CATCH OF THE DAY: A YEAST ONE-HYBRID ASSAY IDENTIFIES A NOVEL DNA-BINDING DOMAIN IN PHYTOPHTHORA SOJAE

Brian D. M. Rutter

A Thesis
Submitted to the Graduate College of Bowling Green State University in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

August 2012

Committee:
Dr. Vipaporn Phuntumart
Dr. Paul F. Morris
Dr. Raymond A. Larsen
ABSTRACT

Vipaporn Phuntumart, Advisor

*Phytophthora sojae* is a devastating soybean pathogen. Each year, the voracious activities of this small, oomycete pathogen cost the world billions of dollars. While a large body of research has delved into how *P. sojae* infects its host, as well as the anatomy and function of effector proteins it uses to promote infection, little is known about the gene regulation coordinating the oomycete’s growth, development and virulence. Automated and manual genome annotation approaches have identified more than 1,000 different transcription factors belonging to approximately 60 families in the genomes of six oomycete species: *Phytophthora sojae, P. ramorum, P. infestans, Pythium ultimum, Hyaloperonospora arabidopsidis* and *Saprolegnia parasitica.* Few if any of these putative transcription factors have been verified experimentally. In order to amend this lack of information, a yeast one-hybrid assay was performed with the intent of identifying known and novel transcription factors in *P. sojae.* Analysis of the promoter regions of the *P. sojae* genome revealed that approximately 20% of predicted genes in the genome contain the motif “GCCGCC,” which is a common binding domain for transcription factors in plants. Apicomplexan parasites also regulate their complex life cycles using transcription factors capable of binding to this motif. The assay captured heretofore-unidentified gene sequences that encode a potentially novel DNA-binding domain.
DEDICATION
ACKNOWLEDGEMENTS

First and foremost, I would like to thank my mentor, Dr. Vipaporn Phuntumart, for being a constant source of guidance and encouragement, as well as for placing more confidence in me than I probably deserve. I am a better scientist and person for having known her and would contend with a hundred more EMSAs just to make her proud. Thanks to Dr. Paul Morris, whose curiosity is as infectious as *P. sojae*, and Dr. Ray Larsen for his ever-gracious support. I am blessed to have had advisors who worked so earnestly to see me succeed.

I am particularly grateful for my peers, the lab members from both Dr. Phuntumart and Morris’ labs: Jennifer Benson, Mike Ludwig, Maribeth Spangler, Linxiao Ge, Jigar Patel, Eric Budge, Rebecca Cull and Patrick Wolf. I cannot think of any other people with whom I would rather work, commiserate and rejoice. I especially want to acknowledge Jennifer and Maribeth for kindly donating *P. sojae* RNA/DNA samples and restoring my sanity every now and again, Mike for help with PCR and Eric for the merciful task of helping me transfer all those damn yeast colonies every month.

Outside of the lab, I had the prayers and support of my family and friends. They have no idea what I actually do all day, but I could not have accomplished anything without their love. I sincerely hope I have made them all proud.

Support for this research came from the National Institute of Food and Agriculture – United Stated Department of Agriculture (Grant#2011-68004-30104). I would also like to thank the Oomycete Molecular Genetic Network. The collaborative efforts of these scientists made this research possible.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Oomycetes: Ancient, Fungus-like Parasites From the Deep</td>
<td>1</td>
</tr>
<tr>
<td>Phytophthora: The Genus Plants Dread</td>
<td>3</td>
</tr>
<tr>
<td><em>Phytophthora sojae</em>: Pretty Much the Only Exciting Thing Going on in America’s Midwest</td>
<td>7</td>
</tr>
<tr>
<td>Effectors: Oomycete Weapon of Choice</td>
<td>10</td>
</tr>
<tr>
<td>Transcription Factors</td>
<td>13</td>
</tr>
<tr>
<td>Hunting Down Regulators: Methods for Studying TFs</td>
<td>15</td>
</tr>
<tr>
<td>OBJECTIVE</td>
<td>18</td>
</tr>
<tr>
<td>METHODS</td>
<td>20</td>
</tr>
<tr>
<td>Analysis of Genomic Data</td>
<td>20</td>
</tr>
<tr>
<td>Generating the Yeast Reporter Strain</td>
<td>20</td>
</tr>
<tr>
<td>Biological Materials</td>
<td>22</td>
</tr>
<tr>
<td>Yeast and <em>E. coli</em> Stocks</td>
<td>23</td>
</tr>
<tr>
<td>cDNA Synthesis</td>
<td>24</td>
</tr>
<tr>
<td>Initial Screening of the One-Hybrid cDNA Library</td>
<td>25</td>
</tr>
<tr>
<td>Isolation and Sequencing of Prey</td>
<td>25</td>
</tr>
<tr>
<td>Confirmation of Positive Interactions</td>
<td>26</td>
</tr>
<tr>
<td>Characterization of Prey</td>
<td>27</td>
</tr>
<tr>
<td>RESULTS</td>
<td>28</td>
</tr>
<tr>
<td>Computational Analysis</td>
<td>28</td>
</tr>
<tr>
<td>Yeast One-Hybrid System</td>
<td>28</td>
</tr>
<tr>
<td>Amino Acid Sequence Analysis of Prey Proteins</td>
<td>34</td>
</tr>
<tr>
<td><em>In silico</em> Analysis of Homologous Proteins in <em>P. infestans</em> and <em>P. ramorum</em></td>
<td>35</td>
</tr>
</tbody>
</table>
Secondary Structure Prediction ................................................................. 35
Signal Peptide/Subcellular Localization Prediction ...................................... 37

DISCUSSION ................................................................................................. 39

REFERENCES ............................................................................................... 43
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. General Life Cycle of P. sojae</td>
<td>8</td>
</tr>
<tr>
<td>2. “GCCGCC” motif identified by Wordseeker</td>
<td>28</td>
</tr>
<tr>
<td>3. Overview of pBait-AbAi vector and its integration into the Y1HG0ld yeast genome</td>
<td>30</td>
</tr>
<tr>
<td>4. Screening for protein-DNA interactions using SMART cDNA and the prey vector pGADT7-Rec</td>
<td>31</td>
</tr>
<tr>
<td>5. Examples of mycelial and zoospore prey amplicons on 0.8% agarose with EtBr</td>
<td>32</td>
</tr>
<tr>
<td>6. Alignment of deduced amino acid sequences from prey fragments</td>
<td>33</td>
</tr>
<tr>
<td>7. Retransformation with selected prey plasmids M2.06 and Z2.32</td>
<td>34</td>
</tr>
<tr>
<td>8. Alignment of conserved region from P. sojae with homologous regions identified in proteins from P. infestans and P. ramorum</td>
<td>36</td>
</tr>
<tr>
<td>9. Secondary structure prediction for homologous prey regions in P. sojae, P. infestans and P. ramorum proteins</td>
<td>37</td>
</tr>
</tbody>
</table>
INTRODUCTION

Oomycetes: Ancient, Fungus-like Parasites From the Deep

The oomycetes are a diverse, ancient class of fungal-like protists that have existed since the Neoproterozoic Period, 1 – 524 billion years ago (Bhattacharya et al., 2009). The planet Earth is home to 500-800 known species of these small, filamentous creatures, some of which creep through damp soil while others lurk in fresh water and marine environments. They grow primarily as masses of thread-like hypha (mycelium) and have adapted to saprophytic and parasitic lifestyles. The oomycetes’ fungus-like morphology and ecological roles, originally led to their classification as fungi. “Oomycota” means “egg fungus,” and several species are still commonly referred to as molds and mildews (Erwin & Ribiero, 1996).

