EUKARYOTIC PLANT PATHOGEN DETECTION THROUGH HIGH THROUGHPUT DNA/RNA SEQUENCING DATA ANALYSIS

By

ANDRES S. ESPINDOLA

Bachelor of Science in Biotechnology Engineering

Escuela Politécnica del Ejército

Quito, Ecuador

2009

Master of Science in Entomology and Plant Pathology

Oklahoma State University

Stillwater, Oklahoma

2013

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 2016

EUKARYOTIC PLANT PATHOGEN DETECTION

THROUGH HIGH THROUGHPUT DNA/RNA

SEQUENCING DATA ANALYSIS

Thesis Approved:

Dr. Carla Garzon

Thesis Adviser

Dr. William Schneider

Dr. Stephen Marek

Dr. Hassan Melouk

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to all the people who made possible the completion of my thesis research. Most importantly to my advisor Dr. Carla Garzon for her continuous support, guidance and motivation. She was a great support throughout my PhD. research and thesis writing.

I would like to express my deepest gratitude to the members of my advisory committee: Dr. William Schneider, Dr. Stephen Maren and Dr. Hassan Melouk, for their guidance, insightful comments and encouragement.

I want to thank the Department of Entomology and Plant Pathology at Oklahoma State University for keeping an excellent environment for the students and professors. This is a great advantage that has helped me to succeed on my research and thesis writing.

Thanks to my fellow lab-mates for keeping a competitive and at the same time very friendly environment full of enriching discussions and productive working hours.

I would like to thank my wife Patricia Acurio, who has been extremely helpful throughout my PhD., her support, patience, love and encouragement were crucial to succeed completing my degree.

Last but not least, I would like to thank my parents, Isabel Camacho & Edison Espindola and my sister, Carolina Espindola for their support and encouragement that helped me to stay strong and productive during this important stage of my life.

Name: ANDRES SEBASTIAN ESPINDOLA CAMACHO

Date of Degree: DECEMBER, 2016

Title of Study: EUKARYOTIC PLANT PATHOGEN DETECTION THROUGH HIGH THROUGHPUT DNA/RNA SEQUENCING DATA ANALYSIS

Major Field: PLANT PATHOLOGY

Abstract:
Plant pathogen detection is crucial for developing appropriate management techniques. A variety of tools are available for rapid plant pathogen detection. Most tools rely on unique features of the pathogen to detect its presence. Immunoassays rely on unique proteins while genetic approaches rely on unique DNA signatures. However, most of these tools can detect a limited number of pathogens at once. E-probe Diagnostics Nucleic acid Analysis (EDNA) is a bioinformatic tool originally designed as a theoretical approach to detect multiple plant pathogens at once. EDNA uses metagenomic databases and bioinformatics to infer the presence/absence of plant pathogens in a given sample. Additionally, EDNA relies on a continuous design and curation of unique signatures termed e-probes. EDNA has been successfully validated in viral, bacterial and eukaryotic plant pathogens. However, most of these validations have been performed solely at the species level and only using DNA sequencing. My thesis involved the refinement of EDNA to increase its detection scope to include plant pathogens at the strain/isolate level. Additional refinements included its increasing EDNA's capacity to use transcriptomic analysis to detect actively infecting plant pathogens and metabolic pathways. Actively infecting/growing plant pathogen detection was performed by using *Slerotinia minor* as an eukaryotic model system. We sequenced and annotated the genome of *S. minor* to be able to use its genome for e-probe generation. *In vitro* detection of actively growing *S. minor* was successfully achieved using EDNA for RNA sequencing analysis. However, actively infecting *S. minor* in peanut was non-detectable. EDNA's capacity to detect the aflatoxin metabolic pathway was also assesed. Actively producing aflatoxin *A. flavus* strains (AF70) were successfully used to differentially detect the production of aflatoxin when *A. flavus* grows in an environment conducive for the production of aflatoxin (maize). Finally, EDNA's detection scope was assesed with eukaryotic strains having very low genetic diversity within its species (*Pythium aphanidermatum*). We were able to successfully discriminate *P. aphanidermatum* P16 strain from *P. aphanidermatum* BR444, concomitantly, these two strains were differentiated from other related species (*Globisporangium irregulare* and *Pythium deliense*) in the same detection run trial.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

I INTRODUCTION AND OBJECTIVES

Genome sequencing studies have grown exponentially since 1977 when the first sequencing by synthesis technology — commonly known as Sanger sequencing — appeared (Sanger, Nicklen & Coulson, 1977). Projects including the human genome (*Homo sapiens*), thale cress (*Arabidopsis thaliana*) and wine grape (*Vitis vinifera*) were initially sequenced using the Sanger method, although the latest assembly releases have been achieved using newer (next generation) sequencing technologies (The Arabidopsis Genome Initiative, 2000; Lander et al., 2001; Jaillon et al., 2007). Sanger sequencing also led to the identification of conserved genetic features within the ribosomal RNA (rRNA) of eukaryotes and prokaryotes solving evolutionary hypotheses of several taxonomical groups (Fox et al., 1980; White, 1990; Weisburg et al., 1991; Kolbert & Persing, 1999; Pereira et al., 2010; Schoch et al., 2012). In 1986, advances in technology incorporated fluorescent dNTPs to Sanger sequencing, introducing some automation to the process (Smith et al., 1986). Alongside, microbial diversity studies started to take advantage of a faster partially-automated sequencing technique and genetic richness found in metagenomes.

Initial microbial diversity studies took advantage of amplicon sequencing (Tringe et al., 2005; Martiny et al., 2006), which relied on isolation and culture of microbes. However, most of those studies introduced a bias to the analysis due to the lack of uncultured microbes. Potential solutions where all microorganisms are included in the analysis was the use of random sequencing in environmental samples also known as "shotgun" metagenomic sequencing (Tyson et al., 2004; Venter et al., 2004). Yet, time consuming laboratory techniques mainly related to sequence purification (cloning), low data yield and the lack of bioinformatic analysis software made microbial diversity studies very limited.

The advent of high-throughput sequencing (HTS) — also named next generation sequencing (NGS), second generation sequencing or massively parallel sequencing (MPS) — has revolutionized genomics and metagenomics studies (Margulies et al., 2005). Microbial diversity studies benefited greatly mainly due to the elimination of the sequence-cloning process and the large data yield. These sequencing technologies also benefited projects like the human microbiome, as well as numerous environmental microbiomes (Sogin et al., 2006; Turnbaugh et al., 2007; Huttenhower et al., 2012). The flexibility of HTS facilitated amplicon sequencing, as well as shotgun metagenome sequencing. Yet, the lack of computing infrastructure, as well as specialized data analysts has created a bottleneck in the analysis of an increased overwhelming amount of data produced by HTS.

Estimating community composition of environmental samples is the backbone task in microbial diversity studies. One advantage of shotgun metagenome sequencing over amplicon sequencing is the increased potential to detect rare or new taxa. Metagenome sequencing takes advantage of whole genomes available in the sample, however its

sensitivity to detect low titer microbes might be compromised and directly related to the sequencing capacity platform. The most common task used for the detection of microbes in metagenomes and perform a read taxon assignment is the use of public databases (i.e., NCBI or EMBL) (Huson et al., 2007). However, other tools have curated their own non-redundant databases with the intent of increasing performance, sensitivity and specificity (Stobbe et al., 2013; Truong et al., 2015; Espindola et al., 2015). Detection sensitivity by using public databases might also be compromised by an excess number of "unknown" sequences present in the metagenomic databases; nonetheless, specificity could be potentially increased due to the continuous public database curations, which eliminates redundant sequences.

Metagenomic analysis tools that incorporate their own curated databases tend to focus on model organisms or organisms that pose some importance to the research community (Truong et al., 2015). E-probe Diagnostic Nucleic Acid Analysis (EDNA) is a metagenomics analysis tool that uses curated databases containing microbe-specific probe databases (e-probes) for read taxon-assignment (Stobbe et al., 2013; Espindola et al., 2015). It was originally created from a plant pathology diagnostic perspective. It takes advantage of fully/partially sequenced plant pathogen genomes to generate its e-probe databases. As a proof of concept it has been shown that it accurately detects plant pathogens on metagenomics datasets (Espindola et al., 2012, 2015; Schneider et al., 2012; Stobbe et al., 2013). EDNA is a tool that was created to take advantage of increasingly lower sequencing costs and it is expected to be advantageous for diagnostics and the study of phytobiomes. As expected, sequencing costs have decreased and multiple phytobiome initiatives have started to populate the literature, particularly aiming to analyze rhizosphere

and phyllosphere of model systems to infer specific hypothesis-driven plant-microbe interactions (Myrold, Zeglin & Jansson, 2014).

EDNA's initial validation processes successfully took advantage of both RNA and DNA sequencing to increase the likelihood of detecting a given organism. Although the tool was flexible enough to detect eukaryotic plant pathogens by using a combined library (dual library) of RNA and DNA (Espindola et al., 2015), EDNA has not been challenged with metatranscriptomic libraries alone. Metatranscriptomic studies create their libraries by using a variety of options, most of them tend to select mRNA via ribosomal RNA depletion or poly(A) capture (Zhao et al., 2014). However, poly(A) capture selection does not contain all available transcripts typically due to mRNA degradation or the presence of non-poly(A) transcripts. Yet, it is the most preferred method due to its lower cost.

It has become crucial to determine if RNA sequencing is in fact contributing valuable reads when the dual library is created. Therefore, EDNA was challenged with the most frequently used RNA sequencing library preparation (Poly(A) capture). Certainly, there are many fully sequenced and annotated eukaryotic organisms having poly(A) capture RNA sequencing libraries. However, most of such model organisms are not eukaryotic plant pathogens.

Proper metagenomic database exploitation can help to address multiple biodiversity hypotheses. Popular biodiversity inferences in metagenomic studies include the detection of microbes and their function in the ecosystem. Microbe detection in metagenomes has been extensively addressed by a variety of binning pipelines and software. Each detection method has different resolution in terms of taxonomic hierarchy. EDNA was originally

developed to detect plant pathogens at the species taxonomic resolution level. However, to effectively determine the source of intentional infections with pathogens for forensic applications will require discrimination and detection of pathogens at strains or isolate level. Although, EDNA's algorithm by itself would not have any issue to detect specific strains, since only a higher stringency in the alignment and parsing parameters are needed. E-probe design becomes a very challenging task that requires meticulous genome examination and selection of target sequences, which can be extremely time demanding unless automated through bioinformatics tools.

Functional approaches aiming to identify putative metabolic pathways that might alter the microbiota interactions have also been developed, where, orthologous genes are searched using well curated databases, such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2004). The elucidation of putative metabolic pathways can also help to infer putative reactions and outcomes from that ecosystem (Myrold, Zeglin & Jansson, 2014; Seo et al., 2014). However, such analysis requires an assembled metagenome, a task that has been proven to be computationally intensive and time consuming. New approaches avoid the use of reference genomes for metagenome assembly (Nielsen et al., 2014). Similarly, the use of RNA sequencing to infer gene expression profiling in microbiomes has started to be used recently without much success. The difficulty of metatranscriptomes assembly is due mainly to reads coming from homologous transcripts in different organisms. Alternatively, raw reads — instead of assembled metatranscriptomes — could potentially be used to infer gene expression patterns in metatranscriptomes. Yet, the association of reads to infer a potential metabolic pathway would require a previous taxon read assignment and filtering to avoid the association of

reads that might not be truly linked in a metabolic pathway reconstruction (Leimena et al., 2013). Certainly, the aforementioned pipeline would eliminate erroneous read association in metabolic pathways, though time consuming due to the requirement of numerous alignments. A new approach to analyze metatranscriptomic data by using unique signatures in genes that change their regulation pattern based on their environment is needed to reduce the time required to analyze metatranscriptomic data and draw functional hypotheses.

My dissertation addresses the issues highlighted above and has the following objectives.

Objectives

1. Sequence, annotate and compare the draft genome of *Sclerotinia minor* with taxonomically related genomes.
2. Use poly(A) capture RNA sequencing for the validation of EDNA's detection capacity of eukaryotic plant pathogens physiologically active in their hosts, using *S. minor* - peanut as a model patho-system.
3. Design a new pipeline for the detection of functional metabolic pathways during the infection of plant pathogens in metatranscriptomes, using aflatoxin-associated genes of *Aspergillus flavus* growing on corn as a model system.
4. Design a pipeline for the development of e-probes with increased taxonomic resolution in EDNA that allow the detection of strains and/or isolates of eukaryotic plant pathogens in metagenomes using *Pythium aphanidermatum* as a model system.

Literature cited

Espindola, A.S., Garzon, C.D., Fletcher, J. and Schneider, W.L. (2012) Validation of EDNA, a newly developed bioinformatics tool, for detection of *Phakopsora*

pachyrhizi from metagenomic samples. In American Phytopathological Society Meeting 2012., p. 35. St Paul, MN.

Espindola, A.S., Garzon, C.D., Schneider, W.L., Hoyt, P.R., Marek, S.M. and Garzon, C.D. (2015) A new approach for detecting Fungal and Oomycete plant pathogens in Next Generation Sequencing metagenome data utilizing Electronic Probes. *Int. J. Data Min. Bioinfomatics* **12**, 115–128.

Fox, G.E., Stackebrandt, E., Hespell, R.B., et al. (1980) The phylogeny of prokaryotes. *Science.* **209**, 457–463.

Huson, D.H., Auch, A.F., Qi, J. and Schuster, S.C. (2007) MEGAN analysis of metagenomic data. *Genome Res.* **17**, 377–386.

Huttenhower, C., Gevers, D., Knight, R., et al. (2012) Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214.

Jaillon, O., Aury, J.M., Noel, B., et al. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**, 463–467.

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. and Hattori, M. (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **32**, D277–D280.

Kolbert, C.P. and Persing, D.H. (1999) Ribosomal DNA sequencing as a tool for identification of bacterial pathogens. *Curr. Opin. Microbiol.* **2**, 299–305.

Lander, E.S., Linton, L.M., Birren, B., et al. (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921.

Leimena, M.M., Ramiro-Garcia, J., Davids, M., et al. (2013) A comprehensive metatranscriptome analysis pipeline and its validation using human small intestine microbiota datasets. *BMC Genomics* **14**, 1–14.

Margulies, M., Egholm, M., Altman, W.E., et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380.

Martiny, J.B.H., Bohannan, B.J.M., Brown, J.H., et al. (2006) Microbial biogeography: putting microorganisms on the map. *Nat. Rev. Microbiol.* **4**, 102–112.

Myrold, D.D., Zeglin, L.H. and Jansson, J.K. (2014) The Potential of Metagenomic Approaches for Understanding Soil Microbial Processes. *Soil Sci. Soc. Am. J.* **78**, 3.

Nielsen, H.B., Almeida, M., Juncker, A.S., et al. (2014) Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat. Biotechnol.* **32**, 822–828.

Pereira, F., Carneiro, J., Matthiesen, R., Asch, B. van, Pinto, N., Gusmão, L. and Amorim, A. (2010) Identification of species by multiplex analysis of variable-length sequences. *Nucleic Acids Res.* **38**, e203.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA Sequencing with Chain-terminating Inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467.

Schneider, W.L., Stobbe, A.H., Daniels, J., et al. (2012) Next-generation diagnostics: Eliminating the excessive sequence processing associated with next-generation sequencing using EDNA. In American Phytopathological Society Meeting 2012., p. 155. St Paul, MN.

Schoch, C.L., Seifert, K. a., Huhndorf, S., et al. (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc. Natl. Acad. Sci.* **109**, 6241–6246.

Seo, K., Rainer, P.P., Hahn, S., et al. (2014) Tackling soil diversity with the assembly of large, complex metagenomes. *Proc. Natl. Acad. Sci.* **111**, 6115–6115.

Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B.H. and Hood, L.E. (1986) Fluorescence detection in automated DNA sequence analysis. *Nature* **321**, 674–679.

Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., Arrieta, J.M. and Herndl, G.J. (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere." *Proc. Natl. Acad. Sci.* **103**, 12115–12120.

Stobbe, A.H., Daniels, J., Espindola, A.S., Verma, R., Melcher, U., Ochoa-Corona, F., Garzon, C., Fletcher, J. and Schneider, W. (2013) E-probe Diagnostic Nucleic acid Analysis (EDNA): A theoretical approach for handling of next generation sequencing

data for diagnostics. *Microbiol. Methods*.

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* **408**, 796–815.

Tringe, S.G., Mering, C. von, Kobayashi, A., et al. (2005) Comparative metagenomics of microbial communities. *Science* **308**, 554–557.

Truong, D.T., Franzosa, E. a, Tickle, T.L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C. and Segata, N. (2015) MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* **12**, 902–903.

Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-liggett, C., Knight, R. and Gordon, J.I. (2007) The human microbiome project: exploring the microbial part of ourselves in a changing world. *Nature* **449**, 804–810.

Tyson, G.W., Chapman, J., Hugenholtz, P., et al. (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**, 37–43.

Venter, J.C., Remington, K., Heidelberg, J.F., et al. (2004) Environmental Genome Shotgun Sequencing of the Sargasso Sea. *Science (80-.).* **304**, 66–74.

Weisburg, W.G., Barns, S.M., Pelletier, D. a. and Lane, D.J. (1991) 16S ribosomal DNA amplification for phylogenetic study. *J. Bacteriol.* **173**, 697–703.

White, T.J., Bruns, T., Lee, S. and Taylor, J. (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protoc. a Guid. to methods Appl.*, 315–322.

Zhao, W., He, X., Hoadley, K.A., Parker, J.S., Hayes, D.N. and Perou, C.M. (2014) Comparison of RNA-Seq by poly(A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC Genomics* **15**, 1–11.

CHAPTER II

II LITERATURE REVIEW

The plant microbiome: microbial diversity and interactions

When ecosystems are in equilibrium and communities in harmony, organisms successfully benefit from their interactions (mutualism) (Engelmoer, Behm & Toby Kiers, 2014). Symbiotic relationships might vary depending on a number of factors. Equilibrium disruption is caused principally by changes in environment which can either benefit or hinder the development of certain species. Such disruptions might not necessarily create a pathogenic relationship, however, commensalism — a relationship among two organisms where one benefits from the other without causing any harm — might occur and a cascade of events like influx/efflux of certain carbonic elements will likely alter other communities. Ultimately, parasitism will occur when one organism is being benefitted from another (host) and causing damage to the host. However, not all dominant species in a community will become parasitic. Microbe interaction is crucial to maintain an equilibrated community. Phyllosphere and rhizosphere microbes maintain an intimate interaction between each other and the plant (Barea et al., 2005).

Therefore, microbial diversity is crucial to maintain a continuously performing system. Understanding their function has become relevant to agriculture since some micro-communities have shown to increase crop yields and decrease pathogen colonization in rhizospheres.

Plants are part of a community shared with aboveground (phyllosphere) and belowground (rhizosphere) microorganisms. Increased effort has been put on rhizosphere research and its association with plant growth, development and diversity (Bar-Ness et al., 1991; Crowley et al., 1991; Derylo & Skorupska, 1992; Bever, Westover & Antonovics, 1997; Van Der Heijden, Bardgett & Van Straalen, 2008). Yet, very little research has been performed on phyllosphere and its impact on plants and the rhizosphere (Wardle et al., 2004). Plant microbiomes have not been extensively studied mainly due to the time consuming process required in microorganism isolation and the difficultness associated with microbial diversity studies (Curtis, Sloan & Scannell, 2002; Torsvik, Øvre\r as & Thingstad, 2002; Gams, 2007). However, the advent of high throughput sequencing (HTS) has increased the research on rhizosphere microbiomes since cloning and isolation of microorganisms was not anymore necessary for this type of sequencing (Buée et al., 2009; Turner, James & Poole, 2013). A variety of sequencing techniques are currently being utilized to infer the microbial diversity in plants, yet, the most utilized is the amplicon HTS using ribosomal RNA (Van Der Heijden, Bardgett & Van Straalen, 2008; Berendsen, Pieterse & Bakker, 2012; Turner, James & Poole, 2013).

DNA sequencing: past, present and insights into the future

First generation sequencing

DNA sequencing has long been utilized for a variety of studies trying to associate phenotypic traits with genotypes. Initial sequencing developments started in the 1970s with Maxam-Gilbert sequencing method (Maxam & Gilbert, 1977) as well as with Frederick Sanger's chain-termination method (Sanger, Nicklen & Coulson, 1977). Maxam-Gilbert sequencing method — also known as chemical-termination sequencing method — relied on radioactive labeling at the 5' end of the DNA fragments. Once the ssDNA was labeled, the sample was divided into four different reaction mixes. One reaction named "G reaction" would cleave the ssDNA strands where Guanines were present. The second reaction named "A+G reaction" would cleave the ssDNA at both adenine and guanine positions. A third reaction named "T+C reaction" would cleave the ssDNA at thymine and cytosine positions. A final reaction named "C reaction" will cleave the ssDNA at cytosine positions only. The four reactions were then loaded separately and run in denaturing acrylamide gels. A final step included the sequence inference by reading the gel and assigning each band a nucleotide (Maxam & Gilbert, 1977). Finally, the bands needed to be visually analyzed using X-ray and a visual base call.

Sanger sequencing is still widely used today, and its approach has long used the chain-terminating dideoxynucleosides (2,3-dideoxynucleoside triphosphate) method. This is a sequencing-by-synthesis method, where primers, polymerase, deoxynucleosides triphosphate (dNTPs) and dideoxynucleoside triphosphate (ddNTPs) are needed for the reaction to occur. The target DNA is divided into four separate reactions where each will contain all dNTPs in excess and only a small amount of a specific ddNTP (the ddNTP

corresponding to one of the four reactions; A, T, G and C). The goal is to incorporate the ddNTPs only occasionally to each of the four reactions therefore, some chains will be terminated and some will not be terminated until later when there is the need to incorporate another similar ddNTP. Thus, at the end of the reaction, amplified sequence products of different lengths will be obtained and visualized in an electrophoresis gel, by loading each reaction in a separate well. Reading the gel is easier than in the Maxam-Gilbert method, since the chains are terminated only at a specific nucleotide. The sequence that is deduced from the gel will be always complementary to the one that was used as a template sequence (Sanger, Nicklen & Coulson, 1977).

Most of the sequencing methods that have been utilized until now for DNA sequencing derived from the original Sanger sequencing. One particular method has been used for nearly 20 year which is the automated dideoxy Sanger sequencing. In this sequencing method, chain-termination is again used by utilizing a high ratio of dNTPs / ddNTPs. However, ddNTPs have been modified with fluorescent tags of different colors that allowed to mixed them together, without the need of separate reactions. The reaction is automatically performed in a PCR machine and then loaded onto capillary electrophoresis systems, where a laser detector will determine which ddNTP terminated the synthesis of the chain. The final result provides a digital chromatogram where the fluorescent peaks of ddNTPs are visualized (Sanger, Nicklen & Coulson, 1977).

Although DNA sequencing has been automated, large genomes, like those of eukaryotic organisms, have been still very challenging to sequence. The human genome was sequenced by using Sanger sequencing during 13 years and only after almost $3 billion dollars in costs. Other human genome projects sequenced the human genome in 3 years

for $300 million dollars (Geoff Spencer, 2001). The human genome sequencing implied the generation of approximately 30,000 BAC clones and sequencing runs in the hundreds of thousands, to finally partially-finish the 2.91-billion base pair human genome (Venter et al., 2001).

Second generation sequencing

Advancements in chemistry of materials has permitted the use of either porous membranes manufactured by anisotropic etching of fiber optic face plates (Leamon et al., 2003) or sequence-bound membranes to create a new type of high-throughput sequencing (HTS), named second-generation sequencing or Next Generation Sequencing (NGS). Most of the NGS platforms follow one concept which is the performance of small (micro) Polymerase chain reactions (PCRs) in specialized membranes or plates and record the incorporation of nucleotides (sequencing by synthesis) by fluorescence and a detector (specialized camera). The pioneer in the NGS technology was Roche with its 454 sequencer, named after the average read lengths that the sequencer yielded (Margulies et al., 2005). Subsequently, life sciences industries started to take advantage of HTS and a variety of NGS brand names appeared. Currently, one of the most utilized platform is Illumina® due to its extremely high yields, although read lengths are not as long as the 454 reads.

Amongst the variety of NGS platforms available, 454 from Roche® and Illumina® (formerly known as Solexa) have been the most utilized in research. Each technology has its benefits and caveats. There are two very important properties that differentiate DNA sequencing platforms, read length (refers to the length of the sequences produced by the platform) and sequencing yield (refers to the amount of sequences produced by the

platform). Illumina sequencing has always produced a very high yield, however, their reads have been short (average of 75bp) which bring problems during sequence assembling. The 454 platform had longer reads (average of 454bp) which made this sequencer the preferred for genome assembly projects, although its yield was low when compared to Illumina sequencing. Both techniques use different approaches to sequence DNA, mainly related to the way that the DNA fragments are sequenced after library preparation.

Library preparation for both 454 and Illumina use similar approaches, initially the DNA has to be randomly fragmented to produce fragments of lengths specific to the sequencing technique. Successively a size selection process in the fragmented DNA, discards reads shorter or larger than a length range. Subsequently, adaptors specific for each technique were ligated to the fragmented DNA. It is important to mention that both sequencing technologies require double stranded DNA (dsDNA) to permit the ligation of adaptors. After adaptors have been properly attached, sequencing is performed. In Illumina sequencing, the dsDNA fragments with its adaptors are denatured to ssDNA and then added to a glass flowcell containing 2 types of bound-oligonucleotides that are complementary to the adaptors ligated to the fragmented DNA. The ssDNA fragments hybridize to two types of oligonucleotides in the glass flowcell due to complementarity (they can attach either from the 5' end or 3' end) and randomly. Once the fragments are hybridized, one elongation step with polymerase forms a dsDNA on each hybridized fragment. A denaturation step eliminates the original template and now only the complement strand is directly bound to the glass flowcell. The next step includes the amplification of the strands trough clonal bridge-amplification, where the complement strand folds over and the adaptor region hybridizes with the second type of oligonucleotide

in the glass flowcell. Once the hybridization has occurred the folded strand is extended with polymerase to form the complementary strand and a final process of denaturation unfolds the bridge and creates two ssDNA strands that are tethered to the glass flowcell. The process is repeated multiple times for hundreds of millions of strands creating clusters of amplification. A final cleavage process eliminates the complementary strands leaving only the forward strands. Sequencing takes place by using sequencing primers in a sequencing-by-synthesis amplification process. Multiple cycles are run, and on each cycle oligonucleotides compete for the elongation of the ssDNA strands, once one fluorescent oligonucleotide is incorporated, its light is emitted and recorded (light wavelengths are different for each oligonucleotide). For a given amplification cluster, the incorporation of nucleotides occurs simultaneously to increase the light emission. The light emission is recorded by a detector and base calling is performed by associating light wavelength to a specific nucleotide and read length is recorded as the number of cycles for each cluster of amplification, with one sequencing read per cluster. If sequencing was finished at this step, only single-end reads would be obtained. However, the Illumina platform can perform sequencing from both ends of the DNA strands, and produce paired-end sequences. If paired-end sequencing occurs, a subsequent bridge amplification is needed where the forward strand is eliminated and only the reverse strand is left tethered to the glass flowcell. Sequencing for the reverse strand will occur in a similar fashion as the forward strand occurred. The final products of Illumina sequencing are either one (single-end) or two (paired-end) files in fastq format. The files contain information regarding base calling quality and the sequence itself.

Similarly, 454 sequencing performs sequencing-by-synthesis with major differences with the Illumina platform. Specifically, 454 uses DNA-capture beads attached to the hybridization oligonucleotides with sequence adaptors that were originally ligated to the dsDNA. The beads containing the hybridized DNA go through a step named emulsion-PCR (emPCR) where the beads are populated with clones of the same sequence, with one read per bead, in place of Illumina's one read per cluster of amplification. Once the beads have gone through emPCR they are loaded onto a PicoTiterPlate™ which contains hundreds of thousands of pico-wells where the beads fit one per well. Sequencing is performed in 454 machines by synthesizing the second strand of the ssDNAs that are hybridized on the beads. The chemistry of 454 sequencing further differs from Illumina sequencing in the methods for addition of oligonucleotides and light emission. While in Illumina sequencing light at different wavelengths is emitted depending on the incorporation of individual oligonucleotides, in 454 sequencing deoxynucleotide triphosphates are incorporated to the synthesizing strand in flows, by flowing T,A,C,G repeatedly. At each step, once a polymerase catalyzes the incorporation of a triphosphate dNTP into the DNA strand, pyrophosphate (PPi) in a quantity that results equimolar to the amount of incorporated nucleotide is emitted. The next step includes the conversion of PPi to ATP by the enzyme ATP sulfurylase in the presence of adenosine 5' phosphosulfate. Finally, the ATP produced converts luciferin to oxyluciferin, emitting light in amounts proportional to the nucleotides incorporated. Finally, the light emissions are recorded by a camera and a pyrogram® is produced. A final flow with apyrase is performed on the PicoTiterPlate™ which degrades unincorporated dNTPs and the excess of ATP.

Subsequently, another dNTP is added to the reaction and the same process occurs until the synthesis is completed.

In 454 sequencing, the addition of dNTPs is performed systematically one after the other as opposed to Illumina, where all are added at the same time, since they have a fluorescent label that emits light at different wavelengths depending on the nucleotide incorporated. Since the signal emission in 454 sequencing is quantifiable, the incorporation of more than one oligonucleotide to the ssDNA tend to be miscalled in the presence of homopolymers larger than eight (Margulies et al., 2005). However, error rates in base calling for losing synchronism on each bead have ranged from 1-2% for carry-forward and 0.1-0.3% for incomplete extension. In contrast, Illumina error rates are mainly presented as substitutions. Nevertheless, a variety of bioinformatic tools have been designed to deal with such errors, most of them developed directly by the sequencing manufacturers (Schirmer et al., 2015).

Third generation sequencing

Sequencing advancements within the last decade have been enormous, which has increased the difficulty to decide which platform best fits individual research needs. Second-generation sequencing took advantage of sequencing short fragments of ssDNA in parallel and helped to solve a variety of questions from genomics to metatranscriptomics. However, one major caveat of second-generation sequencing was read length. Projects that required assembly, as is the case of genome sequencing projects, and in some cases metagenomics and metatranscriptomics, have faced problems mostly due to the difficultness of assembling complex genetic regions usually containing repetitive sequences that cover lengths larger than the read lengths produced by second-generation

sequencing platforms. Single molecule real time sequencing (SMRT™) technologies from Pacific Biosystems have achieved longer reads by sacrificing yield and have been used for filling genome/transcriptome assemblage gaps (Eid et al., 2009). The SMRT™ platform uses what is named zero-mode waveguides (ZMW) which are "nano-wells" that can hold only up to 20 zeptoliters. As opposed to Illumina sequencing which uses a base-linked fluorescent nuclelotide, SMRT™ uses a phosphor-linked fluorophore which is released once the nucleotide is incorporated to the ssDNA via phosphodiester bond formation catalyzed by a polymerase. At the bottom of ZMW, the DNA polymerase is anchored where the nucleotide incorporations occur.

Another single molecule sequencing approach has been created by Oxford Nanopore™ which released the MinION™ sequencer. What makes this sequencing platform unique is the low cost and small size (USB stick size) of the platform itself. The technology behind this sequencer permits the use of nanopores to sequence DNA. Nanopores are proteins placed in an electrical-resistant synthetic polymer membrane in high quantities (approximately 500 nanopores). Buffer contained at both sides of the nanopores as well as electrical potential applied to the system permit the pass of ions through the nanopores. Measurements of current blockage is recorded each time a nucleotide passes through each pore (Stoddart et al., 2009; Olasagasti et al., 2010) and given the fact that each nucleotide will pose a different resistance to current, their identification can be inferred by interpreting the change in electric conductivity on each pore by using a Hidden Markov Model base-calling method. A motor protein is used to limit the speed at which the DNA sequence passes through the pore, such protein is attached to each DNA strand during the library preparation step. There are benefits associated with

the use of the single molecule sequencing through pores and these are mostly related to larger read lengths, but most importantly the elimination of the PCR step which many sequencing-by-synthesis equipment are still utilizing and therefore carrying the sequence error biases associated with PCR amplification (Aird et al., 2011).

DNA/RNA–omics *in plants*

HTS has permitted deeper readings of various aspects of the plant-microbe interaction system. Principally associated with the discovery of new pathogen species or strains. Although genomic and transcriptomic studies are still far from detailed for most plant pathogens, in some cases draft genomes and draft annotations might be sufficient for a variety of studies. Most plant-microbe interaction -omics efforts lately have been focused on gene expression analysis using RNA sequencing, as well as the identification of either pathogenicity factors or resistance genes on the hosts. However, studies related to microbiomes in the rhizosphere and the phyllosphere as a disease affecting factor have been limited and have recently started to be analyzed using HTS (Hale, Broders & Iriarte, 2014). Amplicon sequencing and metagenome sequencing are effective tools to examine the information contained in microbiomes of plants. Similarly, gene expression analysis of microbial communities (metatranscriptomics) explores deeper into the functional interactions of the microbiome and the regulation of potential metabolic pathways. A variety of factors might hinder metatranscriptomic analysis and most of them are related to microbial diversity and common protein coding regions.

Metagenomics and community composition

Metagenomics nowadays refers to high-throughput random sequencing of genomic DNA from an environmental sample. It has been considered a substitute of high-

throughput amplicon sequencing (i.e. rRNA surveys) because no previous gene enrichment is necessary (PCR amplification). Usually, the DNA is extracted and without further treatment it is sequenced by any HTS platform. Certainly, an advantage of metagenomics is that PCR amplification bias is decreased, consequently increasing the discoverability of certain taxa by providing a more general snapshot of the microbial community since bacterial, viruses and eukaryote sequences are retrieved by the technique. Yet, most of the studies related to plant-microbe interactions have been performed using high-throughput amplicon sequencing (Berendsen, Pieterse & Bakker, 2012; Lundberg et al., 2012; Bulgarelli et al., 2013). Amplicon sequencing relies on the amplification of orthologous genes which are shared among a specific group of microbes. Most of the sequences used for high-throughput amplicon sequencing are found in the rRNA region of both eukaryotes and bacteria (16S/18S rRNA). The technique greatly simplifies microbial community analysis since sequence assembly becomes less intensive when compared with metagenomic analyses (Caporaso et al., 2010). The most utilized sequencing platform to perform amplicon sequencing is Illumina®.

Metatranscriptomics and community function

Metatranscriptomics consist in high-throughput random sequencing of mRNA from environmental samples. Contrary to metagenomes that give little insight into microbe community functions, metatranscriptomes provide a snapshot of a microbiome-wide gene expression analysis. The technique involves mRNA extraction by using either a poly(A) tail transcript selection or a ribosomal RNA depletion technique (Zhao et al., 2014). High quality mRNA purification is followed by cDNA generation, cDNA sequencing, and data analysis. Traditional data analysis requires the use of reference genomes to map sequencing

reads and of statistical inference to determine gene expression. However, in the absence of reference genomes —because most of the transcripts found in the metatranscriptome might come from organisms with non-sequenced and non-annotated genomes — gene expression analysis is performed by pairwise alignments against databases of well-annotated genes and/or proteins (Mitra et al., 2011). The aforementioned techniques require an assembled metatranscriptome. Metatranscriptomic assembly is a difficult task mainly due to redundant reads that might be part of two or more completely different species. Yet, there are some metagenomic assemblers that have been used for metatranscriptomic purposes, like Metavelvet (Namiki et al., 2012), Oases (Schulz et al., 2012), and Trinity (Grabherr et al., 2011), which support only RNA sequencing of single species. Additionally, a dedicated metatranscriptomic analysis tool has been created utilizing de Bruijn graph that requires a reference metagenome, but little is known yet about its performance (Ye & Tang, 2016).

EDNA in Plant Pathology

E-probe Diagnostic for Nucleic acid Analysis (EDNA) is a newly developed bioinformatic pipeline that takes advantage of the aforementioned HTS technologies to detect plant pathogens. Its algorithm relies on the use of unique signatures termed e-probes. These unique signatures are designed for each organism of interest by comparing the target organism genome with the nearest neighbor organism genome. EDNA was originally designed to target mainly plant pathogens, although it has been demonstrated that it can be scalable to other types of organisms (Blagden et al., 2016). Although EDNA's detection scope covered plant pathogens like viruses, bacteria and eukaryotic at the species level. Further scalability is necessary mainly at different taxonomic levels. Strain

discrimination as well as the detection of plant pathogens in metatranscriptomes are areas that should be explored since most binning bioinformatic tools have only focused in metagenomic sequences. The importance of targeting plant pathogens in metatranscriptomes relies on the necessity to know if a pathogen is either actively producing specific proteins, actively growing or actively infecting. Similarly, strain discrimination becomes extremely important when there are plant pathogen strains that are known to be more virulent than others and its timely detection might help to create a proper management strategy.

Model organisms in this study

Pythium aphanidermatum

Pythium aphanidermatum (Edson) Fitzpatrick is a necrotroph plant pathogen first described as *Rheosporangium aphanidermatum* by Edson in 1915 and later renamed as *P. aphanidermatum* by Fitzpatrick in 1923 (Plaats-Niterink, Schimmelcultures & van der Plaats-Niterink, 1981). It is a soil-borne pathogen causing root rot and crown rot in a broad number of hosts, principally in juvenile and succulent host tissue like vegetables and horticultural crops (Hendrix & Campbell, 1973). It is the most commonly isolated species from cucumber greenhouses in conjunction with *P. ultimum* (Moulin, Lemanceau & Alabouvette, 1994; Postma, Willemsen-de Klein & van Elsas, 2000; Postma et al., 2005). *P. aphanidermatum* and other *Pythium* spp. are also an important causal agent of pre-emergence damping-off of seedlings, killing its host before emergence. The disease is more prevalent in greenhouse flats, nursery beds and row crops since pathogen dispersal is favored by irrigation systems. The highest number of infection reports come from soilless or hydroponic system, such is the case of nutrient films, rockwool, sawdust and peatbags

(Columbia & English, 1988; Moulin, Lemanceau & Alabouvette, 1994). Usually the systems are pathogen free, however, the infection might be introduced by contaminated tools, seeds, plants, worker's shoes and most importantly from irrigation water. In addition, the lack of competing microorganisms permits *Pythium* spp. to easily establish on the system (Paulitz, Bélanger & Richard, 2001).

P. aphanidermatum has great economic importance due to its wide host range, and disease management costs in conventional and organic crops remains expensive. If disease is not timely managed, *P. aphanidermatum* can cause plant mortality up to 100% in some susceptible host species (Chellemi et al., 2000). Mature crops infected with *P. aphanidermatum* exhibit poor growth, and yield loss is more noticeable in crops where tap roots are the harvested product as is the case of sugar beet or carrots (Martin & Loper, 1999). Mature plants of susceptible crops can be killed by severe root rots. Furthermore, secondary infections by other plant pathogens are highly likely to occur in the sites where *P. aphanidermatum* has infected. Its virulence varies with soil nutritional status, and several studies have shown that soil amendments that increase carbon and nitrogen content influence the prevalence of the pathogen (Abawi & Crosier, 1992; Grünwald, Hu & Van Bruggen, 2000). For example, the use of green manure as a soil enrichment method results in higher disease severity (Manici, Caputo & Babini, 2004).

Pythium diseases are favored by poorly drained soils. Water accumulation in the soil creates a conducive environment for the development of *P. aphanidermatum* and other damping off causal agents. Oospores are the primary survival structures, which can tolerate water stress in dry soils (Stanghellini & Jr, 1972). Oospore dormancy is broken in the presence of a conducive environment. Nonetheless, certain physiological properties

prevent the synchronized germination of all the oospores in a particular environment at the same time, which enhances the ecological fitness of *Pythium* species (Lumsden & Ayers, 1975). Crucial elements that favor oospore germination include the presence of root exudates and organic matter, which induce chemotactic growth of the germ tube (Nelson & Craft, 1989). Due to the ability of the pathogen to produce zoospores, Pythium diseases are more prevalent in greenhouse flats, nursery beds and row crops since pathogen dispersal is favored by irrigation systems. The highest number of infection reports come from soilless or hydroponic system, such is the case of nutrient films, rockwool, sawdust and peatbags (Columbia & English, 1988; Moulin, Lemanceau & Alabouvette, 1994).

Disease cycle

Infected plant roots are rapidly invaded by white mycelium that soon produces sporangia. Sporangia germinate by producing a short hypha that produces a balloon-like structure at its tip called secondary sporangium. The secondary sporangium contains zoospores which are release free in water and soil surrounding roots. Zoospores encyst and germinate by producing germ tubes that penetrate the host to start a new infection. *P. aphanidermatum* is homothallic and the sexual stage of the pathogen includes the formation of spherical oogonia, with smooth surface, and antheridia in the diploid mycelium. Selfing as well as cross fertilization between different strains is possible. The antheridium fertilizes the oogonium and a zygote is formed. The fertilized oogonium undergoes meiosis and karyogamy where a diploid oospore forms, then a thick cell wall is formed for protection. Oospores serve as overwinter or resting structures that germinate when environmental conditions are optimal for either direct germination with an penetration peg, or indirectly by producing a sporangium containing zoospores which later

will encyst and germinate in the presence of host tissue, causing a new infection (Agrios, 2005).

Common disease diagnosis methods

P. aphanidermatum infection is associated with symptoms like wilted plants and seedlings. Signs can be observed as a while mycelia growing either on the soil or the plant roots. Plant roots are observed rotten and completely fragile. However, molecular techniques for plant disease diagnostics has become popular among plant pathologists due to their sensitivity, specificity and rapidness. *P. aphanidermatum* detection has been successfully achieved by using Polymerase chain reaction (PCR). Specifically multiplex PCR and loop-mediated isothermal amplification (LAMP) are molecular detection techniques widely cited in the literature (Wang, Wang & White, 2003; Schroeder et al., 2006; Fukuta et al., 2013; Ishiguro et al., 2013). In oomycete and fungal plant pathogens, disease diagnosis performed by molecular techniques rely on the uniqueness of the rDNA Internal Transcribed Spacer (ITS) which, although being conserved among many *Pythium* spp., it still permits the design of species-specific primers. However, ITS is a highly conserved sequence within *P. aphanidermatum* isolates (Lee, Garzon & Moorman, 2010), thus the design of molecular techniques that can discriminate among isolates/strains might become very challenging. To overcome the low genetic diversity found in the ITS region for *P. aphanidermatum*, other loci should be used for the design of new disease diagnosis molecular techniques.

Management strategies

Usually disease control in ornamental plants is achieved by the use of fungicides. Metalaxyl is the most popular fungicide, which is used as a soil drench, however for

vegetables and horticultural, crops fungicide application is regulated and enforced in compliance with the Pesticide Residues in Food and Feed section, which is regulated by the U.S. Food and Drug Administration (FDA) ("Compliance Policy Guides - CPG Sec. 575.100 Pesticide Residues in Food and Feed - Enforcement Criteria"). The necessity to meet the regulated criteria has permitted the emergence of disease control alternatives like biological control (Hultberg, Alsanius & Sundin, 2000). Biological control has been tested in multiple *Pythium* spp. by using *Bacillus subtilis* for its known antifungal activity, however, the trials showed no antifungal activity against *P. aphanidermatum* (Paulitz, Bélanger & Richard, 2001). On the contrary, strains of *Pseudomomas fluorescence* were found to induce resistance against *P. aphanidermatum* in cucumber roots (Paulitz, Bélanger & Richard, 2001). Control trials have included the use of microbial communities that were shown to successfully suppress the growth of *P. aphanidermatum* in soilless systems (Postma, Willemsen-de Klein & van Elsas, 2000; Postma et al., 2005). Finally the use of arbuscular mycorrhizal fungi for suppressing *P. aphanidermatum* growth in tomato roots has been successfully tested, triggering the expression of pathogenicity related proteins PR-1 genes (Larsen et al., 2011). *P. aphanidermatum* being a soil-borne pathogen may not be eradicated from soil due to its ubiquitous presence, however, disease control is mainly favored by preventative techniques like soilless and hydroponic crop systems as well as biological control.

P. aphanidermatum, being a soil-borne pathogen, is difficult to eradicate from soil due to its ubiquitous presence, however, disease control is mainly favored by preventative techniques like soilless and hydroponic crop systems as well as biological control, and chemical control. Pythium diseases can be managed with fungicides, of which the most

commonly used metalaxyl, which is used as a soil drench and has systemic effects. Unfortunately, resistance to metalaxyl and its enantiomer mefenoxam has been found extensively, hence rotation plans with fungicides containing other active ingredients and alternative modes of action are necessary (Harvey & Lawrence, 2008).

Sclerotinia minor

Sclerotinia minor Jagger produces a survival structure called Sclerotia which is also considered its inoculum source. *S. minor* ascospore production is less frequent than *S. sclerotiorum* (Lib) de Bary under natural circumstances. Therefore, its main dispersion means is usually by rain or irrigation systems. This pathogen is considered a necrotroph, therefore it produces cell-wall degrading emzymes to disrupt host tissue. Factors like soil moisture and temperature affect the survival and germination of sclerotia. Sclerotial germination and mycelial growth occurs from 6 to 30 C, being the optimum temperature at 18 C (Imolehin & Grogan, 1980; Wu & Subbarao, 2008). Optimal sclerotial depths on soil range from 2 to 14 cm in order to maintain either the same or higher number of sclerotia. Sclerotial position and duration in soil, sclerotial shape, activities of other microorganisms and nutrition are also considered to affect the germination of the sclerotia (Coley-Smith et al., 1971; Imolehin & Grogan, 1980; Alexander & Stewart, 1994).

S. minor is considered an important pathogen on several economically important crops, some of them include Soybeans, sunflowers, lettuce, common bean and most importantly peanut in Oklahoma, North Carolina and Texas. The pathogen causing the disease lettuce drop might cause crop losses as high as 75%.

Disease cycle

S. minor disease cycle includes a sclerotial stage (resting/survival structure). Infection of the host is triggered by high humidity causing direct germination of mycelia or indirectly (less common/*in vitro*) by the germination of fruiting bodies (ascocarps) and the posterior production of ascospores (Beach, 1921; Abawi & Grogan, 1979; Adams, P. B., Ayers, 1979; Dillard & Grogan, 1985; Patterson & Grogan, 1985). Secondary inoculum has not been reported for *S. minor,* therefore, dispersal of the inoculum is restricted to sclerotia movement by either animals or agricultural tools (Abawi & Grogan, 1979). Therefore, the incidence of *S. minor* diseases are directly correlated with the density of sclerotia in soil, as opposed to *S. sclerotiorum* disease incidence which is directly correlated with a conducive environment for ascospores dispersal (Hawthorne, 1973; Ekins, Aitken & Goulter, 2002). Infection with *S. minor* to lettuce can occur at any growth stage and irrigation systems are the main dispersal mechanism for the pathogen (Wu & Subbarao, 2006).

Common disease diagnosis methods

Diagnostics of *S. minor* is usually performed by looking at signs and symptoms of the disease. Sclerotial formation on peanut plants of lettuce confirms the presence of *Sclerotinia* spp. The discrimination between *S. minor* and *S. sclerotiorum* is commonly performed visually by comparing sclerotia sizes. *S. sclerotiorum* usually produces larger sclerotia than *S. minor.* However, diagnostics confirmation at species level is done using molecular techniques. Discriminatory simplex and multiplex PCR have been developed to successfully detect these necrotrophic plant pathogens (Abd-Elmagid et al., 2013).

Management strategies

S. minor dispersal is caused principally by irrigation, contaminated tools, contaminated machinery, seed transmission and poor soil movement strategies. Airborne dispersal is less common since the resting structures which are the primary inoculum are dense to be air-borne. Studies have shown that the use of irrigation that un-favors sclerotial movement might be considered as an important management strategy, specifically the use of surface drip irrigation instead of furrow irrigation in lettuce fields (Wu & Subbarao, 2006). Biological control has been tested by using *Coniothyrium minitans*. The organism is integrated to regular disease management with fungicides (Partridge, 2006).

Aspergillus flavus

Aspergillus flavus produces extremely high quantities of conidia which helps to its wide distribution around the world. The spores are easily disseminated by the wind. Environmental factors strongly influence the prevalence of *A. flavus* in the air, consequently having different *Aspergillus* spp. geographically distributed in different climates (Calvo et al., 1980; Guinea et al., 2005). *A. flavus* is a saprophyte organism in the soil and its optimum temperature is 37 °C, however, growth temperatures range from 12 to 48 °C (Hedayati et al., 2007). It overwinters as mycelium or as sclerotia, where sclerotia germinates to produce either hyphae or conidia for dispersion. Usually the field contamination is triggered by high temperatures and drought stress. Its ability to survive extreme environmental conditions allow it to outcompete other microorganisms for nutrients in the soil. It is highly prevalent in regions like Arizona where summer season contains long dry periods (Cotty, 1989). *A. flavus* is considered a mild plant pathogen, principally of corn, peanuts a cotton (Klich, 2007). The prevalence of *A. flavus* on corn

(*Zea mays*) is high and causes a disease called ear rot (Taubenhaus, 1920). In peanuts it affects the seedlings causing the disease yellow mould or aflaroot but, it can also affect mature peanuts by causing rot of peanuts (Pettit, 1984). Another important niche is cotton (*Gossypium hirsutum*) plants by causing the disease called boll rot, however, infection of the fibers is known as yellow spot disease (Marsh et al., 1955).

Maize is one of the world important staple crops in the world after wheat and rice by occupying almost 150 Million Ha. But, it is also the staple crop with the highest production with more than 600 million metric tons in 2003 (Strange & Scott, 2005). Economic losses might not occur due to crop loss, but it occurs due to crop rejection since it does not meet mycotoxin standards. Such crops, contaminated with aflatoxins are usually either redirected or disposed in cases of severe contamination (Binder et al., 2007). In the United States the economic impact of aflatoxin was estimated in $225 million/yr (Schmale & Munkvold, 2009).

Disease cycle

A. flavus reproduces both sexually (teleomorph=*Petromyces*) and asexually (anamorph=*Aspergillus*) (Horn, Moore & Carbone, 2009). Sexual reproduction is heterothallic, therefore it needs a partner to cross and reproduce sexually. The anamorph produces conidiospores which are the main primary inoculum in its life cycle and the most prevalent reproductive stage. However, mycelium and sclerotia contribute to the dispersal of the organism. Sclerotia is the most common source of inoculum on host plants during the following growing season. The primary inoculum are usually located in plant debris and litter on the soil or buried in the soil. Spores are carried by the wind or insects from corn kernels or soil principally. Flowering maize is the most susceptible stage for infection

with *A. flavus* (Jones et al., 1980). Conidial development on silk and kernels occurs during the season (Marsh & Payne, 1984). The pathogen becomes dominant in the host when low humidity and high temperatures are present, favoring conidial development and dispersal. Secondary inoculum has been identified mostly in cotton fields, the main inoculum are conidiospores.

Common disease diagnosis methods

Diagnosis by examination of signs or symptoms does not permit an accurate identification of *A. flavus* due to its multiple overlapping morphological and biochemical characteristics with other close relative in the *Aspergillus* s. str. (Cotty, 1989; Geiser et al., 2000; Hedayati et al., 2007; Samson et al., 2014). *A. flavus* isolates have been classified phenotypically based mainly on sclerotia size and aflatoxin type. In fact there is a study that found a strong relationship between the formation of sclerotia and aflatoxin production suggesting a coregulation (Cotty, 1988). Documented sclerotial morphotypes include L strain and S strain. L strain produce conidiospores and few sclerotia that are usually larger than 400 uM in diameter (Cotty, 1989; Bayman & Cotty, 1993; Horn & Dorner, 1998), whereas S strain produces less conidiospores but multiple sclerotia smaller than 400 uM in diameter and also produces higher amounts of aflatoxin (Bayman & Cotty, 1993). The variability of single isolates and communities of *A. flavus* in the production of aflatoxin (Schroeder & Boller, 1973; Cotty, 1989, 1997; Garber & Cotty, 1997; Horn & Dorner, 1999) has led to the use of strain frequency quantification as a method to determine the presence of toxigenic and/or atoxigenic strains by using Vegetative Compatibility Groups (VCG) (Leslie, 1993). Molecular characterization of toxigenic/atoxigenic strains has been performed in the Aflatoxin gene cluster pathway. Multiple deletions have been found in

atoxigenic strains of *A. flavus* permitting the use of such information for the development of molecular screening tools for rapid identification of atoxigenic strains of *A. flavus* (Chang, Horn & Dorner, 2005; Donner et al., 2010; Callicott & Cotty, 2015).

Management strategies

Aspergillus flavus is a widespread organism due to their capabilities to survive harsh environmental conditions and their ease of dispersal by the production of conidiospores. Cotton and Maize being the most economical important crops attacked by this organism are difficult to manage by application of fungicides due to toxic concerns. Therefore, the main management strategy is the use of atoxigenic strains on the field. Competitive exclusion of toxigenic strains by the occupation of the same niche by atoxigenic strain has been proven to be very effective (Cotty, 1994; Garber & Cotty, 1997).

Literature cited

Abawi, G. and Crosier, D. (1992) Influence of reduced tillage practices on root rot severity and yield of snap beans, 1991. *Biol. Cult. Test* **7**.

Abawi, G.S. and Grogan, R.G. (1979) Epidemiology of Diseases Caused by Sclerotinia Species. *Phytopathology* **69**, 899–904.

Abd-Elmagid, A., Garrido, P.A., Hunger, R., Lyles, J.L., Mansfield, M.A., Gugino, B.K., Smith, D.L., Melouk, H.A. and Garzon, C.D. (2013) Discriminatory simplex and multiplex PCR for four species of the genus Sclerotinia. *J. Microbiol. Methods* **92**, 293–300.

Adams, P. B., Ayers, W. a. (1979) Ecology of Sclerotinia Species. *Phytopathology* **69**, 896–899.

Agrios, G. (2005) Plant Pathology Fifth, ed, Elsevier.

Aird, D., Ross, M.G., Chen, W.-S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C. and Gnirke, A. (2011) Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* **12**, 1–14.

Alexander, B.J.R. and Stewart, a. (1994) Survival of sclerotia of Sclerotinia and Sclerotium spp in New Zealand horticultural soil. *Soil Biol. Biochem.* **26**, 1323–1329.

Barea, J.-M., Pozo, M.J., Azcón, R. and Azcón-Aguilar, C. (2005) Microbial co-operation in the rhizosphere. *J. Exp. Bot.* **56**, 1761–1778.

Bar-Ness, E., Chen, Y., Hadar, Y., Marschner, H. and Römheld, V. (1991) Siderophores of *Pseudomonas putida* as an iron source for dicot and monocot plants. *Plant Soil* **130**, 231–241.

Bayman, P. and Cotty, P.J. (1993) Genetic diversity in *Aspergillus flavus*: association with aflatoxin production and morphology. *Can. J. Bot.* **71**, 23–31.

Beach, W. (1921) The lettuce“ drop” due to *Sclerotinia minor* 165th ed., Pennsylvania State College Agricultural Experiment Station.

Berendsen, R.L., Pieterse, C.M.J. and Bakker, P.A.H.M. (2012) The rhizosphere microbiome and plant health. *Trends Plant Sci.* **17**, 478–486.

Bever, J.D., Westover, K.M. and Antonovics, J. (1997) Incorporating the Soil Community into Plant Population Dynamics: The Utility of the Feedback Approach. *J. Ecol.* **85**, 561–573.

Binder, E.M., Tan, L.M., Chin, L.J., Handl, J. and Richard, J. (2007) Worldwide

occurrence of mycotoxins in commodities, feeds and feed ingredients. *Anim. Feed Sci. