BioMed Central

Research article

Open Access

Expressed sequence tags from the oomycete fish pathogen *Saprolegnia parasitica* reveal putative virulence factors

Trudy Torto-Alalibo[1], Miaoying Tian[1], Kamal Gajendran[2], Mark E Waugh[2], Pieter van West[3] and Sophien Kamoun*[2]

Address: [1]Department of Plant Pathology, The Ohio State University, Ohio Agricultural Research and Development Center, Wooster, Ohio, USA, [2]National Center for Genome Resources, Santa Fe, New Mexico, USA and [3]Aberdeen Oomycete Group, College of Life Sciences and Medicine, University of Aberdeen, Foresterhill, Scotland, United Kingdom

Email: Trudy Torto-Alalibo - trutot@gmail.com; Miaoying Tian - mt326@cornell.edu; Kamal Gajendran - kg@ncgr.org; Mark E Waugh - mark@compbiogen.com; Pieter van West - p.vanwest@abdn.ac.uk; Sophien Kamoun* - kamoun.1@osu.edu

* Corresponding author

Abstract

Background: The oomycete *Saprolegnia parasitica* is one of the most economically important fish pathogens. There is a dramatic recrudescence of *Saprolegnia* infections in aquaculture since the use of the toxic organic dye malachite green was banned in 2002. Little is known about the molecular mechanisms underlying pathogenicity in *S. parasitica* and other animal pathogenic oomycetes. In this study we used a genomics approach to gain a first insight into the transcriptome of *S. parasitica*.

Results: We generated 1510 expressed sequence tags (ESTs) from a mycelial cDNA library of *S. parasitica*. A total of 1279 consensus sequences corresponding to 525944 base pairs were assembled. About half of the unigenes showed similarities to known protein sequences or motifs. The *S. parasitica* sequences tended to be relatively divergent from *Phytophthora* sequences. Based on the sequence alignments of 18 conserved proteins, the average amino acid identity between *S. parasitica* and three *Phytophthora* species was 77% compared to 93% within *Phytophthora*. Several *S. parasitica* cDNAs, such as those with similarity to fungal type I cellulose binding domain proteins, PAN/Apple module proteins, glycosyl hydrolases, proteases, as well as serine and cysteine protease inhibitors, were predicted to encode secreted proteins that could function in virulence. Some of these cDNAs were more similar to fungal proteins than to other eukaryotic proteins confirming that oomycetes and fungi share some virulence components despite their evolutionary distance

Conclusion: We provide a first glimpse into the gene content of *S. parasitica*, a reemerging oomycete fish pathogen. These resources will greatly accelerate research on this important pathogen. The data is available online through the Oomycete Genomics Database [1].

Background

Water molds such as *Saprolegnia* and *Aphanomyces* species are responsible for devastating infections on fish in aquaculture, fish farms and hobby fish tanks [2,3]. Members of the genus *Saprolegnia* cause saprolegniosis, a disease that is characterized by visible white or grey patches of filamentous mycelium on the body or fins of freshwater fish [2]. The oomycete *Saprolegnia parasitica* is economically one of the most important fish pathogens, especially on salmon and trout species. It causes tens of million dollar

losses to aquaculture business worldwide, notably in Scotland, Scandinavia, Chile, Japan, Canada, and the USA [4,5]. *S. parasitica* infections are second only to bacterial diseases. In Japan, there is an annual mortality rate of 50% in coho salmon and elver due to *S. parasitica* infections [5-8]. In the United States, "winter kill" in catfish caused by *Saprolegnia* results in financial loses of up to 50%, which represents an economic loss of $40 million [6]. In Scotland, saprolegniosis also causes significant losses with the main problem occurring in salmon hatcheries.

Previously, *Saprolegnia* infections were kept under control with malachite green, an organic dye that is very efficient at killing the pathogen. However, since 2002 the use of malachite green has been banned around the world, due to its carcinogenic and toxicological effects. This has resulted in dramatic recrudescence of *Saprolegnia* infections. Therefore, there is an urgent need for novel alternative methods of management of Saprolegniosis.

Saprolegnia is often considered an opportunistic pathogen that is saprotrophic and necrotrophic [6]. However, it has become apparent that some *S. parasitica* strains are highly virulent and able to cause primary infections on salmon [3,9,10]. Infections occur on both eggs and fish. On eggs the disease is manifested by profuse mycelial growth on the egg surface resulting in rapid death. On fish, *Saprolegnia* invades epidermal tissues and can infect the entire surface of the body [11]. It causes cellular necrosis as well as dermal and epidermal damage, which ultimately leads to death by heamodilution [5,12]. Severe *Saprolegnia* infections result in lethargic behaviour, loss of equilibrium and commonly death of the fish [12,13].

Oomycete species can be pathogenic on plants, insects, crustaceans, fish, vertebrate animals, and various microorganisms [14,15]. Oomycetes, including *Saprolegnia*, have many fungus-like characteristics, but are not true fungi. A number of studies have indicated that they should be classified with the golden-brown algae and diatoms as stramenopiles [16-18]. This implies that oomycetes evolved genetic and biochemical mechanisms for interaction with animals and plants that are different from those of true fungi [14]. Indeed, oomycetes have several clearly defined developmental stages that are not found in fungal pathogens. For example, *Saprolegnia* species have a complex life cycle that includes both sexual and asexual reproduction. The asexual spore or sporangium is formed at the end of hyphal cells and can release many motile primary zoospores [6]. The primary zoospores swim only for a short time before they encyst and release a secondary zoospore. Secondary zoospores are motile for a longer period and are the main infection spore [5,11]. Secondary zoospores are able to encyst and release new zoospores several times. This process is called "polyplanetism" [19],

and may have evolved to allow the zoospores to have several attempts to locate and infect a host [6]. Uniquely within the class of oomycetes, secondary zoospores of *Saprolegnia* can possess hairs that are thought to be required for attachment to the host [11,19]. For example zoospores of *S. parasitica* have long hooked hairs that are believed to increase efficiency of the attachment to the fish hosts [12,19].

Although different in their selection of host organisms, plant and fish pathogenic oomycetes have many features in common. Evidently, the formation of specialised spore structures including zoospores, sporangia and oospores are similar. Also, infection strategies are comparable to some extent, involving encystment and attachment of zoospores on host surfaces, and penetration of host tissues. Furthermore, it is hypothesized that similar to biotrophic plant pathogenic oomycetes, such as *Peronospora* and several *Phytophthora* species, suppression of host defenses is likely to play a critical role in *Saprolegnia* pathogenesis. Host defense suppression by oomycetes remains poorly understood and only a few pathogen molecules that suppress host defenses have been identified in pathogenic oomycetes [20-22]. There is intriguing evidence that *Saprolegnia*-infected fish appear to be immuno-compromised [23,24]. Possibly, virulence factors secreted by the pathogen might account for the immuno-suppression and the lack of an effective response to pathogen infection.

Despite the huge economic importance of animal pathogenic oomycetes, such as *S. parasitica*, very little is known about the fundamental molecular mechanisms underlying development, pathogenicity and host specificity [14]. A thorough understanding of the basic molecular processes in *Saprolegnia*, the nature of the interactions with its hosts, and the identification of genes and proteins involved in these processes, could lead to novel control strategies that increase fish health, reduce disease losses and increase profits. In this study we used a genomics approach to gain a first insight into the transcriptome of *S. parasitica*. We generated random cDNA sequences (expressed sequence tags or ESTs) from a cDNA library of *S. parasitica* to identify genes that inform us about the biology of this organism and that could be involved in pathogenicity. We provide an overview of the identified sequences as well as a detailed description of a number of notable cDNAs. The data is available through a publicly accessible website as part of the Oomycete Genomics Database (OGD) [1].

Results
cDNA library and sequencing
We constructed a unidirectional cDNA library using mRNA isolated from mycelium of ATCC90214, a *S.*

parasitica strain isolated from diseased salmon [25]. Mycelium was obtained from nutrient deprived 29-day-old *in vitro* culture. In other oomycetes, such as *P. infestans*, similar treatments promote the expression of stress-related genes and possibly mimic infection conditions [26,27]. A total of 2296 sequencing reactions corresponding to the 5' end of the cDNA insert were performed. Of these, 2102 gave readable sequences. The sequences and the quality (phred) scores were fed into NCGR's X Genome Initiative (XGI) [28] annotation pipeline and subjected to further quality controls [29]. 1510 ESTs remained after vector and low quality sequences were removed. Of these, 5% were assessed to be in reverse orientation based on the occurrence of at least eight consecutive A residues within the first 38 bp. Following additional quality screening and assembly, 1279 consensus sequences (so-called unigenes) were obtained consisting of 1146 singletons and 133 consensus with two or more ESTs. In total, 525,944 bp of assembled sequences were obtained corresponding to an average consensus sequence length of 411 bp. At 61%, the GC content of the assembled sequences was relatively high and similar to the GC content reported for *Phytophthora* spp. (57–58%) [14,30].

Sequence annotation

The 1279 consensus sequences were annotated using the methods implemented in the XGI pipeline (see methods). A total of 609 sequences (48%) showed significant similarities to known protein sequences (E value < 10⁻⁵) based on BLASTX searches, 398 (31%) gave significant hits to protein motifs in the BLOCKS+ database, and 600 (47%) gave hits to the InterPro protein motif database with at least one of the 12 algorithms implemented in InterProScan. Among these, InterPro database searches with HmmPfam revealed 340 hits. In total, 585 consensus sequences (46%) could be assigned identities based on the Gene Ontology (GO) Consortium (see methods). The differences between the different analyses are expected for such bioinformatics annotations and reflect, among other things, differences in sensitivity between the various programs. A total of 70 sequences were positive with the PexFinder algorithm and are candidate for carrying signal peptides and encoding extracellular proteins.

Taxonomic identity of the homologs of S. parasitica cDNAs

Considering the classification of oomycetes as stramenopiles, it was interesting to systematically examine the taxonomic identity of the homologs of *S. parasitica* cDNAs. To this end, we took advantage of the availability of several eukaryotic genomes to compile a data set of 270334 proteins covering six major phyla of eukaryotes: fungi, animals, plants, alveolates, discicristates, and heterokonts. The data included the complete proteomes of at least one species for all phyla except for discicristates (see
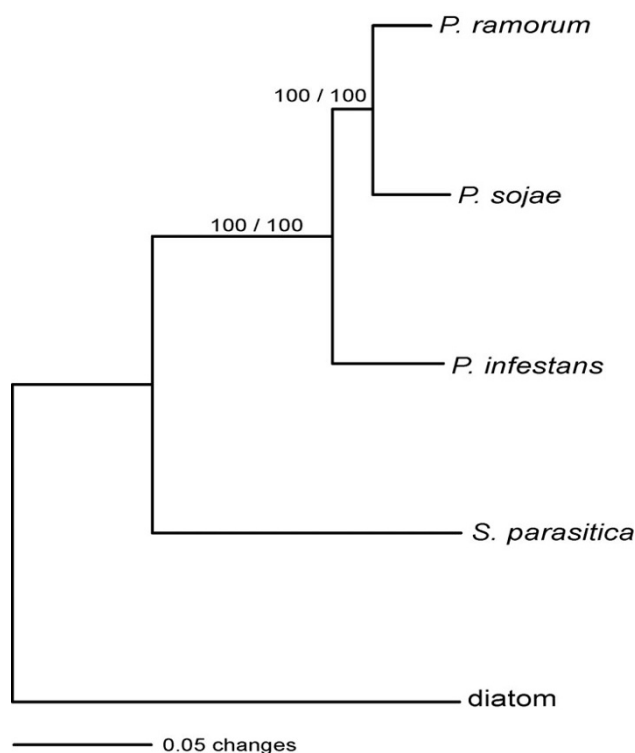

Figure 1
Phylogenetic relationships between *Saprolegnia parasitica* and four other stramenopiles. The phylogenetic tree was constructed using the neighbor joining method based on concatenated alignments from 18 conserved proteins (2533 amino acids). Percentile bootstrap values based on 1000 replications and obtained with the neighbor joining methods/maximum parsimony methods are indicated at the nodes. The scale bar represents 5% weighted amino acid sequence divergence.

methods). We used BLASTX to compare the 1279 *S. parasitica* unigenes to these eukaryotic proteins. In total, 715 sequences showed no significant hits (E value > e-5). Of the 582 sequences that showed significant hits, 32% (185) had the top hit to a diatom protein confirming the affinity between diatoms and oomycetes as stramenopiles. In contrast, only 56 (about 10%) sequences had a fungal protein as a top hit.

Phylogenetic analyses

We also exploited the sequence data to examine phylogenetic affinities between *S. parasitica*, three *Phytophthora* spp., and the diatom *Thalassiosira pseudonana*. We performed reciprocal BLAST searches to identify a common set of conserved protein sequences between the five species. Multiple alignments of the conserved portions of 18 different proteins covering 2533 amino acids were con-

Table 1: Most represented protein domains in the *Saprolegnia parasitica* EST contigs

InterPro ID	Description	Number of contigs
IPR000719	Eukaryotic protein kinase	22
IPR004000	Actin and actin-like	11
IPR001993	Mitochondrial energy transfer proteins (carrier protein)	9
IPR000626	Ubiquitin domain	6
IPR002110	Ankyrin-repeat	5
IPR001680	G-protein beta WD-40 repeats	5
IPR000795	GTP-binding elongation factor	5
IPR001353	Multispecific proteasome proteases	5
IPR003008	Tubulin/FtsZ family	5
IPR000873	AMP-dependent synthetase and ligase	4
IPR000254	Cellulose-binding domain, fungal type	4
IPR002130	Cyclophilin-type peptidyl-prolyl cis-trans isomerase	4
IPR002048	EF-hand	4
IPR000173	Glyceraldehyde 3-phosphate dehydrogenase	4
IPR000232	Heat shock factor (HSF)-type DNA-binding domain	4
IPR001806	Ras GTPase superfamily	4
IPR000504	RNA-binding region RNP-1 (RNA recognition motif)	4
IPR001464	Annexin family	3
IPR001623	DnaJ N-terminal domain	3
IPR001753	Enoyl-CoA hydratase/isomerase	3
IPR001179	FKBP-type peptidyl-prolyl cis-trans isomerase (PPIase)	3
IPR001005	Myb DNA binding domain	3
IPR003014	PAN domain	3
IPR000668	Papain cysteine protease (C1) family	3
IPR001272	Phosphoenolpyruvate carboxykinase (ATP)	3
IPR001849	Pleckstrin homology (PH) domain	3
IPR001232	SKP1-like	3
IPR000063	Thioredoxin	3
IPR001440	TPR repeat	3

catenated and used in phylogenetic analyses. The obtained tree clearly supported a monophyletic relationship between the four oomycetes, and consistent with published phylogenies of *Phytophthora* [31,32] suggested that *P. sojae* and *P. ramorum* are more closely related (Fig. 1). Average amino acid identity among sequences of the *Phytophthora* spp. was 93% (range, 91.1–95.0%). In contrast, average amino acid identity between *S. parasitica* and the three *Phytophthora* spp. was 77% (range, 76.4–77.9%). Average amino acid identity between the diatom and the four oomycetes was 66.7% (range, 66.2–67.5%).

Most represented protein domains
As described above, searches of the InterPro database with HmmPfam revealed 340 hits among the 1297 consensus sequences. We classified the InterPro domains based on their abundance (Table 1). The 340 hits corresponded to 198 different InterPro domains. Of these, 29 domains were represented three times or more (range, 3–22). By far the most abundant domain was IPR000719 for eukaryotic protein kinase, which was represented 22 times. Other domains that function in signal transduction, such as

ankyrin-repeat, G-protein, Ras GTPase, myb DNA binding, were also relatively abundant. Other well-represented domains comprised fungal type I cellulose-binding domain (IPR000254), PAN domain (IPR003014), and papain cysteine protease family (IPR000668).

Fungal type I cellulose binding domain
Four sequences showed similarities to fungal-type I cellulose binding domain (CBD) (InterPro domain IPB000254). Three of these occurred in cDNAs predicted to encode extracellular proteins. Two cDNAs, Sp_002_00594 and Sp_001_01439, encoded putative proteins with two CBDs. The full length sequence of cDNA SPM5F8 (Sp_001_01439) was obtained (GenBank accession number DQ143887). This cDNA contained an ORF of 306 bp corresponding to a protein of 101 amino acids. SignalP [33] analysis of the predicted protein identified a 20-amino acid signal peptide with a significant mean S value of 0.93. Domain IPB000254 has been mainly reported in fungi [34], but also occurs in one protein from *Ectocarpus siliculosus* Virus EsV-1 (*Phycodnaviridae*), a viral pathogen of brown algae [35]. To determine

the extent to which the CBD occurs in oomycetes, we performed iterative BLAST searches of all publicly available *Phytophthora* sequences using the *S. parasitica* CBD sequences. In total, 12 different sequences similar to the *S. parasitica* CBDs were recovered from *P. infestans* (3), *P. sojae* (4), and *P. ramorum* (5). The 18 oomycete CBD sequences aligned perfectly over a 34 amino acid region. Multiple alignments of the oomycete CBDs with the well-studied CBD of cellobiohydrolase I (Cel6A) of *Trichoderma resei* [36,37] suggested that the major features of the domain are conserved in oomycetes (Fig. 2). Amino acid residues defining the Cel6A domain including the four cysteine backbone, as well as a glutamine (Gln32) and the three tyrosines (Tyr3, Tyr29 and Tyr30) that are important for binding to cellulose, were frequently conserved. Nonetheless, tyrosines, particularly Tyr3 and Tyr29, were often replaced by other aromatic residues, such as tryptophane and phenylalanine, as observed for various fungal CBDs, such as endoglucanase I of *T. resei* [38].

CBEL-like sequences

Seven distinct cDNA sequences showed similarity to the <u>C</u>ellulose <u>B</u>inding, <u>E</u>licitor and <u>L</u>ectin-like protein (CBEL), a 34-kDa extracellular glycoprotein that was first isolated from *Phytophthora parasitica* [39,40]. Five of these CBEL-like cDNAs were also predicted to encode extracellular proteins based on PexFinder analyses. CBEL has a dual function: (1) it is required for attachment to plant surfaces, (2) it elicits necrosis and defense gene expression in tobacco plants [39,40]. CBEL contains two regions with similarity to the PAN module (InterPro: IPR000177), a conserved domain that includes the Apple domain and functions in protein-protein or protein-carbohydrate interactions [41]. The similarity to CBEL in the identified *S. parasitica* cDNAs centered on the PAN module regions. Three of the seven *S. parasitica* cDNAs contained two PAN-like domains while the other four had a single domain. We used these sequences to survey the available *Phytophthora* sequences for additional PAN/CBEL-like domains using iterative BLAST searches. In total, 42 PAN/CBEL-like domains were identified in 28 putative proteins of *P. infestans*, *P. sojae*, *P. ramorum*, and *P. parasitica* suggesting that this domain is widely distributed in oomycetes (Fig. 3). Multiple alignment of the 52 oomycete CBEL-like domains revealed a conserved pattern centered around a conserved core of six cysteines (Fig. 3B).

