FUNCTIONAL GENOMICS OF EXTRACELLULAR PROTEINS OF PHYTOPHTHORA INFESTANS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of the Ohio State University

By

Gertrude Ayerchoo Torto, B.Sc.(Hons), MS

The Ohio State University

2003

Dissertation Committee:

Dr. Sophien Kamoun, Adviser
Dr. Anne Dorrance
Dr. David Francis
Dr. Terry Graham

Approved by

_______________________
Adviser
Department of Plant Pathology
ABSTRACT

Genomics offers new perspectives for the discovery of novel genes in the oomycetes, a unique group of eukaryotic microbes comprising saprophytes as well as pathogens of plants and animals. Among the oomycetes, the most devastating plant pathogen is the potato and tomato late blight pathogen Phytophthora infestans. Currently tens of thousands of expressed sequence tags (ESTs) are available for P. infestans. The challenge of the post-genomic era is to link a sequence to a phenotype (functional genomics) with as little experimental effort as possible. In the first study, we combined EST data mining with functional assays to identify extracellular effector proteins from P. infestans, which can manipulate plant molecular and cellular processes to induce defense or disease like responses. We developed the PexFinder algorithm for automated identification of Phytophthora extracellular proteins (Pex). Using a virus based high throughput functional assay to express Pex cDNAs in Nicotiana benthamiana and tomato, we identified two novel necrosis inducing effectors crn1 and crn2, which are members of
large gene family. *crn1* and *crn2* were demonstrated to be expressed during *P. infestans* infection on tomato. In addition, *crn2* induced defense related genes in tomato.

In a second study, we characterized a *Pex* cDNA, *pipg1* which represents the first endopolygalacturonase (endoPG) described from *P. infestans*. Generally, endopolygalacturonases are plant cell wall degrading enzymes primarily found and characterized extensively in plants and fungi. We showed that *pipg1* is a member of a gene family and is expressed in both preinfection and infection stages. Phylogenetic analyses showed an affinity of *pipg1* to fungal endoPGs, a feature that contrasts with phylogenies obtained with ribosomal sequences or compiled mitochondrial and chromosomal genes.

In a third study, we generated ESTs from a cDNA library constructed from mRNA of *Saprolegnia parasitica*, an oomycete pathogen that causes saprolegniasis in fish. We performed preliminary comparisons of abundance of selected ESTs between *S. parasitica* and *P. infestans* and assessed the implications for *S. parasitica*. The long-term objective of this project is to develop genomic resources from diverse oomycete pathogens for comparative genomics.
To

Mom and Dad, Ruth and Amaziah Torto

My husband Phil Tam-Al Alalibo
ACKNOWLEDGEMENTS

My utmost thanks go to my major advisor who initiated me into the field of genomics and provided the financial support for my studies. His guidance, encouragement and never-ending persuasion to attain the best have been instrumental to the successful completion of this work. I am very grateful to my advisory committee members for their encouragement and valuable suggestions, especially to Dr. David Francis for helping me with statistics.

My gratitude goes most especially to the research assistants in the lab, Shujing Dong and Diane Kinney for their valuable technical assistance. To my colleagues past and present, Dr. Bill Morgan, Dr. Walid Hamada, Dr. Antonino Testa, Miaoying Tian, Edgar Huitema and Zhenyu Liu, thanks for the useful discussions.

To the many visiting students and interns, especially Allison Styer and Laura Rauser for your expert technical assistance. To Mildred Ochwo, Jorunn Bos and Luis da Cunha, it was nice working with you during your internship.

To my husband Phil thank you very much for your long
lasting support, for listening to my woes and accomplishments. To Mom and Dad, I did this for you. To the rest of my family, both Tortos and Alalibos, thank you for your endless prayers, it certainly worked in carrying me through all these years.

I also wish to express my gratitude to the staff and faculty of the Department of Plant Pathology for moral support and friendship. To my fellow graduate students especially Massimo Merighi, Richard Edema and Miguel Vega-Sanchez, thanks for the friendship and useful discussions.

To the library staff, thank you for putting up with me in the three months that I worked on my dissertation.

Last but not the least my gratitude goes to my friends especially the Edema family who made sure I at least had a good dinner, To Danny Dorman, for your endless words of encouragement. To Lucille Houser, Kristin and Jhonny Mera, Sonia Pereira, Aparna Gazula and all the friends I had here on campus, thanks for providing an outlet for relaxation.

To God be all the glory.
VITA

Name.............Gertrude (Trudy) Ayerchoo Torto.

PREVIOUS EDUCATION AND WORK EXPERIENCE


1998-present.....Graduate Student. Department of Plant Pathology/Plant Molecular Biology and Biotechnology Interdisciplinary Program. The Ohio State University.
AWARDS AND HONORS

2002....C.C. Allison Award for high achievement in research and dedicated service to the Plant Pathology Dept.
The Ohio State Univ.

2002....International Society for Plant Molecular Biology Student Travel Award.

2001....Edward Hayes Award for Outstanding Research - The Ohio State Univ.

2001....SIGMA XI Certificate of Distinction for Outstanding Research.

2001....Patricia Ngwira Travel Grant Award.

1998....Research Assistantship: Plant Pathology Dept. The Ohio State Univ.

PUBLICATIONS


Trudy A. Torto, Laura Rauser and Sophien Kamoun. 2002. The pipg1 gene of the oomycete Phytophthora infestans encodes a


**FIELD OF STUDY**

Major Field: Plant Pathology (Plant Molecular Biology and Biotechnology Interdisciplinary Program).
# TABLE OF CONTENTS

Abstract........................................................................................................................................................................ ii  
Dedication..................................................................................................................................................................... iv 
Acknowledgements......................................................................................................................................................... v  
Vita.............................................................................................................................................................................. vii  
List of tables............................................................................................................................................................... xii  
List of figures............................................................................................................................................................... xiii  

Chapters
1. Literature Review....................................................................................................................................................... 1  
EST Mining and Functional Expression Assays Identify Extracellular Effector Proteins from the Plant Pathogen Phytophthora infestans
Abstract........................................................................................................................................................................ 27  
Introduction................................................................................................................................................................. 29  
Materials and Methods.................................................................................................................................................. 33  
Results .......................................................................................................................................................................... 41  
Discussions.................................................................................................................................................................... 51  
Acknowledgement.......................................................................................................................................................... 61  
References ...................................................................................................................................................................... 62  
The Pipg1 Gene of the Oomycete Phytophthora infestans Encodes A Fungal-Like Endopolygalacturonase
Abstract........................................................................................................................................................................ 81  
Introduction................................................................................................................................................................. 82  
Materials and Methods.................................................................................................................................................. 85  
Results .......................................................................................................................................................................... 88  
Discussions.................................................................................................................................................................... 92  
Acknowledgement.......................................................................................................................................................... 97  
References...................................................................................................................................................................... 98
Analysis of Expressed Sequence Tags from the Oomycete Pathogen, *Saprolegnia parasitica*

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>109</td>
</tr>
<tr>
<td>Introduction</td>
<td>110</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>113</td>
</tr>
<tr>
<td>Results</td>
<td>116</td>
</tr>
<tr>
<td>Discussions</td>
<td>119</td>
</tr>
<tr>
<td>Acknowledgement</td>
<td>125</td>
</tr>
<tr>
<td>References</td>
<td>125</td>
</tr>
<tr>
<td>General Conclusion</td>
<td>136</td>
</tr>
<tr>
<td>Bibliography</td>
<td>141</td>
</tr>
</tbody>
</table>
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>List of endopolygalacturonases from different taxonomic groups used in the phylogenetic analysis</td>
</tr>
<tr>
<td>4.1</td>
<td>List of most abundant sequences in <em>Saprolegnia parasitica</em> EST dataset</td>
</tr>
<tr>
<td>4.2</td>
<td>List of endo-1,3-β-glucanases from different taxonomic groups used in phylogenetic analysis</td>
</tr>
<tr>
<td>4.3</td>
<td>Notable transcripts in <em>Saprolegnia parasitica</em> EST dataset present or absent in <em>Phytophthora infestans</em> EST datasets</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Schematic illustration of the PexFinder algorithm for identification of extracellular proteins from expressed sequence tags………………..</td>
</tr>
<tr>
<td>2.2</td>
<td>Scatter diagrams of output scores obtained with SignalP v2.0………………………………………………………………………………………………</td>
</tr>
<tr>
<td>2.3</td>
<td>Two dimensional gel electrophoresis analysis of secreted proteins of Phytophthora infestans………………</td>
</tr>
<tr>
<td>2.4</td>
<td>Strategy for functional screening of Phytophthora infestans cDNAs using agroinfection with a binary potato virus X (PVX) vector………………….</td>
</tr>
<tr>
<td>2.5</td>
<td>Symptoms observed on Nicotiana benthamiana plants after inoculation with Agrobacterium tumefaciens containing the binary potato virus X (PVX) vector expressing Phytophthora infestans crn1 and crn2…………………………………………………………………………………………………………</td>
</tr>
<tr>
<td>2.6</td>
<td>Alignment of Phytophthora infestans CRN1 and CRN2 predicted amino acid sequences…………………………………………………………………………………………………………</td>
</tr>
<tr>
<td>2.7</td>
<td>Time course of expression of crn1, crn2 and actA of Phytophthora infestans during colonization of tomato…………………………………………………………………………………………………………</td>
</tr>
<tr>
<td>2.8</td>
<td>Symptoms observed on Nicotiana benthamiana plants after inoculation with Agrobacterium tumefaciens containing binary potato virus X (PVX) derivatives expressing crn1, crn2 and inf1 from Phytophthora infestans or pSOJNIP from Phytophthora sojae…………………………………………………………………………………………………………</td>
</tr>
<tr>
<td>2.9</td>
<td>Induction of tomato defense genes by crn genes…</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>3.1</td>
<td>Multiple alignment of the deduced endopolygalacturonase amino acid sequences from fungi, oomycete and plants</td>
</tr>
<tr>
<td>3.2</td>
<td>Phylogenetic analysis of 39 representative endopolygalacturonases from fungi, plants, insects, bacteria and the oomycete <em>Phytophthora infestans</em></td>
</tr>
<tr>
<td>3.3</td>
<td>Presence of <em>pipg1</em>-ike sequences in <em>Phytophthora</em> genome</td>
</tr>
<tr>
<td>3.4</td>
<td>Expression of <em>pipg1</em> in various developmental and infection stages of <em>Phytophthora infestans</em> and tomato</td>
</tr>
<tr>
<td>4.1</td>
<td>Graph showing percentage frequency of redundant sequences</td>
</tr>
<tr>
<td>4.2</td>
<td>Pie chart representing output from similarity searches against public databases using the BLAST algorithm</td>
</tr>
<tr>
<td>4.3</td>
<td>Phylogenetic reconstruction of the endoglucanase family of proteins from plants, fungi and oomycetes</td>
</tr>
</tbody>
</table>
CHAPTER 1

LITERATURE REVIEW

MOLECULAR GENETICS AND GENOMICS OF PHYTOPHTHORA INFESTANS

Introduction

*Phytophthora infestans*, (Mont.) de Bary is the causative agent of the late blight disease of tomato and potato and is by far the most devastating disease of potato worldwide (Fry and Goodwin, 1997b). *P. infestans*, which has caused the Irish potato famine in the mid nineteenth century (de Bary, 1876), continues to cause multi-billion dollar losses annually in potato and tomato production (Fry and Goodwin, 1997a; Fry and Goodwin, 1997b). The havoc that *P. infestans* wreaks on potato and tomato is yet to be effectively controlled, and the problem worsened with the recent emergence of highly aggressive and fungicide-insensitive strains (Fry and Goodwin, 1997a; Fry and Goodwin, 1997b). In fact, recent reports warned that potato blight might cause catastrophic losses, and possibly famine, in Eastern Europe, and recent epidemics in that region resulted in as much as 70% losses in yield (Schiermeier, 2001; Garelik, 2002).
*P. infestans* belongs to a unique taxonomic group of organisms called the oomycetes. This group includes various plant and animal pathogens as well as saprophytic species (Margulis and Schwartz, 2000). Historically, based on their fungal-like morphology and physiology, the oomycetes have been referred to as fungi. Increasing biochemical (Bartnicki-Garcia and Wang, 1983; Pfyffer et al., 1990) and molecular (Paquin et al., 1997; Sogin and Silberman, 1998) evidence has shown that oomycetes are not fungi, but are more related to heterokont algae. Their unique phylogenetic position suggests that molecular mechanisms underlying host infection and interaction could be unique. Invariably, fungal pathogens, for which molecular studies are more advanced, cannot serve as models to study oomycetes. Also, in light of the different evolutionary history of the fungi, the unique biochemical features of oomycetes render them insensitive to many of the fungicides available (Griffith et al., 1992; Kirk et al., 1999). Effective management of diseases caused by the oomycetes, will come from a thorough understanding of the mechanisms underlying pathogenicity and plant responses to the pathogen and the development of specific fungicides.

In this review, I discuss the life cycle of *P. infestans*, pathogenicity, elicitors and host/nonhost resistance, and finally I discuss recent genomic resources and functional genomic systems available for *P. infestans*. 2
**Phytophthora infestans infection cycle**

The *P. infestans* infection cycle is well known (Pristou and Gallegly, 1954; Coffey and Wilson, 1983; Agrios, 1988; Erwin and Ribeiro, 1996). Infection is initiated when sporangia come into contact with a moist leaf surface. The sporangia will either germinate directly at temperatures above 15°C or release biflagellate zoospores at temperatures below 15°C. The motile biflagellated zoospores then germinate after encystment on the surface of the plant. Following appressorium formation, infection tubes emerge and penetrate epidermal cells. In susceptible plants (compatible interactions), hyphae spread into the mesophyll layer, occasionally forming haustorium-like feeding structures. After colonization, sporangiophores are formed at the tip of emerging hyphae from the stomata. These become inocula for subsequent aerial spread of the pathogen. Infected foliage becomes yellow, water soaked and ultimately turns black. In resistant plants (incompatible interactions), a form of programmed cell death known as the hypersensitive response (HR) is induced. Cytological studies demonstrated that the hypersensitive response is associated with all forms of resistance to *P. infestans*, albeit at different rates of induction (Vleeshouwers et al., 2000). In race specific resistant hosts and in nonhost
plants, induction of the HR is limited to one or a few cells and results in the arrest of pathogen growth in the early stages of infection (Kamoun et al., 1999c; Vleeshouwers et al., 2000). Other types of resistance, such as partial or rate-limiting resistance, also involve the HR, which can occur as a trailing type of necrosis (Kamoun et al., 1999c; Vleeshouwers et al., 2000).

**Up-regulated genes in preinfection and infection stages of P. infestans**

Microbial pathogens and plants communicate through the exchange of signal molecules (Baker et al., 1997). To exploit the host for nutrients, plant microbial pathogens have to overcome the barriers of the plant surface formed by the cuticle and the epidermis and colonize plant tissue. The gene products that facilitate the infection process and mediate pathogenicity are not well known in P. infestans. A few genes that are induced during infection (*ipi* or *in planta* induced genes) have been identified and were hypothesized to be more likely involved in pathogenesis than constitutive genes. The first of such *ipi* genes was described by Pieterse et al., (1993). In this study, the authors used a differential hybridization procedure, which involved the screening of a genomic library with labeled cDNA from a susceptible potato cultivar infected with *P.*
Amongst the nine in planta induced genes identified, the *ipiO* gene is the best characterized. *IpiO* belongs to a small clustered gene family (Pieterse et al., 1994) and was shown to be expressed early in the infection process, precisely during the biotrophic stages of the disease cycle (van West et al., 1998). Whether or not *ipiO* is involved or essential for pathogenicity remains unknown.

In the *Cladosporium fulvum*-tomato pathosystem, in planta induced genes have been shown to function in virulence, for example, mutants in the in planta induced genes *ecp1* and *ecp2* have been shown to have reduced fitness on tomato (Lauge et al., 1997).

Three other genes, named *car* (cyst-germination-specific acidic repeat) were cloned from in vitro grown *P. infestans* germinating cysts and appresoria, with the assumption that products from genes activated in these stages which directly precedes the start of the infection process may be involved in host cell invasion (Görnhardt et al., 2000). These genes are also clustered, a feature reminiscent of pathogenicity islands in phytopathogenic bacteria. The *car* gene products bear 51% sequence homology with human mucins. Based on this similarity, the Car proteins are presumed to form a mucous cover around the germling to aid in adhesion, as well as protecting it from desiccation, physical damage and defense responses of the plant.
Hydrolytic enzymes in *P. infestans*

Most phytopathogenic fungi are able to penetrate and colonize host tissues by secreting hydrolytic enzymes. Such enzymes include cutinases (Kolattukudy et al., 1995), pectinases (Johnston and Williamson, 1992; Reignault et al., 1994; ten Have, 2001) and glycosidases (Henrissat and Davies, 1997). Mechanical penetration is usually facilitated by the appresorium (Rijkenberg et al., 1980) and in some cases penetration is facilitated by both mechanical means and hydrolytic enzymes. The *P. infestans* endopolygalacturonase, *pipg1* has been cloned and characterized and suggested to belong to a gene family (Torto et al., 2002) (see chapter 3).

Elicitors and host resistance

*P. infestans*, like other plant pathogens, produces effector molecules that modulate plant biochemical and cellular responses to promote infection and delay host responses in compatible plant genotypes (Collmer et al., 2000; Knogge, 1996; Lauge and De Wit, 1998). Some effectors (avirulence) are recognized by incompatible host genotypes, triggering a genetically encoded set of defense responses, leading often to the hypersensitive response (HR), a form of programmed cell death (Collmer et al., 2000; Kjemtrup et al., 2000; Knogge, 1996; Lauge and De Wit, 1998; Staskawicz
et al., 2001). Over the last years, avirulence (Avr) genes have been cloned and characterized in bacteria, nematodes and fungi, however no Avr gene has been cloned from oomycetes yet. In a bid to locate and eventually clone Avr genes from *P. infestans*, a comprehensive genetic map was constructed (van der Lee et al., 1997). This was built with 183 AFLP and seven RFLP markers on the F1 progeny from the cross between two parental field isolates, which are homokaryotic and diploid. The linkage map encompasses 10 major and 7 minor linkage groups representing ca 1200 cM of the genome. In a later study, bulked segregant analysis facilitated the positioning of six dominant Avr genes on the map (Van der Lee et al., 2001). A cluster of three tightly linked genes, *Avr3*, *Avr10* and *Avr11* was located on linkage group VIII. *Avr4*, *Avr1* and *Avr2* were located on linkage groups A2-a, IV and VI respectively. The clustering of genes on one linkage group presents a perfect scenario for positional cloning to facilitate the discovery of multiple Avr genes. A bacterial artificial chromosome (BAC) library representing about 10 genome equivalents has been constructed (Birch and Whisson, 2001; Whisson et al., 2001). This library was constructed from a selected individual within the F1 population used for the genetic map, thus it contains all the six Avr genes described by van der Lee et al., 2001. Using a pooling strategy, the BAC library was screened for AFLP markers. Information from the
pools allowed the construction of a contig of 11 BAC clones in the region of Avr3, Avr10, Avr11, gene cluster. Thus genome maps will aid in efficient map-based cloning of Avr genes as well as other genes of interest in P. infestans.

Recently the first resistance gene in a host of P. infestans, namely the Solanum demissum R1 gene, was cloned. Based on complementation analysis, the R1 gene was shown to elicit the HR upon infection with the P. infestans race 4 carrying the Avr1 gene. This is the first of 11 R genes introgressed from wild species into the cultivated S. tuberosum to be cloned (Ballvora et al., 2002). The isolation and cloning of the Avr1 gene would facilitate a complete understanding of R1-Avr1 interaction.

Race specific resistance is generally not considered to be durable in this pathosystem. Deployment of single R genes in host cultivars selects for virulent races capable of rendering the resistance ineffective (Watsie, 1991; Fry and Goodwin, 1997). In P. infestans, new races have been reported to appear 2-3 years after the release of race specific resistant cultivars (Erwin and Ribeiro, 1996). Furthermore, the spread in P. infestans populations of the A2 mating types have created the possibility of sexual reproduction resulting in sexual recombinants with increased virulence and variability (Knapova and Gisi, 2002). There is thus an urgent need to develop new sources
of resistance that may prove durable.

**Elicitors and nonhost resistance**

Nonhost resistance is defined as resistance to pathogens at the species or genus level (Kamoun et al., 1999c). Cytological studies have shown that *P. infestans* penetrates the epidermal cells of solanaceous plants like *Solanum nigrum*, tobacco as well as non solanaceous plants such as parsley and *Arabidopsis* (Vleeshouwers et al., 2000). Upon penetration of the epidermal cell only the epidermal cell or a few cells around the penetrating hyphae are sacrificed. Given what is known so far about this type of resistance, it is thought to involve an arsenal of *R* genes recognizing multiple or essential avirulence genes (Kamoun, 2001).

A family of extracellular protein elicitors termed elicitins, have been identified in *P. infestans* and other *Phytophthora* species (Huet et al., 1995; Kamoun et al., 1993; Pernollet et al., 1993). In the *P. infestans*- *Nicotiana* pathosystem, a species-specific Avr factor, INF1 was cloned and extensively characterized (Kamoun et al., 1997). Resistance of *N. benthamiana* to *P. infestans* resulting in HR was shown to be mediated by recognition of INF1 (Kamoun et al., 1998). The extent of the HR in terms of timing and severity varies with different *Nicotiana*
species. Also, INF1-deficient mutants produced disease lesions in *N. benthamiana* but were unable to infect *Nicotiana tabacum*. Cytological examinations show that a wild type INF1 producing strain is restricted in growth in *N. benthamiana* whereas INF1-deficient mutants can colonize the leaf tissue. In *Nicotiana tabacum*, however, the ingress of both wild type and INF1 deficient mutants is stopped immediately upon penetration. This indicates that other avirulence factors besides INF1 trigger the nonhost HR response in *Nicotiana tabacum*. Ultimately, resistance mediated by elicitors may be very complex. Unraveling the molecular and biochemical basis underlying elicitin mediated resistance may prove useful in understanding nonhost resistance as a whole.

**Genomic resources for Phytophthora infestans**

*P. infestans* was not left out of the genomic wave that started in the mid to late nineties. A community based initiative, the *Phytophthora* Genome Initiative (http://www.ncgr.org.pgc) rolled the ball with an EST project resulting in the generation of ca 2,500 ESTs. (Kamoun et al., 1999b). This initiative led to more sequencing efforts, and in 2001, a USDA funded project aimed at generating an additional 14,000 *P. infestans* ESTs including both *in vitro* and infected potato libraries.
Another industry-funded project generated more than 70,000 *P. infestans* ESTs from various developmental as well as infection stages (Lam, 2001).

The genome size of *P. infestans* is 250Mb with a haploid chromosome count estimated by microscopy to be n=8-10 (Sansome and Brasier, 1973). Currently, only a 1X draft genome sequence is available (Lam, 2001) for *P. infestans*. However if funding for the complete genome sequence of *P. infestans* is provided, existing genetic and genomic resources such as the comprehensive genetic map (van der Lee et al., 1997), BAC contigs (Whisson et al., 2001) and the large annotated ESTs (Lam, 2001) will facilitate efficient annotation.

**Data mining for candidate genes**

To efficiently annotate sequences generated from large sequencing projects, automated data analysis pipelines have been developed. These pipelines have often integrated existing programs and bioinformatic tools. Public *P. infestans* ESTs are housed in the *Phytophthora* Genome Consortium (PGC) database. Automated analysis involves a systematic approach starting with a gatherer that examines incoming data files to eliminate errors such as improper file headers or contents that are not sequencing data (Waugh et al., 2000, Kamoun et al., 1999b). This is
followed in the pipeline by vector screening to remove vector sequences. Subsequently, an interface to BLAST (Altschul et al., 1990) and motif searching programs identifies putative functions and possible motifs for each EST sequence input into the pipeline (Waugh et al., 2000). As new programs are developed, the curators add on to the pipeline. Similarity searches against public databases to assign putative functions provide useful information in selecting candidate genes for further studies. This approach has limitations, since about 45% of sequences cannot be assigned functions in this manner. Other strategies are thus needed to choose candidate genes.

The PexFinder algorithm was developed to meet such a need (See Chapter 2). PexFinder integrates a series of sequence analysis and data manipulation programs with SignalP v2.0, a program that predicts the presence of signal peptides and their cleavage sites (Nielsen et al., 1997). The algorithm was then used as a tool to identify P. infestans cDNAs encoding extracellular proteins. Candidate cDNAs encoding extracellular proteins are likely to be involved in the interaction with the plant since P. infestans comes in contact first with the surface of the plant and subsequently invades the intercellular space. Moreover, several classes of oomycete and fungal effector molecules are known to require secretion (Jia et al., 2000; Lauge and De Wit, 1998). A dataset of P. infestans encoding
extracellular proteins have been generated and functional studies revealed novel necrosis inducing peptides from this dataset (See chapter 2).

The interaction transcriptome represents a combination of the entire set of transcripts of both plant and microorganism expressed during the interaction (Birch and Kamoun, 2000). ESTs generated from cDNA of infected tissues, a so called ‘interaction library’, may pose a challenge when it comes to distinguishing between the pathogen and the plant sequences. In Phytophthora-plant interaction libraries, specifically for the two economically important ones, *P. sojae* and *P. infestans*, this problem has been resolved (Birch and Kamoun, 2000; Hraber and Weller, 2001; Qutob et al., 2000). In a bid to distinguish *P. sojae* and soybean ESTs in an interaction library, Qutob et al., (2000) compared the GC content of ESTs from individual EST libraries of *P. sojae* and soybean. This revealed that the mean GC content of soybean is 46% and that of *P. sojae* peaked at 58%. Graphical representation of the GC content of ESTs in the *P. sojae* and soybean libraries produced two distinct but slightly overlapping normal distribution curves. Whereupon similar analysis of the *P. sojae* infected library resulted in two peaks, which corresponded, to the uninfected soybean and the axenically grown *P. sojae*. Most importantly this has been proven for *P. infestans*-tomato interaction libraries
as well (Kamoun et al., unpublished data). This concept is useful for identifying *Phytophthora* and plant genes involved in the interaction for further functional analysis.

**Functional genomics**

The exponential accumulation of sequence data has created a challenge for the *Phytophthora* research community to develop genome-wide experimental approaches to determine gene function. Stable transformation is a requirement for any functional analysis. The standard transformation technique for *Phytophthora* is the liposome-PEG mediation transformation of protoplasts, followed by regeneration of the protoplast and antibiotic selection on agar medium (Judelson et al., 1991; Judelson et al., 1992). High frequency rates of co-transformation (up to 50%) were observed especially if the two plasmids are linearized with restriction enzymes with compatible ends (Judelson, 1993). The efficiency of transformation of *P. infestans* with BAC clones of about 102 kb was shown to be higher than with smaller plasmids (Randall and Judelson, 1999). This will be advantageous in allowing the detection of important genes in large regions of the genome. The limiting step in this procedure, however appears to be in the heterologous integration of the introduced plasmids and the low
regeneration rates of the protoplasts. Currently, other transformation procedures are being developed that seem to bypass some of the limitations mentioned above. These include Agrobacterium transformation and electroporation of zoospores (Tyler, 2001).

Gene targeting is the generation of specific mutations in a genome by homologous recombination-mediated integration of foreign DNA sequences. Gene targeting represents the ultimate genetic tool for functional genomics and has been used for functional genetic studies in bacteria, yeast and other filamentous fungi (Balhadere and Talbot, 2001; Rothstein, 1991; Schaefer, 2002). In Phytophthora, which is diploid in almost all its developmental stages, direct disruption of a target gene has not been reported owing to an apparent low frequency of homologous recombination (Judelson et al., 1992). Currently gene silencing is the only proven technique to obtain mutants in Phytophthora (Kamoun et al., 1998; van West et al., 1999). The gene silencing technique has been reported only for one gene inf1 and has to be proven successful for a reasonable number of genes before considering it for systematic application. Generally gene silencing in P. infestans is sequence specific, similar to what has been described in other systems (Matzke and Matzke, 1995). In the inf1 silenced transformants only the inf1 mRNA was affected but not the related inf2b mRNA. However, it is
expected that genes with similarity higher than 80% at the nucleotide sequence can be silenced using a single construct. It should also be noted that the observed spread of the silenced state from silenced transgenic nuclei to non-transformed nuclei when mixed together in a heterokaryotic strain (van West et al., 1999) forms a good prospect for functional genomics.

Given that there is no systematic approach to study the functional role of genes in *Phytophthora*, plant gene expression systems can probably be adapted to systematically screen *Phytophthora* genes for responses in plants. The PVX based expression system has been employed for this purpose in different pathosystems. In a targeted search for HR based resistance in a diverse selection of tomato lines, (Lauge et al., 1998) used the PVX system to systemically express the elicitor, ECP2, in tomato plants. Identification of lines responding with an HR led to the successful isolation of the corresponding resistant factor designated Cf-ECP2. In another work the secreted elicitors Avr9 and infl from *C. fulvum* and *P. infestans* respectively were expressed in plants via the PVX system. As a result, HR responses were observed in Cf9 tomato or Cf-9 transgenic tobacco and all cultivars of tobacco respectively (Kamoun et al., 1999a). In fact, Takken et al., (2000) exploited the PVX system in a high throughput screening of cDNAs from *C. fulvum* grown under nutrient-starved conditions. The
authors reported the identification of HR responses induced by the elicitor AVR4 in resistant Cf-4 tomato plants and three other cDNAs inducing non-specific HR. More recently, this system has also been used in Phytophthora-N. benthamiana pathosystem. Qutob et al., (2002) expressed the PsojNIP cDNA from P. sojae infected soybean tissue in N. benthamiana using the PVX system. This study confirmed the necrosis inducing activity of PsojNIP as determined earlier for orthologs from fungal phytopathogens.

A collection of cDNAs encoding secreted proteins from P. infestans were expressed in N. benthamiana using the PVX system in a semi-high throughput way. This led to the identification of two novel necrosis inducing peptides CRN1 and CRN2 which were reported to induce defense responses in Nicotiana (Torto et al., 2001, see chapter 2). Clearly the PVX expression system holds promise for the high throughput screening of pathogen genes in plants to elucidate plant responses. In the future this system will complement gene disruption techniques to elucidate the functional roles of Phytophthora genes.

Conclusions and future perspectives
Since the recognition of *P. infestans* as a plant pathogen and also as the cause of famine and population displacement in Ireland in the mid 1800s, the biology, genetics and pathogenic strategies have been studied. However, relatively little research has explored *Phytophthora* diseases at the molecular level. The wealth of data comprising sequences involved in different developmental stages as well as infection stages should help in establishing expression profiles to understand mechanisms underlying different phenomena. There is a need however, to develop systematic gene disruption techniques to help elucidate the functional role of these genes. Evidently sequencing the genome of *P. infestans* will prove a valuable resource. Annotation of the genome of *P. infestans* will be facilitated by large unigene set generated from the ESTs as well as available physical and genetic maps. A complete genome sequence can be used for comparative analysis amongst *Phytophthora* species or with other organisms to decipher, for example the basic set of genes that makes *P. infestans* a plant pathogen.

There is much hope that in the near future these resources will help us understand mechanisms underlying pathogenicity and host responses and ultimately lead to the development of improved strategies to control *P. infestans*. 
REFERENCES


Reignault, P., Mercier, M., Bompeix, G., and Boccara, M. (1994). Pectin methylesterase from Botrytis cinerea:


CHAPTER 2

EST Mining and Functional Expression Assays Identify Extracellular Effector Proteins from the Plant Pathogen Phytophthora infestans

ABSTRACT

Plant pathogenic microbes have the remarkable ability to manipulate biochemical, physiological and morphological processes in their host plants. These manipulations are achieved through a diverse array of effector molecules that can either promote infection or trigger defense responses. We describe a general functional genomics approach aimed at identifying extracellular effector proteins from plant pathogenic microorganisms by combining data mining of expressed sequence tags (ESTs) with virus-based high throughput functional expression assays in plants. PexFinder, an algorithm for automated identification of extracellular proteins from EST data sets, was developed and applied to 2147 ESTs from the oomycete plant pathogen.
Phytophthora infestans. The program identified 261 ESTs (12.2%) corresponding to a set of 142 nonredundant Pex (Phytophthora extracellular proteins) cDNAs. Of these, 78 (55%) Pex cDNAs were novel with no significant matches in public databases. Validation of PexFinder was performed using proteomic analysis of secreted proteins of P. infestans. To identify which of the Pex cDNAs encode effector proteins that manipulate plant processes, high throughput functional expression assays in plants were performed on 63 of the identified cDNAs using an Agrobacterium tumefaciens binary vector carrying the potato virus X (PVX) genome. This led to the discovery of two novel necrosis-inducing cDNAs, crn1 and crn2, encoding extracellular proteins that belong to a large and complex protein family in Phytophthora. Further characterization of the crn genes indicated that they are both expressed in P. infestans during colonization of the host plant, tomato and that crn2 induced defense response genes in tomato. Our results indicate that combining data mining using PexFinder with PVX-based functional assays can facilitate the discovery of novel pathogen effector proteins. In principle, this strategy can be applied to a variety of eukaryotic plant pathogens, including oomycetes, fungi and
nematodes. [The sequence data reported in this paper have been submitted to the GenBank data library under accession nos.AF424638-AF424690. Supplemental Research Data is available on-line at http://www.oardc.ohio-state.edu/phytophthora/supp.htm]

INTRODUCTION

Interactions between plants and microbial pathogens involve complex signal exchanges at the plant surface and intercellular space interface (Baker et al., 1997; Parniske 2000; Hahn and Mendgen 2001). For example, plant pathogens have the remarkable ability to manipulate biochemical, physiological and morphological processes in their host plants through a diverse array of extracellular effector molecules that can either promote infection or trigger defense responses (Knogge 1996; Lauge and De Wit 1998; Collmer et al., 2000; Kjemtrup et al., 2000; Staskawicz et al., 2001). Typically, such molecules are secreted into the intercellular interface between the pathogen and the plant or delivered inside the host cell to reach their cellular target. Thus, discovery programs that target genes encoding extracellular proteins can be expected to increase the probability of identifying genes involved in virulence.
This approach has been taken successfully in the study of bacterial pathogens and symbionts. For example, an early study showed that *Sinorhizobium meliloti* mutants deficient in extracellular proteins were five times more likely to be affected in symbiosis than random mutants (Long et al., 1988). More recently, the characterization of effector proteins secreted through the type III secretion system of animal and plant associated bacteria has emerged as a key strategy for understanding mechanisms of virulence (Collmer et al., 2000; Kjemtrup et al., 2000; Staskawicz et al., 2001). In eukaryotic plant pathogens, genomic studies that focus systematically on extracellular proteins remain limited to nematodes, in which secretions from the esophageal gland cells are thought to play critical roles in infection (Wang et al., 2001). However, several classes of oomycete and fungal effector molecules, such as elicitor proteins that induce plant defense responses and a programmed cell death response termed the hypersensitive response (HR), are known to require secretion (Lauge and De Wit 1998; Jia et al., 2000). Therefore, secretion is an essential mechanism for delivery of virulence factors by eukaryotic plant pathogens to their appropriate site in infected plant tissue.
In eukaryotic cells, most secreted and membrane proteins are exported through the general secretory pathway (also known as type II secretion system) via short, N-terminal, amino-acid sequences known as signal peptides (von Heijne 1985; Rapoport 1992). Typically, signal peptides contain one or two charged amino acids followed by a hydrophobic core, and the signal peptidase cleavage site is defined by a pair of small uncharged amino acids (von Heijne 1985). Although most of these features can be identified in known extracellular proteins, the particular amino acid sequences are highly degenerate, and cannot be identified using DNA hybridization or PCR based techniques (Klein et al., 1996). However, with the advent of genomics, large sets of sequence data became available creating the opportunity to develop and test predictive software to identify extracellular proteins. For example, SignalP v2.0, a program that was developed using machine learning methods, assigns signal peptide prediction scores and putative cleavage sites to unknown amino acid sequences with a high level of accuracy (Nielsen et al., 1997; Nielsen and Krogh 1998; Menne et al., 2000). The Irish famine pathogen, *Phytophthora infestans*, is an eukaryotic oomycete microorganism that causes late blight, a worldwide
devastating disease of potato and tomato (Fry and Goodwin 1997a; Fry and Goodwin 1997b). Although a pathogen of great economic importance, little is known about the molecular mechanisms involved in pathogenicity and host specificity of \textit{P. infestans} and only a handful of genes have been implicated in the interaction with host plants (Kamoun 2000; Kamoun 2001). Structural genomics of \textit{Phytophthora} is underway. Pilot cDNA sequencing projects were performed for \textit{P. infestans} and another species \textit{Phytophthora sojae} (Kamoun et al., 1999b; Qutob et al., 2000) resulting in a database of expressed sequence tags (ESTs) (Waugh et al., 2000). With the accumulation of sequence data for \textit{Phytophthora}, the challenge is shifting to data mining and functional analyses. One important goal is to be able to associate a biological function with sequences with no significant similarity to known genes. With this objective in mind, we set up to identify systematically \textit{P. infestans} cDNAs encoding extracellular proteins from EST databases. Here, we describe PexFinder, an algorithm for the automated identification of putative extracellular proteins from ESTs. We applied PexFinder to a \textit{P. infestans} EST data set and selected 63 candidate Pex (\textit{Phytophthora} extracellular proteins) cDNAs for functional expression in plants using a
viral vector. This functional genomics strategy resulted in the discovery of a novel family of necrosis-inducing genes that are predicted to encode extracellular proteins with no similarity to sequences in public databases.

MATERIALS AND METHODS

PexFinder program

PexFinder was written as a series of programs/scripts in C++ and Perl using a Linux platform. The program runs in conjunction with SignalP v2.0b2 and is available on the web (http://www.cbs.dtu.dk/services/SignalP-2.0) or as a standalone application. The programs as well as detailed instructions for use can be downloaded from http://www.oardc.ohiostate.edu/phytophthora/pexfinder.

Data sets

The processed ESTs (GenBank accessions BE775444-BE777460) were generated as 5' reads of random clones picked from a cDNA library constructed from mycelial RNA of P. infestans (Kamoun et al., 1999b). A control data set of annotated sequences corresponding to 59 oomycete genes (online supplementary material Table S2) was compiled from GenBank (23 sequences) and the Phytophthora Genome Initiative (PGI) database (36 sequences) (Waugh et al.,
The 36 full length cDNAs were selected from a list of nonredundant ESTs with highly significant matches (E values <10-30) to known proteins (Kamoun et al., 1999b). The inserts of these cDNAs were fully sequenced using a primer walking approach and deposited in GenBank (see online supplementary material Table S2 for accession numbers). Subcellular localizations were determined based on information obtained from the InterPro collection of databases of protein families, domains, and functional sites (Apweiler et al., 2001), GenBank (Karsch-Mizrachi and Ouellette 2001), or was based on literature review of the examined protein.

**Strains and culture conditions**

The ESTs were obtained from *P. infestans* strain isolate DDR7602 (US-1 genotype, A1 mating type). This isolate is avirulent on *N. benthamiana*. *P. infestans* isolate 90128 (A2 mating type, race 1.3.4.7.8.9.10.11) was used in the infection assay. The isolates were routinely grown on rye agar medium supplemented with 2% sucrose (Caten and Jinks 1968). For RNA extraction, plugs of mycelium were transferred to modified Plich medium (Kamoun et al., 1993) and grown for 2-3 weeks before harvesting. For proteomics, the synthetic medium of Henniger (1959) was
used. *A. tumefaciens* strain GV3101 (Holsters et al., 1980) was used in molecular cloning experiments and was routinely grown at 28°C in Luria-Bertani (LB) media using the appropriate antibiotics (Sambrook et al.1989). Prior to storage at 80°C, cultures were grown without shaking in microplates containing 60 µl LB freezing buffer (36 mM K2HPO4, 13.2 mM K2HPO4, 1.7 mM citrate, 0.4 mM MgSO4, 6.8 mM (NH4)2SO4, 4.4 %v/v glycerol in 1x LB).

**Proteomic analysis of secreted proteins**

We cultured *P. infestans* in the synthetic medium of Henniger (1959). Seven to ten days after inoculation, culture filtrate was collected. Care was taken to keep mycelia untouched. A 0.2 µm filter (Sartorius, AG Göttingen, Germany) was used to remove any loose cell debris before the collections were stored at -20°C. For protein precipitation, we added acetone to the culture filtrates up to 60%(v/v) and kept the solution at -20°C for 1 hour to reduce the amount of INF1, an abundantly secreted protein in *P. infestans* (Kamoun et al., 1998). A protein pellet was obtained by centrifugation (10,000 rpm, 10 min) and air-dried at room temperature before storage at -20°C. Protein samples were dissolved in 2-D lysis buffer (0.01 M Tris-HCl, pH 7.4, 1 mM EDTA, 8 M urea, 0.05 M DTT, 10% (v/v)
glycerol, 5%(v/v) CHAPS and 6%(w/v) ampholytes (Resolyte pH 3-10)) and kept at -70 °C. Up to 60 µl of protein sample was run on a 7 cm pH 3-10 non-linear pH gradient strip (IPG)(Amersham-Pharmacia Biotech.) and run as described by Martin et al., (2001). For the second dimension electrophoresis, a mini 2-D gel system was employed using the protocol described by Martin et al., (2001). Proteins were visualized by colloidal Coomassie Brilliant Blue G-250 staining as described by Cash et al., (1997), or silver stained using the Plus-One Protein staining kit from Amersham Biosciences, according to the manufacturer’s instructions. Molecular masses of the proteins were determined by co-electrophoresis with standard protein markers. Iso-electric points were determined based on the linearity of the IPG strip. Peptide-mass mapping was conducted on several protein spots from the Coomassie-stained gels. Excised proteins were in-gel-digested with trypsin (sequencing -grade modified trypsin; Promega) and prepared as described by Shevchenko et al., (1996) and Wilm et al., (1996). Mass spectra were acquired on a PerSeptive Biosystems Voyager-DESTR MALDI-TOF mass spectrometer. The masses of the tryptic fragments were entered into the MS-Fit program (MS-Fit 3.1.1, Protein Prospector 3.2.1,
University of California) to search the Pex cDNA sequences for protein identification. Search parameters were as follows: maximum peptide mass tolerance was 200ppm; cysteines were modified by carbamidomethylation and maximum allowed missed cleavages were 2.

**Construction of cDNA mini-libraries**

A mini-library of the 63 cDNAs was constructed in the binary potato virus X (PVX vector) pGR106 (Jones et al.1999; http://www.jic.bbsrc.ac.uk/Sainsbury-Lab/david-baulcombe/Services/vigsprotocol.htm) as follows. The 63 cDNA inserts were PCR amplified using the high fidelity pfu DNA polymerase (Strategene, La Jolla, CA) and primers PSPORT-ASC (5'-GGCCGCGCGCCCTCCCCGATTCGACCCACGCGTCC-3') and SP6 (5'-TACGATTTAGTGACACTATAG-3') that correspond to vector sequences. The approximate size of the cDNA inserts was estimated using gel electrophoresis, and the 63 cDNAs were divided into 8 pools of 4-10 clones based on fragment size. Following pooling, the cDNAs were digested with Asc I and Not I, ligated to pGR106, and electroporated into *A. tumefaciens* GV3101 to generate an agroinfection-ready mini-library of ca. 760 clones with each pool represented by at least an 8-fold coverage.
**Plant assays**

Young *N. benthamiana* plants at approximately the 3-4-leaf stage (ca.3-week-old) were used for the agroinfection screening assays. Plants were cultured and maintained in a greenhouse with an ambient temperature of 22-25°C and high light intensity. Under these conditions, plants grew vigorously, which resulted in rapid PVX infection. Inoculations were performed on two lower leaves by dipping a wooden sterile toothpick in a recombinant *A. tumefaciens* GV3101 (PVX::cDNA) cultures grown on solid agar medium and wounding each leaf twice around the main vein and near the base of the leaf. An excess of bacteria was used for the inoculations. Mosaic, crinkling, local and systemic necrotic symptoms were scored daily and typically started developing within 5-7 days after inoculation. Inserts of clones of interest were PCR amplified with PVX primers PVX-F (5'-AATCAATCACAGTGTTGGCTTGC-3') and PVX-R (5'-AGTTGACCTATGGGCTGTGTG-3') and sequenced using the same primers. PVX::cDNA constructs that induced necrotic symptoms were retested at least five times on 3-4 leaf plants and also in side-by-side inoculation on older plants with more than 6 leaves to allow comparative analyses. Tomato and tobacco plants were inoculated using similar
methods. For the time course northern blot of *P. infestans* colonization of tomato, fully expanded tomato leaves were detached and inoculated as described earlier (Kamoun et al., 1998). A 10 µl sample containing ca. 1000 zoospores of *P. infestans* was used to inoculate the underside of the detached leaves. Leaf discs of similar sizes were dissected from the inoculated regions while making sure that the inoculation spot is in the center. The leaf disks were frozen in liquid nitrogen and stored at –80°C for later use. For the northern blot experiment on plant defense responses, cotyledons of young tomato plants with fully expanded leaves were toothpick inoculated with the PVX derivatives. Leaves were harvested from systemically infected leaves 12 days post inoculation, frozen in liquid nitrogen, and stored at –80°C for later use. DNA manipulations, sequencing and sequence were conducted essentially as described elsewhere (Sambrook et al., 1989). cDNAs were sequenced by primer walking using an ABI Prism 377 automated sequencer (PE Applied Biosystems). Similarity searches were performed locally on an Intel Linux workstation using BLAST (Altschul et al., 1997). Local databases included GenBank nonredundant and dBEST (Karsch-Mizrachi and Ouellette, 2001), PGI (Waugh et al., 2000), and
SPC, a proprietary database of Syngenta Inc. containing ca. 34,000 ESTs from *P. infestans* (courtesy of S. Lam, Syngenta, Research Triangle Park, NC). Multiple alignment of the *crn*-like sequences from *P. infestans*, *P. sojae*, and *Phytophthora medicaginis* was conducted using the program CLUSTAL-X (Thompson et al., 1997).

**Northern blot hybridizations**

Total RNA from *P. infestans* mycelium and from infected tomato leaves was isolated using the Trizol reagent (Gibco-BRL, Bethesda, MD) following the manufacturer’s instructions. To prepare the Northern blot, 15 µg RNA was denatured in NorthernMax Gel Loading Solution (Glyoxal) (Ambion, Austin, TX USA) according to the manufacturer’s instructions. RNA was separated by gel electrophoresis on a 1% agarose gel in 1X BPTE buffer (10mM PIPES- piperazine-N,N'-bis(ethanesulfonic acid), 30mM Bis-Tris, 1mM EDTA (pH 8.0) (Sambrook et al., 1989). Transfer was done onto Hybond N+ membrane (Amersham, Piscataway, NJ USA) as described by manufacturer. Hybridizations were done with radiolabelled probes at 65°C in Modified Church Buffer (0.36M Na2HPO4, 0.14 NaH2PO4, 1mM EDTA and 7%SDS). Membranes were washed at 55°C for 15 mins. each in 1XSSC/0.5%SDS and 0.5X SSC/0.5%SDS. Membranes with RNA side
up were exposed to a phosphor imager screen (Molecular Dynamics Storm 840 Phosphor Imager for 2-48 hours. Probes for the crn and actin genes were gel-purified fragments digested from the corresponding cDNA clones (this study, Unkles et al 1991). Probes for the tomato genes were gel purified DNA fragments obtained from cDNA clones (E. Huitema and S. Kamoun, unpublished) and corresponded to the PR1a gene (van Kan et al., 1992), the hsr201 gene (Czernic et al.1996) and tubulin (GenBank accession CAD13178). All probes were radiolabelled with $\alpha$ -32 P-dATP using a random primer labeling kit (Gibco-BRL, Bethesda, MD).

RESULTS

**PexFinder: An algorithm for the automated identification of extracellular proteins from ESTs**

To identify signal peptides from ESTs, we developed an algorithm, PexFinder, which integrates a series of sequence analysis and data manipulation programs with SignalP v2.0, a program that predicts the presence of signal peptides and their cleavage sites using artificial neural networks (SignalP-NN) and hidden Markov models (SignalP-HMM)(Nielsen et al.,1997; Nielsen and Krogh 1998). The PexFinder pipeline is illustrated in Fig. 2.1. Inputs to the pipeline
were edited nucleotide sequences (5'ESTs) in the FASTA format (Pearson 1990). These sequences were scanned for start codons (ATG), the first 70 codons were translated, and an output file was generated in a format that can be read by SignalP v2.0. Following SignalP v2.0 analysis, the output files were manipulated to select for signal peptides containing sequences using the following two criteria: (1) SignalP-HMM prediction is "Signal Peptide", and (2) SignalP-NN predicts a cleavage site between 10 and 40 amino acids in length. An output file containing only those ESTs that were identified as signal peptide positives was generated. These ESTs were then queried against the original EST database using the BLASTN algorithm (Altschul et al., 1997) to check for and eliminate redundant sequences. Nonredundant sequences predicted to encode extracellular proteins were queried against GenBank nonredundant database using BLASTX (Altschul et al., 1997) for annotation.

**PexFinder processing of P. infestans ESTs**

We analyzed a total of 2147 ESTs from *P. infestans* (GenBank accessions BE775444-BE777460) using PexFinder. 261 ESTs corresponding to 142 nonredundant sequences were predicted to contain signal peptide-like sequences (on-line
supplementary material Table S1). A scatter diagram of the predicted mean S scores of SignalP-NN and the scores of SignalP-HMM for the 142 selected ESTs showed clustering at the higher quadrant of the graph (Fig. 2.2A). For comparison, we used SignalP v2.0 to analyze a set of 59 known and annotated oomycete proteins (on-line supplementary material Table S2). A scatter diagram of the SignalP-NN and SignalP-HMM scores showed a clear separation of the cytoplasmic oomycete proteins from the extracellular ones (Fig. 2.2B). This suggests that the SignalP v2.0 program can identify unambiguously a signal peptide in oomycete sequences and that the selected ESTs are likely to encode putative extracellular proteins.

Analysis of selected Pex cDNAs

Of the 142 nonredundant Pex cDNAs, 78 (55%) were novel and showed no significant matches (E value >10^{-5}) to sequences in public databases, whereas the remaining 64 (45%) showed significant hits (E value <10^{-5}) and 56 (39%) of these could be assigned a putative function (on-line supplementary material Table S1). To evaluate the performance of PexFinder, we assigned the most likely subcellular localization to 55 of the Pex cDNAs based on the functional identity of their best matches and
independently of the SignalP scores (on-line supplementary material Table S1). Of the 55 cDNAs, 40 (73%) were predicted to encode extracellular or membrane-associated proteins as expected for proteins carrying functional signal peptides. Of these, 21 were predicted to encode extracellular proteins, 13 integral membrane proteins, and 6 proteins localized in the lumen of the endoplasmic reticulum. On the other hand, 8 of the cDNAs were predicted to encode cytosolic protein, 6 mitochondrial protein, and 1 peroxisomal protein, suggesting that 15 (17%) of the cDNAs were likely false positives. Eight ESTs corresponded to full length cDNAs for known *Phytophthora* extracellular proteins, such as INF1 elicitin, elicitin-like proteins, cyst germination specific protein, and necrosis-inducing protein. Ten ESTs were predicted to encode secreted enzymes or other products that may function in virulence. These include chitinase, arabinofuranosidase/β-xylosidase, endo-glucanase, exo-glucanase, glycosyl hydrolase, peroxidase, protease, and thaumatin. A total of 6 cDNAs with no matches in public databases were predicted to encode cysteine-rich protein less than 150 amino acids in length, a feature typical of effector proteins involved in plant-microbe interactions (Templeton et al. 1994).
Validation of PexFinder using proteomics

We used proteomics of secreted proteins of *P. infestans* to further validate PexFinder. We collected filtrates from *P. infestans* cultures grown in a synthetic medium, precipitated the proteins with acetone, and subjected them to two-dimensional gel electrophoresis (Fig. 2.3). A total of 30 spots were excised from the gel, digested with trypsin, and subjected to peptide mass mapping using a MALDI-TOF mass spectrometer. Taking into account unsuccessful analyses and redundant spots, a total of 22 different peptide mass spectra were obtained. We then matched the observed masses of the peptides with predicted masses deducted from the 142 Pex cDNAs using the program Protein Prospector 3.2.1 (University of California). A total of 9 proteins could be matched unambiguously to cDNAs identified using PexFinder (Fig. 2.3) and on-line supplementary material Table S1). These results provide experimental validation of the PexFinder algorithm.

Functional expression of Pex cDNAs in plants

Ectopic expression of single pathogen genes in plant cells often lead to phenotypic effects, such as induction of HR-like necrosis (Lauge and De Wit 1998; Kjemtrup et
To identify Pex cDNAs that induce necrotic symptoms in plants, we developed a strategy for functional expression screening of selected *P. infestans* cDNAs in *Nicotiana benthamiana* using an *A. tumefaciens* binary potato virus X (PVX)-based vector (Fig. 2.4). *N. benthamiana* was selected because it is highly susceptible to PVX infection and exhibits HR-based resistance to *P. infestans* (Kamoun et al. 1998). We selected 63 Pex cDNAs, most of which correspond to sequences with no significant matches in GenBank (on-line supplementary material Table S1), for cloning in the binary PVX expression vector pGR106 (Jones et al. 1999; http://www.jic.bbsrc.ac.uk/Sainsbury-Lab/david baulcombe/Services/vigsprotocol.htm). The inserts of the 63 cDNAs were divided in 8 pools of 4-10 clones and ligated to pGR106 resulting in an *A. tumefaciens* infection-ready mini-library. A total of 760 clones were then inoculated in *N. benthamiana*, a plant species that is highly amenable to high throughput PVX assays (Baulcombe 1999) and that is known to display resistance responses to *P. infestans* (Kamoun et al., 1998; Kamoun 2001). Phenotypic evaluation of plants infected with the recombinant PVX identified 8 clones that induced necrotic symptoms at the inoculation site, as well as necrosis and crinkling in systemic leaves.
Partial DNA sequencing and restriction fragment analyses of the inserts of the 8 clones revealed that they correspond to 2 cDNAs, named crn1 (2 clones) and crn2 (6 clones) for crinkling and necrosis. Additional infection assays with PVX::crn1 and PVX::crn2 confirmed that both cDNAs consistently induced necrotic symptoms in N. benthamiana, whereas only mosaic symptoms were observed following inoculation with the empty vector (Fig. 2.5). Symptoms induced by PVX::crn2 started at about 5-7 days after inoculation depending on the age of the plant and leaves, and generally developed 7-8 days faster than symptoms induced by PVX::crn1. Additional infection assays with PVX::crn1 and PVX::crn2 on solanaceous plants showed that the induction of necrosis was not specific to N. benthamiana, but also occurred on the nonhost plant tobacco (Nicotiana tabacum) and the host plant tomato (Lycopersicon esculentum) (data not shown). On these plants, symptoms also varied in intensity with crn2 always causing faster and more severe necrosis than crn1.

**crn1 and crn2 are members of a large and complex gene family in Phytophthora**

The full sequence of the crn1 and crn2 cDNAs was determined. crn1 sequence revealed a 1296 bp ORF encoding a
predicted protein composed of 431 amino acids, whereas *crn2* revealed a 1371 bp ORF encoding a predicted protein composed of 456 amino acids. BLASTP searches against public databases revealed no similarity to proteins from other organisms. However, CRN1 and CRN2 shared significant similarity to each other (Fig 2.6) and TBLASTN searches against current EST databases of *Phytophthora* revealed that both proteins correspond to a large family in *Phytophthora* with at least 34 different members (on-line supplementary material Table S3). Multiple alignment of the 34 CRN-like sequences from *P. infestans*, *P. sojae* and *P. medicaginis* indicated a significant degree of homology between the N-terminal region of the different proteins (on-line supplementary material Fig.S1). 28 of the 33 full length CRN-like proteins were predicted to contain signal peptides using SignalP v2.0 (on-line supplementary material Table S3).

The *crn* genes are expressed during colonization of tomato by *P. infestans*

To determine whether the *crn* genes are expressed during colonization of plants, we analyzed the expression of *crn1* and *crn2* during the interaction of *P. infestans* with its host plant tomato. Total RNA was isolated from
leaves of tomato 0,1,2,3 and 4 days after inoculation with
P. infestans 90128 and from P. infestans mycelium grown in
liquid rye-sucrose medium. A northern blot containing these
samples was hybridized with probes of the P. infestans crn
and actA genes (Fig. 2.7). Because the total RNA extracted
from infected leaves consists of a mixture of Phytophthora
and plant RNA, the signals obtained on northern blots with
probes of differentially expressed genes should be
normalized to actual Phytophthora RNA levels as determined
by the signals obtained with the probe of a constitutively
expressed actin gene. Consistent with increases in mycelial
biomass during the infection, actA transcripts were first
detected at day 2 and reached maximal levels at days 3 and
4. An identical pattern of expression to actA was observed
for both crn genes suggesting that these genes are
constitutively expressed during colonization of tomato and
in in vitro grown mycelium.

Comparison of necrosis induction by crn to known
Phytophthora elicitors

We compared the timing and appearance of the necrotic
symptoms induced by the crn genes to those induced by known
Phytophthora defense response elicitors, such as P.
infestans INF1 (Kamoun et al., 1998; Kamoun et al., 1999a)
and *P. sojae* sojNIP (Qutob et al., 2002). PVX recombinants expressing *crn1*, *crn2*, *inf1*, and *PsojNIP* were inoculated side-by-side on mature leaves of *N. benthamiana* and examined for necrotic symptoms (Fig. 2.8). The first symptoms were observed with PVX::*inf1* and PVX::*PsojNIP* as early as 5 days after inoculation. In contrast PVX::*crn2* and PVX::*crn1* inoculated plants started developing necrosis about 7 and 12 days after inoculation, respectively. Compared to PVX::*inf1* and PVX::*PsojNIP* which induced dark necrotic lesions around the site of inoculation, the necrotic lesions induced by PVX::*crn2* started off as light brown lesions which developed over time to have darker margins.

*crn2* induces the expression of defense response genes in tomato

To determine whether the *crn* genes induce defense responses in plants, we inoculated tomato plants with PVX::*crn1*, PVX::*crn2*, and used as negative controls: mock inoculation, the empty PVX vector pGR106, and PVX carrying the *P. infestans* *inf1* gene, which does not induce necrosis on tomato (Kamoun et al., 1999a). Twelve days after inoculation, total RNA was isolated from tomato leaves systemically infected with PVX virions or the equivalent
leaves from mock inoculated plants. At this stage, only crn2 showed necrotic symptoms. A northern blot containing these samples was hybridized with probes of the tomato defense and pathogen-induced genes PR1a (van Kan et al., 1992) and hsr201 (Czernic et al., 1996), as well as a probe for the constitutive tomato gene tubulin. In contrast to the other treatments, PVX::crn2 treatment resembled pathogen infection and induced significant expression of both tomato defense response genes PR1a and hsr201 (Fig. 2.9).

DISCUSSION

Single pass cDNA sequencing is a relatively rapid and inexpensive procedure that has become a popular approach for large scale gene discovery in eukaryotic plant pathogens, such as oomycetes, fungi, and nematodes (Kamoun et al., 1999b; Popeijus et al., 2000; Qutob et al., 2000; Dautova et al., 2001; Skinner et al., 2001). However, due to the inherent low quality and partial nature of single pass cDNA sequences, gene annotation remains challenging, particularly the assignment of a putative function to sequences with no significant similarity to known genes. Here, we describe a general approach towards the
identification of novel effector proteins from EST databases of eukaryotic plant pathogens by combining data mining for extracellular proteins with high throughput functional expression assays in plants. This strategy led to the discovery of two novel necrosis and defense response-inducing cDNAs, crn1 and crn2, that encode related extracellular proteins and that are members of a large, newly discovered, gene family in Phytophthora.

We developed the algorithm PexFinder that scans ESTs for signal peptide sequences by integrating a series of scripts with SignalP, a program that distinguishes between cytoplasmic and extracellular oomycete proteins (Fig. 2.2B). We evaluated PexFinder performance by assigning a subcellular localization of the products of 55 of the Pex cDNAs based on functional annotation of their best matches and independently of SignalP scores (on-line supplementary material Table S1). PexFinder performance was satisfactory since 73% of the examined cDNAs were similar to proteins known to be extracellular or membrane-associated in Phytophthora or in other organisms. However, some limitations in PexFinder and SignalP predictions were apparent. For example, some of the signal sequences identified in ESTs predicted to encode integral membrane
proteins may correspond to internal transmembrane domains rather than N-terminal secretory signal peptides. In addition, six cDNAs were predicted to encode mitochondrial proteins suggesting that SignalP may not discriminate well between mitochondrial and secretory signal sequences. Future improvements in the SignalP program as well as the integration of additional steps to the algorithm should help address these limitations and further improve PexFinder performance. Nevertheless, despite these limitations, PexFinder was validated convincingly using proteomic analyses of secreted proteins of *P. infestans*. A total of 22 different and randomly picked secreted proteins were fingerprinted using mass spectra of tryptic peptides. Of these, 9 proteins (41%) matched sequences in the selected set of 142 Pex cDNAs (Fig. 2.3 and on-line supplementary material Table S1). This may seem like a fairly high percentage considering that we only examined a relatively small number of ESTs. However, significant overlap between the two data sets is not surprising considering that both EST and proteomic analyses enrich for highly expressed genes and proteins.

Annotation of EST function is done ordinarily using similarity searches to sequences in public databases. Our
results indicate that the PexFinder algorithm can help annotate those ESTs with no significant matches to known proteins resulting in a manageable number of candidate genes that can then be tested using functional assays. The PVX expression system has emerged as a reliable and robust system for transient expression of genes in plant cells and has been widely used to express virulence and avirulence genes from plant pathogens (Hammond-Kosack et al., 1995; Kamoun et al., 1999a; Qutob et al., 2002; Tobias et al., 1999; Tampakaki and Panopoulos 2000). More recently, PVX has been successfully used to assay random cDNA clones from the fungal pathogen Cladosporium fulvum for HR-inducing activity. In this study, we used the PVX expression system to assay 63 of the identified Pex cDNAs, including 54 cDNAs of unknown function. We constructed agroinfection-ready cDNA libraries by cloning pools of candidate full length Pex cDNAs into the binary PVX vector, thus reducing the time required to handle a large number of cDNAs. An alternative approach would have been to screen random clones as performed by Takken et al., (2000). In fact, we conducted a similar screen with 2400 random cDNAs from P. infestans (A. Testa and S. Kamoun, unpublished) and identified five necrosis-inducing cDNAs. This suggests that
selecting for candidate cDNAs using PexFinder resulted in a 15-fold increase in identifying necrosis-inducing cDNAs (2/63 or 3% compared to 5/2400 or 0.2%). Preselecting Pex cDNAs also takes full advantage of existing genomic resources and avoids potential cloning biases against large transcripts. On the other hand, screening for random cDNAs is a hypothesis-free approach that has the potential of revealing effector molecules of an unexpected nature. In summary, we view both approaches as complementary strategies that can be equally pursued for the discovery of novel effector proteins from plant pathogens.

We selected to perform functional screenings in *N. benthamiana*, a plant species that is highly susceptible to PVX, is amenable to high throughput PVX assays, and exhibits virus induced gene silencing (VIGS) (Baulcombe 1999). In addition, *N. benthamiana* is known to display HR-based resistance responses to *P. infestans* and holds great promise for dissecting elicitor response and resistance to *Phytophthora* (Kamoun et al., 1998; Kamoun 2001). Using PVX expression assays in *N. benthamiana*, we discovered two novel necrosis-inducing cDNAs, *crn1* and *crn2*, that encode related and novel extracellular elicitor proteins of *Phytophthora*. Recombinant PVX carrying the *crn* genes were
compared in side-by-side inoculations to PVX constructs expressing the HR-elicitor inf1 elicitin gene of P. infestans (Kamoun et al., 1998; Kamoun et al., 1999b), and the defense and necrosis-inducing gene PsojNIP of P. sojae (Qutob et al., 2002). Overall, the responses induced by the crn genes differed from those induced by inf1 and PsojNIP. First, necrotic symptoms were delayed compared to the other two elicitors, and occurred 1-2 days later for crn2 and ca. 7 days later for crn1 (Fig. 2.8). The exact significance of this difference is unclear at this stage but could be due to a lower specific activity for the CRN products thus requiring more virus accumulation before an effective concentration is reached. The appearance of the necrotic lesions also varied between the crn genes and the inf1 and PsojNIP genes. Both INF1 and PsojNIP induced dark black lesions in Nicotiana, whereas CRN1 and CRN2 induced light brown lesions that later developed dark black edges. This difference could reflect the induction of secondary metabolite pathways by CRN that vary from those induced by the other elicitors. However, the ability of CRN2 to induce defense and pathogenesis-related genes was reminiscent of INF1 and other elicitors. In tobacco, similar to INF1, the necrosis induced by CRN2 was associated with the activation
of the tobacco defense response gene PR2 (data not shown). In tomato, CRN2 induced the defense genes PR1a and hsr201 (Fig. 2.9) that are known to be up-regulated during infection by microbial pathogens. In this respect, CRN2 appeared to mimic pathogen infection of tomato. In summary, it appears that CRN induced defense responses in plants similar to those induced by known defense elicitors, but with different timing and appearance of the necrotic response. This indicates that CRN may induce different defense pathways in plants from those induced by INF1 and PsojNIP. The crn genes may thus be useful in helping dissect plant defense pathways. Future work will include extensive studies of plant response to CRN and identification of plant mutants deficient in CRN response.

Molecular characterization of the crn genes suggests that they are widely distributed and form a complex gene family in Phytophthora similar to the ubiquitous INF elicitin family (Kamoun et al., 1997; Kamoun 2000). However, the predicted CRN proteins are structurally unrelated to INF elicitins. The large size (ca. 450 amino-acids) of CRN1 and CRN2, and the presence of CRN conserved motifs suggest that these proteins may function as secreted enzymes similar to the 42-kDa glycoprotein elicitor of
*P. sojae* (Sacks et al. 1995), which was recently found to possess transglutaminase activity (T. Nürnberg, pers. comm.). No similarity was also noted to the recently characterized necrosis-inducing protein PsojNIP of *P. sojae* (Qutob et al., 2002). However, unlike the *inf1* and *PsojNIP* genes, which were reported to be down-regulated in the early stages of infection, the crn genes appeared to be constitutively expressed both *in planta* and *in vitro*. The observation that the *crn* genes are expressed in *P. infestans* during colonization of tomato (Fig. 2.7) suggests that the CRN products could be secreted and thus delivered to plant cells during infection. Analyses of the response induced by the *crn* cDNAs in various plants suggest that they are general elicitors that trigger necrotic responses non-specifically in both resistant *Nicotiana* species and the susceptible host plant tomato. In this respect, the CRN differ from specific elicitors, such as INF1, which induce defense responses only in specific plant genotypes (Kamoun et al., 1998; Kamoun et al., 1999a), but resemble *PsojNIP*, which functions in several dicotyledonous plants (Qutob et al., 2002). Recently, the general elicitors of plant pathogens have been likened to pathogen-associated molecular patterns (PAMPs), which are surface-derived
molecules that induce the expression of defense response genes and the production of antimicrobial compounds in both animal and plant cells (Nürnberg and Brunner 2002; Gomez-Gomez and Boller 2002). Whether the CRN proteins function as PAMPs remains to be determined, but is supported by the observation that the crn genes were found in several Phytophthora species. Alternatively, the CRN could be aiding in colonization of plant tissue during the late necrotrophic phase of infection as proposed for the PsojNIP protein (Qutob et al., 2002). In the future, additional functional analyses of the crn genes in P. infestans and other Phytophthora species will help determine the nature of the contribution of these genes to the infection process.

In principle, the two-step functional genomics strategy we developed can be applied to a variety of eukaryotic plant pathogens, including fungi and nematodes. Preliminary analyses of PexFinder processed ESTs from plant parasitic nematodes suggest that the algorithm performs well in that system (T. Maier and T. Baum, pers. comm.). In the future, accumulation of ESTs from plant pathogens and further improvements in virus vector technology should
allow abroad application of this strategy to the study of plant-microbe interactions.
ACKNOWLEDGMENTS

The proteomics part is solely the work of Shaung Li, Pieter van West and Neil A.R. Gow with assistance from the following members of the Aberdeen Proteome Facility, Scotland UK: specifically Phil Cash, Evelyn Argo, and Audrey Innes for running the 2 D gels, and Liz Stewart and Ian Davidson for Mass spectrometry analysis. I am grateful to Allison Styer, Ian Holford, Peter Hraber, and Callum Bell for automation of the PexFinder algorithm. Thanks to Shujing Dong, Diane Kinney, Caitlin Cardina and Edgar Huitema for technical assistance, Jesse Ewing for help with figure design, Abdelhafid Bendahmane for the recipe of the LB freezing medium, and Isabelle Malcuit and David Baulcombe for providing pGR106 and help with the PVX system. I also thank Tea Meulia and the staff of the OARDC Molecular and Cellular Imaging Center for help with DNA sequencing.

This work was supported by the OARDC Research Enhancement Grant Program, Syngenta Biotechnology, and NSF grant DBI-0211659. The proteomics part performed by the lab of Peter van West and Neil Gow was supported by the Royal Society. Salaries and research support were provided, in part, by State and Federal Funds appropriated to the Ohio
Agricultural Research and Development Center, the Ohio State University.

A modified version of this chapter has been submitted for publication as:


REFERENCES


Nürnberg, T., and Brunner, F. (2002). Innate immunity in plants and animals: emerging parallels between the


Figure 2.1. Schematic illustration of the PexFinder algorithm for identification of extracellular proteins from expressed sequence tags. The various steps of the algorithm are explained in the text. The programs as well as detailed instructions can be obtained from http://www.oardc.ohio-state.edu/phytophthora/pexfinder.
Figure 2.1.
Figure 2.2. Scatter diagrams of output scores obtained with SignalP v2.0 for (A) 142 putative extracellular proteins identified from *Phytophthora infestans* using PexFinder, and (B) 59 known and annotated oomycete proteins. HMM score refers to the prediction score obtained with the SignalP V2.0 hidden Markov model program, and mean S score refers to the score obtained with the SignalP V2.0 neural network program. Scores for both programs range from 0 to 1, with higher scores indicating a higher probability for a signal peptide. Sequences predicted to correspond to extracellular proteins based on annotation using the InterPro database (Apweiler et al. 2001), GenBank (Karsch-Mizrachi and Ouellette 2001), or literature review are shown as black boxes, and other sequences are shown as gray boxes.
Figure 2.2.
Figure 2.3. Two-dimensional gel electrophoresis analysis of secreted proteins of *Phytophthora infestans*. Circled spots highlighted by a number correspond to 9 identified Pex proteins from a total of 30 spots analyzed using mass spectrometry of tryptic peptides. (1) BE777050: arabinofuranosidase/β-xylosidase; (2) BE776562: acidic chitinase; (3) BE776874: enoyl Coenzyme A hydratase; (4, 5) BE777263: glutathione-S-transferase; (6) BE776368: unknown protein; (7) BE776395: similar to *Phytophthora sojae* avr1b; (8) BE777082: unknown protein; (9) BE776294: peptidylprolyl isomerase.
Figure 2.4. Strategy for functional screening of Phytophthora infestans cDNAs using agroinfection with a binary potato virus X (PVX) vector. Pools of candidate cDNAs were ligated to the binary PVX vector, and electroporated into Agrobacterium tumefaciens to generate infection ready mini-libraries. Leaves of Nicotiana benthamiana plants were then toothpick inoculated with individual clones. Clones inducing necrotic symptoms were identified by polymerase chain reaction amplification amplification and sequencing.
Figure 2.5. Symptoms observed on *Nicotiana benthamiana* plants after inoculation with *Agrobacterium tumefaciens* containing the binary potato virus X (PVX) vector expressing *Phytophthora infestans* crn1 and crn2 cDNAs. Inoculated leaves and plantlets of *N. benthamiana* were photographed 12 (A, C, E) or 16 (B, D, F) days after inoculation with *A. tumefaciens* containing the binary PVX vector pGR106 (Jones et al. 1999) (A-B), pGR106 derivative expressing crn1 cDNA (C-D), or pGR106 derivative expressing crn2 cDNA (E-F). Panels B, D, and E correspond to an independent experiment from panels A, C, and E, in which smaller plants were inoculated resulting in faster systemic spread of the virus. Panels A-B show no symptoms at the inoculation site and systemic mosaic symptoms typical of PVX infection. Panels C-E show localized necrosis at site of inoculations, and panels D-F show systemic necrotic symptoms. Note that the symptoms obtained with the crn2 construct were observed as early as 5 days after inoculation, and the symptoms induced by crn1 occurred later and started around 12 days after inoculation.
Figure 2.5.
Figure 2.6. Alignment of Phytophthora infestans CRN1 and CRN2 predicted amino acid sequences. The amino acid sequences of CRN1 and CRN2 were predicted from full open reading frame (ORF) sequences. Identical amino acids are shaded in dark gray and similar ones in light gray. The predicted mature CRN1 protein starts at position 18 (VDI...), and the mature CRN2 protein at position 23 (SQL...). A multiple alignment of all known CRN-like sequences is available as a Web Supplement.
Figure 2.7. Time course of expression of crn1, crn2 and actA of Phytophthora infestans during colonization of tomato. Total RNA isolated from infected leaves of tomato, 0, 1, 2, 3, or 4 days after inoculation, from non-infected leaves (TO), and from P. infestans mycelium grown in synthetic medium (MY) was sequentially hybridized with probes from the crn1, crn2, and actA genes. The approximate sizes of the transcripts are approximately 1700 nucleotides for crn1 and crn2, and 1600 nucleotides for actA.
Figure 2.8. Symptoms observed on *Nicotiana benthamiana* plants after inoculation with *Agrobacterium tumefaciens* containing binary potato virus X (PVX) derivatives expressing *crn1*, *crn2*, and *inf1* from *Phytophthora infestans* or *PsojNIP* from *Phytophthora sojae*. Side-by-side toothpick inoculations were performed on leaves of 5-week old *N. benthamiana*. The binary PVX vector vector pGR106 was used as a negative control. Photographs were taken 5, 7, 9 and 11 days post inoculation (dpi).
Figure 2.9. Induction of tomato defense genes by crn genes. Total RNA (15 μg) isolated from tomato 12 days after mock inoculation with a toothpick (W), or with the binary potato virus X pGR106 (V), PVX::crn1 (crn1), PVX::crn2 (crn2), and PVX::inf1 was sequentially hybridized with probes from the tomato PR1a and hsr201 defense response genes. A probe from the tomato tubulin (tub) gene was used as a loading control. Note in tomato and under these experimental conditions, only PVX::crn2 induced necrotic lesions.
CHAPTER 3

The Pipg1 Gene of the Oomycete Phytophthora infestans
Encodes a Fungal-Like Endopolygalacturonase

ABSTRACT

Endopolygalacturonases (endoPGs) are plant cell wall degrading enzymes that have been implicated in the invasion of plant tissue by pathogenic microbes. EndoPGs have been described from bacteria, plants, insects and numerous species of phytopathogenic fungi. In this study, we describe the first endoPG sequence from oomycetes, a unique group of eukaryotic plant pathogens that exhibit fungal-like filamentous growth but share little taxonomic affinity to fungi. The characterized gene, pipg1, was identified from the potato late blight pathogen Phytophthora infestans, and was predicted to encode a secreted glycoprotein with all the signature sequences of endoPGs. Multiple pipg1-like sequences were identified in the P. infestans genome. Pipg1 was expressed during preinfection and infection stages. Phylogenetic analysis of endoPGs indicated that pipg1 forms a unique class that was
significantly more similar to fungal endoPGs than to plant or bacterial ones. This unexpected affinity between PIPG1 and fungal endoPGs contrasts with phylogenies obtained using ribosomal sequences or compiled protein sequences from mitochondrial and chromosomal genes, raising interesting questions about the evolution of these enzymes in oomycetes.

INTRODUCTION

Most microbial plant pathogens produce an array of extracellular enzymes that degrade plant cell walls and aid in penetration and colonization of plant tissue. One of the first enzymes secreted by invading pathogens is endopolygalacturonase (endoPG, E.C. 3.2.1.15), which breaks up the $\alpha, 1,4$ polygalacturonic acid of pectins, a complex polysaccharide found in the middle lamella and primary cell wall of higher plants (Esquerre-Tugaye et al., 2000; ten Have, 2001). EndoPGs have been described from plants, bacteria, as well as from a large number of phytopathogenic fungi (Lang and Dornenburg, 2000). The role of endoPGs in plant-fungal interactions centers around a complex cross talk of signals and responses. Whereas it is generally accepted that endoPG degradation of plant cell wall is an
important step in penetration and invasion of host tissue, some of the pectic fragments released through the enzymatic activity of endoPGs can activate defense responses in plants (Esquerre-Tugaye et al., 2000; Lang and Dornenburg, 2000; ten Have, 2001). In addition, plants produce a class of proteins, known as polygalacturonase-inhibiting proteins (PGIPs), that can specifically bind and inactivate endoPGs (Esquerre-Tugaye et al., 2000; Lang and Dornenburg, 2000; ten Have, 2001). Therefore, the exact contribution of endoPGs to virulence remains unclear and is likely to vary depending on the pathosystem examined.

Oomycetes represent a diverse group of organisms that includes pathogens of plants, insects, crustaceans, fish and vertebrate animals, as well as saprophytic species. Traditionally and due to their filamentous growth habit, oomycetes have been classified in the Kingdom Fungi. However, modern molecular and biochemical analyses suggest that oomycetes share little taxonomic affinity to filamentous fungi, but are more closely related to brown algae (heterokonts) in the Kingdom Stramenopiles (Baldauf et al., 2000; Kumar and Rzhetsky, 1996; Lang et al., 1999; Paquin et al., 1997; Van de Peer and De Wachter, 1997). Among the oomycetes, Phytophthora spp. cause some of the
most destructive plant diseases in the world and are arguably the most devastating pathogens of dicot plants (Erwin and Ribeiro, 1996). For example, *Phytophthora infestans*, also known as the cause of the Irish potato famine, remains a destructive pathogen resulting in multibillion-dollar losses in potato and tomato production (Fry and Goodwin, 1997a; Fry and Goodwin, 1997b). Despite its economic importance and singular phylogenetic position, little is known about the molecular mechanisms that control pathogenicity in *P. infestans*. In recent years, structural genomics, such as, large scale sequencing of random cDNAs, emerged as a promising strategy for the identification of novel virulence genes (Kamoun et al., 1999; Qutob et al., 2000; Skinner et al., 2001). We and others in the oomycete research community generated thousands of expressed sequence tags (ESTs) from *Phytophthora* that were compiled into searchable databases (Kamoun et al., 1999; Qutob et al., 2000; Waugh et al., 2000).

In this study, we describe and characterize *pipg1*, an endopolypgalacturonase gene from *P. infestans*, and one of the first putative virulence genes of *Phytophthora* identified using EST databases. Characterization of endoPGs from oomycetes should bring a new perspective to
understanding the general role of these enzymes in interactions between plants and microbes.

MATERIALS AND METHODS

Phytophthora strains and culture conditions

Two *P. infestans* isolates from the Netherlands, 88069 (A1 mating type, race 1.3.4.7) and 90128 (A2 mating type, race 1.3.4.7.8.9.10.11) were used throughout the study and in the infection assays. The two isolates infect tomato plants at different rates. 88069 produces slowly expanding necrotic lesions, whereas 90128 produces rapidly expanding and greenish (biotrophic) lesions. *P. infestans* T30-4, a hybrid between two Dutch isolates, as well as three isolates from Ohio, OH3, OH10 and OH23, were used in the DNA blot experiment. *P. infestans* strains were routinely grown on rye agar medium supplemented with 2% sucrose (Caten and Jinks, 1968). For DNA and RNA extraction, plugs of mycelium were transferred to modified Plich medium (Kamoun *et al*., 1993) and grown for 2-3 weeks before harvesting. Germinating cysts and the time courses of *P. infestans* infection of tomato tissue were obtained exactly as described earlier (Kamoun *et al*., 1997).
DNA manipulations

DNA manipulations were conducted essentially as described elsewhere (Sambrook et al., 1989). The pipg1 cDNA was sequenced by primer walking using an ABI Prism 377 automated sequencer (PE Applied Biosystems). In addition to vector primers, the following primers were used: 11B10F (5′-ACTGTGACTGGACCTGGAAC-3′), 11B10F2 (5’GCCTCGCATGAGCTCCAGGC-3′), and 11B10R (5’-CTCCCTGACCACGTAGCC-3′). Total DNA of P. infestans was isolated using a Genomic DNA kit (Qiagen, Valencia, CA). DNA was transferred to Hybond N+ (Amersham, Arlington Heights, IL) using the alkaline procedure and Southern hybridizations were performed at 65°C as described elsewhere (Sambrook et al., 1989). Filters were washed at 55°C for 15 minutes in 1x SSC (75 mM NaCl and 7.5 mM sodium citrate)/0.5% sodium dodecyl sulfate (SDS) (wt/vol), and subsequently for another 15 minutes in 0.5x SSC/0.1x SDS. Hybridization signals were detected and analyzed using a Molecular Dynamics Storm 840 phosphor imager. Gel purified DNA fragments containing the pipg1 open reading frame (ORF) were used as probes and radiolabelled with α-32P-dCTP using a random primer labeling kit (Gibco-BRL, Bethesda, MD).
Reverse transcriptase-polymerase chain reaction (RT-PCR) analysis

Total RNA from *P. infestans* and from infected tomato was isolated using the Trizol reagent (Gibco-BRL, Bethesda, MD) according to the manufacturer’s instructions. First strand cDNAs were synthesized from 4.5 μg of total RNA with a universal oligo(dT) primer and the ThermoScript reverse transcriptase from the ThermoScript RT-PCR System (Gibco-BRL, Bethesda, MD). Reactions were carried out according to the manufacturer’s instructions and incubated at 50°C for 1 hour. PCR amplifications were carried out with 1/250th of the cDNA product using the primer pair PIPG-RTF1 (5’-TCCGCACCTTCAGCATTCTC-3’) and PIPG-RTR1 (5’-CGCCCTTGGTCTTGTTGTAATCTG-3’) targeted towards a 450 bp region at the 3’ end of the pipg1 ORF. Integrity of the mRNA and cDNA were controlled with primers EF2-F1 (5’-TGACGCTATCGCGCCAAGGAATC-3’) and EF2-R1 (5’-TAACGCTGAGCGCCAAGGAATC-3’) that are specific for the constitutive elongation factor 2 (*ef2*) gene of *P. infestans* (W.R. Morgan and S. Kamoun, unpublished data).
Phylogenetic analysis

Multiple alignment of 39 (Table 3.1) endoPG amino acid sequences was conducted using the program CLUSTAL-X (Thompson et al., 1997). PAUP v4.0b8 (Swofford, 2001) was used to reconstruct phylogenetic trees of the endoPG family using the neighbor joining method and maximum parsimony with 1000 bootstrap replications.

RESULTS

Molecular analysis of a P. infestans cDNA encoding an endopolygalacturonase

We scanned the P. infestans EST database described by Kamoun et al., Kamoun et al., (1999). for sequences encoding cell wall degrading enzymes. One EST, MY-11-B-10, was identified by similarity searches to show significant similarity to endopolygalacturonases. DNA sequencing of the full cDNA revealed an open reading frame (ORF) of 1125 bp corresponding to a predicted translated product of 374 amino acids. SignalP (Nielsen et al., 1997) analysis of the predicted protein identified a 20-amino acid signal peptide with a significant mean S value of 0.88. Ten potential N-glycosylation sites, corresponding to the consensus sequence NX(S/T), were also identified. Similarity searches
of the predicted protein against the nonredundant database of GenBank using the BLASTP program (Altschul et al., 1997) revealed highly significant matches to fungal endoPGs (E value = $10^{-44}$ for best hit), insect endoPGs (E value = $10^{-31}$), plant endoPGs (E value = $10^{-17}$), and bacterial endoPGs (E value = $10^{-10}$). Searches against the InterPro database (Apweiler et al., 2001) revealed similarity to InterPro domain IPR000743 for polygalacturonase. We then used Clustal X (Thompson et al., 1997) to generate a multiple alignment of the *P. infestans* endoPG with 38 endoPGs from fungi, insects, plants, and bacteria (Fig. 3.1). The length of the *P. infestans* endoPG (374 amino acids) was within the range of known endoPGs (361 to 452 amino acids). Five conserved amino acid blocks defining the endoPG family signature and catalytic sites were conserved and included NTD, DD, GHGXSIGS, RIK and Y (Stratilova et al., 1996; Stratilová et al., 1993). The only substitution in these motifs that was unique to the *P. infestans* sequence was a serine in the first residue of the GHGXSIGS motif. However, histidine, the key catalytic residue in this motif (Rao et al., 1996; Stratilová et al., 1993), remains conserved. Based on these results, we propose that the analyzed *P. infestans* cDNA is likely to encode a functional endoPG and
we designated the cDNA *pipg1*. The identified signal peptide and glycosylation sites suggest that PIPG1 is likely to be a secreted glycoprotein.

**Phylogenetic analyses**

To further investigate the relationship between PIPG1 and endoPGs, we reconstructed the phylogeny of the endoPG family using both the neighbor-joining distance matrix method and maximum parsimony (Fig. 3.2). A total of 1,000 bootstrap replications were conducted with each method to determine the statistical significance of the obtained branches. The phylogenetic trees generated were overall similar. When considering bootstrap values higher than 85%, three main branches grouping the bacterial, plant, and fungal/insect endoPGs were observed. PIPG1 clustered outside these main branches suggesting that *Phytophthora* endoPGs may form a new class of these enzymes. Interestingly, PIPG1 was significantly more closely related to the fungal/insect clade than to the plant or bacterial clades.

**Pipg1 is a member of a multigene family**

In order to examine the occurrence of *pipg1*-like sequences in the *P. infestans* genome, we performed a low stringency Southern blot hybridization of total DNA
isolated from European and US isolates of *P. infestans* with a probe from the *pipgl* cDNA. All isolates examined showed a similar pattern of at least eight hybridizing fragments Fig. 3.3, suggesting the presence of multiple endoPG-like sequences in *P. infestans*. We then hybridized the *pipgl* probe to a genomic library of *P. infestans*. Preliminary restriction fragment and sequence analyses of hybridizing clones confirmed that the endoPG gene family of *P. infestans* is complex. (T. Torto and S. Kamoun, unpublished data).

**Expression of *pipgl* during preinfection and infection stages of *P. infestans***

The *pipgl* cDNA was identified from a library constructed from mRNA isolated from mycelium grown in synthetic medium (Kamoun *et al.*, 1999). To determine whether *pipgl* is expressed under conditions relevant to the interaction with plants, we used RT-PCR analysis to detect *pipgl* mRNA in a preinfection stage (germinating cysts) and in a time course of *P. infestans* infection of tomato. PCR amplifications were carried out with equal amounts of cDNA from the various stages using primers specific for *pipgl*. Transcripts for *pipgl* were detected in germinating cysts, in mycelium of both *P. infestans* isolates 88069 and 90128,
in tomato leaves infected with 88069 2-5 days post inoculation, and in tomato leaves infected with 90128 1-5 days post inoculation. No pipg1 mRNA could be detected in 88069 infected leaves 1 day post-inoculation, in uninfected tomato leaves, and in a water control. Control RT-PCR amplifications were conducted on the same cDNAs with primers specific for the constitutively expressed elongation factor 2 (ef2) gene.

**DISCUSSION**

EndoPGs have been described from bacteria, plants, insects and numerous species of phytopathogenic fungi. In this study, we describe and characterize the first endoPG gene from oomycetes, a unique group of eukaryotic plant pathogens that exhibit fungal-like filamentous growth but share little taxonomic affinity to fungi. The characterized gene, pipg1, was identified from the potato late-blight pathogen *P. infestans* and was predicted to encode a secreted glycoprotein with all the signature sequences of endoPGs. The recent taxonomic positioning of oomycetes as stramenopiles suggest a certain level of evolutionary relationship to brown algae and possibly to plants (Baldauf et al., 2000). Based on systematic analyses of cDNA
sequences, Kamoun et al. (1999) concluded that *P. infestans* genes are not particularly related to fungal sequences and a bias towards plant sequences was apparent. This has been further confirmed using a series of phylogenetic analyses of *Phytophthora* genes encoding metabolic enzymes and other conserved proteins (A. LaLumia, W.R. Morgan and S. Kamoun, unpublished results). Therefore, the affinity observed here between PIPG1 and fungal endoPGs is unexpected and contrasts with phylogenies obtained using ribosomal sequences or compiled protein sequences from mitochondrial and chromosomal genes (Baldauf et al., 2000; Kumar and Rzhetsky 1996; Lang et al., 1999; Paquin et al., 1997; Van de Peer and De Wachter 1997). There are several possible explanations for the exceptional phylogeny of PIPG1. For example, the observed affinity to fungal endoPGs may reflect convergent evolution through which phylogenetically distinct enzymes have evolved to share significant similarity. This is supported by the fact that both plant pathogenic fungi and oomycetes occupy comparable ecological niches and their endoPGs may target similar substrates, host pectins, and face similar selective forces. Alternatively, the endoPG phylogenetic tree could reflect horizontal gene transfer events, in which plant pathogenic
oomycetes acquired endoPG genes from fungi. An exhaustive survey and phylogenetic analysis of endoPG sequences from plant-pathogenic, animal pathogenic, saprophytic oomycetes and other stramenopile species should help test the horizontal gene transfer hypothesis and distinguish between these two explanations. Multigene families of endoPGs have been described in both filamentous fungi and plants (Frassinet-Tachet et al., 1995; ten Have, 2001; Torki et al., 2000; Wubben et al., 1999). Some factors that could account for multiple endoPGs are substrate specificity and coevolution with plant PGIPs. In this study, experiments involving DNA hybridizations and suggest that multiple pipg1-like sequences occur in the P. infestans genome and that endoPGs form a gene family in P. infestans Further experimentation will be required to determine the role of the pipg multigene family in Phytophthora.

In a recent study, ten Have et al., (2001) demonstrated that the endoPG genes of Botrytis cinerea are differentially expressed in various hosts and stages of infection. The emerging view is that endoPGs are not always expressed in a particular host plant and are likely to also contribute to saprophytic survival (ten Have 2001; ten Have et al., 2001). To determine whether pipg1 is expressed
during host infection, expression studies were conducted using RT-PCR and two *P. infestans* isolates, 88069 and 90128, that infect tomato plants at different rates. These experiments showed that *pipg1* is expressed during both preinfection and infection stages and thus could play a role in the penetration and invasion of plant tissue by *P. infestans*. Even though RT-PCR is at best a semi-quantitative assay, the levels of *ef2* mRNA detected appeared to correlate with the expected differences in infection rate between the two *P. infestans* isolates tested. Messenger RNAs for *ef2* and *pipg1* were detected earlier in the interaction between 90128 and tomato than in the 88069 interaction, consistent with the differences in symptom development between the two interactions. Thus, the *pipg1* expression pattern appeared to resemble that of the constitutive gene *ef2* and *pipg1* was expressed in both types of interaction.

*Phytophthora* spp. and other oomycetes have been reported to secrete a variety of extracellular enzymes, including plant cell wall-degrading enzymes, such as endoPGs (Jarvis et al., 1981; McIntyre and Hankin, 1978; Moreau and Seibles, 1985). In this study, we initiated the molecular genetic characterization of plant cell wall-
degrading enzymes in Phytophthora through the identification and characterization of pipg1, the first oomycete cDNA encoding an endoPG. Future studies on the endoPG gene family of P. infestans should provide additional insights into the mechanisms of pathogenicity of this important and distinct group of plant pathogens and should improve our general understanding of endoPGs in interactions between plants and microbes.
Acknowledgements

I am grateful to Laura Rauser for her earlier contribution to this study. I also thank Shujing Dong and Melissa Barty and for technical assistance, Arjen ten Have and Jan van Kan for useful discussions as well as Steve Whisson and Paul Birch for *P. infestans* strain T30-4. I thank Tea Meulia and the staff of the OARDC Molecular and Cellular Imaging Center for help with DNA sequencing and use of the phosphor imager.

This work was supported by the OARDC Research Enhancement Grant Program. Salaries and research support were provided by State and Federal Funds appropriated to the Ohio Agricultural Research and Development Center, the Ohio State University.

A modified version of this chapter is published as:
REFERENCES


Figure 3.1. Multiple alignment of the deduced endopolygalacturonase amino acid sequences of *Aspergillus niger* (CAA74744), *Botrytis cinerea* (AAC64374), *Kluyveromyces marxianus* (CAA03900), *Phaedon cochleariae* (CAA76930), *Sitophilus oryzae* (AAG35693), *Fusarium oxysporum* (BAA20555), *Phytophthora infestans* (AY052571), *Arabidopsis thaliana* (CAA20037), *Lycopersicon esculentum* (AAD17250), *Brassica napus* (CAA65702) and *Nicotiana tabacum* (S32008). Identical amino acids are shaded in dark gray and similar amino acids shaded in light gray. The conserved endoPG family signatures (I) NTD, (II) DD, (III) GHGXSIGS, (IV) RIK and (V) are underlined.
Figure 3.2. Phylogenetic analysis of 39 representative endopolygalacturonases from fungi, plants, insects, bacteria and the oomycete Phytophthora infestans. The tree was constructed using the neighbor-joining method in CLUSTAL-X. Values at the nodes represent percentage occurrence obtained after bootstrap analysis (1000 iterations) of the phylogenetic analysis. The genus name and the GenBank accession number is given for all the endopolygalacturonases represented on the tree. The species represented on the tree as well as the amino acid length of the different endopolygalacturonases are shown in Table 3.1.
Figure 3.2.
Table 3.1. List of endopolygalacturonases from different taxonomic groups used in the phylogenetic analysis.
<table>
<thead>
<tr>
<th>Group</th>
<th>Species</th>
<th>GenBank Accession number</th>
<th>Length (amino acids)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plants</td>
<td>Arabidopsis thaliana</td>
<td>CAA20037</td>
<td>374</td>
</tr>
<tr>
<td></td>
<td>Arabidopsis thaliana</td>
<td>AAC04907</td>
<td>664</td>
</tr>
<tr>
<td></td>
<td>Prunus persica</td>
<td>S71523</td>
<td>458</td>
</tr>
<tr>
<td></td>
<td>Brassica napus</td>
<td>CAA65702</td>
<td>433</td>
</tr>
<tr>
<td></td>
<td>Lycopersicon esculentum</td>
<td>AAD17250</td>
<td>367</td>
</tr>
<tr>
<td></td>
<td>Medicago sativum</td>
<td>CAA72003</td>
<td>395</td>
</tr>
<tr>
<td></td>
<td>Cucumis melo</td>
<td>AAC26511</td>
<td>423</td>
</tr>
<tr>
<td></td>
<td>Nicotiana tabacum</td>
<td>S32008</td>
<td>385</td>
</tr>
<tr>
<td></td>
<td>Zea mays</td>
<td>CAA40851</td>
<td>408</td>
</tr>
<tr>
<td></td>
<td>Lycopersicon esculentum</td>
<td>AAC28905</td>
<td>387</td>
</tr>
<tr>
<td>Fungi</td>
<td>Kluyveromyces marxianus</td>
<td>CAA03900</td>
<td>361</td>
</tr>
<tr>
<td></td>
<td>Chondrostereum purpureum</td>
<td>BAA08102</td>
<td>403</td>
</tr>
<tr>
<td></td>
<td>Botrytis cinerea</td>
<td>AAC24956</td>
<td>371</td>
</tr>
<tr>
<td></td>
<td>Fusarium oxysporium</td>
<td>BAA20555</td>
<td>371</td>
</tr>
<tr>
<td></td>
<td>Gibberella fujikuroi</td>
<td>AAA74586</td>
<td>373</td>
</tr>
<tr>
<td></td>
<td>Aspergillus niger</td>
<td>CAA74744</td>
<td>378</td>
</tr>
<tr>
<td></td>
<td>Botrytis cinerea</td>
<td>AAC24953</td>
<td>397</td>
</tr>
<tr>
<td></td>
<td>Sclerotinia sclerotiorum</td>
<td>CAA74019</td>
<td>387</td>
</tr>
<tr>
<td></td>
<td>Botrytis cinerea</td>
<td>AAC24955</td>
<td>380</td>
</tr>
<tr>
<td></td>
<td>Botrytis cinerea</td>
<td>AAC24950</td>
<td>374</td>
</tr>
<tr>
<td></td>
<td>Botrytis cinerea</td>
<td>AAC64374</td>
<td>382</td>
</tr>
<tr>
<td></td>
<td>Sclerotinia sclerotiorum</td>
<td>AAC14466</td>
<td>380</td>
</tr>
<tr>
<td></td>
<td>Cryphonectria parasitica</td>
<td>AAB36616</td>
<td>369</td>
</tr>
<tr>
<td></td>
<td>Fusarium oxysporum</td>
<td>AAC27792</td>
<td>361</td>
</tr>
<tr>
<td></td>
<td>Colletotrichum</td>
<td>Q00446</td>
<td>363</td>
</tr>
<tr>
<td></td>
<td>lindemuthianum</td>
<td>CAA64727</td>
<td>365</td>
</tr>
<tr>
<td>Insects</td>
<td>Phaedon cochleariae</td>
<td>CAA76930</td>
<td>367</td>
</tr>
<tr>
<td></td>
<td>Sitophillus oryzae</td>
<td>AAG35693</td>
<td>365</td>
</tr>
<tr>
<td></td>
<td>Penicillium griseoroseum</td>
<td>AAC83692</td>
<td>376</td>
</tr>
<tr>
<td>Oomycetes</td>
<td>Phytophthora infestans</td>
<td>AYO52571</td>
<td>374</td>
</tr>
<tr>
<td>Eubacteria</td>
<td>Erwinia carotovorum</td>
<td>AAA57139</td>
<td>402</td>
</tr>
<tr>
<td></td>
<td>Burkholderia cepacia</td>
<td>AAB4698</td>
<td>458</td>
</tr>
<tr>
<td></td>
<td>Erwinia carotovorum</td>
<td>P26509</td>
<td>402</td>
</tr>
<tr>
<td></td>
<td>Rhizobium viti</td>
<td>AAB51699</td>
<td>549</td>
</tr>
<tr>
<td></td>
<td>Bradyrhizobium japonicum</td>
<td>AAG60962</td>
<td>510</td>
</tr>
<tr>
<td></td>
<td>Thermotoga maritima</td>
<td>H72376</td>
<td>448</td>
</tr>
<tr>
<td></td>
<td>Pseudomonas solanacearum</td>
<td>A44508</td>
<td>529</td>
</tr>
<tr>
<td></td>
<td>Clostridium acetobutylicum</td>
<td>AAK78335</td>
<td>513</td>
</tr>
</tbody>
</table>

**Table 3.1.**
**Figure 3.3.** Presence of *pipg*-like sequences in *Phytophthora infestans* genome. Total DNA (5 μg) from Ohio strains, OH 3, OH 10, OH 23 (lanes 1-3) and European strains 88069, 90128, T30-4 (lanes 4-6) of *P. infestans* were digested with HindIII and hybridized with a probe containing the full ORF of the pipg cDNA. The approximate sizes of the fragments are indicated on the left.
Figure 3.4. Expression of pipg1 in various developmental, and infection stages of Phytophthora infestans and tomato. Total RNA from P. infestans (88069) mycelium (1), germinating cyst (2) uninfected tomato (3), infected leaves of tomato 1 (4), 2 (5), 3 (6), 4 (7) and 5 (8), days after inoculation, P. infestans (90128) mycelium (9), uninfected tomato (10), infected leaves of tomato 1 (11), 2 (12), 3 (13), 4 (14) 5 (15) days after inoculation and a water control were used for RT-PCR analysis as described in text. Amplification of the P. infestans elongation factor (ef2) was used as a control to determine the integrity of the cDNA.
CHAPTER 4

Analysis of expressed sequence tags from the oomycete fish pathogen Saprolegnia parasitica

ABSTRACT

Saprolegnia parasitica is an oomycete pathogen of great economic importance in the fishery industry. Currently there is a need to develop effective chemicals to control this pathogen. Expressed sequence tags (ESTs) represent a relatively inexpensive approach of discovering novel genes that can serve as targets for chemical control. In order to create a resource base for gene discovery in S. parasitica we constructed a cDNA library and sequenced 2296 cDNAs corresponding to a set of 1338 unique sequences. The EST dataset was fully annotated based on similarity searches to public databases, and by Gene Ontology and motif designations. 44% of the total ESTs had matches to public databases. Some of the most notable genes identified include an endo-1,3-ß-glucanase, serine and cysteine proteases, and several genes with similarity to the
Phytophthora cellulose binding and elicitor protein (CBEL). Phylogentic analyses of the *S. parasitica* endo-1,3-β-glucanase indicated affinity to fungal homologs and may support the recent speculation of convergent evolution between fungal and oomycete hydrolytic enzymes.

**INTRODUCTION**

*Saprolegnia* is the main genus of water molds responsible for significant infections on fish in aquaculture facilities and freshwater bodies (Neish and Hughes, 1980; Willoughby and Pickering, 1977). Members of the genus *Saprolegnia* cause the disease Saprolegniasis. Saprolegniasis is characterized by visible white or gray patches of filamentous mycelium on the head or fins of freshwater fish (Neish and Hughes, 1980). *Saprolegnia parasitica* is an important pathogen on various types of fish including salmon and causes millions of dollars loss to aquaculture business worldwide, notably Japan (Hatai and Hoshiai, 1992) and the USA (Bly et al., 1994).

*S. parasitica* is a member of the oomycetes, a group of unique eukaryotic microbes that are phylogenetically unrelated to the fungi (Bartnicki-Garcia and Wang, 1983; Sogin and Silberman, 1998). Owing to their distinct
taxonomic differences, the oomycetes may possess unique biochemical characteristics (Bartnicki-Garcia and Wang, 1983; Griffith et al., 1992) which render them insensitive to many fungicides. Malachite green was the effective choice of chemical treatment of *S. parasitica* until it was banned because of its teratogenic properties (Clemensen et al., 1984). In light of this, there is an urgent need for new chemical agents targeted against essential components of *S. parasitica* to prevent it from proliferation and causing diseases.

The life cycle of *S. parasitica* is typical of most oomycete pathogens except that in most genera within the Saprolegniales, two morphologically distinct zoospore and cysts (primary and secondary) are observed [Holloway, 1977 #685; Beakes, 1983 #686].

The molecular basis of pathogenesis in *S. parasitica* is not understood. To date only a few *S. parasitica* sequences are deposited in GenBank, corresponding to genes encoding ribosomal RNA (rRNA), a beta-amylase cDNA, three unknown transcripts, and the putative RNA binding protein *puf1* (Andersson and Cerenius, 2002). The *puf1* transcript purported to be a post-transcriptional regulator was found using the differential-display reverse transcription-PCR.
(ddRT-PCR). Even though this allows the identification of differentially expressed fragments from spore and mycelium life stages, transcripts corresponding to the fragments have to be identified from cDNA libraries and the differential expression verified with Northern blots. Alternatively, a myriad of informative transcripts, including stage specific transcripts can be generated from cDNA libraries. Expressed sequence tags (ESTs) thus constitute a rapid and informative strategy for gene discovery and for studying gene profiles in economically important pathogens.

Projects that generated EST datasets for the oomycete plant pathogens, *Phytophthora infestans* (Kamoun et al., 1999) and *Phytophthora sojae* (Qutob et al., 2000), identified a number of candidate genes involved in virulence, pathogenicity and various biochemical and physiological processes. Generating and sequencing ESTs from a cDNA library of *S. parasitica* would be useful in identifying genes involved in virulence which may serve as possible targets for the design of novel chemical agents. Additionally, sequences from the library will serve as a resource for comparative analysis among the oomycetes.
In this study, we report the construction of a cDNA library prepared from \textit{S. parasitica} and provide some preliminary analysis of random cDNA sequences from this library.

**MATERIALS AND METHODS**

**Strains and growth conditions**

\textit{Saprolegnia parasitica} ATCC90214 was used in this study. This is an isolate from lesions on coho salmon (\textit{Oncorhynchus kisutch}) (Hatai \textit{et al.}, 1990). Working stocks of this strain were routinely maintained on cornmeal agar (Difco Lab. Detroit, MI) prepared according to the manufacturer’s instructions. To obtain axenically prepared mycelium, ATCC90214 was grown in GY broth (5 g glucose, 2.5g yeast extract/L) (Griffin, 1978) for 29 days at which time the broth was completely used up to mimic a stressed situation. Cultures were routinely incubated at $18^\circ$C.

**cDNA construction**

Total RNA from \textit{S. parasitica} mycelium was isolated using a 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),
as stipulated by the manufacturers. The cDNA library was prepared in plasmid pSPORT1 using the Superscript™ plasmid system for cDNA synthesis and cloning (Invitrogen Life Technologies, Carlsbad, CA) according to the manufacturer's instructions. Polyadenylated mRNA was used to synthesize oligo (dT) primed cDNAs, which were cloned unidirectionally in NotI/SalI digested vector pSPORT1. Plasmid ligations were transformed into *Escherichia coli* ElectroMax-DH10B™ cells (Invitrogen Life Technologies, Carlsbad, CA). Selection was done on Luria-Bertani (LB) agar plates containing kanamycin (25mg/L) (Sambrook and Russel, 2001). 2304 individual colonies were picked randomly with the Qpix robot (Genetix, Hampshire, UK) into LB freeze media in a 384 well plate. 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 for sequencing. Identification codes for the EST sequences 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

DNA from bacterial cultures was purified at GTSF using Qiagen 3000 or Autogen 850 robots. Fluorescently labeled sequencing products were generated by PCR amplification 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. A dataset representing 2296 EST sequences and the corresponding electropherograms were then made available through the Geospiza Finch web interface.

Phylogenetic analysis

Plasmid containing the full insert of a representative of CON_001_00550 was sequenced by a primer walking approach with primers SPM16A2-F1: 5’-CAACCAGACGCTCGGCCTCC-3’ and SPM16A2-F2: 5’- CGACTTTTACTTCCTCGCC-3’. Sequences of endo-1,3-ß-glucanase from other organisms were obtained from GenBank and are shown in Table 4.3 with their accession numbers. Multiple alignment of 14 endo-1,3-ß-glucanase from plants, fungi and the oomycetes were performed using CLUSTAL-X program (Thompson et al., 1997). The phylogeny of the endo-1,3-ß-glucanases was reconstructed using the neighbor joining method with 1000 bootstrap replications.
RESULTS

Analysis of EST sequences

A directional cDNA library was constructed from in vitro grown mycelium. Quality analysis (phred scores) and base calling of the DNA sequences were done at the GSTF using a Phred algorithm. 2102/2296 sequences passed the quality analysis.

The average number of bases >Q20 per read was 369. The sequences and the quality (phred) scores were downloaded and fed into the X Genome Initiative (XGI) annotation pipeline housed at the National Center for Genome Research. After vector and low quality sequences were removed, 1570 (68%) sequences remained. Out of this, only 5.7% were oriented in the reverse direction, this being assessed as sequences with at least eight consecutive As (AAAAA) within the first thirty-eight bases. For further quality screening, an initial window of 60 nucleotides with an average quality of Q=15 was selected, This was followed by a 20 nucleotide sliding window with a minimum average quality over the window of Q=15. Subsequently, a data set of 1,338 sequences with an average insert size of 408bp was generated and used for contig assembly. This dataset included 1206 (90%) singletons and 132 (10%) consensus with
more than one EST (Fig. 4.1). The consensus was generated using Phrap and the criterion was a minscore (represents quality scores of the overlapping sequence) of 300 and a minmatch (minimum length of the overlap between two sequences within a cluster) of 75. Similarity and motif searches were done for each sequence in the dataset. Where available these sequences were further annotated with a Gene Ontology identity. The annotated sequence dataset is housed in the XGI database.

**Annotation of *S. parasitica* ESTs**

The *S. parasitica* mycelium EST dataset was annotated by three different methods. Sequences were automatically queried against the NCBI nonredundant protein database using the BLASTX algorithm (Altschul et al., 1990). Sequence similarities with an E value < $10^{-5}$ were considered statistically significant. From this analysis, 750 (54%) had no hits or known homology to sequences in the public databases whereas 588 (44%) had homology (Fig. 4.2). Another form of classification was based on using the Gene Ontology classification. Here sequences were grouped based on the process they are involved in, the molecular function, and the cellular components, as defined by the Gene Ontology (GO) Consortium (Ashburner, 2000). Here 576
(43%) did not match with any GO identities while 762 (57%) were assigned GO identities. These results overlapped with results from the similarity search with BLAST, even though 12 more sequences were annotated with GO.

**Identification of most abundant cDNA clones**

The most abundant cDNAs represent the consensus sequences with ESTs ranging from 4-15 (Fig. 4.1, Table 4.1). The most abundant sequence, CON_015_00140 had 15 ESTs in its consensus and it formed only 1.1% of the total number of sequences. This consensus did not match any sequence in public databases. The next abundant sets of sequences comprised a cluster of six in each set. These sets each represented 0.4% of the total sequences. Similarity searches identified one group CON_006_00135 as actin, CON_006_00136 as polyubiquitin, CON_006_00137 matched a putative initiation factor 5A-4, CON_006_00139 matched an unknown sequence whereas CON_006_00138 did not match any sequence in the public database. Six different consensus of 5 ESTs were identified, each representing 0.4% of total sequences. Of these, two represented a putative 60S ribosomal protein and a L28-like ribosomal protein, while another one represented actin, three consensus had no matches to public databases. The last group of abundant
sequences comprised 4 ESTs in a consensus group each representing 0.3%. In brief, sequences in this category showed matches to several genes (Table 4.1).

**Phylogenetic analysis of Endo-1,3-β-glucanases**

The phylogeny of the endo-1,3-β-glucanases was reconstructed from 15 different organisms representing fungi, plants and the oomycetes (Table 4.2). Three distinct clades were identified corresponding to plant endoglucanases, fungal/oomycete endoglucanases and yeast endoglucanases (Fig 4.3).

**Notable transcripts**

Selected transcripts of interest based on annotation or comparison to *Phytophthora* sequences are shown in Table 4.3 and were analyzed in more details. These include, serine and cysteine proteases, several sequences with similarity to the *Phytophthora* cellulose binding and elicitor protein (CBEL), and an EST with similarity to a thiamine biosynthetic enzyme.

**DISCUSSION**

**Analysis of EST sequences**

In this study we describe the first EST dataset from *S. parasitica*, an economically important oomycete which
causes huge losses in revenue in the aquaculture industry. The ESTs were generated from a cDNA library constructed from a mycelium grown in liquid medium. The annotated EST dataset consists of 1338 sequences representing a high percentage (90.1%) of singletons. The number of singletons was relatively higher than observed for other published oomycete EST datasets including the ESTs generated from *P. infestans* mycelium (Kamoun et al., 1999) and *P. sojae* zoospore and mycelium (Qutob et al., 2002). The most abundant cluster of ESTs was present only at a frequency of 1.1%. This low frequency of redundancy is likely due to the parameters used for the cluster analysis.

Similarity searches revealed that 44% of the total EST sequences had matches to protein sequences in the public databases (Fig. 4.2.) This high percentage of sequences with no match, often termed orphan sequences, is recognized in large scale and EST sequencing of various organisms. However, the percentage of sequences with no match to public databases is quite high compared to the 33% observed for *P. sojae* mycelium ESTs (Qutob et al., 2000) and 37% in *P. infestans* mycelium ESTs (Kamoun et al., 1999).

Abundant cDNAs found in the *S. parasitica* library included transcripts involved in activities of the cell.
cycle, protein biosynthesis, electron transfer, metabolism, heat shock response and proteolysis. While the first processes may represent general morphological development and physiological states of the pathogen, heat shock responses and proteolysis may represent sequences involved in pathogen response to a stressed state and pathogenesis respectively.

**Phylogenetic analysis of Endo-1,3-β-glucanase**

Extensive studies of endo-1,3-β-glucanases in fungi and plants have shown them to be involved in development and signaling (Hrmova and Fincher, 2001; Pitson et al., 1993). Recently, the cloning and characterization of the first oomycete endo-1,3-β-glucanase Piendo1 was reported (McLeod et al., 2002). According to the authors, Piendo1 was more related to fungal endo-1,3-β-glucanases than to plants. A similar relational pattern to fungus was obtained with CON_001_00550, a S. parasitica homolog of endo-1,3-β-glucanase identified in this study. This phylogeny signifies a deviation from the expected norm obtained using ribosomal sequences or compiled protein sequences from mitochondrial genes (Baldauf et al., 2000; Paquin et al., 1997). The phylogenetic affinity of the endo-1,3-β-glucanase of S. parasitica to fungal endo-1,3-β-glucanases
are likely to reflect convergent evolution of enzymes acting on the same substrate or with the same function. Similar reports have been made from phylogenetic studies on endopolygalacturonases from *P. infestans* and *P. cinnamomi* (Gotesson *et al.*, 2002; Torto *et al.*, 2002). Alternatively, a horizontal gene transfer event may have taken place between a common ancestor of *Phytophthora*, *Saprolegnia* and fungi.

**Proteolytic Enzymes**

Various types of proteases were found in the annotated dataset. These included vacuolar, sulfhydryl, serine and cysteine proteases. The serine proteases include chymotrypsin. Early in the 1970s, Peduzzi and Bizzozero, (1977) showed that the mycelium of certain *Saprolegnia* species pathogenic on fish exhibited chymotrypsin activity and speculated that this enzymatic activity contributed to pathogenesis. Furthermore, a serine protease, AaSP2 from a related crayfish pathogen, *Aphanomyces astacus* was recently cloned and characterized (Bangyeekhun *et al.*, 2001). The authors showed that AaSP2 was expressed at high levels during growth in crayfish plasma membrane. With these genes cloned, it will be easier to study their expression during the interaction with the host.
Cellulose Binding, Elicitor and Lectin (CBEL)

Several consensus sequences matched the Cellulose Binding, Elicitor and Lectin-like protein (CBEL), a 34-kDa glycoprotein localized to the cell wall of Phytophthora parasitica. The P. parasitica CBEL binds to cellulose, elicits necrosis and defense gene expression in tobacco and functions in the agglutination of red blood cells (Mateos et al., 1997). It remains to be seen what role each of these domains play in the S. parasitica homologs during interaction with the fish host considering that the host does not possess cellulose and progressive penetration of the epidermal tissue by hyphae causes hemorrhage.

Thiamine Biosynthesis

Among the oomycetes, members of the genus Phytophthora require exogenous sources of thiamine (Leonian and Lilly, 1938). However, members of the genus Pythium can synthesize thiamine and do not require an exogenous source for growth. The presence of a thiamine biosynthetic enzyme CON_001_00275 is an indication that S. parasitica, like members of the genus Pythium, may synthesize thiamine de novo for their growth.
Elicitins

Elicitins are a group of 10-kDa proteins found in all *Phytophthora* and some *Pythium* species (Huet et al., 1995; Kamoun et al., 1993; Pernollet et al., 1993). Interestingly none was found in the *S. parasitica* dataset. Elicitins have been reported to function as sterol carriers (Mikes et al., 1998; Mikes et al., 1997). Studies over the years have shown that among the oomycetes members of the Saprolegniales (example *S. parasitica*), Leptomitales and Lagenidiales are able to synthesize sterols de novo whereas members of the Pythiaceae (*Phytophthora* and *Pythium*) lack the ability to synthesize sterols. Possibly, *S. parasitica* may not need sterol carriers to obtain sterols exogenously for proper growth. This finding however needs to be supported with further studies including hybridization.

The scope of this work is limited to describing putative functions of selected transcripts, however, molecular and cellular characterization of the transcripts described as well as other candidates from the *S. parasitica* dataset especially those involved in development, physiology and pathogenesis should help in defining essential chemical targets for effective control of *S. parasitica*. Furthermore, additional sequencing to
generate a large unigene set will be useful for comparative studies among the oomycetes.

ACKNOWLEDGEMENTS

I am grateful to Annette Thelen and her team at the Genomics Technology Support Facility (GTSF) at the Michigan State University for sequencing. I also thank Mark Waugh and John Sullivan at the NCGR for automated analysis and annotation of the EST dataset. Thanks to Miaoying Tian, Shujing Dong and Diane Kinney for technical assistance.

REFERENCES


Figure 4.1. Graph showing percentage frequency of redundant ESTS. Contigs formed with 1 sequence are singletons and contigs comprising 2 or more sequences in a consensus are redundant.
Figure 4.2. Pie chart representing output from similarity searches against public databases using the BLAST algorithm. White sector represents sequences with matches (E value < 10^{-5}) to the public databases and the gray shaded portion represents sequences with no significant matches (E value > 10^{-5})
<table>
<thead>
<tr>
<th>XGI Identity</th>
<th>Gene Ontology Identity</th>
<th>BLASTX Best Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>ON_015_00140</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ON_006_00138</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ON_006_00135</td>
<td>GO:0007049-cellcycle</td>
<td>actin</td>
</tr>
<tr>
<td>ON_006_00139</td>
<td>-</td>
<td>unknown</td>
</tr>
<tr>
<td>ON_006_00136</td>
<td>GO:0007046-proteinbiosynthesis</td>
<td>Polyubiquitin</td>
</tr>
<tr>
<td>ON_006_00137</td>
<td>GO:0006412-ribosomebiogenesis</td>
<td>Putative initiation factor 5A-4</td>
</tr>
<tr>
<td>ON_005_00130</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ON_005_00134</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ON_005_00132</td>
<td>GO:0007049-cellcycle</td>
<td>actin</td>
</tr>
<tr>
<td>ON_005_00131</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ON_005_00133</td>
<td>GO:0006412-proteinbiosynthesis</td>
<td>putative 60S ribosomal protein L6</td>
</tr>
<tr>
<td>ON_005_00129</td>
<td>GO:0006412-proteinbiosynthesis</td>
<td>ribosomal protein L28-like</td>
</tr>
<tr>
<td>ON_004_00119</td>
<td>-</td>
<td>pepsinogen</td>
</tr>
<tr>
<td>ON_004_00120</td>
<td>-</td>
<td>unknown</td>
</tr>
<tr>
<td>ON_004_00121</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ON_004_00122</td>
<td>-</td>
<td>Electron transfer flavoprotein</td>
</tr>
<tr>
<td>ON_004_00123</td>
<td>-</td>
<td>ribosome</td>
</tr>
<tr>
<td>ON_004_00124</td>
<td>-</td>
<td>Serine palmitoyltransferase-like protein</td>
</tr>
<tr>
<td>ON_004_00125</td>
<td>-</td>
<td>Expressed sequence C80529</td>
</tr>
<tr>
<td>ON_004_00126</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ON_004_00127</td>
<td>-</td>
<td>Heat shock protein Hsp90</td>
</tr>
<tr>
<td>ON_004_00128</td>
<td>-</td>
<td>Manganese resistance protein</td>
</tr>
</tbody>
</table>

**Table 4.1.** List of most abundant sequences in *S. parasitica* EST dataset. The number of sequences in each consensus is shown within the XGI identity CON-015-00140 as underlined.
<table>
<thead>
<tr>
<th>Group</th>
<th>Species</th>
<th>GenBank Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plants</td>
<td>Glycine max</td>
<td>U41323P</td>
</tr>
<tr>
<td></td>
<td>Solanum tuberosum</td>
<td>U01901</td>
</tr>
<tr>
<td></td>
<td>Hordeum vulgare</td>
<td>M13237</td>
</tr>
<tr>
<td></td>
<td>Triticum aestivum</td>
<td>Z22873</td>
</tr>
<tr>
<td></td>
<td>Oryza sativa</td>
<td>AF323610</td>
</tr>
<tr>
<td></td>
<td>Hordeum vulgare</td>
<td>M96938</td>
</tr>
<tr>
<td>Fungi</td>
<td>Candida albicans</td>
<td>CAB62579</td>
</tr>
<tr>
<td></td>
<td>Saccharomyces cerevisiae</td>
<td>Z72550</td>
</tr>
<tr>
<td></td>
<td>Schizosaccharomyces pombe</td>
<td>CAB57443</td>
</tr>
<tr>
<td></td>
<td>Aspergillus fumigatus</td>
<td>AFO38596</td>
</tr>
<tr>
<td></td>
<td>Candida albicans</td>
<td>U12975</td>
</tr>
<tr>
<td></td>
<td>Saccharomyces cerevisiae</td>
<td>M31072</td>
</tr>
<tr>
<td></td>
<td>Pichia jadinii</td>
<td>AF091241</td>
</tr>
<tr>
<td>Oomycete</td>
<td>Phytophthora infestans</td>
<td>AF494013</td>
</tr>
<tr>
<td></td>
<td>Saprolegnia parasitica</td>
<td>Not yet assigned</td>
</tr>
</tbody>
</table>

**Table 4.2.** List of endo-1,3-ß-glucanases from different taxonomic groups used in the phylogenetic analysis.
<table>
<thead>
<tr>
<th>Consensus identity</th>
<th>Putative function</th>
</tr>
</thead>
<tbody>
<tr>
<td>CON_001_00550</td>
<td>endo-1,3-ß-glucanase</td>
</tr>
<tr>
<td>CON_001_00275</td>
<td>thiamine biosynthetic enzyme</td>
</tr>
<tr>
<td>CON_001_00001</td>
<td>vacuolar protease A</td>
</tr>
<tr>
<td>CON_001_01257</td>
<td>cysteine protease</td>
</tr>
<tr>
<td>CON_001_00168</td>
<td>cysteine proteinase</td>
</tr>
<tr>
<td>CON_001_00168</td>
<td>cysteine proteinase</td>
</tr>
<tr>
<td>CON_002_00078</td>
<td>serine protease-like</td>
</tr>
<tr>
<td>CON_001_01336</td>
<td>serine protease-like</td>
</tr>
<tr>
<td>CON_002_00059</td>
<td>sulphydryl proteinase</td>
</tr>
<tr>
<td>CON_001_00003</td>
<td>CBEL</td>
</tr>
<tr>
<td>CON_001_00006</td>
<td>CBEL</td>
</tr>
<tr>
<td>CON_001_00478</td>
<td>CBEL</td>
</tr>
<tr>
<td>CON_001_01231</td>
<td>CBEL</td>
</tr>
<tr>
<td>CON_002_00072</td>
<td>CBEL</td>
</tr>
<tr>
<td>CON_003_00090</td>
<td>CBEL</td>
</tr>
<tr>
<td>CON_003_00091</td>
<td>CBEL</td>
</tr>
<tr>
<td>CON_003_00099</td>
<td>CBEL</td>
</tr>
</tbody>
</table>

**Table 4.3.** Notable transcripts in *Saprolegnia parasitica* EST dataset
Figure 4.3. Phylogenetic reconstruction of the endoglucanase family from plants, fungi and the oomycetes. The phylogenetic tree was constructed using the neighbor joining method on multiple alignments of amino acid sequences of the ORFs. Bootstrap values above 50% from 1000 replications are indicated at the nodes.
GENERAL CONCLUSIONS

The first part of the research (Chapter 2) focuses on the functional genomics of extracellular proteins of *Phytophthora infestans*. This study describes a general approach of identifying novel effectors from EST databases using a combination of data mining for extracellular proteins and high throughput functional expression assays in plants. We developed an algorithm (PexFinder) for the automated identification of extracellular proteins of *Phytophthora*. There were some limitations to the algorithm inherent to the SignalP V2.0 program that it was based on. For example six of the sequences selected by the algorithm were mitochondrial proteins indicating that the SignalP program does not specifically select genes encoding extracellular proteins. However PexFinder was validated convincingly using proteomic analysis of secreted proteins of *P. infestans* (Pieter van West lab). Of the 22 randomly picked secreted proteins from the proteomics study, 9 (41%) matched sequences in the selected set of 142 Pex cDNAs. The set of 142 sequences encoding extracellular
proteins included small extracellular cysteine rich proteins (typical features of most fungal avirulence genes) (Templeton et al., 1994) and cell wall degrading enzymes (implicated in facilitating the penetration of pathogens into the plant) (ten Have, 2001) as well as novel cDNAs.

We employed a PVX-based assay to identify effector proteins in the dataset which can cause disease-like responses (virulence) or induce plant defenses (avirulence) and identified two novel necrosis inducing effectors, crn1 and crn2. These effectors were found to belong to a large gene family. The crn genes were constitutively expressed both in vitro and in planta suggesting that CRN proteins could be secreted and delivered to plant cells during infection. In addition, crn2 inoculated tomato plants were shown to induce defense related responses. The CRNs could be likened to pathogen-associated molecular patterns (PAMPs) which are surface derived molecules that induce the expression of surface response genes and the production of antimicrobial compounds in both animal and plant cells (Nürnberger and Brunner, 2002). This finding however needs to be corroborated with further studies. In the future, functional studies of the crn genes in P. infestans should be conducted to determine the nature of these genes in the
infection process. In addition, studies directed at plant responses to crn and the identification of plant mutants deficient of crn responses should help dissect defense pathways underlying the crn induced responses in plants.

In the second study (Chapter 3), we described the characterization of the first endopolygalacturonase (endoPGs) cDNA pipg1 from P. infestans. We showed that pipg1 displays the key signatures of endoPGs identified in plants and fungi. Pipg1 clustered with the fungal endoPGs contrasting with phylogenies obtained using ribosomal sequences or compiled sequences from mitochondrial and chromosomal genes (Baldauf et al., 2000; Paquin et al., 1997). This may represent convergent evolution through which phylogenetically distinct enzymes have evolved to share significant similarity. Alternatively the evolutionary affinity to fungi may reflect horizontal transfer whereby plant pathogenic oomycetes acquired endoPGs from fungi. Our studies also showed that multiple pipg1 like sequences occurred in the P. infestans genome a feature observed with other endoPGs. Presence of multiple genes may infer different substrate specificities or reflect co-evolution with plant PGIPs. Furthermore we showed pipg1 to be expressed in both the preinfection and
infection stages. Future studies should focus on an exhaustive survey and phylogenetic analysis of endoPG sequences from plant pathogenic, animal pathogenic and saprophytic oomycetes and other stramenopile species to help test the horizontal gene transfer hypothesis. In addition, further studies to understand the mechanisms of pathogenicity should improve our general understanding of endoPGs in interactions between plants and microbes.

The last study (Chapter 4) describes a cDNA library constructed from mycelium of the fish pathogenic oomycete *Saprolegnia parasitica*. Of the 1,338 sequences that passed quality control checks, 44% of the were assigned putative functions (E < 10^{-5}) based on similarity searches to public databases. On the other hand 56% had no matches (E > 10^{-5}). Phylogenetic analysis of an endo-1,3-ß-glucanase from *S. parasitica* EST dataset together with other endo-1,3-ß-glucanases from plants, fungi and oomycete showed a closer relation to fungi than to plants. This is similar to the phylogenetic relation observed for other *Phytophthora* degradative enzymes, for example, the endoPGs mentioned above for *P. infestans* (Torto et al., 2002) and also for *P. cinnamomi* (Gotesson et al., 2002). Moreover this result supports the speculation of convergent evolution propounded
for some oomycete degradative enzymes and also indicates that horizontal gene transfer may have occurred between the ancestors of *Phytophthora*, *Saprolegnia* and fungi. In the long term, generating more sequences from the *S. parasitica* cDNA library to ensure a large set of unigenes should be useful for comparative studies among the oomycetes.


Hatai, K., and Hoshiai, G. (1992). Mass mortality in cultured coho salmon (Oncorhynchus kisutch) due to


and tomato in France and Switzerland. Plant Pathol. 51:641-653.


during transition from biotrophy to necrotrophy. Plant J. 32:361-373