Subtle differences in morphology and biochemistry, however, set the oomycetes apart from true fungi. These differences include the release of biflagellated, heterokont zoospores, sexual reproduction involving gametangial meiosis, tubular mitochondrial cristae, aseptate mycelium and cell walls built primarily out of cellulose rather than chitin. The oomycetes also possess anabolic pathways for the synthesis of lysine and tryptophan that differ markedly from those found in fungi (reviewed in Hardham et al., 1994; Erwin & Ribiero, 1996).

Similarities between the oomycetes and the fungi are not due to any phylogenetic relationship. On the contrary, the oomycetes’ hyphal growth and often saprotrophic lifestyles are likely an example of convergent evolution (Money et al., 2004). These fungus-like characteristics may also have arisen due to multiple horizontal gene transfers.
between fungi and early ancestors of the oomycetes (Richards et al., 2006). In any event, the true ancestry of the oomycetes lies elsewhere. Modern phylogenetic tools have shown that oomycete rRNA subunits group more readily with sequences belonging to chromalveolates than with fungi (Van der Auwera et al., 1994; Saunders et al., 1995; Ben Ali et al., 2001, 2002; Lara & Belbahri, 2011). The same is true of sequences encoding oomycete cytoplasmic proteins, such as heat shock proteins, elongation factors, and enzymes involved in the synthesis of actin and tubulin (Harper et al., 2005).

Chromalveolata is a supergroup, the beginnings of which stretch back 1,300 million years ago to a secondary endosymbiotic event between a common ancestor and a red algae plastid (Cavalier-Smith, 2002; Yoon et al., 2002). This supergroup is composed of a diverse number of algae and protists divided into four subgroups: the Cryptophyceae, Haptophyta, Alveolata and Stramenopiles. The oomycetes fall into the latter group, making its home among brown algae, diatoms, and golden-brown algae. Although commonly grouped with the stramenopiles, the oomycetes are also closely related to the alveolates, particularly the apicomplexan parasites, the most infamous of which are responsible for causing malaria (Ald et al., 2005; Harper et al., 2005).

There is little doubt that the oomycetes descended from photosynthetic organisms. They possess hundreds of genes similar to those found in red alga or cyanobacteria, as well as a number of biosynthetic genes that were presumably acquired from a plastid (Tyler et al., 2006). Despite their ancestry, the oomycetes have long since lost the ability to photosynthesize. The majority are saprophytes, leeching nutrients out of dead and decaying organic matter. Fossils interpreted as early oomycetes dating back to the Carboniferous Period are often found in association with plant matter, seeds and cortical
tissue, but rarely are there any signs indicative of disease or resistance. The earliest oomycetes may have therefore been saprophytes (reviewed in Krings et al., 2011).

Over the years, however, and from very early on, the oomycetes saw an increasing trend towards parasitism (reviewed in Beakes & Sekimoto, 2009). Göker et al. (2007) proposed that early saprophytic oomycetes adapted to grow on stressed or wounded organisms and eventually developed the machinery to infect healthy hosts, thereby giving rise to opportunistic and then obligate parasites. These important events most likely occurred in a marine environment, where some of the earliest divergent genera of oomycetes preyed upon seaweeds and nematodes. Nematodes that ventured to squirm into terrestrial soil may have taken their parasites with them, allowing the oomycetes to transition from sea to soil and aiding the evolution of oomycete plant pathogens (reviewed in Beakes & Sekimoto, 2009).

**Phytophthora: The Genus Plants Dread**

The oomycetes have adapted to parasitize a wide range of creatures, including various microorganisms, crustaceans, fish and some mammals. At least one species, *Pythium insidiosum*, has known clinical importance in animals and humans. Of all the appetites found in the oomycota, however, the most voracious and destructive are those directed against plants. There are several species of plant-pathogenic oomycetes, but none have caused more damage than those belonging to the genus *Phytophthora* (Erwin & Ribiero, 1996).

“*Phytophthora*” is Greek for “plant destroyer.” It is a name over 60 pathogenic species have earned. If plants had nightmares, these tiny, creeping organisms would
surely haunt them. Species of *Phytophthora* are found in every corner of the world and affect virtually all dicot plants, in addition to several monocots. These miniature destroyers are responsible for causing a number of diseases in ecologically and agriculturally important plants. Consequently, the members of this genus have had and continue to have profound environmental and economic impacts wherever they are found (Erwin & Ribiero, 1996).

Forests are excellent witnesses to the might of *Phytophthora*. Temperate forest soils are home to a number of indigenous *Phytophthora* species, some of which are completely new to science. These forest-dwelling oomycetes infect and kill the fine root hairs of trees. Under normal environmental conditions, the fine root hairs are replaced at a rate equal to their destruction, and the trees never develop any outward signs of disease. The introduction of an invasive *Phytophthora* species can upset this peaceful balance with disastrous consequences (Rizzo & Garbelotto, 2003).

In California and other portions of the western United States, *Phytophthora ramorum* is wreaking havoc among the oaks. *P. ramorum* has an ever-expanding host range but it is best known for causing sudden oak death. Since 1995, millions of oaks (*Quercus* spp.) and tanoaks (*Lithocarpus densiflorus*) have withered under its creeping expansion. The tanoaks, *P. ramorum*’s most susceptible host, are essentially being eliminated from portions of the state (Rizzo, 2005; Hansen, 2008).

The loss of trees takes a toll on vertebrate and invertebrate creatures alike that rely on oaks and tanoaks as a source of food and shelter, as well as microbial communities in association with roots. The wake of lifeless timber *P. ramorum* leaves behind also creates a fire hazard (Rizzo & Garbelotto, 2003; Rizzo, 2005; Hansen, 2008). As if that were not
enough, the persistence of *P. ramorum* in reservoir hosts, such as the California bay laurel (*Umbellularia californica*), may hamper future restoration efforts (DeLio *et al.*, 2009).

Elsewhere in the world, *Phytophthora cinnamomi* poses a serious risk to several different species of trees. This oomycete has a broad host range and a long track record of causing ecological damage. It helped wipe out the southern chestnut forests in America, decimated populations of shortleaf pine and is the terror of cork oak woodlands in Europe. In Australia, *P. cinnamomi* is considered the fifth most destructive force threatening the continent. It destroys entire plant communities, eliminating 50-75% of the species present. The Jarrah eucalyptus trees (*Eucalyptus marginata*) and heathland vegetation have suffered the most casualties. *P. cinnamomi* can devour entire forests, converting them to grass or shrubland. In the process, it has displaced native fauna, ravaged national parks and threatened the existence of many rare endemic plant communities (Weste & Marks, 1987; Erwin & Ribiero, 1996).

Economically speaking, species of *Phytophthora* have been tugging at the world’s purse strings for decades. They cost the United States tens of millions of dollars each year in damages to crops, in addition to control measures. The annual global costs are incalculably higher (Erwin & Ribiero, 1996; Wrather & Koenning, 2006; Tyler, 2007).

*Phytophthora infestans* is perhaps the best-known species of *Phytophthora*. It is the cause of late blight in potatoes (*Solanum tuberosum*), tomatoes (*Lycopersicon esculentum*) and other members of the *Solanaceae*. In the 1840s, this oomycete ransacked Ireland’s potato crops, causing a famine that displaced millions of people overseas to Europe and America and killed approximately one million others (Erwin & Ribiero,
1996). *P. infestans* has maintained its ferocity through migrations and mutations. Its destructive appetite costs the world four billion dollars each year (Haverkort et al., 2008).

*Phytophthora palmivora* and *Phytophthora megakarya*, are two oomycetes responsible for causing black pod disease, an affliction that prevents 10-20% of the world’s cocoa from ever making it into candy bars (McMahon & Purwantara, 2004). *P. megakarya*’s only known host is the cocoa plant (*Theobroma cacao*). It is endemic to Africa and can cause losses in cocoa pods up to 90% (Opoku et al., 2000). Infected trees also run the risk of developing cankers, which can result in the death of entire branches or the whole tree. *P. palmivora* has a much wider host range. In addition to causing black pod disease on a global level, it can infect several tropical fruits, black pepper, rubber, and a wide array of ornamental plants (Erwin & Ribiero, 1996).

A third costly oomycete, the organism featured in this study, is the soybean pathogen *Phytophthora sojae*. *P. sojae* is particularly aggressive in its pathogenic activities, making it an endless vexation to those countries leading the world in soybean production: Brazil, the USA and China. It has a very narrow host range, feeding only on soybeans (*Glycine max*) and some species of lupines (Erwin & Ribiero, 1996). Despite its picky eating habits, *P. sojae*’s annual global activities amount to $1-2 billion dollars in damages (Tyler, 2007). In America’s northern Midwest alone, hundreds of millions of dollars are lost each year in damages to soybean crops.
**Phytophthora sojae:** Pretty Much the Only Exciting Thing Going on in America’s Midwest

*P. sojae* is strictly a soil pathogen. Its entire complex life cycle takes place underground (Fig. 1). As is true for all the oomycetes, *P. sojae* exists primarily as aseptate mycelium. Under certain conditions, such as aging or when nutrients are limited, the tips of non-mating hyphae will swell, producing asexual spores known as sporangiophores (Fabritius *et al.*, 2002). Sporangia are multinucleated and, while mitosis no longer takes place in sporangia, their cytoplasm remains active and undesiccated (Hardham & Hyde, 1997). At higher temperatures, sporangia can directly germinate in order to produce more hyphae, but when the conditions are wet and cold (below 12°C), single nuclei within the sporangia will compartmentalize, resulting in the formation of zoospores (Judelson & Blanco, 2005; Walker & Van West, 2007).

Zoospores are critical to the process of infection. They form rapidly within a sporangium and are expelled by turgor pressure, which builds due to the accumulation of proline (Ambikapathy *et al.*, 2001). A single sporangium can release 10-30 zoospores (Tyler, 2007). The spores are bean-shaped, single-nucleated, lack cell walls and possess two flagella of unequal lengths, with which they can swim freely through sodden earth. A longer whiplash flagellum is used for steering, while a shorter tinsel flagellum creates enough thrust to propel the zoospores forward (Walker & Van West, 2007).

Zoospores are a brief stage in *P. sojae*’s life cycle, typically lasting only hours (Tyler, 2007). Swimming spores are attracted chemotactically and electrostatically to the roots of potential host plants (Morris *et al.*, 1992; Tyler, 2002). Once contact is made with a root, physical and chemical stimuli cause a zoospore to rapidly form a primary cell
Fig. 1 General Life Cycle of P. sojae. A) An infection begins when zoospores approach a soybean root and B) encyst on its surface. C) The Cysts germinate in order to produce mycelium. D) The mycelium gradually spreads over and into the plant causing mass cell death by 48 hours post infection. E) Tips of some of the hyphae on the surface of the plant will swell and develop into sporangia. F) Sporangia can release 10-30 zoospores. G) Within the infected plant tissue, hyphae can develop into anthridium and oogonium. H) Self fertilization will occur resulting in an oospore. I) The oospore can germinate in order to produce more mycelium and sporangia, J) which can generate and release more zoospores, beginning the process anew.
wall, shed its flagella, and secrete an adhesive material that helps it bind to the surface of
the plant (Robold & Hardham 2005). This process is known as encystment.

Although the cysts attach primarily to the roots, all parts of a soybean plant are
susceptible to infection. Once fixed to a plant, the cysts germinate, extending germ tubes
and mycelium over and into the surface of their host, colonizing both the cortex and the
vascular tissue (Erwin & Ribiero, 1996). *P. sojae* is hemibiotrophic, capable of sustaining
itself on both living and dead tissues. Within the first twenty-four hours post-infection
(hpi), *P. sojae* behaves as a biotroph. Hyphae pervade deep into the host, and nutrients
are obtained from the living tissues through the formation of haustoria, specialized
structures that invaginate into the host tissues while remaining enveloped in the host cell
membrane. By forty-eight hpi, the pathogen has largely transitioned to a necrotroph.
Large water-soaked lesions appear as a result of the maceration and collapse of host
tissues (Moy et al., 2004).

Successfully colonized hyphae will produce more sporangia at the surface of the
plant and oospores within the plant tissue (Hardham, 2007). Sexual oospores represent a
fusion of male and female gametangia. *P. sojae* is homothallic (self-fertilizing), and
oospores form readily, even in culture. They are long-lived and germinate directly to
produce hyphae (Tyler, 2007). Thick-walled, asexual chlamydospores can also be
produced in necrotic plant tissues. These hardy spores function as survival structures and
preserve the organism through unfavorable conditions (Tyler 2007).
**Effectors: Oomycete Weapon of Choice**

The process of infection is made possible through the secretion of pathogen effector proteins. Effectors are molecules that manipulate host cell structure and function in order to facilitate infection and/or trigger host cell responses (Kamoun, 2006). Oomycete effectors are secreted from haustoria (Whisson *et al.*, 2007) and target one of two locations in a plant cell: the apoplast or the cytoplasm (Kamoun, 2006).

Apoplastic effectors are secreted into the extracellular spaces of host cells and commonly protect against hydrolytic enzymes, such as endoglucanases and chitinases. For example, *P. sojae* produces a number of glucanase inhibitor proteins (GIPs). These proteins are secreted into the apoplast, where they bind and inhibit the activity of plant-secreted endoglucanases (Ham *et al.*, 1997; Rose *et al.*, 2002). Without these effectors, endoglucanases would degrade polysaccharides present in *P. sojae*’s cell wall, impairing its growth and making it more susceptible to lysis. Oligosaccharides released from the cell wall can also serve as signal molecules that induce further defense responses in a plant (van Loon *et al.*, 2006).

Cytoplasmic effectors are secreted into plant cells and manipulate host responses. There are currently two known classes of cytoplasmic effectors in the oomycetes. The first class, RXLR effectors, is named after a conserved four amino acid motif found in the N-terminals of its members. The RXLR motif is crucial for delivery of the effectors into host cells. Any mutations in this region prevent the accumulation of RXLR effectors in host cell cytoplasm (Whisson *et al.*, 2007; Dou *et al.*, 2008). The RXLR motif is sufficient for translocation of effectors into host cells; no pathogen machinery is required (Dou *et al.*, 2008). It binds to phatidylinositol-3-phosphate (PI3P) molecules covering the
plasma membrane of plant cells and most likely causes the effector to be taken up by lipid raft-mediated endocytosis (Kale, et al., 2010).

Interestingly, the RXLR motif is strikingly similar to a host-translocation signal found in malarial effectors. The RXLX motif of malaria parasites, such as *Plasmodium falciparum*, allows the translocation of effectors inside host erythrocytes. The two motifs are so similar in sequence and location an RXLR motif from *P. infestans* can replace a *Plasmodium* motif without disrupting the effector’s ability to invade erythrocytes (Bhattacharjee et al., 2006).

The genomes of *Phytophthora* species are littered with RXLR effector genes. *P. sojae* alone encodes over 350 such genes (Jiang et al., 2008). They function mainly in suppressing host immune responses, particularly cell death. Plant cells typically respond to the presence of a pathogen with a hypersensitive response (HR), a form of programmed cell death that limits the growth and spread of a pathogen. RXLR effectors can interfere with a cell’s ability to detect or respond to an invading pathogen, ending the threat of an HR and allowing the pathogen to grow uninterrupted (reviewed in Jones & Dangl, 2006). *P. sojae* effectors, such as Avr1b (Shan et al., 2004), Avr1a, Avr3a (Qutob et al., 2009), Avr4/6 (Dou et al., 2010) and Avr3a/5 (Dong et al., 2011) all stifle immune responses in susceptible interactions.