Glycosyl hydrolases

Six cDNAs showed similarity to various classes of glucanases. One of these, Sp_001_01488, showed significant similarity to microbial endo-1,3-β-glucanases (glycosyl hydrolase family 17), as well as high similarity to the recently described gene, *piendo1*, from *P. infestans* [42]. The 1258 bp insert of a full-length cDNA (SPM16A2) corresponding to this endo-1,3-β-glucanase was fully
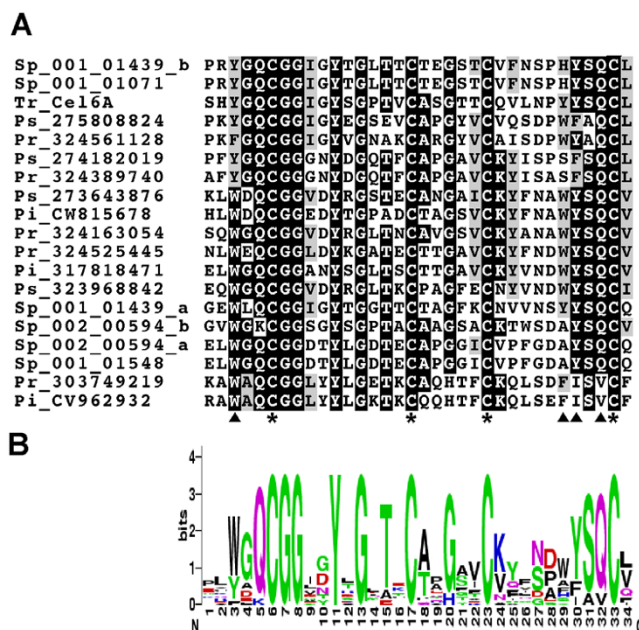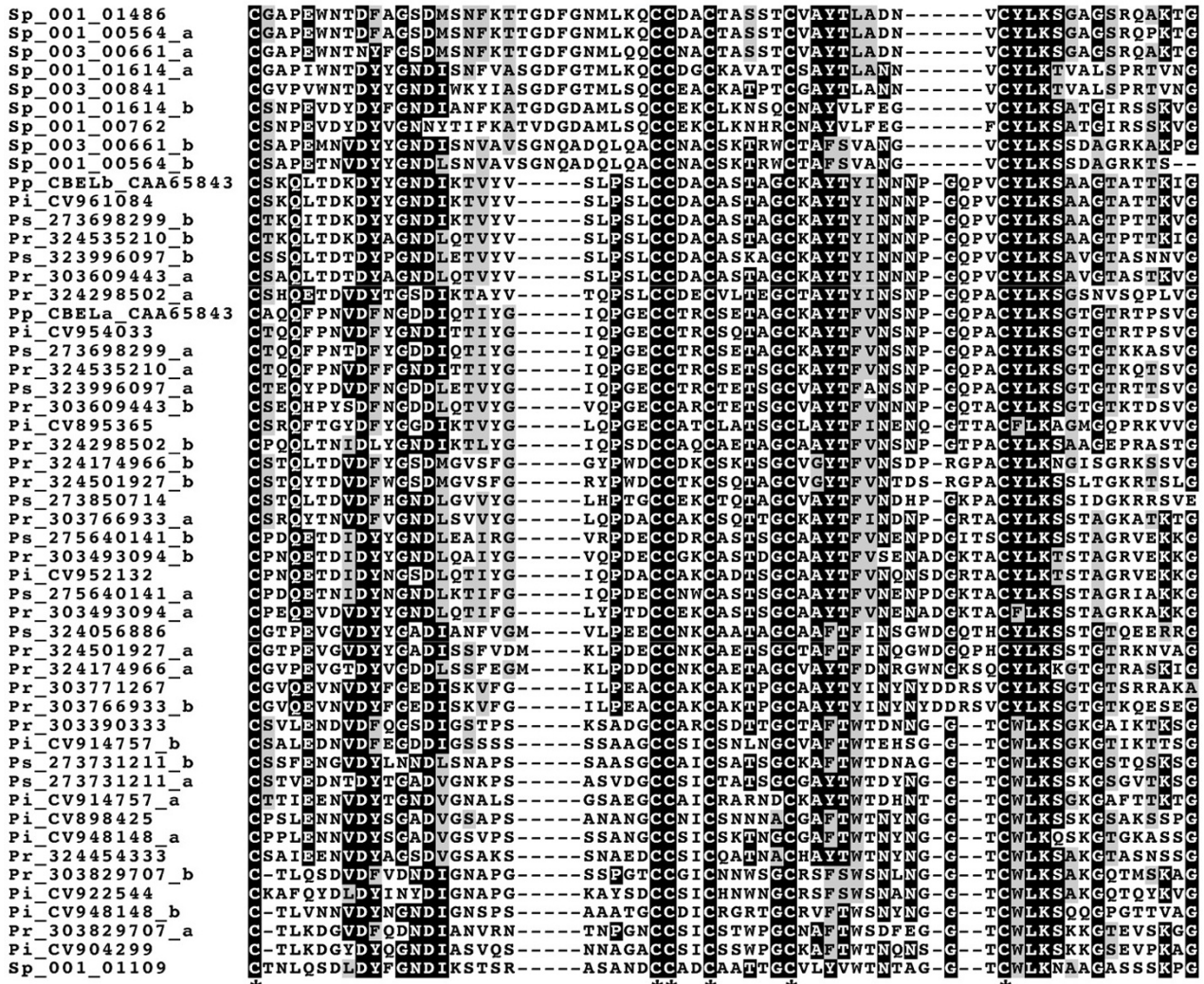

Figure 2
Fungal type I cellulose binding domain (CBD) is widespread in oomycetes. (A) Multiple sequence alignment of 18 oomycete type I CBDs with the domain of *Trichoderma resei* Cel6A (Tr_Cel6A). The four conserved cysteine residues are marked with asterisks. The glutamine and three aromatic residues that are known to be important for binding the carbohydrate substrate are shown by arrows. Sequence names refer to the *Saprolegnia parasitica* unigene (Sp_), *Phytophthora infestans* (Pi_), *Phytophthora sojae* (Ps_), and *Phytophthora ramorum* (Pr_) followed by the OGD accession number, GenBank accession number, or NCBI Trace Archive identifier (Ti number). Multiple domains originating from the same sequence are marked with the letters a or b at the end of the sequence name. (B) Consensus sequence pattern of the oomycete type I CBD. Consensus sequence was calculated using WebLogo. The bigger the letter, the more conserved the amino acid site. The positions of amino acids in the consensus sequence correspond to the positions in the sequence alignment in panel A.

sequenced (GenBank accession number <u>AY974332</u>). The sequence revealed a single ORF of 1197 bp encoding a predicted protein of 398 amino acids with 38% identity to PIENDO1. SignalP [33] analysis of the predicted protein identified a 19-amino acid signal peptide with a significant mean S value of 0.94. Previously, phylogenetic analyses of several *Phytophthora* hydrolytic enzymes, including PIENDO1, revealed unexpected affinity to fungal proteins [27,42-44]. Indeed, BLASTP searches of the protein encoded by SPM16A2 showed significant similarities to fungal and bacterial proteins (top hits with E value = 2e-
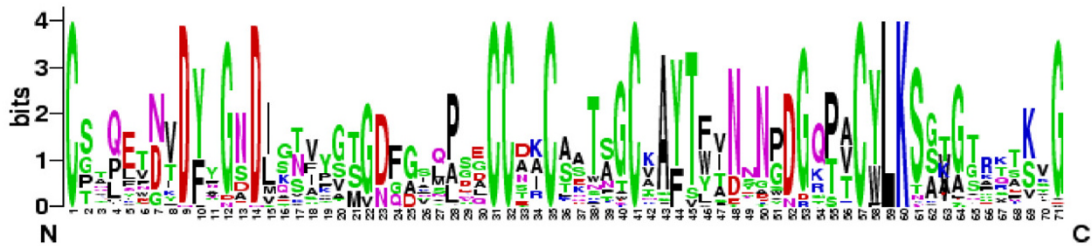
A

B

Figure 3
PAN module/Apple domain is widespread in oomycetes. (A) Multiple sequence alignment of 52 oomycete PAN module/Apple domain including the two domains of the previously described CBEL protein of *Phytophthora parasitica* (Pp_CBELa and Pp_CBELb). The six conserved cysteine residues are marked with asterisks. Sequence names refer to the *Saprolegnia parasitica* unigene (Sp_), *Phytophthora infestans* (Pi_), *Phytophthora sojae* (Ps_), and *Phytophthora ramorum* (Pr_) followed by the OGD accession number, GenBank accession number, or NCBI Trace Archive identifier (Ti number). Multiple domains originating from the same sequence are marked with the letters a or b at the end of the sequence name. (B) Consensus sequence pattern of the oomycete oomycete PAN module/Apple domain. Consensus sequence was calculated using WebLogo. The bigger the letter, the more conserved the amino acid site. The positions of amino acids in the consensus sequence correspond to the positions in the sequence alignment in panel A.

