a combination of Blast2GO mapping (Conesa and Götz 2008), BlastP against the refseq_protein database (E-value cut-off, 1.0 E-6) and the PFAM batch server database (E-value cut-off, 1.0 E-5) (Punta et al. 2012). Homologues of genes potentially involved in pathogenicity were further manually curated, using a combination of reciprocal Blast analyses, database and literature searches.

De novo prediction of putative translocation motifs

Proteins targeted to the secretory pathway, i.e., either containing a signal peptide (SP) or a signal anchor (SA) were identified with HECTAR (Gschloessl et al. 2008). The SP cleavage sites were predicted with PredSL (Petsalaki et al. 2006). Gene clusters containing secreted (SP) or signal anchor-containing (SA) proteins were defined based on the TRIBE-MCL clustering described above. Similar to the strategy of Saunders et al. (2012), a *de novo* motif search was performed with MEME (Bailey et al. 2009) to identify ab initio potential RxLR-type host translocation motifs. Since our EST sequences might be truncated at the 5′-terminus, we chose not to impose any *a priori* positional constraint. MEME was run on the 194 SP proteins contained in the 43 clusters encompassing more than 4 records. The program was set to report the 25 most robust motifs of 4 to 40 amino acids, occurring zero or once per sequence. The resulting MEME motifs were visually scored according to the following criteria: (1) how many times the motifs are present; (2) their dispersion of their position along the protein sequence sites; and (3) whether they were found across different tribes (as would be expected for a conserved translocation motif). Because subcellular targeting algorithms sometimes struggle to discriminate SP from SA, we also ran as a control the same analysis on all SA-containing clusters

Sequence similarity networks for oomycete elicitin and elicitin-like genes

P. porphyrae elicitins and elicitin-like genes were identified using similarity and PFAM searches, aligned in CLC Genomics Workbench and the alignment was visually inspected and corrected for misalignments. The resulting alignment was trimmed, exported in Clustal (.aln) format and used to generate an hmm model with HMMER3 (Eddy 2011). The resulting hmm model was used to screen all *Pythium* datasets mentioned in the section below (E-value cut-off, 1.0 E-5), yielding 321 *Pythium* sp. proteins. Additionally, 1,931 oomycete proteins containing elicitin or elicitin-like domains were obtained from the NCBI by keyword search, as well as 31 proteins annotated as elicitins or elicitin-like on the AphanoDB repository (https://www.polebio.lrsv.ups-tlse.fr/aphanoDB/). All 2,219 proteins were loaded on the EFI-EST (Enzyme Function Initiative-Enzyme Similarity Tool) online server (Gerlt et al. 2015) and used for the generation of sequence similarity networks (SSNs) using default parameters for
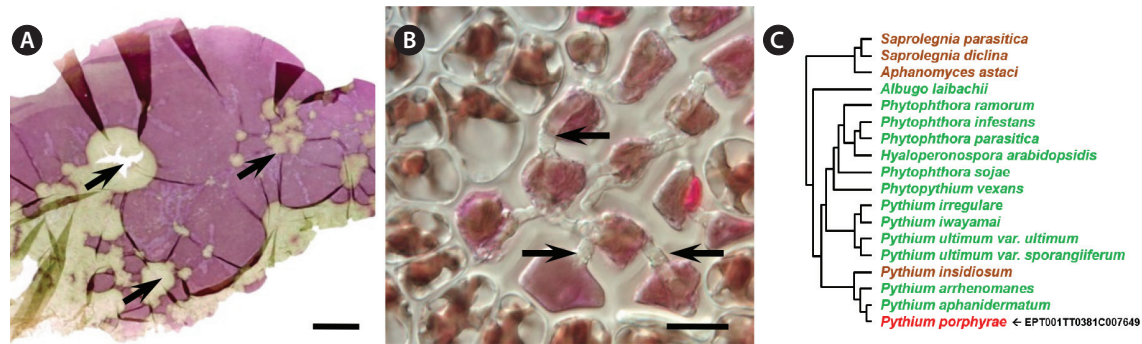
Fig. 1. Symptoms and morphology of *Pythium porphyrae*, the oomycete causing the Red Rot Disease of *Pyropia* sp. (A) *Pyropia* blade with symptomatic areas infected by *P. porphyrae* (arrows). (B) Enlarged image of the infected blade with visible fungal mycelium (arrows) growing through the host cells. (C) Schematic summary tree of oomycete phylogeny. Animal, plant, and red algal parasites are indicated by brown, green, and red fonts, respectively. Scale bars represent: A, 20 mm; B, 10 μm.

edge calculation. To account for diversity of domains and protein length present in the dataset, an edge score threshold of 7 (equivalent to an E-value cut-off, 1.0 E-7) was selected for the definition of similarity clusters. Resulting SSNs were visualized and analyzed in Cytoscape (Demchak et al. 2014).

Clustering with additional known *Pythium* genomes

To define *P. porphyrae* specific proteome, the genomes of *P. ultimum* var. *ultimum*, *P. ultimum* var. *sporangiferum*, *P. iwayamai*, *P. aphanidermatum*, *P. irregulare*, *P. arrhenomanes*, *Phytopythium vexans* (formerly *Pythium vexans*) and the genome of the animal pathogen *P. insidiosum* were obtained from the *Pythium* Genome Database (http://pythium.plantbiology.msu.edu/) and the NCBI (www.ncbi.nlm.nih.gov/genome/34156), respectively. Gene clustering was performed through the get-homologue software (Contreras-Moreira and Vinuesa 2013) using the implemented OrthoMCL (option -M) algorithm to report all possible gene clusters, including species-specific clusters (parameter -t = 0). A Pangenome Matrix was computed (compare_clusters.pl script available through get_homologues) and used to define the *P. porphyrae* specific proteome.

RESULTS

Assembly and initial annotation of the *Pythium porphyrae* gene repertoire

A clonal strain of *P. porphyrae* was isolated in 2012 from infected *Pyropia* sp. blades growing in Seochon

(36°12′ N, 126°50′ E). Clear symptoms of red rot disease could be observed, starting from small red patches on the blades where fungal spores germinated. Areas of dead host tissue spread and ultimately lead to the disintegration of the entire blades (Fig. 1A). Classical optical microscopy confirmed the typical intracellular growth of *P. porphyrae* through host algal cells (Fig. 1B). Three EST libraries were generated from *P. porphyrae* after 0, 4, and 30 h of exposure to *Pyropia* blades, and corresponded to 630,388, 667,490, and 575,530 reads, respectively. The EST corresponding to the 18S ribosomal RNA (EPT001TT0381C007649) was 100% identical to known *P. porphyrae* sequences (AY598673 and JQ898472), thus validating the taxonomic assignment of this isolate as *P. porphyrae* (Fig. 1C). An overview of our EST analysis strategy is illustrated in Fig. 2. The reference transcriptome pooling all ESTs libraries contained 10,819 unigenes, representing 11.1 Mb of non-redundant sequence data. The corresponding partial proteome extracted by OrfPredictor consisted of 10,807 predicted ORFs. Further filtering (i.e., excluding sequences under 41 amino acid, and CAST quality filtering) led to a final ORF dataset of 10,554 sequences (hereafter referred to as the "*P. porphyrae* proteome"), of which 77% (8,150 proteins) could be assigned to PFAM hits (Supplementary Table S1). Assessment of completeness using the BUSCO software (Simão et al. 2015) reported 89.9% of 429 core eukaryotic genes (Supplementary Table S2). As can be expected for an EST dataset, a significant proportion of the detected genes (14%) were fragmented, and the overall completeness was less than for most fully sequenced *Pythium*. For the definition of multigenic gene families, TRIBE-MCL clustering retrieved 1,216 gene clusters of two or more ORFs (Fig. 3, Supplementary Table S1).
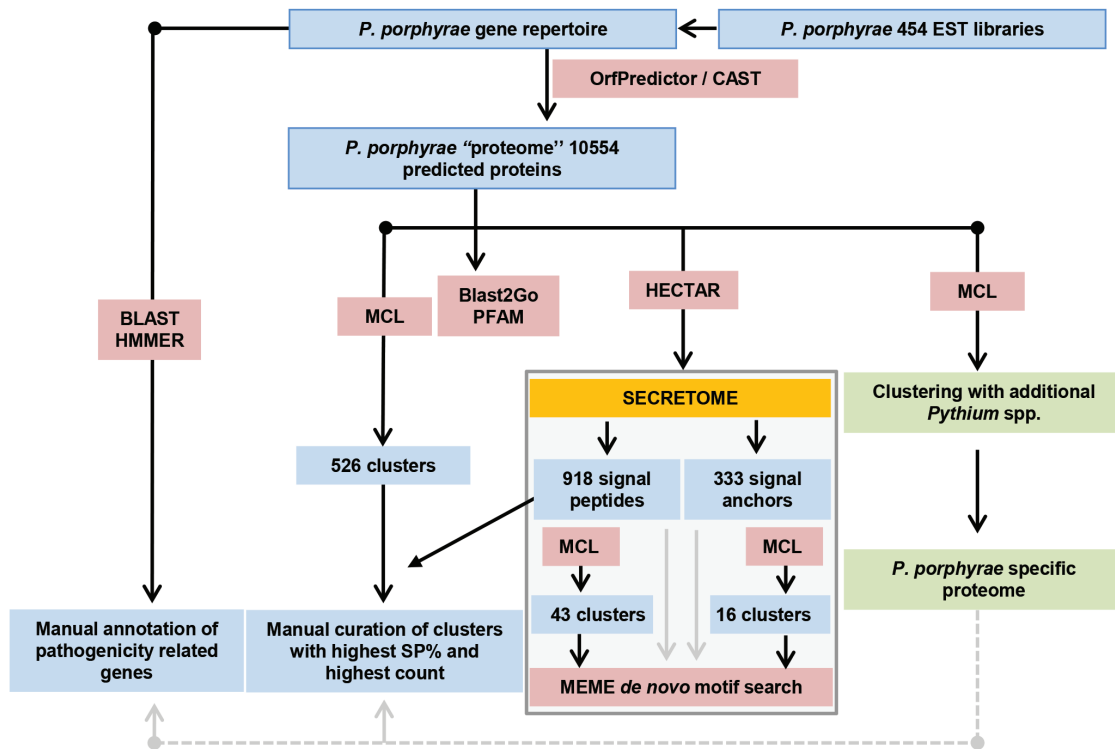
Fig. 2. Annotation strategy and clustering of *Pythium porphyrae* gene repertoire. Counts are given for gene clusters containing 3 or more records.
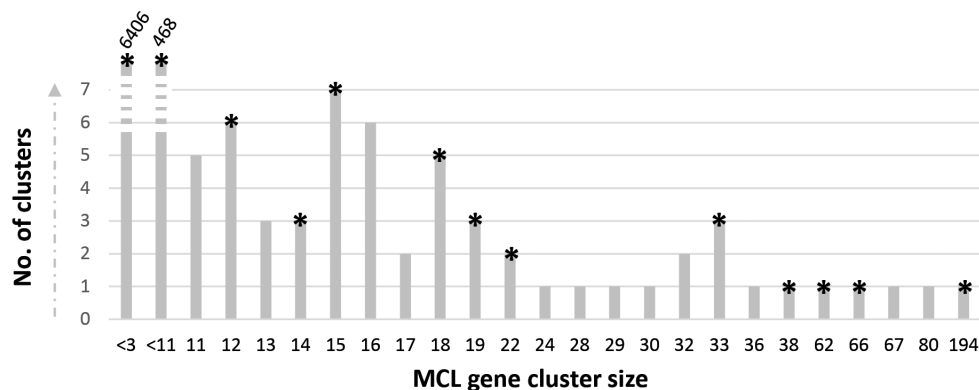

Fig. 3. Distribution of gene cluster size according to the MCL clustering algorithm. Stars indicate the occurrence of clusters containing at least 3 secreted proteins.

Annotation of putative pathogenicity-related genes

Over the last decades, several protein families have been implicated in oomycete pathogenicity. The most studied families are listed in Table 1 together with an account of their size in *P. ultimum* and *Phytophthora infes-*

tans. Among these, we manually curated the potential *P. porphyrae* orthologues with a focus on RxLR effectors, crinklers (CRN), elicitin / elicitin-like proteins, protease inhibitors, cellulose-binding elicitor lectins (CBELs), and other proteins that could relate to cell-wall binding or hydrolysis. A full list of the genes manually inspected in this study is provided in Supplementary Table S3.

RxLR effectors and crinklers

Despite extensive searches with BLAST or regular expressions, no expanded RxLR family in *P. porphyrae* could be identified. Only two proteins, EPT001TT0381C005379 and EPT001TT0381C004327, displayed significant homology to putative RxLR-like proteins conserved across Peronosporales and Pythiales when blasted against the nr database, with the highest similarity to *P. insidiosum* (57.7% to GAY02831.1 and 60.4% to GAX97019.1, respectively). A SP could be detected for both proteins but only EPT001TT0381C004327 carried an RxLR motif. When blasted against the broad spectrum of RxLR-like proteins defined by McGowan & Fitzpatrick (McGowan and Fitzpatrick 2017), EPT001TT0381C005379 was most similar to a *Phytophthora cryptogea* (PHCR|05525, 58.9% identity) protein that did not cluster with any other omycete species. On the contrary, EPT001TT0381C004327 was most similar to a *P. arrhenomanes* protein (PPYAR|006221, 60.9% identity) belonging to a conserved multispecies cluster of putative RxLR-like proteins (https://oomycetes.github.io/rxlrs.html). We also identified 9 sequences homologous to the *P. ultimum* YxSL[RK] proteins, which was hypothesized as a new family of RxLR-like candidate effectors (Lévesque et al. 2010). As in *P. ultimum*, the gene cluster detected for *P. porphyrae* family was especially rich in secreted proteins (eight out of nine members). However, the YxSL[RK] motif was only poorly conserved, replaced by Y[KS][TS]L[RK] in *P. porphyrae*, and in one instance interrupted by an insertion (Supplementary Fig. S1).

In total, 13 CRNs were identified by similarity search with other known oomycete CRNs (Supplementary Table S3). All contigs that spanned the protein N-terminus had an N-terminal LxLYLAR consensus (Supplementary Fig. S2, red triangles) clearly related to the *P. ultimum* LxLYLAR/K motif (Lévesque et al. 2010). The LxLYLAR motif was followed by two conserved WL residues (Supplementary Fig. S2, yellow triangles) and a stretch of variable amino acids. As in *P. infestans*, a highly conserved HVLVxxP motif (Supplementary Fig. S2, orange triangles) linked the conserved N-terminus to the variable C-terminus. Eleven of the 13 identified *P. porphyrae* CRNs contain a domain already reported in a *P. infestans* CRN, including the DC domain, which was shown to stimulate cell death when expressed *in planta* (Haas et al. 2009).

Cellulose-binding proteins and cell wall hydrolases

While looking for determinants of pathogenicity on red algae, we searched for enzymes that might be specifically involved in the degradation of red algal carbohydrates, especially sulphated polysaccharides. We found about twenty sulphatases and sulphate permeases that were well conserved with *Phytophthora*, and therefore likely irrelevant in the context of red algal cell wall degradation (data not shown). Extensive searches for enzymes bearing homology to bacterial agarases, carrageenases, and porphyranases (families GH16, 50, 86, 96, and 118) also remained unsuccessful. On the other hand, we identified three cellulases (EPT001TT0381C001561, EPT001TT0381C001658, EPT001TT0381C001964), as well as a putative cutinase (EPT001TT0381C004750) most closely related to *Phytophthora* cutinases. Finally, we found seven CBEL-like (cellulose-binding elicitor lectins) proteins carrying both carbohydrate binding modules (CBMs) and PAN/APPLE domains (Supplementary Table S3). The gene cluster containing these CBEL-like proteins contained 11 additional proteins with PAN/APPLE domains alone or in combination with other domains (Supplementary Fig. S3).

Table 1. Candidate gene families related to pathogenesis identified in *Pythium porphyrae* and comparison with *P. ultimum* and *P. infestans*

Gene family annotation	*P. porphyrae*	*P. ultimum*[a]	*P. infestans*[a]
Crinklers	14[b]	26	196
Elicitin-like	27[b]	37[b]	40
PcF/SCR-like	-[b]	3	16
NPP1-like	-[b]	7	27
RxLR candidate effectors	-[b]	0	563
YXL[KR] candidate effectors	-[b]	0	0
CBEL	9[b]	17	14
Protease inhibitors	13	43	38
Serine protease	52[c]	85	60
Aspartate protease	5[c]	29	16
Cysteine protease	56[c]	42	38
Cutinase (CAZy)	1[b]	0	4
Glycoside hydrolase (CAZy)	63[c]	180	277
Pectin esterase	-[b]	0	13
Lipase	19[c]	31	19
Phospholipase	2[c]	20	36
Cytochromes P450	20[c]	41	28
ABC transporter	126[c]	140	137

[a]Gene counts follow Lévesque et al. (2010).
[b]Manual sequence curation; see individual gene annotations in Supplementary Table S3.
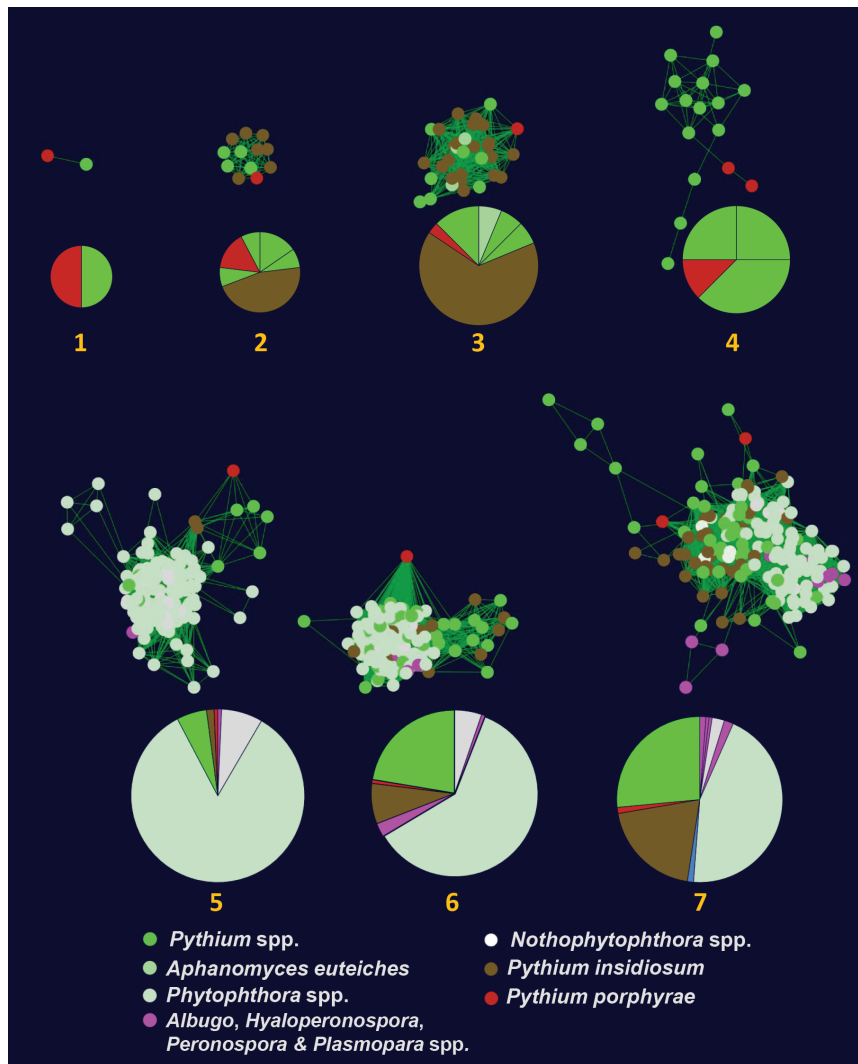[c]Based on automated PFAM/IPR analysis; see Supplementary Table S1.

Fig. 4. Sequence similarity networks computed for oomycete elicitin-like genes. Only the seven networks containing *Pythium porphyrae* elicitins are represented. Each node represents an elicitin-like gene color coded according to each oomycete lifestyle: plant parasite (green), obligate plant parasite (purple), saprobe (white), animal parasite (brown), and red algal parasite (red). Pie charts illustrate the repartition of oomycete species within each network. A full account of oomycete elicitin-like sequence similarity networks is available in Supplementary Fig. S4.

Elicitins

We identified 26 elicitin-like proteins in *P. porphyrae* (Supplementary Table S3). Eighteen proteins had a predicted SP, whereas the lack thereof in the others was likely due to incomplete coverage of the N-termini. We aligned and visually inspected elicitin-like domain structure of *P. porphyrae*, validating the conservation and spacing of the 6 cysteine residues for 22 proteins (Supplementary Table S3). No known C-terminal domains could be detected on any of the 26 proteins. An hmm model derived from *P. porphyrae* elicitins was used to screen all putative elicitin-like proteins from the genomes of 8 additional *Pythium* species and coupled to all available elicitin-like proteins from public databases to generate a SSN (see methods). Only 9 *P. porphyrae* elicitins were assigned to 7 oomycete elicitin clusters (Fig. 4, Supplementary Fig. S3) using an EFI-EST edge score threshold of 7. Note that increasing this threshold to 10 and above abolished most clustering of the *P. porphyrae* proteins, thus reflecting the overall divergence of this elicitin-like repertoire. Three clusters were specific to the Pythiales (Clusters 1, 2, 4 on Fig. 4) while the others overlapped the Peronosporales and / or Albuginales. The 17 remaining *P. porphyrae* elicitins could not relate to any known oomycete elicitins (Supplementary Fig. S3).
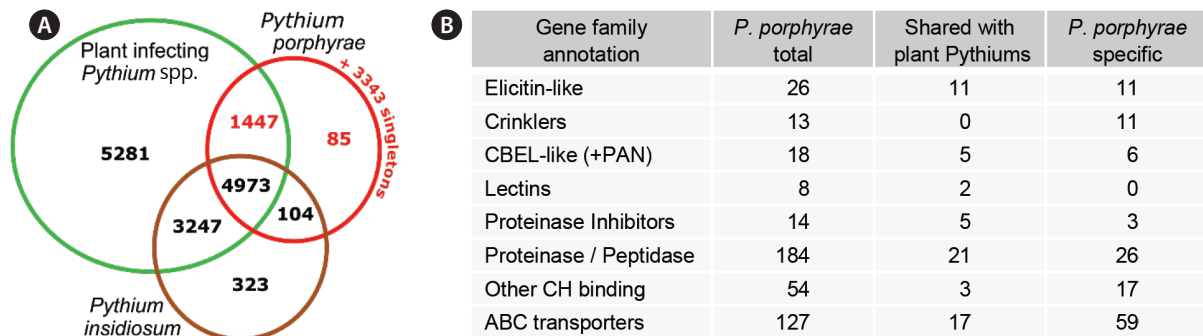
Fig. 5. Orthologous gene family clustering with additional *Pythium* species. (A) Venn diagram illustrating the number of shared and specific gene clusters between *Pythium porphyrae*, the animal-infecting *Pythium insidiosum* and seven phytopathogenic *Pythium* species. The 7 plant *Pythium* datasets have been thoroughly analyzed elsewhere (Lévesque et al. 2010), and are merged here for clarity. The detailed cluster composition available in Supplementary Table S5. (B) Pathogenesis-related candidates amongst genes specific to *Pythium porphyrae* or shared with at least one phytopathogenic *Pythium* species (red fonts on A).

Ab initio prediction of secreted candidate pathogenicity effectors

To detect putative novel pathogenicity effectors, we screened the *P. porphyrae* secretome for candidate translocation motifs. Supplementary Table S4 contains the results of the subcellular targeting software HECTAR. 918 ORFs (8.7% of the proteome) were predicted to contain a SP and 333 to contain a SA (3.2% of the proteome). This result is in the range of other known oomycetes (e.g., 7.7 and 7.5% of secreted proteins in *Phytophthora sojae* and *P. ramorum*) (Tyler et al. 2006), but higher than predicted in *P. ultimum* (4.9%) (Lévesque et al. 2010). The PFAM analysis retrieved 1,476 and 558 hits, mapped against 724 SP and 284 SA proteins (78.9 and 85.3% of all SP and SA ORFs, respectively) (Supplementary Table S5). We defined clusters of secreted proteins following the original TRIBE-MCL clustering (Fig. 3) and ran MEME on these subsets. Generally, amino-acid motifs found in the secretome were highly tribe-dependent, and essential characteristic of specific protein families. Despite our extensive unbiased searches, we have been unable to identify any known or novel candidate translocation motif in *P. porphyrae*.

Overlap of *Pythium porphyrae* gene repertoire with other *Pythium* species

To identify candidate genes that may account for the algal host specificity, we used a standard OrthoMCL clustering to compare our gene repertoire to the proteomes of 7 plant-associated *Pythium* species (analyzed by Lévesque et al. 2010) and to the proteome of the animal

pathogen *P. insidiosum* (see methods). A core proteome of 4,973 clusters (5,299 genes) was shared between all included *Pythium* species (Fig. 5A). Of the 6,609 *P. porphyrae* clusters, 104 were shared only between *P. porphyrae* and *P. insidiosum*, while a total of 1,447 clusters (representing 1,507 *P. porphyrae* proteins) were shared with at least one plant-pathogenic *Pythium* species and not detected in *P. insidiosum*. Altogether, 34% of the proteome was specific to *P. porphyrae*, with 200 proteins in clusters and 3,443 singletons. Interestingly, numerous putative pathogenesis-related candidates could be found in the specific *P. porphyrae* proteome defined by OrthoMCL (Fig. 5B), notably 11 elicitin-like and the majority of CRNs. For CBEL-like proteins, the association of CBM and PAN/APPLE was not specific to phytopathogenic *Pythium* species as homologs were also detected in *P. insidiosum* (Supplementary Fig. S3).

DISCUSSION

Here we present a first account of the gene repertoire of *Pythium porphyrae* the genome of which has not been sequenced to date. Thus, our analysis of this EST dataset provides an interesting insight into the molecular determinants of pathogenicity of the most destructive disease-causing organism in *Pyropia* farms.

The *Pythium porphyrae* gene repertoire conforms to expectations

In terms of gene content, *P. porphyrae* mostly conforms to the trends described in other *Pythium* species

and across oomycetes (Table 1). We detected all the classic families of proteins reported in other *Pythium* species as candidate pathogenicity proteins, including CRNs, Elicitings, CBELs, protease inhibitors. Despite extensive searches, and consistent with other *Pythium* species, we failed to evidence any expanded RxLR effector family or new candidate translocation motifs with similar structural properties (Lévesque et al. 2010, Horner et al. 2012, McGowan and Fitzpatrick 2017, Krajaejun et al. 2018). Finally, only one of the 14 CRNs identified in *P. porphyrae* is predicted as secreted. Despite the caveat that some of our sequences may not be full length, this resonates with the earlier finding that only two among the 26 *P. ultimum* CRN proteins contained a SP (Lévesque et al. 2010). Our observations suggest that the primary function of *Pythium* CRNs may not be to act as virulence effectors translocated into the cytoplasm of their host. Altogether, and similar to other described *Pythium* species, *P. porphyrae* pathogenicity is unlikely to reside in fast-evolving, expanded families of host-translocated pathogenicity effectors similar to those described in *Phytophthora* and other (hemi)biotroph oomycete pathogens.

The enzymatic arsenal of a reformed plant pathogen?

In our attempts to identify molecular determinants of host specificity, we were unable to identify any degradative enzyme that would specifically target red algal polysaccharides. It is worth noting that while such enzymes have been identified in prokaryotes (Fu and Kim 2010, Schultz-Johansen et al. 2018), eukaryotic counterparts have yet to be characterized. Thus, we cannot exclude the possibility that some *P. porphyrae* specific hydrolases reported here may act as red algal degradation enzymes.

Although *P. porphyrae* has not been found on any other organism than *Pyropia*, a recent report demonstrated that several plant species inoculated with *P. porphyrae* develop symptoms including root rot and, occasionally, elicit a hypersensitive response (Klochkova et al. 2016). Here, we evidenced a cutinase, an enzyme specialized in the degradation of plant cuticles. Several cellulases are also present, although cellulose is notably absent from the gametophytic (i.e., blade) stage of *Pyropia* infected by *P. porphyrae* (Mukai et al. 1981). These elements suggest that *P. porphyrae* has evolved from and retained the metabolic capacity of a saprophyte or pathogen exploiting terrestrial plants. We hypothesize that the plant defence response or symptoms triggered by *P. porphyrae* lab-controlled inoculation results from the detection by

the hosts of conserved microbial signatures such as the elicitin-like and CBEL-like proteins reported in this study.

In conclusion, we only identified gene families shared with known oomycete parasites. Although our comparative approaches suggest some degree of specificity for several classic pathogenicity proteins of *P. porphyrae* (Fig. 5B), the molecular determinants of adaptation to red algal hosts remain elusive. Savory et al. (2018) speculated that the broad host range of *P. aphanidermatum*, the closest relative of *P. porphyrae* within clade A, could be linked with a striking expansion of substrate range for a transporter acquired through horizontal gene transfer. Using the gene of *P. aphanidermatum* (pag1_scaffold_175-fgenesh-gene-0.5) as bait, we could not to retrieve a credible ortholog for *P. porphyrae.*

Keeping in mind the inherent limitations of an EST dataset (completeness, truncated transcripts) for gene clustering, we hope that accessing the full genomic context and detailed gene expression patterns in a time course of infection will improve our understanding the diversity and functions of *P. porphyrae* pathogenicity determinants.

SUPPLEMENTARY MATERIALS

Supplementary Table S1. Automated annotation, clustering, and subcellular targeting for the predicted *Pythium porphyrae* proteome (https://www.e-algae.org).
Supplementary Table S2. Comparative assessment of

completeness Benchmarking Universal Single-Copy Orthologs (BUSCO v1.8.1) (https://www.e-algae.org).

Supplementary Table S3. Manually curated list of candidate pathogenicity related genes in *Pythium porphyrae* (https://www.e-algae.org).

Supplementary Table S4. Subcellular targeting prediction by the HECTAR online tool (https://webtools.sb-roscoff.fr/) (https://www.e-algae.org).

Supplementary Table S5. Comparative OrthoMCL clustering of the *Pythium porphyrae* proteome with known *Pythium* molecular data (https://www.e-algae.org).

Supplementary Fig. S1. Multiple alignment of candidate YxSL[RK] proteins of *Pythium porphyrae*. Multiple alignment performed in the CLC workbench software, using the implemented Clustal algorithm with default parameters. The sequence logo illustrates the poorly conserved YxSL[RK] motif, replaced by Y[KS][TS]L[RK] in *P. prophyrae* and highlighted in red (https://www.e-algae.org).

Supplementary Fig. S2. Domain and motif composition of *Pythium porphyrae* crinklers. Only one sequence of each structural category is presented (https://www.e-algae.org).

Supplementary Fig. S3. Diversity of domain combinations in cellulose-binding elicitor lectins-like and PAN-containing proteins of *Pythium porphyrae*. The fonts used for *P. porphyrae* proteins reflects the OrthoMCL clustering performed on 9 *Pythium* spp. proteomes, with proteins specific to *P. porphyrae* (red) or shared only with plant *Pythuym* spp. (green); all other proteins have homologs additionally shared with the animal pathogen *P. insidiosum* (https://www.e-algae.org).

Supplementary Fig. S4. Sequence similarity networks (SSNs) for 2,219 oomycete proteins containing elicitin or elicitin-like domains, gathered from (1) the NCBI protein portal, (2) the AphanoDB platform, (3) custom hmm search amongst *Pythium* spp. genomes, and (4) the *Pythium porphyrae* gene repertoire. Numbers indicate the clusters containing *P. porphyrae* elicitin-like proteins. Note the 17 nodes corresponding to *P. porphyrae* specific proteins (upper right corner) (https://www.e-algae.org).

REFERENCES

Adhikari, B. N., Hamilton, J. P., Zerillo, M. M., Tisserat, N., Lévesque, C. A. & Buell, C. R. 2013. Comparative genomics reveals insight into virulence strategies of plant pathogenic oomycetes. PLoS ONE 8:e75072.

Ah-Fong, A. M. V., Shrivastava, J. & Judelson, H. S. 2017. Lifestyle, gene gain and loss, and transcriptional remodeling cause divergence in the transcriptomes of *Phytophthora infestans* and *Pythium ultimum* during potato tuber colonization. BMC Genomics 18:764.

Arasaki, S., Akino, K. & Tomiyama, T. 1968. A comparison of some physiological aspects in a marine *Pythium* on the host and on the artificial medium. Bull. Misaki Mar. Biol. Inst. Kyoto Univ. 12:203-206.

Ascunce, M. S., Huguet-Tapia, J. C., Ortiz-Urquiza, A., Keyhani, N. O., Braun, E. L. & Goss, E. M. 2017. Phylogenomic analysis supports multiple instances of polyphyly in the oomycete peronosporalean lineage. Mol. Phylogenet. Evol. 114:199-211.

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W. & Noble, W. S. 2009. MEME Suite: tools for motif discovery and searching. Nucleic Acids Res. 37:W202-W208.

Berger, H., Yacoub, A., Gerbore, J., Grizard, D., Rey, P., Sessitsch, A. & Compant, S. 2016. Draft genome sequence of biocontrol agent *Pythium oligandrum* strain Po37, an Ooomycota. Genome Announc. 4:e00215-16.

Conesa, A. & Götz, S. 2008. Blast2GO: a comprehensive suite for functional analysis in plant genomics. Int. J. Plant Genomics 2008:619832.

Contreras-Moreira, B. & Vinuesa, P. 2013. GET_HOMO-LOGUES, a versatile software package for scalable and robust microbial pangenome analysis. Appl. Environ. Microbiol. 79:7696-7701.

de Cock, A. W. A. M., Lodhi, A. M., Rintoul, T. L., Bala, K., Robideau, G. P., Abad, Z. G., Coffey, M. D., Shahzad, S. & Levesque, C. A. 2015. *Phytopythium*: molecular phylogeny and systematics. Persoonia 34:25-39.

Demchak, B., Hull, T., Reich, M., Liefeld, T., Smoot, M., Ideker, T. & Mesirov, J. P. 2014. Cytoscape: the network visualization tool for GenomeSpace workflows. F1000Res. 3:151.

Eddy, S. R. 2011. Accelerated profile HMM searches. PLoS Comput. Biol. 7:e1002195.

Ellis, J. G. & Dodds, P. N. 2011. Showdown at the RXLR motif: serious differences of opinion in how effector proteins from filamentous eukaryotic pathogens enter plant cells. Proc. Natl. Acad. Sci. U. S. A. 108:14381-14382.

Enright, A. J., Van Dongen, S. & Ouzounis, C. A. 2002. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 30:1575-1584.

Fu, X. T. & Kim, S. M. 2010. Agarase: review of major sources, categories, purification method, enzyme characteristics and applications. Mar. Drugs 8:200-218.

Gachon, C. M. M., Sime-Ngando, T., Strittmatter, M., Cham-

bouvet, A. & Kim, G. H. 2010. Algal diseases: spotlight on a black box. Trends Plant Sci. 15:633-640.

Gerlt, J. A., Bouvier, J. T., Davidson, D. B., Imker, H. J., Sadkhin, B., Slater, D. R. & Whalen, K. L. 2015. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): a web tool for generating protein sequence similarity networks. Biochim. Biophys. Acta 1854:1019-1037.

Gschloessl, B., Guermeur, Y. & Cock, J. M. 2008. HECTAR: a method to predict subcellular targeting in heterokonts. BMC Bioinformatics 9:393.

Haas, B. J., Kamoun, S., Zody, M. C., Jiang, R. H. Y., Handsaker, R. E., Cano, L. M., Grabherr, M., Kodira, C. D., Raffaele, S., Torto-Alalibo, T., Bozkurt, T. O., Ah-Fong, A. M. V., Alvarado, L., Anderson, V. L., Armstrong, M. R., Avrova, A., Baxter, L., Beynon, J., Boevink, P. C., Bollmann, S. R., Bos, J. I. B., Bulone, V., Cai, G., Cakir, C., Carrington, J. C., Chawner, M., Conti, L., Costanzo, S., Ewan, R., Fahlgren, N., Fischbach, M. A., Fugelstad, J., Gilroy, E. M., Gnerre, S., Green, P. J., Grenville-Briggs, L. J., Griffith, J., Grünwald, N. J., Horn, K., Horner, N. R., Hu, C. -H., Huitema, E., Jeong, D. -H., Jones, A. M. E., Jones, J. D. G., Jones, R. W., Karlsson, E. K., Kunjeti, S. G., Lamour, K., Liu, Z., Ma, L., MacLean, D., Chibucos, M. C., McDonald, H., McWalters, J., Meijer, H. J. G., Morgan, W., Morris, P. F., Munro, C. A., O'Neill, K., Ospina-Giraldo, M., Pinzón, A., Pritchard, L., Ramsahoye, B., Ren, Q., Restrepo, S., Roy, S., Sadanandom, A., Savidor, A., Schornack, S., Schwartz, D. C., Schumann, U. D., Schwessinger, B., Seyer, L., Sharpe, T., Silvar, C., Song, J., Studholme, D. J., Sykes, S., Thines, M., van de Vondervoort, P. J. I., Phuntumart, V., Wawra, S., Weide, R., Win, J., Young, C., Zhou, S., Fry, W., Meyers, B. C., van West, P., Ristaino, J., Govers, F., Birch, P. R. J., Whisson, S. C., Judelson, H. S. & Nusbaum, C. 2009. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. Nature 461:393-398.

Horner, N. R., Grenville-Briggs, L. J. & Van West, P. 2012. The oomycete *Pythium oligandrum* expresses putative effectors during mycoparasitism of *Phytophthora infestans* and is amenable to transformation. Fungal Biol. 116:24-41.

Jiang, R. H. Y. & Tyler, B. M. 2012. Mechanisms and evolution of virulence in oomycetes. Annu. Rev. Phytopathol. 50:295-318.

Kim, G. H., Moon, K. -H., Kim, J. -Y., Shim, J. & Klochkova, T. A. 2014. A revaluation of algal diseases in Korean *Pyropia* (*Porphyra*) sea farms and their economic impact. Algae 29:249-265.

Klochkova, T. A., Jung, S. & Kim, G. -H. 2016. Host range and salinity tolerance of *Pythium porphyrae* may indicate its terrestrial origin. J. Appl. Phycol. 29:371-379.

Klochkova, T. A., Shim, J. B., Hwang, M. S. & Kim, G. H. 2012. Host-parasite interactions and host species susceptibility of the marine oomycete parasite, *Olpidiopsis* sp., from Korea that infects red algae. J. Appl. Phycol. 24:135-144.

Krajaejun, T., Khositnithikul, R., Lerksuthirat, T., Lowhnoo, T., Rujirawat, T., Petchthong, T., Yingyong, W., Suriyaphol, P., Smittipat, N., Juthayothin, T., Phuntumart, V. & Sullivan, T. D. 2011. Expressed sequence tags reveal genetic diversity and putative virulence factors of the pathogenic oomycete *Pythium insidiosum*. Fungal Biol. 115:683-696.

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

Kushwaha, S. K., Vetukuri, R. R. & Grenville-Briggs, L. J. 2017. Draft genome sequence of the mycoparasitic oomycete *Pythium oligandrum* strain CBS 530.74. Genome Announc. 5:e00346-17.

Lee, S. J., Hwang, M. S., Park, M. A., Baek, J. M., Ha, D. -S., Lee, J. E. & Lee, S. -R. 2015. Molecular identification of the algal pathogen *Pythium chondricola* (Oomycetes) from *Pyropia yezoensis* (Rhodophyta) using ITS and *cox*1 markers. Algae 30:217-222.

Lee, S. J., Jee, B. Y., Son, M. -H. & Lee, S. -R. 2017. Infection and *cox*2 sequence of *Pythium chondricola* (Oomycetes) causing red rot disease in *Pyropia yezoensis* (Rhodophyta) in Korea. Algae 32:155-160.

Lévesque, C. A., Brouwer, H., Cano, L., Hamilton, J. P., Holt, C., Huitema, E., Raffaele, S., Robideau, G. P., Thines, M., Win, J., Zerillo, M. M., Beakes, G. W., Boore, J. L., Busam, D., Dumas, B., Ferriera, S., Fuerstenberg, S. I., Gachon, C. M. M., Gaulin, E., Govers, F., Grenville-Briggs, L., Horner, N., Hostetler, J., Jiang, R. H. Y., Johnson, J., Krajaejun, T., Lin, H., Meijer, H. J. G., Moore, B., Morris, P., Phuntmart, V., Puiu, D., Shetty, J., Stajich, J. E., Tripathy, S., Wawra, S., van West, P., Whitty, B. R., Coutinho, P. M., Henrissat, B., Martin, F., Thomas, P. D., Tyler, B. M., De Vries, R. P., Kamoun, S., Yandell, M., Tisserat, N. & Buell, C. R. 2010. Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. Genome Biol. 11:R73.

Lévesque, C. A. & de Cock, A. W. A. M. 2004. Molecular phylogeny and taxonomy of the genus *Pythium*. Mycol. Res. 108:1363-1383.

McGowan, J. & Fitzpatrick, D. A. 2017. Genomic, network, and phylogenetic analysis of the oomycete effector arsenal. mSphere 2:e00408-17.

Min, X. J., Butler, G., Storms, R. & Tsang, A. 2005. OrfPredictor: predicting protein-coding regions in EST-derived sequences. Nucleic Acids Res. 33:W677-W680.

Morgan, W. & Kamoun, S. 2007. RXLR effectors of plant pathogenic oomycetes. Curr. Opin. Microbiol. 10:332-338.

Mukai, L. S., Craigie, J. S. & Brown, R. G. 1981. Chemical composition and structure of the cell walls of the conchocelis and thallus phases of *Porphyra tenera* (Rhodophyceae). J. Phycol. 17:192-198.

Petsalaki, E. I., Bagos, P. G., Litou, Z. I. & Hamodrakas, S. J. 2006. PredSL: a tool for the N-terminal sequence-based prediction of protein subcellular localization. Genomics Proteomics Bioinformatics 4:48-55.

Promponas, V. J., Enright, A. J., Tsoka, S., Kreil, D. P., Leroy, C., Hamodrakas, S., Sander, C. & Ouzounis, C. A. 2000. CAST: an iterative algorithm for the complexity analysis of sequence tracts. Bioinformatics 16:915-922.

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E. L. L., Eddy, S. R., Bateman, A. & Finn, R. D. 2012. The Pfam protein families database. Nucleic Acids Res. 40:D290-D301.

Qiu, L., Mao, Y., Tang, L., Tang, X. & Mo, Z. 2019. Characterization of *Pythium chondricola* associated with red rot disease of *Pyropia yezoensis* (Ueda) (Bangiales, Rhodophyta) from Lianyungang, China. J. Oceanol. Limnol. 37:1102-1112.

Saunders, D. G. O., Win, J., Cano, L. M., Szabo, L. J., Kamoun, S. & Raffaele, S. 2012. Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. PLoS ONE 7:e29847.

Savory, F. R., Milner, D. S., Miles, D. C. & Richards, T. A. 2018. Ancestral function and diversification of a horizontally acquired oomycete carboxylic acid transporter. Mol. Biol. Evol. 35:1887-1900.

Schultz-Johansen, M., Bech, P. K., Hennessy, R. C., Glaring, M. A., Barbeyron, T., Czjzek, M. & Stougaard, P. 2018. A novel enzyme portfolio for red algal polysaccharide degradation in the marine bacterium *Paraglaciecola hydrolytica* S66^T encoded in a sizeable polysaccharide utilization locus. Front. Microbiol. 9:839.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210-3212.

Sparrow, F. K. 1931. Two new species of *Pythium* parasitic on green algae. Ann. Bot. 45:257-277.

Sun, F., Kale, S. D., Azurmendi, H. F., Li, D., Tyler, B. M. & Capelluto, D. G. S. 2013. Structural basis for interactions of the *Phytophthora sojae* RxLR effector Avh5 with phosphatidylinositol 3-phosphate and for host cell entry. Mol. Plant-Microbe Interact. 26:330-344.

Tyler, B. M., Tripathy, S., Zhang, X., Dehal, P., Jiang, R. H. Y., Aerts, A., Arredondo, F. D., Baxter, L., Bensasson, D., Beynon, J. L., Chapman, J., Damasceno, C. M. B., Dorrance, A. E., Dou, D., Dickerman, A. W., Dubchak, I. L., Garbelotto, M., Gijzen, M., Gordon, S. G., Govers, F., Grunwald, N. J., Huang, W., Ivors, K. L., Jones, R. W., Kamoun, S., Krampis, K., Lamour, K. H., Lee, M. -K., McDonald, W. H., Medina, M., Meijer, H. J. G., Nordberg, E. K., Maclean, D. J., Ospina-Giraldo, M. D., Morris, P. F., Phuntumart, V., Putnam, N. H., Rash, S., Rose, J. K. C., Sakihama, Y., Salamov, A. A., Savidor, A., Scheuring, C. F., Smith, B. M., Sobral, B. W. S., Terry, A., Torto-Alalibo, T. A., Win, J., Xu, Z., Zhang, H., Grigoriev, I. V., Rokhsar, D. S. & Boore, J. L. 2006. *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science 313:1261-1266.

Van der Plaats-Niterink, A. J. 1981. Monograph of the genus *Pythium*. Stud. Mycol. 21:1-244.

Wawra, S., Trusch, F., Matena, A., Apostolakis, K., Linne, U., Zhukov, I., Stanek, J., Koźmiński, W., Davidson, I., Secombes, C. J., Bayer, P. & van West, P. 2017. The RxLR motif of the host targeting effector AVR3a of *Phytophthora* infestans is cleaved before secretion. Plant Cell 29:1184-1195.