A second class of effectors is known as crinklers (CRNs). Similar to RXLR effectors, CRNs possess a conserved N-terminal motif, LXLFLAK, that functions in host translocation (Schornack et al., 2010). Unlike RXLR effectors, CRNs are thought to function as necrosis-inducing proteins by analogy to bacterial effectors that cause cell death, chlorosis, and tissue browning (Torto et al., 2003). Their exact functions and
mechanisms are still poorly understood, though many seem crucial to the process of infection. For example, \textit{P. sojae} constitutively expresses two CRNs, PsCRN63/115, during infection. Silencing both genes impairs the pathogen’s ability to suppress host defenses (Liu et al., 2011).

It should be noted that not all soybean cultivars are equally susceptible to \textit{P. sojae}’s arsenal of effectors. Millions of years of selective pressure have resulted in a form of effector-triggered immunity (ETI). ETI involves a number of nucleotide-binding site and leucine-rich repeats (NBS-LRR) proteins that patrol the cytoplasm of the cell and are capable of triggering defensive pathways upon the direct or indirect detection of an effector (reviewed in Jones & Dangl, 2006).

NBS-LRR proteins are encoded by resistance (R) genes. There are at least fourteen known soybean R genes that provide protection against \textit{P. sojae} (Burnham et al., 2003). These genes have been used with some success in producing transgenic crops. This strategy, however, is only effective when an invading strain of \textit{P. sojae} possesses effector genes that correspond to the crop’s resistance genes. Furthermore, pathogen effectors and other proteins associated with infection are subject to selection by diversification and are constantly changing due to the pressures presented by plant resistance genes (Tyler et al., 2006). A single allele can only provide protection for eight to fifteen years before a crop’s defenses are once again surmounted (Schmitthenner, 1985). The limitations of resistance genes have fueled a search for newer, better methods of control.
Transcription Factors

The study of oomycetes has progressed significantly over the years. Where once the oomycetes were lumped together with fungi, they are now recognized as distinctly separate creatures more closely related to algae, diatoms and apicomplexan parasites. Their life cycles have been carefully observed and recorded, and the methods by which they infect a plant and circumvent its immune system have been scrutinized with advanced molecular techniques. Still, there are many aspects of oomycete biology that remain unstudied and obscure.

A very large black box concerns the molecular mechanisms that regulate an oomycete’s development and coordinate the process of infection and virulence. Most genes within an organism are regulated, ensuring that the organism produces the correct amount of proteins at the right time and rate, as well as in the proper location. The majority of regulation occurs at the level of transcription and is conducted by proteins known as transcription factors (TFs). These proteins are capable of binding upstream of, within or else a considerable distance away from a gene in order to increase or decrease the gene’s rate of transcription (reviewed in Latchman, 1990, 1997; reviewed in Pabo & Sauer, 1992).

Most TFs possess two basic domains, a DNA-binding domain that allows for attachment to DNA sequences known as cis-elements, and an activation domain that binds to other factors and promotes the formation of an initiation complex. TFs are typically grouped into families based on their DNA-binding motifs. While there are many large, established families of TFs, it is important to remember that a number of important DNA-binding proteins, such as the SV40 large T antigen and human p53 tumor
suppressor gene, do not belong to any known families. Of course, there is also no reason
to assume that all DNA-binding domains have been discovered (reviewed in Pabo &

Identifying the molecules involved in gene expression in the oomycetes can lead
to a better understanding of how these pathogens regulate their development and
pathogenesis. This, in turn, can lead to better means of control. On the whole, however,
gene regulation in stramenopiles is poorly understood. Studies, such as Iyer et al. (2008)
and Rayko et al. (2010), have computationally identified some of the most abundant TF
families in the oomycetes using comparative genomic analysis. These studies suggest that
the majority of TFs encoded by oomycetes belong to Myb and C2H2-type zinc finger
families. The genomes could only be analyzed for known TF families, and, while Iyer et
al. (2008) was content with the total number of TFs identified in the oomycetes given
their complexity, Rayko et al. (2010) suggested that their own estimates might be low
owing to unidentified, lineage-specific TFs.

A small number of studies have cloned actual *Phytophthora* transcription factors
and observed their phenotypic effects upon silencing (Blanco & Judelson, 2005; Wang et
al., 2009; Xiang & Judelson, 2010). For example, Wang et al. (2009) identified a C2H2-
type zinc finger protein in *P. sojae*, PsCZF1, that is highly expressed in oospores and the
tips of hyphae. Silencing PsCZF1 markedly reduced the production of oospores and
caused a loss of virulence. For this and similar studies, DNA binding was assumed but
never proven. Targeted *cis*-elements were not identified.
Hunting Down Regulators: Methods for Studying TFs

A variety of molecular and biochemical techniques exist for studying transcription. For example, microarrays are commonly used to identify potential cis-elements in a genome. A microarray can determine genome-wide gene expression. The promoter regions of genes with similar expression patterns can be aligned in order to discover putative cis-elements (Tavazoie et al., 1999). This strategy can suggest which family of TFs regulates certain genes, but it cannot identify specific proteins nor can it identify uncharacterized TFs (Deplancke et al., 2004). Microarrays can also be used to test the affinity of known TFs to whole-genome intergenic regions. This approach can identify new target genes a TF may regulate (Mukherjee et al., 2004).

Mobility shift assays are often used to study specific DNA-protein interactions in vitro. This technology is based on the fact that unbound DNA travels farther in a gel than DNA attached to a protein. Such an assay can test the ability of cloned proteins or domains to bind and impede specific oligos in polyacrylamide gel. It can therefore identify particular factors and the specific regions of DNA to which they bind (Latchman, 2008). Mobility shift assays are an excellent way to determine DNA-protein interactions, however, this method is in vitro and only suggest what may occur within a living cell.

DNase I footprinting can go one step further than mobility shift assays and identify specific areas of contact between a TF and its cis-element. In this method, oligos are incubated with a protein and then exposed to DNA-degrading enzymes. The resulting DNA fragments are separated on a gel and observed. Regions to which the protein bound will be protected from degradation (Latchman, 2008). DNase I footprinting can be carried
out both *in vitro* and *in vivo*. It represents a powerful tool for investigating specific DNA-protein interactions throughout the life of a cell.

Other methods, such as chromatin immunoprecipitation (ChIP), can be used to study TFs *in vivo*. ChIP involves fixing a living cell with formaldehyde and then precipitating transcription complexes using antibodies. Once the cross-link between TF and DNA is broken, DNA fragments can be subjected to PCR in order to identify specific genes or cloned and sequenced in order to discover all the genes with which a TF interacts. ChIP can also be combined with microarray technology in order to locate genome-wide binding sites (Latchman, 2008).

Unfortunately, the generation of antibodies and microchips can be expensive and time consuming. ChIP cannot accurately detect interactions for TFs that are expressed at very low levels, in a small number of cells, or during a brief stage of development. Furthermore, this method has trouble distinguishing among different isoforms, and although ChIP methods are excellent at identifying particular regions to which TFs bind, they are less helpful in uncovering which TFs actually do the binding (Deplancke *et al.*, 2004). This problem is better addressed using a powerful *in vivo* technique known as the yeast one-hybrid (Y1H) system.

Y1H systems can detect protein-DNA interactions in the nucleus of a living cell. This approach has two main components, bait and prey. The bait is a sequence of DNA containing tandem repeats of a *cis*-motif. It is inserted upstream of a reporter gene in a vector, which is then introduced into the genome of a yeast strain *via* homologous recombination. The prey consists of potential DNA binding proteins. Prey cDNA molecules are expressed in the yeast reporter strain as fusion proteins containing
constitutively expressed yeast activation domains. If a prey fragment encodes a protein capable of binding the bait, activation of the reporter gene will occur, generally resulting in a color change or antibiotic resistance (Deplancke et al., 2004).

This system has two major advantages. First, it tests DNA binding in the natural environment of the nucleus. This verifies that successful protein-DNA interaction occur in nature and minimizes the need for optimization, a normally frustrating component of in vitro techniques. Second, the system allows for the screening of uncharacterized promoter elements, as well as entire cDNA libraries. It therefore has an incredible potential for identifying novel cis-elements and transcription factors (Lopato et al., 2006).
OBJECTIVE

The purpose of this study was to identify known and novel transcription factors in *P. sojae* using a Y1H system. Prey for the Y1H assay was generated from mycelium and zoospore cDNA libraries. The bait for the assay consisted of three tandem copies of “GCCGCC.” This motif was identified as an overrepresented word present in the upstream genomic regions of *P. sojae*. Approximately 20% of all genes in *P. sojae* have this motif within 1,000 bp upstream of their start sequences.

The GCCGCC motif is a common binding site for the majority of transcription factors belonging to the AP2/ethylene response motif binding protein (EREBP) (Ohme-Takagi & Shinshi, 1995; Fujimoto *et al.*, 2000). Members of the EREBP subfamily possess a single AP2 domain and are involved in responses to pathogenesis, as well as environmental stresses, such as heat and drought (Ohme-Takagi & Shinshi, 1995). Members of the AP2 subfamily possess two AP2 domains and help regulate various aspects of plant development (Jofuku *et al.*, 1994).

Although AP2 domains were once thought to exist exclusively in plants, they have also been discovered in the Apicomplexan parasites, including the malaria parasite *Plasmodium falciparum*. Much like plant AP2 proteins, ApiAP2 proteins regulate growth and development as the parasites transition through the various stages of their life cycles (Balaji *et al.*, 2005; Behnke *et al.*, 2005; Yuda *et al.*, 2009; Yuda *et al.*, 2010). Although ApiAp2 domains differ from plant AP2 domains in binding specificity, they fold into similar structures (Lindner *et al.*, 2010) and possess 7 out of 11 conserved residues required for binding “GCCGCC” motifs (Balaji *et al.*, 2005).
*P. sojae* has a intimate association with plants and is closely related to the apicomplexans, parasites with which it shares a number of different strategies. In light of *P. sojae*’s lifestyle and phylogeny, it seems reasonable to suspect that the “GCCGCC” motif may also function as an important *cis*-motif in the oomycetes.

Although this study did not capture an AP2-like domain using the GCCGCC motif as bait, it was able to capture a conserved, novel domain present in thirty-one *P. sojae* proteins, with homologs in *P. infestans* and *P. ramorum*. The identification and future characterization of this domain will shed much needed light on transcriptional regulation in *Phytophthora* and related genera, such as *Pythium* and the downy mildews. Such a protein may also serve as a new and better target for the control of a destructive pathogen.
METHODS

Analysis of Genomic Data

Wordseeker (Lichtenberg et al., 2010) was used in order to confirm the significant presence of the “GCCG GCC” motif in the upstream genomic sequences of \( P. \) sojae and related species. Genomic data from \( P. \) sojae, as well as \( P. \) infestans and \( P. \) ramorum, was downloaded from the Broad Institute (http://www.broadinstitute.org/annotation/genome/Saprolegnia_parasitica/MultiHome.html). The data for each organism included: all genes, all transcripts, 1000 bp upstream of each gene, 1000 bp upstream of each gene including the UTR, 1000 bp downstream of each gene and 1000 bp downstream of each gene including the UTR. Each data file was uploaded to Wordseeker 2.0 (http://wordseeker.org/) and analyzed under the default settings.

Alternatively, genomic upstream data was filtered into separate files based on the presence of chosen motifs. Upstream regions for RXLR genes were also included. These files were then analyzed using Wordseeker.

Generating the Yeast Reporter Strain

The reporter construct was generated using the Matchmaker Gold Yeast One-Hybrid (Y1H) Library Screening System (Clontech Laboratories, Inc., Mountain View, CA) according to the manufacturer’s instructions. The bait sequence consisted of three tandem copies of “GCCG GCC.” The forward and reverse strands of the bait sequence were designed to contain a 5’ TCGA overhang for cloning into a vector digested by HindIII and XhoI. The use of two restriction enzymes protects against the reannealing of linearized vectors and ensures that an insert enters into its proper orientation. The two
complementary bait sequences were synthesized by Integrated DNA Technologies (IDT Inc., Coralville, IA), resuspended in 1X TE buffer to a concentration of 100 µM, mixed together in equal portions and annealed to create a double-stranded sequence.

The bait vector, pAbAi (Clontech Laboratories, Inc., Mountain View, CA), was double digested with HindIII and XhoI. The linear plasmid was purified on a 1.5% agarose gel with Ethidium Bromide (EtBr) and extracted using a DNA Clean and Concentrator™ Kit (Zymo Research Corporation, Irving, CA). The double stranded bait sequence was ligated into the linear pAbAi using a Quick Ligation™ Kit (New England Biolabs® Inc., Ipswich, MA). The resulting plasmid (pBait-AbAi) was purified using a MiniElute PCR Purification Kit (Qiagen Inc., Germantown, MD).

The pBait-AbAi vector was transformed into E. coli and grown at 37°C on LB agar containing 100 µg/ml ampicillin. Plasmids from the positive colonies were extracted using a QIAprep® Spin Mini Prep Kit (Qiagen Inc., Germantown, MD). The concentrations of plasmid were measured using a Thermo Scientific Nano Drop 1000 (Thermo Fisher Scientific Inc.), and the integrity of the plasmids was observed on a 2% agarose gel stained with EtBr. In order to check for the presence of the bait insert, a sample of the suspected pBait-AbAi and the empty pAbAi vector were digested with SacI and run on 2% agarose with EtBr. The vector, pAbAi, has a single SacI restriction site that is removed after the insertion of a bait sequence. SacI cannot digest pBait-AbAi, but it can linearize pAbAi.

The confirmed pBait-AbAi was digested using BstBI and purified using a MiniElute PCR Purification Kit (Qiagen Inc., Germantown, MD). The concentration of linear pBait-AbAi was measured using a Thermo Scientific Nano Drop 1000 (Thermo
Fisher Scientific Inc.), and the integrity of the linear plasmid was observed on a 2% agarose gel with EtBr. In an attempt to save time, the purified, linear pBait-AbAi was transformed into *S. cerevisiae* Y1HGold using the “Quick and Dirty’ Plasmid Transformation of Yeast Colonies” protocol, as described in Amberg *et al.*, 2005. Transformation reactions were screened on SD/-Ura agar at 30°C. Resulting positive colonies were analyzed using the Matchmaker Insert Check PCR Mix 1 (Clontech Laboratories, Inc., Mountain View, CA) in order to confirm the successful integration of pBait-AbAi into the yeast genome.

Positive DNA-protein interactions using the Matchmaker Gold Y1H System are indicated by resistance to Aureobasidin A (AbA). The minimal inhibitory concentration of Aureobasidin A for each reporter strain was determined by plating 100 µl of yeast cells suspended in 0.9% NaCl at an OD$_{600}$ of ~0.002 on SD/-Ura plates containing 0, 100, 150, and 200 ng/ml of Aureobasidin A. Colonies resistant to concentration higher than 200 ng/ml were not considered as candidates for a reporter strain. Confirmed reporter strains with a minimal resistance to Aureobasidin A ≤ 200 ng/ml were grown overnight in YPDA broth at 30°C and 250 rpm, mixed with glycerol to a final concentration of 25%, and stored as 1 ml aliquots at -80°C for future use. A concentration of Aureobasidin A 50 ng/ml higher than the minimal concentration was used during the actual process of screening the cDNA library.

**Biological Materials**

*P. sojae* strain R2 was grown on V8 agar plates in order to collect mycelium and zoospores for total RNA extraction. In order to produce mycelium, *P. sojae* was grown
on nylon transfer membranes placed over V8 agar plates at room temperature for 5-7 days. The mycelium was scraped off the membranes, frozen in liquid nitrogen, crushed to a fine powder and stored at -80°C until further processing.

Zoospores were produced by growing *P. sojae* for three days, after which time they were flooded with cold tap water and stored overnight at room temperature. The following morning, the plates were washed with tap water every 20-30 minutes and placed back in the dark between washes. Once the sporangia burst, the water from the plates was collected in 50 ml falcon tubes. The plates were washed one last time with 2-3 ml of tap water. Any mycelium in the water was removed by filtering the water through a Miracloth (Calbiochem®, San Diego, CA). The tubes were centrifuged for 2 minutes at 2,000 rpm and room temperature, and 300 µl were immediately pipetted directly from the bottom of the tube into a 1.5 ml microcentrifuge tube. The number of zoospores was counted using a haemacytometer. The zoospores were then flash frozen and stored at -80°C until further analysis.

One Shot® TOP10 Chemically Competent *E. coli* (Invitrogen™, Carlsbad, CA) was grown in LB broth and LB agar plates with ampicillin for selection. The Y1H Gold yeast strain (Clontech Laboratories, Inc., Mountain View, CA) was grown in a variety of media, including YPDA, SD/-Ura, and SD/-Leu, depending the experiments. Aureobasidin A and ampicillin were both used as selecting agents.

**Yeast and *E. coli* Stocks**

Yeast colonies carrying a prey sequence with a confirmed positive interaction were grown overnight in YPDA broth with 200 ng/ml AbA and 100 ug/ml ampicillin at
30°C and 250 rpm. Cultures were spun down at high speed, decanted, and resuspended with equal portions of fresh YPDA broth and 50% glycerol to a final concentration of 25%. All samples were stored as 1 ml aliquots at -80°C.

Stock samples of *E. coli* colonies containing the empty pAbAi and pGADT7-Rec vectors were produced. Stocks were also made of colonies containing prey vectors with confirmed positive interactions in the reporter strain. All *E. coli* colonies were grown in LB with 100 µg/ml ampicillin overnight at 37°C. Cultures were spun down at high speed, decanted, and resuspended with equal portions of fresh LB broth and 50% glycerol to a final concentration of 25%. All samples were stored as 1 ml aliquots at -80°C.

cDNA Synthesis

RNA was extracted from the powdered sample of *P. sojae* mycelium and 300 µl of zoospores (4 X 10⁴ zoospores/ml) using an RNeasy Plant Mini Kit (Qiagen Inc., Germantown, MD). During the extraction of RNA from zoospores, the buffers and samples were kept on ice at all times, and the centrifuge was cooled to 17°C. The concentrations of RNA were measured using a Thermo Scientific Nano Drop 1000 (Thermo Fisher Scientific Inc.). Each sample of RNA was also run on a 2% agarose, RNase-free gel with EtBr in order to confirm its integrity. Approximately 0.10-2.0 µg of total RNA from each sample were converted into cDNA using First Strand SMART cDNA Synthesis (Clontech Laboratories, Inc., Mountain View, CA), as described in the Matchmaker™ Gold Yeast One-Hybrid Library Screening System Manual (Clontech Laboratories, Inc., Mountain View, CA). Each sample of cDNA was run on a 2% agarose
gel with EtBr in order ensure proper band lengths. A smear of DNA above approximately 400 bp should increase the likelihood of having full-length cDNA sequences.

Initial Screening of the One-Hybrid cDNA Library

Following the protocol outlined in the Yeastmaker™ Yeast Transformation System 2 User Manual (Clontech Laboratories, Inc., Mountain View, CA), 3 µg of pGADT7-Rec were cotransformed with 2-5 µg of mycelium or zoospore cDNA into the reporter strain. The transformation reactions were spread on SD/-Leu plates containing 200 ng/ml Aureobasidin A and grown at 30°C for 4 days. Aureobasidin A is a cyclic depsipeptide produced by the fungus, *Aureobasidium pullulans R106*. AbA disrupts yeast membrane integrity by inhibiting the synthesis of inositol phosphorylceramides, major components of the yeast plasma membrane (Nagiec *et al.*, 1997). Plates lacking Aureobasidin A were used as positive controls. Positive colonies were isolated and maintained on SD/-Leu with 200 ng/ml Aureobasidin A at 4°C. The colonies were transferred to fresh plates every month.

Isolation and Sequencing of Prey

It was assumed that colonies exhibiting strong growth (i.e. large, “puffy” colonies appearing after 2-3 days) were indicative of a strong interaction between bait and prey. The prey sequences from such colonies were amplified using the Matchmaker Insert Check PCR Mix 2 according to the manufacturer’s instructions (Clontech Laboratories, Inc., Mountain View, CA). Five µl of each 50 µl reaction were run on 0.8% agarose gels with EtBr. Colonies that produced a clean, single band greater than 400 bp in length were
chosen for further analysis. PCR amplicons were purified using the MiniElute PCR Purification Kit (Qiagen Inc., Germantown, MD) and sent to the University of Cincinnati and Children's Hospital DNA Core Facility (Cincinnati, Oh) for sequencing using the T7 Sequencing Primer.

The sequencing data was edited manually in order to remove regions corresponding to pGADT7-Rec. Sequences were translated into amino acids using the ExPASy Translate Tool (http://web.expasy.org/translate/), aligned using ClustalW software and grouped according to similarity.

**Confirmation of Positive Interactions**

Library plasmids containing prey sequences chosen for further screening were rescued from their respective colonies using an Easy Yeast Plasmid Isolation Kit (Clontech Laboratories, Inc., Mountain View, CA) and cloned in *E. coli*. Positive colonies were selected for ampicillin resistance on LB plates. Cloned plasmids were isolated from the *E. coli* using a QIAprep® Spin Mini Prep Kit (Qiagen Inc., Germantown, MD) and 100 ng of each were transformed into the yeast reporter strain according to the methods outlined in the Yeastmaker™ Yeast Transformation System 2 User Manual (Clontech Laboratories, Inc., Mountain View, CA). Selection occurred on SD/-Leu plates containing 200 ng/ml Aureobasidin A. The plates were incubated at 30°C for 3 days.
Characterization of Prey

The Joint Genome Institute’s database of *P. sojae* V3.0 filtered model proteins was searched using the BLASTP algorithm (http://genome.jgi-psf.org/pages/blast.jsf?db=Physo1_1). Protein models matching the sequenced prey fragments were cataloged and analyzed on several levels, including domain, secondary structure and signal peptide/subcellular localization predictions. Protein domains were predicted using Interpro (Hunter et al., 2009). Secondary structure prediction was performed using Jpred3 (Cole et al., 2008) and NetSurfP V1.1 (Petersen et al., 2009). Signal peptide/subcellular localization predictions were made using Phobius (Kall et al, 2007), SignalP V4.0 (Kahsay et al., 2011), HECTAR (Gschloessl et al., 2008), Wolf PSORT (Horton et al., 2007), TargetP (Emanuelsson et al., 2000), CELLO V2.5 (Yu et al., 2004,2006), Euk-PLoc V2.0 (Chou et al., 2010), iPSORT (Bannai et al, 2001, 2002), and Yloc (Briesemeister et al., 2010).

Homologous proteins in related species of oomycetes were searched for in the Broad Institute’s Saprolegnia Database using the BLASTP program (http://www.broadinstitute.org/annotation/genome/Saprolegnia_parasitica/Blast.html;jsessionid=178C645B4B792C7BAF7161E3A4EB3D12.route980?sp=Sblastp). These protein sequences were subjected to the same gamut of analysis.
RESULTS

Computational Analysis

Wordseeker 2.0 (http://word-seeker.org/) was used to discover putative transcription factor binding sites within the genome of *P. sojae*. The program identified a “GCCGCC” motif with p-values of $0 \cdot 1.11 \cdot 10^{-16}$ overrepresented only in the genomic region 1,000 bp upstream of genes (Fig. 2). An analysis of the *P. sojae* genome revealed that approximately 20% of all predicted genes contain the motif “GCCGCC” within 1,000 bp upstream of their start sequence. The same is true for 12% of all predicted RxLR genes in *P. sojae*.

![Cluster Logo](image)

**Fig. 2** “GCCGCC” motif identified by Wordseeker. “GCCGCC” is an overrepresented motif in the upstream regions of *P. sojae* genes. The motif is reported to have respective p-values of $0 \cdot 1.11 \cdot 10^{-16}$ in the genomic regions 1,000 bp upstream of genes with and without UTRs.

Yeast One-Hybrid System

The Matchmaker Gold Yeast One-Hybrid (Y1H) Library Screening System (Clontech, CA) was used to isolate cDNAs encoding domains capable of interacting with the “GCCGCC” motif. The reporter strain for the Y1H carried the integrated bait vector pAbAi, which possesses two selectable markers, *URA3* and *AUR1-C*. *URA3* is a nutritional marker that allows for growth on media lacking uracil. The *AUR-1C* gene...
confers resistance to the antibiotic Aureobasidin A (AbA) and is driven by the yeast iso-
1-cytochrome C minimal promoter. A bait sequence containing three tandem repeats of
“GCCGCC” was inserted upstream of the promoter, ensuring that expression of the
AUR1-C gene and subsequent resistance to AbA occurred only as a result of interactions
between the bait and prey (Fig. 3).

The reporter strain was cotransformed with the prey vector, pGADT7-Rec, and
SMART cDNA from either P. sojae mycelium or zoospores. Within the yeast cells,
SMART cDNA inserts into the prey vector via homologous recombination and is
expressed as a fusion protein linked to the yeast GAL4 activation domain (Fig. 4). The
reporter strain was shown to tolerate levels of AbA below 150 ng/ml. Therefore,
transformants were screened for resistance to 200 ng/ml AbA and growth on SD media
lacking leucine, a deficit overcome by the prey plasmid’s LEU2 gene.

Transformation with mycelial (M) cDNA resulted in the screening of 3.75 x 10^6
transformants and the isolation of over 300 colonies. Transformation with zoospore (Z)
cDNA resulted in the screening of 5.85 x 10^6 transformants and the isolation of
approximately 250 colonies. The cDNA prey fragments of 145 colonies (78 from M and
67 from Z) were amplified using the Matchmaker Insert Check PCR Mix 2 (Clontech,
CA). At total of 87 colonies (41 from M and 46 from Z) produced clean, single bands
over 400 bp in length and were sent for sequencing (Fig. 5).

The sequencing data was edited manually in order to remove regions
corresponding to pGADT7-Rec and translated into amino acid sequences. The majority
of all prey fragments analyzed from both M and Z encoded the same conserved 50 amino
acid sequence with slight variations (Fig. 6).
Fig. 3 Overview of pBait-AbAi Vector and its Integration into the Y1HGolor Yeast Genome. (A) The pBait-AbAi vector contained three tandem repeats of the bait sequence, “GCCGCC,” inserted upstream of the AUR-1C gene that codes for resistance to AbA. The vector also contained the nutritional marker, URA3, for growth on media lacking uracil, as well as necessary genes for selective growth in E. coli. The pBait-AbAi vector was digested with BstBI and transformed into Y1HGolor yeast. (B) Within the yeast, the linear pBait-AbAi vector was integrated into the yeast genome via homologous recombination within the coding region of URA3 and the inactive URA3-52.
Fig. 4 Screening for Protein-DNA Interactions Using SMART cDNA and the Prey Vector pGADT7-Rec. (A) Total RNA was isolated from mycelium and zoospores. SMART cDNA synthesis created cDNA fragments flanked by regions homologous to the prey vector, pGADT7-Rec. (B) Both the SMART cDNA and linear pGADT7-Rec were cotransformed into the reporter yeast strain. The SMART cDNA fragments integrated into the prey vectors via homologous recombination within the yeast cells and were expressed as fusion proteins, attached to the yeast GAL4 activation domain. (C) In a successful protein-DNA interaction, the expressed prey cDNA binds the bait. The fused GAL4 activation domain recruits yeast transcription machinery, leading to expression of AUR-1C.
Fig 5 Examples of mycelial and zoospore prey amplicons on 0.8% agarose with EtBr. Prey sequences were amplified from transformed colonies exhibiting strong growth on selection agar plates using the Matchmaker Insert Check PCR Mix 2 (Clontech, CA). Five µl of each 50 µl reaction was run on 0.8% agarose gels stained with EtBr. Colonies that produced a clean, single band greater than 400 bp in length were chosen for sequencing and further analysis.

Two representative prey sequences (M2.06 and Z2.32) were chosen for further screening. The fragments were chosen in order to confirm the prey’s ability to bind DNA, as well as to test whether variations in amino acid sequence affected the prey’s binding activity. Both M2.06 and Z2.32 encode peptides containing the same basic 50 amino acid sequence, but they differ from each other within this region by 6 amino acids. The translated fragments also vary in length, 117 aa and 133 aa for M2.06 and Z2.32, respectively, and were isolated from different cDNA libraries.
Fig. 6 Alignment of deduced amino acid sequences from prey fragments. A large number of the isolated prey fragments encoded the same amino acid sequence and share a conserved region approximately 50 amino acids in length (boxed in red). Two prey fragments, 1365 (M2.06) and 1350 (Z2.32) (boxed in black), were chosen for further analysis. These representative sequences where chosen because of the variation existing between the two in amino acid sequence as well as length and source.
The plasmids harboring these prey sequences were isolated from their respective yeast colonies, cloned and transformed into the reporter strain. Transformants produced healthy colonies on SD/-Leu containing 200 ng/ml AbA thereby confirming the interaction between the conserved motif and the “GCCGCC” binding site (Fig. 7).

Amino Acid Sequence Analysis of Prey Proteins

The conserved sequence was BLASTed against the *P. sojae* genome from the Joint Genome Institute, VMD and Broad Institute. Analysis showed that the *P. sojae* genome contains 31 proteins carrying the conserved sequence. Only one gene is predicted to have a ubiquitin-like domain, while the rest are classified as putative proteins with no known domains or motifs.
In *silia* Analysis of Homologous Proteins in *P. infestans* and *P. ramorum*

The conserved prey sequence captured in the Y1H was used to search for homologous proteins using BLASTP against the Broad Institute Genome Database (http://www.broadinstitute.org/annotation/genome/Saprolegnia_parasitica/Blast.html;jsessionid=178C645B4B792C7BAF7161E3A4EB3D12.route980?sp=Sblastp). The search yielded six proteins in *P. infestans* and one in *P. ramorum* possessing homologous regions sharing 35-54% identity (Fig. 8).

Secondary Structure Prediction

All 31 *P. sojae* proteins, as well as the homologous proteins identified in *P. infestans* and *P. ramorum*, were analyzed for predicted secondary structures, signal peptides and subcellular localization. Secondary structures were predicted using Jpred3 (Cole *et al.*, 2008) and NetSurfP V1.1 (Petersen *et al.*, 2009). The conserved region corresponding to the prey in *P. sojae* is predicted to encode two α-helices separated by a loop or turn (Fig. 9). A similar prediction was made for the corresponding regions in *P. infestans* and *P. ramorum*. 
Fig. 8 Alignment of conserved region from *P. sojae* with homologous regions identified in proteins from *P. infestans* and *P. ramorum*. The conserved regions from the 31 proteins identified in *P. sojae* were aligned with homologous regions from six proteins found in *P. infestans* and one protein found in *P. ramorum*. The conserved regions of these homologous proteins share 35-54% identity. Regions of the N-terminals of the proteins (not shown) were also highly conserved among all species.
Fig. 9 Secondary structure prediction for homologous prey regions in *P. sojae*, *P. infestans* and *P. ramorum* proteins.

The prey region in each of the 31 *P. sojae* proteins is predicted to encode two α-helices separated by a loop. Although the sequences were not as well conserved, the corresponding regions in the *P. infestans* proteins and *P. ramorum* protein are predicted to fold into similar structures.

**Signal Peptide/Subcellular Localization Prediction**

Signal peptide/subcellular localization was predicted by a number of different programs, including Phobius (Kall et al., 2007), SignalP V4.0 (Petersen et al., 2011), HECTAR (Gschloessl et al., 2008), Wolf PSORT (Horton et al., 2007), TargetP (Emanuelsson et al., 2000), CELLO V2.5 (Yu et al., 2006), Euk-PLoc V2.0 (Chou & Shen, 2010), iPSORT (Bannai et al., 2002), and Yloc (Briesemeister et al., 2010a, 2010b). On average, the proteins from *P. sojae* were not predicted to have signal peptides. In fact, only one of the proteins was consistently predicted to possess a signal
peptide. In terms of subcellular localization, the programs predicted that 44% of all 31 proteins localize to the nucleus, 41% localize to the cytoplasmic and 15% localize to the mitochondria.

Similar to the *P. sojae* proteins, the homologous proteins in *P. infestans* and *P. ramorum* were not predicted to have signal peptides. The *P. ramorum* protein was consistently predicted to localize to the cytoplasm. While the programs were in less agreement about where the *P. infestans* proteins should be targeted, the results seem to suggest that they are most likely localized to the cytoplasm.
DISCUSSION

This study attempted to identify known and novel transcription factors within the soybean pathogen *P. sojae* using a Y1H assay. A “GCCGCC” motif identified as an overrepresented word in the upstream sequences of *P. sojae* was used as bait. In plants, the “GCCGCC” motif is a common binding site for members of the AP2/EREBP (Ohme-Takagi & Shinshi, 1995; Fujimoto *et al.*, 2000). Apicomplexan parasites also possess proteins with AP2 binding domains. Although the amino acid sequences of these domains differ from plant AP2 domains, they fold into similar structures (Lindner *et al.*, 2010) and possess 7 out of 11 conserved residues used in binding to “GCCGCC” motifs (Balaji *et al.*, 2005). Given the presence of “GCCGCC” as a significant upstream motif in *P. sojae*, it seems reasonable to predict that this motif may serve as an important *cis*-element within the pathogen. Moreover, if the “GCCGCC” motif acts as a *cis*-element, then *P. sojae* AP2 proteins may exist, as suggested by computational studies (Iyer *et al.*, 2008; Rayko *et al.*, 2010).

The Y1H assay was able to capture proteins containing a 50 aa sequence capable of binding the “GCCGCC” motif. Surprisingly, this motif bears no resemblance to the canonical three β-sheets and one α-helix typical of AP2 domains (Allen *et al.*, 1998). Instead it is predicted to fold into two α-helices. Alpha helices are so widely employed in site-specific DNA-binding, it is difficult to imagine them doing anything else (Pabo & Sauer, 1992). Moreover, the predicted structure is reminiscent of a helix-loop-helix (HLH) domain. Proteins possessing HLH domains comprise a family of over 240 identified transcriptional regulators present in a wide range of eukaryotic organisms (reviewed in Massari & Murre, 2000). These proteins coordinate a myriad of metabolic
and developmental processes, including phosphate uptake in yeast, sex determination in fruit flies, myogenesis, heart and pancreatic development, neurogenesis and hematopoiesis (reviewed in Massari & Murre, 2000).

BLAST analysis identified 31 putative proteins in *P. sojae* that possess the captured domain, which was present in both mycelial and zoospore cDNA libraries. It is likely that these proteins are common to both stages of the life cycle. Indeed, RNAseq data available through the Virginia Bioinformatics Institute’s microbial database (http://vmd.vbi.vt.edu/) suggests that all 31 proteins are expressed during vegetative growth and the process of infection. Interestingly, with the exception of one model, there are no known functional domains predicted for these proteins.

*P. ramorum* and *P. infestans* encode 1 and 6 homologous proteins, respectively. These proteins possess conserved regions corresponding to the domain identified in *P. sojae* and do not have any known functional domains. The conserved regions of these homologous proteins share 35-54% identity and are predicted to fold into roughly the same structures. Variations in the sequences of these homologs may reflect different binding specificities. “GCCGCC” was not identified as an overrepresented motif in sequences belonging to *P. infestans* or *P. ramorum*. Furthermore, the divergence and expansion of genes encoding the identified motif in *P. sojae* suggests that these genes are under selective pressure and may therefore serve some crucial function in the organism’s pathogenicity.

Of all the proteins in *P. sojae* and related species, only one was consistently predicted to have a signal peptide. It therefore seems as though these proteins function within the cell. Subcellular prediction software identified a majority of the proteins as
targeted to the nucleus. Current prediction software, however, is not optimized for protist sequences and possesses a limited degree of accuracy (Gschloessl et al., 2008; Min, 2010).

It will be interesting to see if any of these proteins are secreted by the pathogen. In plants, the “GCCGCC” motif is found upstream of many pathogenesis-related genes and is bound by ethylene responsive factor (ERF) transcription factors in response to biotic stress. Interactions between ERF transcription factors and the “GCCGCC” motif are thought to coordinate resistance to several pathogens in *Arabidopsis* (Fujimoto *et al.*, 2000), tomato (Chakravarthy *et al.*, 2003), soybean (Mazarei *et al.*, 2002) wheat (Zhang *et al.*, 2007) and rice (Cao *et al.*, 2006). *P. sojae* proteins capable of binding the “GCCGCC” motif could serve as a powerful weapon if they could be delivered from *P. sojae* into a host’s nucleus. Such proteins could potentially manipulate host pathogenesis-related genes for the benefit of the pathogen.

Further studies are needed to address the binding specificity of the domain. While “GCCGCC” may represent the best binding site for this new domain, it may also only fulfill the minimum conditions for binding. Identifying the most favorable sequence for binding can facilitate the discovery of genes it may regulate. Should the domain represent a new variety of HLH, its preferred binding site may very well differ from “GCCGCC.” HLH domains typically bind to E-boxes, DNA elements possessing a core sequence of CANNTG (reviewed in Massari & Murre, 2000).

Knock-out studies can also be used to assess the importance of these 31 proteins identified in *P. sojae* on the pathogen’s growth, development and virulence. Before any such studies can be performed, however, it will be necessary to correct inaccurate gene
models. The *P. sojae* genome is a work-in-progress, and many of the models have regions of unidentified DNA sequences directly upstream. Therefore, predicted N-terminals for these proteins may be incorrect. A cursory glance at RNAseq data also suggests that several predicted introns may be incorrect.

A particularly interesting question concerns the N-terminals of these proteins, portions of which are strongly conserved across *P. sojae, P. infestans* and *P. ramorum*. These regions may represent unidentified activation domains. In the future, it will be interesting to assay conserved N-terminal regions for protein-protein interactions. HLH proteins commonly form heterodimers, which are crucial to their functioning as TFs (reviewed in Pabo & Sauer, 1992; reviewed in Massari & Murre, 2000).

The identification of a potentially new DNA binding domain is both exciting and expected. Nature possesses a wealth of DNA-binding domains, but there is certainly room for discovery. At the same time, very little is known about the inner workings of *P. sojae*. The 31 proteins identified in this study represent new, potential DNA-binding proteins and may yet provide clues to protein-protein interactions in *P. sojae*. Understanding how *P. sojae* regulates gene expression is a step towards newer methods of control of this destructive crop pathogen.
REFERENCES