21 and 3e-22, respectively) but none to other eukaryotic proteins.

Proteases

We found a set of 12 cDNAs with similarity to aspartyl (2), serine (3), and cysteine (7) proteases among the annotated sequences of *S. parasitica*. The sequence of SPM3B2, a full length cDNA corresponding to unigene Sp_004_00851 was obtained (GenBank accession number AY974331). This cDNA encoded a putative protein of 379 amino acids. BLASTP searches of the MEROPS database [45] revealed significant similarity to pepsin aspartic proteases such as cathepsin D (MEROPS Family A01, E value = 1e-72 for best hit). SignalP [33] analysis of the predicted protein identified a 17-amino acid signal peptide with a significant mean S value of 0.73. We also determined the full length sequence of cDNA SPM9F1 (Sp_001_01152) (GenBank accession number AY974330). This cDNA encoded a putative protein of 524 amino acids with significant similarity to papain cysteine proteases (MEROPS Family C01A, E value = 1e-58 for best hit). SignalP [33] analysis of the predicted protein identified a 22-amino acid signal peptide with a significant mean S value of 0.75. BLASTP searches against GenBank NR and the *Phytophthora* data sets revealed that both proteases are widely distributed among eukaryotes and oomycetes.

Protease inhibitors

We have also identified two *S. parasitica* cDNAs with similarity to protease inhibitor domains of two structural classes: (1) Kazal-like serine protease inhibitor (InterPro domain IPR002350, MEROPS family I1), (2) cysteine protease inhibitor (InterPro IPR000010, MEROPS family I25). We further analyzed these two cDNAs by aligning their putative inhibitor domains to those of known protease inhibitors (Fig. 4). Sp_001_01027 showed significant similarity to the Kazal-like inhibitors recently described by Tian *et al.* [21] from *P. infestans* and other plant pathogenic oomycetes. Amino acid residues defining the Kazal motif, including the six cysteine backbone, tyrosine and asparagine residues, were conserved in Sp_001_01027 (Fig. 4A). The predicted active site P1, which is central to the specificity of Kazal inhibitors [46,47], consisted of a proline, and therefore differed from all previously reported oomycete Kazal domains [21]. Sp_001_01374 is predicted to encode a secreted protein that bears the hallmark of the cystatin class of cysteine protease inhibitors including the highly conserved QXVXG motif in the first binding loop (L1) [48] (Fig. 4B). These findings suggest that secretion of protease inhibitors is a common feature of oomycetes.
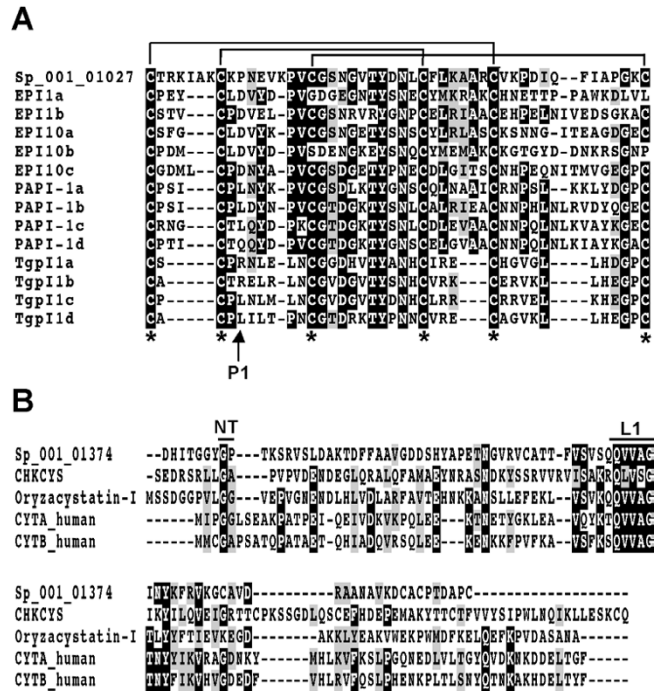

Figure 4
Protease inhibitors in *Saprolegnia parasitica*. (A) Sequence alignment of Sp_001_01027 predicted amino acid sequence with representative Kazal family inhibitor domains. Protein names correspond to protease inhibitors of *Saprolegnia parasitica* Sp_001_01027, *Phytophthora infestans* EPI1 (EPI1a-b, AY586273) and EPI10 (EPI10a-c, AY586282), the crayfish *Pacifastacus leniusculus* (PAPI-1a-d, CAA56043), and the apicomplexan *Toxoplasma gondii* (TgPI-1a-d, AF121778). The conserved cysteine residues that define the Kazal family protease inhibitor domain are marked with asterisks. The putative disulfide linkages formed by cysteine residues within the predicted Kazal domains are shown. The position of the predicted P1 residues is shown by an arrow. (B) Sequence alignment of *Saprolegnia parasitica* Sp_001_01374 (N-terminal fragment of the mature protein), chicken egg white cystatin (CHKCYS, P01038, mature protein), rice oryzacystatin-I (Oryzacystatin-I, P09229), human cystatin A (CYTA_human, P01040) and cystatin B (CYTB_human, P04080). The proposed active-site residues in cystatins, forming the N-terminal trunk (NT) and first binding loop (L1), are indicated.

Thiamine biosynthetic enzyme

We identified one sequence (Sp_001_00801) with significant similarity to a thiamine biosynthetic enzyme from plants (top hit to protein AAV92556 from the conifer *Pseudotsuga menziesii*, E value = 2e-63) and fungi (*Schizosaccharomyces pombe* protein CAA21093, E value = 3e-53). Unlike *S. parasitica* and other oomycetes, members of the genus *Phytophthora* are thiamine auxotrophs, they require exogenous sources of thiamine for growth

[49,50]. Interestingly, BLAST searches of the genome sequence reads of *P. sojae* and *P. ramorum*, as well as all available sequences of *P. infestans*, failed to reveal sequences with similarity to Sp_001_00801 or to the plant and fungal enzymes. These findings suggest that this thiamine biosynthetic enzyme may have been lost in the *Phytophthora* lineage and could be related to thiamine auxotrophy in this genus.

Discussion

In this study we generated 1510 high quality ESTs from *S. parasitica*, an economically important and reemerging oomycete pathogen that causes multimillion dollar losses in the aquaculture industry. The ESTs were generated from a cDNA library constructed from one-month old nutrient deprived mycelium cultures. So far significant data sets of oomycete ESTs have been described for three plant pathogenic species, *P. infestans* [26,27], *P. sojae* [30] and *P. parasitica* [51]. Therefore, the *S. parasitica* ESTs offer some insights into the transcriptome of animal pathogenic oomycetes, which have been extremely understudied. Prior to this work, only 13 nucleotide and 2 protein sequences of *S. parasitica* could be retrieved from Gen-Bank (March 2005 release). The sequence data and the corresponding annotations described in this study are accessible through an interactive public resource, the Oomycete Genomics Database (OGD). We hope that this pilot genomics project will accelerate research on this important pathogen and lays the foundation for more significant genome and cDNA sequencing initiatives of animal pathogenic oomycetes.

We used the *S. parasitica* ESTs to confirm the phylogenetic affinities between oomycetes and diatoms [16-18]. About 32% of the *S. parasitica* sequences that showed significant similarities to eukaryotic proteins matched a protein of the diatom *T. pseudonana* as a top hit. Within the oomycetes, *S. parasitica* is classified with other water molds, such as *Achlya* and *Aphanomyces*, in the order Saprolegniales [52-54]. These species are morphologically very distinct from the great majority of plant pathogens, such as the Peronosporales *Phytophthora* and downy mildews, or the Pythiales *Pythium* [55]. Indeed, the *S. parasitica* sequences tended to be relatively divergent from *Phytophthora* sequences. For example, based on the sequence alignments of 18 different conserved proteins, the average amino acid identity between *S. parasitica* and three *Phytophthora* spp., *P. infestans*, *P. sojae*, and *P. ramorum*, was 77% compared to 93% within *Phytophthora*. Differences in transcript content were also noted. cDNAs with similarity to elicitins, a group of 10-kDa proteins that occur in all *Phytophthora* and some *Pythium* species [56-61] and form 1–2% of mycelial ESTs in *Phytophthora* [26,27,62], were not found in the *S. parasitica* dataset. Although, elicitin-like genes could very well occur in the genome of *S. parasitica*, they do not seem to be abundantly expressed in mycelium. Elicitins were shown to function as sterol carriers [63,64]. Among the oomycetes, members of the Saprolegniales are able to synthesize sterols de novo whereas *Phytophthora* and *Pythium* spp. are sterol auxotrophs [50]. Possibly, *S. parasitica* may not require sterol carriers, such as elicitins, for optimal hyphal growth. Another difference between *Phytophthora* and *Saprolegnia* involves thiamine metabolism. Members of the genus *Phytophthora* are thiamine auxotrophs and require exogenous sources of thiamine for growth [49,50]. We identified one sequence in *S. parasitica* (Sp_001_00801) that shows significant similarity to a thiamine biosynthetic enzyme from plants and fungi but that is absent in the draft genome sequences of *P. sojae* and *P. ramorum*. This finding suggests that this thiamine biosynthetic enzyme may have been lost in the *Phytophthora* lineage and could be related to thiamine auxotrophy in this genus.

We searched the annotated data set for *S. parasitica* sequences that show similarities to known proteins and protein motifs that could inform us about the biology and pathology of this microbe. About half of the unigenes showed similarities to known protein sequences and could be assigned a putative function. A number of sequences showed particularly interesting similarities. cDNAs with similarity to signal transduction proteins, such as kinases and transcription factors, were particularly abundant. In total, 70 cDNAs encoded proteins with a putative signal peptide that are potentially secreted to the extracellular space. Secretion is an essential mechanism for delivery of virulence factors by eukaryotic pathogens to their appropriate site in infected host tissue. Therefore, several putative secreted proteins of *S. parasitica*, such as CBD proteins, CBEL-like proteins, glycosyl hydrolases, proteases, and protease inhibitors could function in virulence and will be worthy of additional studies.

Phylogenetic analyses indicated that several *Phytophthora* proteins, particularly hydrolytic enzymes such endopolygalacturonases, pectate lyases, exo-1,3-beta-glucanases, and an endo-1,3-beta-glucanase, are more similar to fungal proteins than to their counterparts in other eukaryotes [27,42-44]. These observations are in sharp contrast with phylogenies constructed from ribosomal sequences or compiled protein sequences from mitochondrial and housekeeping chromosomal genes, which indicate considerable evolutionary distance between oomycetes and fungi [16-18,65,66]. The apparent discrepancies between these phylogenies could reflect convergent evolution in the arsenal of hydrolytic enzymes between these pathogens, perhaps as a result of common mechanisms of infection among filamentous microbes [27,67]. The *S. parasitica* sequences allowed us to evaluate whether the similarity to fungal proteins extends to oomycetes other

than *Phytophthora* and to animal pathogenic oomycetes. Although no *S. parasitica* sequences similar to the cell wall degrading enzymes endopolygalacturonases and pectate lyases were found, we identified a cDNA, SPM9F1, that encodes a 524 amino acid protein with high similarity to endo-1,3-β-glucanases, including the recently described PIENDO1 of *P. infestans* [42]. Similar to PIENDO1, SPM9F1 was most similar to fungal glucanases and no significant BLASTP hits (E value > 0.01) were observed to non-fungal eukaryotic proteins. Therefore, conservation in the arsenal of hydrolytic enzymes appears to extend beyond *Phytophthora* spp. to the Saprolegniales and animal pathogenic oomycetes.

Domain annotation of the *S. parasitica* sequences revealed the occurrence of a protein domain typically associated with fungi. Type I CBDs (InterPro domain IPB000254) are thought to be unique to fungi [34], although a related domain also occurs in the brown algae viral pathogen *Ectocarpus siliculosus* Virus EsV-1 (*Phycodnaviridae*) [35,68]. In this study, we found that this domain is widespread and diverse in *S. parasitica* and other oomycetes. A total of 18 domains from four oomycete species were found to share a 34 amino acid region that aligns perfectly with the canonical *T. resei* Cel6A CBD highlighting a core of conserved four cysteines and aromatic residues known to bind the cellulosic substrate [36,38]. Interestingly, the occurrence of this CBD in a virus of brown algae, which are related to oomycetes, suggests that type I CBDs might be more widespread in stramenopiles although we did not detect them in the draft genome sequence of the diatom *T. pseudonana*. In fungi, the CBDs are usually located in the N- or C-terminal regions of hydrolytic enzymes, such as cellulases and xylanases, and function by concentrating the catalytic domains on the surface of the insoluble cellulose substrate [34]. One of the *S. parasitica* cDNAs, SPM5F8, encodes a small 101 amino acid protein with two CBDs. Such a protein could function as a scaffolding component of the multienzyme complex known as cellulosome [34]. The function of this and other CBD proteins in *S. parasitica* may relate to attachment to organic debris on the host surface or during saprophytic growth. Alternatively, since cellulose is a major component of the cell wall of oomycetes, these proteins may play endogenous function in cell wall biogenesis.

Seven *S. parasitica* cDNAs showed similarity to CBEL, a 34-kDa cell wall glycoprotein of *P. parasitica* that binds to cellulose and host surfaces, functions in the agglutination of red blood cells, and elicits necrosis and defense gene expression in tobacco [39,40]. The similarity centered mainly on two regions of CBEL that match the PAN module/Apple domain (InterPro IPR000177). The CBEL-like PAN module, which is thought to function in protein-protein or protein-carbohydrate interactions [41], appeared

to be particularly diverse in oomycetes with 52 different sequences identified in five species. The PAN module was found in proteins with diverse functions, such as the blood coagulation factor XI and the plasma protein pre-kallikrein [41]. Recently, several secreted proteins from apicomplexan mammalian parasites were found to contain Apple-like domains and are thought to play a role during parasite attachment and invasion of host cells [69-72]. For example, MIC4, an adhesin secreted by the apicomplexan *Toxoplasma gondi*, contains six Apple domains [69]. It remains to be determined the extent to which the secreted PAN/CBEL-like proteins of *S. parasitica* play a role in attachment and invasion during interaction with the fish host. Nonetheless, it appears that in oomycetes, similar to the apicomplexan parasites, some adhesins are secreted PAN module proteins.

Proteolytic enzymes are considered important virulence factors that aid in host colonization and release of nutrients by animal pathogenic microbes. It has long been known that the *Saprolegnia* spp. pathogenic on fish exhibit significant extracellular protease activity and it was postulated that this enzymatic activity contributes to pathogenesis [73]. A serine protease gene, *AaSP2* from the related crayfish pathogen *Aphanomyces astacus*, was recently characterized and shown to be highly expressed during *in vivo* growth [74]. However, besides *AaSP2*, genes for secreted proteases of animal pathogenic oomycetes have not been reported. In this study, we identified a diverse set of 12 cDNAs of *S. parasitica* with similarity to the major catalytic classes of proteases. A number of the identified proteases had a signal peptide that would predict them to be localised at the interface between pathogen and host and suggests that they are candidate virulence factors.

Tian et al. [21] recently reported that plant pathogenic oomycetes secrete a diverse family of Kazal-like serine protease inhibitors with at least 35 members identified from *P. infestans*, *P. sojae*, *P. ramorum*, *P. brassicae*, and the downy mildew *Plasmopara halstedii*. Among these, the two-domain EPI1 protein and the three domain EPI10 of *P. infestans* were found to inhibit and interact with P69B, a defense subtilase of tomato, and were suggested to play a role in counterdefense [21]. Inhibitors of serine protease might be ubiquitous among eukaryotic parasites. For instance, the apicomplexan obligate parasite *Toxoplasma gondii* secretes TgPI-1 and TgPI-2, four-domain serine protease inhibitors of the Kazal family [75-77], and the intestinal hookworm *Ancylostoma ceylanicum* secretes an 8-kDa broad spectrum serine protease inhibitor of the Kunitz family [78]. Here we found that Kazal-like motifs also occur in Saprolegniales proteins.

In addition to Kazal-like motifs, we also discovered a cDNA that encodes a secreted protein with similarity to

the cystatin class of cysteine protease inhibitors [79]. Cysteine protease inhibitors, such as chagasin, have been reported in animal parasites, mainly trypanosomids, and are thought to target proteases of the insect vector or the mammalian host [80-82]. Perhaps, inhibition of host proteases is a widespread counterdefense strategy in animal and plant pathogenic eukaryotes. Future studies will help to address whether the discovered protease inhibitors play a role in *S. parasitica*-fish interactions.

Conclusion
This pilot cDNA sequencing project provides a first look into the gene content of *S. parasitica* and sets the basis for genomics research in this reemerging animal pathogen. Annotation of the ESTs revealed a number of genes that could function in virulence. Future work will focus on developing molecular tools for functional analysis of *S. parasitica* genes. In this regards, stable transformation of *Saprolegnia monoica* has been reported [83], and the RNAi protocol recently developed for *P. infestans* [84] should be adaptable to *S. parasitica*. Gene expression profiling will also be applied to investigate transcriptome changes during *S. parasitica*-fish interactions. Overall, these resources will greatly accelerate research on this important pathogen and could lead to novel perspectives for controlling saprolegniosis.

Methods
Strains and growth conditions
Saprolegnia parasitica ATCC90214, an isolate from lesions on coho salmon (*Oncorhynchus kisutch*) [25], was used in this study. Working stocks of this strain were routinely maintained on cornmeal agar (Difco Lab. Detroit, MI) at 18°C. To obtain axenically prepared mycelium, ATCC90214 was grown in GY broth (5 g glucose, 2.5 g yeast extract/L) for 29 days, which corresponds to stationary phase. Mycelium was harvested by filtration and immediately frozen prior to RNA extraction.

cDNA construction
Total RNA from *S. parasitica* mycelium was isolated using the phenol-guanidine isothiocyanate based reagent Trizol, (Life Technologies Carlsbad, CA) according to the manufacturer's instructions. PolyA+ mRNA was isolated using the oligotex mRNA purification kit (Qiagen, Valencia, CA). The cDNA library was synthesized and cloned in plasmid pSPORT1 using the Superscript™ plasmid system for cDNA synthesis and cloning (Invitrogen Life Technologies, Carlsbad, CA). Polyadenylated mRNA was used to synthesize oligo (dT) primed cDNAs, which were cloned unidirectionally in *Not*I/*Sal*I digested vector pSPORT1. Plasmid ligations were transformed into *Escherichia coli* ElectroMax-DH10B ™ cells (Invitrogen Life Technologies, Carsbad, CA). Selection was done on Luria-Bertani (LB) agar plates containing ampicillin (50 mg/L) [85]. Individ-

ual colonies were picked randomly with the Qpix robot (Genetix, Hampshire, UK) into 384 well plates containing LB freezing buffer (36 mM K_2HPO_4, 13.2 mM K_2HPO_4, 1.7 mM citrate, 0.4 mM $MgSO_4$, 6.8 mM $(NH4)_2SO_4$, 4.4 % v/v glycerol in 1 × LB), incubated overnight without shaking, and stored at -80°C. Subsequently clones were transferred from the 384 well plate to 96 well plate for shipment to the Genomics Technology Support Facility (GTSF) at the Michigan State University where they were sequenced following manufacturer's recommendations using an ABI Prism 3700 DNA Analyzer. Identification codes for the cDNAs/ESTs were derived from the position of the corresponding cDNA clone in the microtiter plates preceded by SPM (for *Saprolegnia parasitica* mycelial) and the successive number of the microtiter plate.

DNA sequencing
For the ESTs, DNA from bacterial cultures was purified at GTSF using Qiagen 3000 or Autogen 850 robots. Fluorescently labeled sequencing products were generated using the universal T7 primer resulting in 5' cDNA sequences. The sequencing products were separated by capillary electrophoresis on an ABI Prism 3700 DNA Analyzer (PE Applied Biosystems). A dataset representing 2296 EST sequences and the corresponding electropherograms were then made available through the Geospiza Finch web interface of GTSF. The complete inserts of selected cDNAs were sequenced by primer walking at the OARDC Molecular and Cellular Imaging Center (MCIC), Wooster, Ohio, using an ABI Prism 377 automated sequencer (PE Applied Biosystems).

Bioinformatics
The sequences were processed using the XGI pipeline [28]. The assembly described in this paper is known as the May 2004 assembly. The consensus sequences (unigenes) were named Sp_N1_N2_May04 with N1 referring to the number of ESTs in the contig, and N2 the contig number. The consensus sequences were annotated using the methods implemented in the XGI pipeline [29]. These include BLASTX [86] searches against NCBI non-redundant (nr) protein library; BLIMPS search against Blocks+ protein motif database [87,88]; searches with the 12 algorithms of InterProScan [89] against the InterPro database [90]; and identification of signal peptides for extracellular secretion with PexFinder [91], an algorithm based on SignalP 2.0 [33,92]. Automated post-analysis annotation links BLAST and Blocks+ hits to their cognate Gene Ontology entries [93,94], whereas InterPro hits are automatically linked to GO annotations.

Additional similarity searches using BLAST [86] and other bioinformatics analyses were also performed locally on Mac OSX G4/G5 workstations. BLAST E-value lower than 0.01 were retained, and searches were conducted with the

low-complexity filter on. Local databases were compiled from GenBank nonredundant (NR), dBEST, and TraceDB databases [95] and the Broad Institute [96]. They included "darwin_270334.faa" a curated dataset of 270334 eukaryotic proteins that we compiled. The data covers six major phyla: fungi, animals, plants, alveolates, discicristates, and heterokonts. It includes the complete proteomes of 17 species and at least the complete proteome of one species for all phyla except for discicristates. The MEROPS database of proteases and protease inhibitors was also queried [45]. Multiple alignments were conducted using the program Clustal-X [97], adjusted manually as necessary, and visualized with BOXSHADE [98]. Consensus sequences were visualized with weblogo [99]. A cDNA was deemed likely to be full length when it was the most 5' proximal EST among assemblies and gave hits to the N-terminal portion of known proteins following similarity searches.

Phylogenetic analysis

A data set of concatenated protein sequences was developed to perform phylogenetic comparisons of four oomycete species and the diatom *Thalassosira pseudonana* [100]. First, BLASTX searches of the *S. parasitica* unigenes against the diatom proteome were performed. Matching sequences with E value < 1e-20 were extracted and then used to search WGS and EST reads of *P. infestans*, *P. sojae*, and *P. ramorum*. A total of 18 sequences that were conserved among all five species were identified and were aligned individually with Clustal-X. Poorly aligned edges were then trimmed and the alignments were concatenated. PAUP v4.0b8 (Sinauer Associates Inc., Sunderland, MA) was used to reconstruct phylogenetic trees using the neighbor joining method and maximum parsimony with 1000 bootstrap replications.

Data dissemination

The DNA sequences, assemblies, and annotations are publicly available through the Oomycete Genomics Database (OGD) [1]. The 1510 high quality ESTs were also deposited in NCBI's GenBank under accession numbers DN615772-DN617281.

Authors' contributions

TTA, performance of majority of wet lab experiments including construction of cDNA library, annotation and analyses of specific sequences, writing of manuscript. MT, sequencing of select cDNAs, annotation and analyses of specific sequences. KG, performance of bioinformatics analyses including OGD pipeline. MEW, performance of bioinformatics analyses including OGD pipeline, writing of manuscript. PvW, annotation and analyses of specific sequences, writing of manuscript. SK, supervision of experimental work, annotation and analyses of specific sequences, writing of manuscript.

References

1. **Oomycete Genomics Database (OGD)**   [http://www.oomycete.org]
2. Neish GA, Hughes GC: **Fungal diseases of fishes, Book 6.** Neptune, New Jersey, USA, T.W.F. Publications; 1980.
3. Willoughby LG, Pickering AD: **Viable Saprolegniaceae spores on the epidermis of the salmonid fish Salmo trutta and Salmo alpinus.** *Trans Br Mycol Soc* 1977, **68**:91-95.
4. Bly JE, Lawson LA, Abdel-Aziz ES, Clem LW: **Channel catfish, Ictalurus punctatus, immunity to Saprolegnia sp.** *J Appl Aquacult* 1994, **3**:35-50.
5. Hatai K, Hoshiai GI: **Pathogenicity of Saprolegnia parasitica Coker.** In *Salmon Saprolegniasis* Edited by: Mueller GJ. Portland, Oregon, U.S. Department of Energy, Bonneville Power Administration; 1994:87-98.
6. Bruno DW, Wood BP: **Saprolegnia and other oomycetes.** In *Fish Diseases and Disorders Volume 3.* Edited by: Woo PTK and Bruno DW. Wallingford, UK, CAB International; 1999:599-659.
7. Hussein MM, Hatai K, Nomura T: **Saprolegniosis in salmonids and their eggs in Japan.** *J Wildl Dis* 2001, **37**:204-7.
8. Hatai K, Hoshiai G: **Mass mortality in cultured coho salmon (Oncorhynchus kisutch) due to Saprolegnia parasitica coker.** *J Wildl Dis* 1992, **28**:532-6.
9. Neish GA: **Observations on saprolegniasis of adult sockeye salmon, Oncorhynchus nerka (Walbaum).** *J Fish Biol* 1977, **10**:513-522.
10. Whisler HC: **Identification of Saprolegnia spp. Pathogenic in Chinook Salmon.** Washington, D.C., US Department of Energy; 1996:43 p..
11. Willoughby LG: **Fungi and Fish Diseases.** Stirling, Scotland, Pisces Press; 1994:57 p..
12. Pickering AD, Willoughby LG: **Saprolegnia infections of salmonid fish.** In *Microbial Diseases of Fish* Edited by: Roberts RJ. London, England, Academic Press; 1982:271-297.
13. Bruno DW, Poppe TT: **A Color Atlas of Salmonid Diseases.** London, England, Academic Press; 1996:189 p..
14. Kamoun S: **Molecular genetics of pathogenic oomycetes.** *Eukaryot Cell* 2003, **2**:191-9.
15. Margulis L, Schwartz KV: **Five Kingdoms: An Illustrated Guide to the Phyla of Life on Earth.** New York, W.H. Freeman and Co.; 2000.
16. Sogin ML, Siiberman JD: **Evolution of the protists and protistan parasites from the perspective of molecular systematics.** *Int J Parasitol* 1998, **28**:11-20.
17. Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF: **A kingdom-level phylogeny of eukaryotes based on combined protein data.** *Science* 2000, **290**:972-7.
18. Baldauf SL: **The deep roots of eukaryotes.** *Science* 2003, **300**:1703-6.
19. Beakes G: **A comparative account of cyst coat ontogeny in saprophytic and fish-lesion (pathogenic) isolates of the Saprolegnia declina-parasitica complex.** *Can J Bot* 1982, **61**:603-625.
20. Rose JK, Ham KS, Darvill AG, Albersheim P: **Molecular cloning and characterization of glucanase inhibitor proteins: coevolution of a counterdefense mechanism by plant pathogens.** *Plant Cell* 2002, **14**:1329-45.
21. Tian M, Huitema E, da Cunha L, Torto-Alalibo T, Kamoun S: **A Kazal-like extracellular serine protease inhibitor from Phytophthora infestans targets the tomato pathogenesis-related protease P69B.** *J Biol Chem* 2004, **279**:26370-7.
22. Sanchez LM, Ohno Y, Miura Y, Kawakita K, Doke N: **Host selective suppression by water-soluble glucans from Phytophthora spp. of hypersensitive cell death of suspension-cultured cells**

from some solanaceous plants caused by hyphal wall elicitors of the fungi. *Ann Phytopathol Soc Japan* 1992, **58:**664-670.

23. Alvarez F, Villena A, Zapata A, Razquin B: **Histopathology of the thymus in Saprolegnia-infected wild brown trout, Salmo trutta L.** *Vet Immunol Immunopathol* 1995, **47:**163-72.

24. Alvarez F, Razquin B, Villena A, Lopez Fierro P, Zapata A: **Alterations in the peripheral lymphoid organs and differential leukocyte counts in Saprolegnia-infected brown trout, Salmo trutta fario.** *Vet Immunol Immunopathol* 1988, **18:**181-193.

25. Hatai K, Willoughby LG, Beakes GW: **Some characteristics of Saprolegnia obtained from fish hatcheries in Japan.** *Mycol Res* 1990, **94:**182-190.

26. Kamoun S, Hraber P, Sobral B, Nuss D, Govers F: **Initial assessment of gene diversity for the oomycete pathogen Phytophthora infestans based on expressed sequences.** *Fungal Genet Biol* 1999, **28:**94-106.

27. Randall TA, Dwyer RA, Huitema E, Beyer K, Cvitanich C, Kelkar H, Fong AM, Gates K, Roberts S, Yatzkan E, Gaffney T, Law M, Testa A, Torto-Alalibo T, Zhang M, Zheng L, Mueller E, Windass J, Binder A, Birch PR, Gisi U, Govers F, Gow NA, Mauch F, van West P, Waugh ME, Yu J, Boller T, Kamoun S, Lam ST, Judelson HS: **Large-scale gene discovery in the oomycete Phytophthora infestans reveals likely components of phytopathogenicity shared with true fungi.** *Mol Plant Microbe Interact* 2005, **18:**229-243.

28. **X Genome Initiative (XGI)** [http://www.ncgr.org/xgi]

29. Waugh M, Hraber P, Weller J, Wu Y, Chen G, Inman J, Kiphart D, Sobral B: **The Phytophthora genome initiative database: informatics and analysis for distributed pathogenomic research.** *Nucleic Acids Res* 2000, **28:**87-90.

30. Qutob D, Hraber PT, Sobral BW, Gijzen M: **Comparative analysis of expressed sequences in Phytophthora sojae.** *Plant Physiol* 2000, **123:**243-54.

31. Cooke DE, Drenth A, Duncan JM, Wagels G, Brasier CM: **A molecular phylogeny of Phytophthora and related oomycetes.** *Fungal Genet Biol* 2000, **30:**17-32.

32. Kroon LP, Bakker FT, van den Bosch GB, Bonants PJ, Flier WG: **Phylogenetic analysis of Phytophthora species based on mitochondrial and nuclear DNA sequences.** *Fungal Genet Biol* 2004, **41:**766-82.

33. Nielsen H, Engelbrecht J, Brunak S, von Heijne G: **A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites.** *Int J Neural Syst* 1997, **8:**581-99.

34. Levy I, Shani Z, Shoseyov O: **Modification of polysaccharides and plant cell wall by endo-1,4-beta-glucanase and cellulose-binding domains.** *Biomol Eng* 2002, **19:**17-30.

35. Delaroque N, Muller DG, Bothe G, Pohl T, Knippers R, Boland W: **The complete DNA sequence of the Ectocarpus siliculosus Virus EsV-1 genome.** *Virology* 2001, **287:**112-32.

36. Mattinen ML, Kontteli M, Kerovuo J, Linder M, Annila A, Lindeberg G, Reinikainen T, Drakenberg T: **Three-dimensional structures of three engineered cellulose-binding domains of cellobiohydrolase I from Trichoderma reesei.** *Protein Sci* 1997, **6:**294-303.

37. Srisodsuk M, Lehtio J, Linder M, Margolles-Clark E, Reinikainen T, Teeri TT: **Trichoderma reesei cellobiohydrolase I with an endoglucanase cellulose-binding domain: action on bacterial microcrystalline cellulose.** *J Biotechnol* 1997, **57:**49-57.

38. Mattinen ML, Linder M, Drakenberg T, Annila A: **Solution structure of the cellulose-binding domain of endoglucanase I from Trichoderma reesei and its interaction with cello-oligosaccharides.** *Eur J Biochem* 1998, **256:**279-286.

39. Gaulin E, Jauneau A, Villalba F, Rickauer M, Esquerré-Tugayé MT, Bottin A: **The CBEL glycoprotein of Phytophthora parasitica var. nicotianae is involved in cell wall deposition and adhesion to cellulosic substrates.** *J Cell Sci* 2002, **115:**4565-4575.

40. Villalba Mateos F, Rickauer M, Esquerre-Tugaye MT: **Cloning and characterization of a cDNA encoding an elicitor of Phytophthora parasitica var. nicotianae that shows cellulose-binding and lectin-like activities.** *Mol Plant-Microbe Interact* 1997, **10:**1045-1053.

41. Tordai H, Banyai L, Patthy L: **The PAN module: the N-terminal domains of plasminogen and hepatocyte growth factor are homologous with the apple domains of the prekallikrein family and with a novel domain found in numerous nematode proteins.** *FEBS Lett* 1999, **461:**63-67.

42. McLeod A, Smart CD, Fry WE: **Characterization of 1,3-b glucanase and 1,3;1,4-b glucanase genes from Phytophthora infestans.** *Fungal Genet Biol* 2002, **38:**250-263.

43. Gotesson A, Marshall JS, Jones DA, Hardham AR: **Characterization and evolutionary analysis of a large polygalacturonase gene family in the oomycete plant pathogen Phytophthora cinnamomi.** *Mol Plant-Microbe Interact* 2002, **15:**907-921.

44. Torto TA, Rauser L, Kamoun S: **The pipg1 gene of the oomycete Phytophthora infestans encodes a fungal-like endopolygalacturonase.** *Curr Genet* 2002, **40:**385-390.

45. Rawlings ND, Tolle DP, Barrett AJ: **MEROPS: the peptidase database.** *Nucleic Acids Res* 2004, **32:**D160-D164.

46. Lu W, Apostol I, Qasim MA, Warne N, Wynn R, Zhang WL, Anderson S, Chiang YW, Ogin E, Rothberg I, Ryan K, Laskowski MJ: **Binding of amino acid side-chains to S1 cavities of serine proteinases.** *J Mol Biol* 1997, **266:**441-461.

47. Lu SM, Lu W, Qasim MA, Anderson S, Apostol I, Ardelt W, Bigler T, Chiang YW, Cook J, James MN, Kato I, Kelly C, Kohr W, Komiyama T, Lin TY, Ogawa M, Otlewski J, Park SJ, Qasim S, Ranjbar M, Tashiro M, Warne N, Whatley H, Wieczorek A, Wieczorek M, Wilusz T, Wynn R, Zhang W, Laskowski MJ: **Predicting the reactivity of proteins from their sequence alone: Kazal family of protein inhibitors of serine proteinases.** *Proc Natl Acad Sci U S A* 2001, **98:**1410-1415.

48. Margis R, Reis EM, Villeret V: **Structural and Phylogenetic Relationships among Plant and Animal Cystatins.** *Arch Biochem Biophys* 1998, **359:**24–30.

49. Leonian LH, Lilly VG: **Studies on the nutrition of fungi. I. Thiamin, its constituents, and source of nitrogen.** *Phytopathol* 1938, **28:**531-548.

50. Erwin DC, Ribeiro OK: **Phytophthora Diseases Worldwide.** St. Paul, Minnesota, APS Press; 1996.

51. Panabieres F, Amselem J, Galiana E, Le Berre JY: **Gene identification in the oomycete pathogen Phytophthora parasitica during in vitro vegetative growth through expressed sequence tags.** *Fungal Genet Biol* 2005, **42:**611-623.

52. Dick MW: **Phylum Oomycota.** In *Handbook of Protoctista* Edited by: Margulis L, Corliss JO, Melkonian M and Chapman DJ. Boston, Jones & Bartlett; 1990:661-685.

53. Leclerc MC, Guillot J, Deville M: **Taxonomic and phylogenetic analysis of Saprolegniaceae (Oomycetes) inferred from LSU rDNA and ITS sequence and ITS sequence comparisons.** *Antonie van Leeuwenhoek* 2000, **77:**369-377.

54. Riethmuller A, Weiss M, Oberwinkler F: **Phylogenetic studies of Saprolegniomycetidae and related groups based on nuclear large subunit ribosomal DNA sequences.** *Can J Bot* 1999, **77:**1790-1800.

55. Riethmuller A, Voglmayr H, Goker M, Weiss M, Oberwinkler F: **Phylogenetic relationships of the downy mildews (Peronosporales) and related groups based on nuclear large subunit ribosomal DNA sequences.** *Mycologia* 2002, **94:**834-849.

56. Kamoun S, Young M, Glascock C, Tyler BM: **Extracellular protein elicitors from Phytophthora: Host-specificity and induction of resistance to fungal and bacterial phytopathogens.** *Mol Plant Microbe Interact* 1993, **6:**15-25.

57. Kamoun S, Klucher KM, Coffey MD, Tyler BM: **A gene encoding a host-specific elicitor protein of Phytophthora parasitica.** *Mol Plant Microbe Interact* 1993, **6:**573-581.

58. Kamoun S, Young M, Forster H, Coffey MD, Tyler BM: **Potential role of elicitins in the interaction between Phytophthora: species and tobacco.** *App Env Microbiol* 1994, **60:**1593-1598.

59. Kamoun S, van West P, de Jong AJ, de Groot K, Vleeshouwers V, Govers F: **A gene encoding a protein elicitor of Phytophthora infestans is down-regulated during infection of potato.** *Mol Plant-Microbe Interact* 1997, **10:**13-20.

60. Kamoun S, Lindqvist H, Govers F: **A novel class of elicitin-like genes from Phytophthora infestans.** *Mol Plant-Microbe Interact* 1997, **10:**1028-1030.

61. Ponchet M, Panabieres F, Milat ML, Mikes V, Montillet JL, Suty L, Triantaphylides C, Tirilly Y, Blein JP: **Are elicitins cryptograms in plant-Oomycete communications?** *Cell Mol Life Sci* 1999, **56:**1020-1047.

62. Qutob D, Huitema E, Gijzen M, Kamoun S: **Variation in structure and activity among elicitins from Phytophthora sojae.** *Mol Plant Pathol* 2003, **4:**119-124.

63. Mikes V, Milat ML, Ponchet M, Ricci P, Blein JP: **The fungal elicitor cryptogein is a sterol carrier protein.** *FEBS Lett* 1997, **416:**190-192.

64. Mikes V, Milat ML, Ponchet M, Panabieres F, Ricci P, Blein JP: **Elicitins, proteinaceous elicitors of plant defense, are a new class of sterol carrier proteins.** *Biochem Biophys Res Commun* 1998, **245:**133-139.

65. Paquin B, Laforest MJ, Forget L, Roewer I, Wang Z, Longcore J, Lang BF: **The fungal mitochondrial genome project: evolution of fungal mitochondrial genomes and their gene expression.** *Curr Genet* 1997, **31:**380-395.

66. Lang BF, Seif E, Gray MW, O'Kelly CJ, Burger G: **A comparative genomics approach to the evolution of eukaryotes and their mitochondria.** *J Eukaryot Microbiol* 1999, **46:**320-326.

67. Latijnhouwers M, de Wit PJGM, Govers F: **Oomycetes and fungi: similar weaponry to attack plants.** *Trends Microbiol* 2003, **11:**462-469.

68. Van Etten JL, Graves MV, Muller DG, Boland W, Delaroque N: **Phycodnaviridae-large DNA algal viruses.** *Arch Virol* 2002, **147:**1479-1516.

69. Brecht S, Carruthers VB, Ferguson DJ, Giddings OK, Wang G, Jakle U, Harper JM, Sibley LD, Soldati D: **The toxoplasma micronemal protein MIC4 is an adhesin composed of six conserved apple domains.** *J Biol Chem* 2001, **276:**4119-4127.

70. Brown PJ, Billington KJ, Bumstead JM, Clark JD, Tomley FM: **A microneme protein from Eimeria tenella with homology to the Apple domains of coagulation factor XI and plasma prekallikrein.** *Mol Biochem Parasitol* 2000, **107:**91-102.

71. Brown PJ, Gill AC, Nugent PG, McVey JH, Tomley FM: **Domains of invasion organelle proteins from apicomplexan parasites are homologous with the Apple domains of blood coagulation factor XI and plasma pre-kallikrein and are members of the PAN module superfamily.** *FEBS Lett* 2001, **497:**31-38.

72. Deng M, Templeton TJ, London NR, Bauer C, Schroeder AA, Abrahamsen MS: **Cryptosporidium parvum genes containing thrombospondin type 1 domains.** *Infect Immun* 2002, **70:**6987-6995.

73. Peduzzi R, Bizzozero S: **Immunochemical investigation of four Saprolegnia species with parasitic activity in fish: serological and kinetic charcterization of a chymotrypsin-like activity.** *Microbial Ecol* 1977, **3:**107-119.

74. Bangyeekhun E, Cerenius L, Söderhäll K: **Molecular cloning and characterization of two serine proteinase genes from the crayfish plague fungus, Aphanomyces astaci.** *J Invert Pathol* 2001, **77:**206-216.

75. Morris MT, Coppin A, Tomavo S, Carruthers VB: **Functional analysis of Toxoplasma gondii protease inhibitor 1.** *J Biol Chem* 2002, **277:**45259-45266.

76. Morris MT, Carruthers VB: **Identification and partial characterization of a second Kazal inhibitor in Toxoplasma gondii.** *Mol Biochem Parasitol* 2003, **128:**119-122.

77. Pszenny V, Angel SO, Duschak VG, Paulino M, Ledesma B, Yabo MI, Guarnera E, Ruiz AM, Bontempi EJ: **Molecular cloning, sequencing and expression of a serine proteinase inhibitor gene from Toxoplasma gondii.** *Mol Biochem Parasitol* 2000, **107:**241-249.

78. Milstone AM, Harrison LM, Bungiro RD, Kuzmic P, Cappello M: **A broad spectrum Kunitz type serine protease inhibitor secreted by the hookworm Ancylostoma ceylanicum.** *J Biol Chem* 2000, **275:**29391-29399.

79. Rawlings ND, Tolle DP, Barrett AJ: **Evolutionary families of peptidase inhibitors.** *Biochem J* 2004, **378:**705-716.

80. Santos CC, Sant'anna C, Terres A, Cunha ESNL, Scharfstein J, de ALAP: **Chagasin, the endogenous cysteine-protease inhibitor of Trypanosoma cruzi, modulates parasite differentiation and invasion of mammalian cells.** *J Cell Sci* 2005, **118:**901-915.

81. Besteiro S, Coombs GH, Mottram JC: **A potential role for ICP, a Leishmanial inhibitor of cysteine peptidases, in the interaction between host and parasite.** *Mol Microbiol* 2004, **54:**1224-1236.

82. Rigden DJ, Mosolov VV, Galperin MY: **Sequence conservation in the chagasin family suggests a common trend in cysteine proteinase binding by unrelated protein inhibitors.** *Protein Sci* 2002, **11:**1971-1977.

83. Mort-Bontemps M, Fevre M: **Transformation of the oomycete Saprolegnia monoica to hygromycin-B resistance.** *Curr Genet* 1997, **31:**272-275.

84. Whisson SC, Avrova AO, van West P, Jones JT: **A method for double-stranded RNA mediated transient gene silencing in Phytophthora infestans.** *Mol Plant Pathol* 2005, **6:**153-163.

85. Sambrook J, Fritsch EF, Maniatis T: **Molecular Cloning: A Laboratory Manual.** 2nd edition. Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory; 1989.

86. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nuc Acids Res* 1997, **17:**3389-3402.

87. Henikoff JG, Greene EA, Pietrokovski S, Henikoff S: **Increased coverage of protein families with the blocks database servers.** *Nucleic Acids Res* 2000, **28:**228-230.

88. Henikoff JG, Pietrokovski S, McCallum CM, Henikoff S: **Blocks-based methods for detecting protein homology.** *Electrophoresis* 2000, **21:**1700-1706.

89. Zdobnov EM, Apweiler R: **InterProScan--an integration platform for the signature-recognition methods in InterPro.** *Bioinformatics* 2001, **17:**847-848.

90. Apweiler R, Attwood TK, Bairoch A, Bateman A, Birney E, Biswas M, Bucher P, Cerutti L, Corpet F, Croning MD, Durbin R, Falquet L, Fleischmann W, Gouzy J, Hermjakob H, Hulo N, Jonassen I, Kahn D, Kanapin A, Karavidopoulou Y, Lopez R, Marx B, Mulder NJ, Oinn TM, Pagni M, Servant F: **The InterPro database, an integrated documentation resource for protein families, domains and functional sites.** *Nucleic Acids Res* 2001, **29:**37-40.

91. Torto T, Li S, Styer A, Huitema E, Testa A, Gow NA, van West P, Kamoun S: **EST mining and functional expression assays identify extracellular effector proteins from plant pathogen Phytophthora.** *Genome Res* 2003, **13:**1675-1685.

92. Nielsen H, Brunak S, von Heijne G: **Machine learning approaches for the prediction of signal peptides and other protein sorting signals.** *Protein Eng* 1999, **12:**3-9.

93. **The Gene Ontology** [http://geneontology.org]

94. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25:**25-29.

95. **TraceDB databases** [http://www.ncbi.nlm.nih.gov]

96. **Broad Institute** [http://www.broad.mit.edu/seq]

97. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic Acids Res* 1997, **25:**4876-4882.

98. **BOXSHADE** [http://bioweb.pasteur.fr/seqanal/interfaces/boxshade.html]

99. **WebLogo** [http://weblogo.berkeley.edu]

100. Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Bechner M, Brzezinski MA, Chaal BK, Chiovitti A, Davis AK, Demarest MS, Detter JC, Glavina T, Goodstein D, Hadi MZ, Hellsten U, Hildebrand M, Jenkins BD, Jurka J, Kapitonov VV, Kroger N, Lau WW, Lane TW, Larimer FW, Lippmeier JC, Lucas S, Medina M, Montsant A, Obornik M, Parker MS, Palenik B, Pazour GJ, Richardson PM, Rynearson TA, Saito MA, Schwartz DC, Thamatrakoln K, Valentin K, Vardi A, Wilkerson FP, Rokhsar DS: **The genome of the diatom Thalassiosira pseudonana: ecology, evolution, and metabolism.** *Science* 2004, **306:**79-86.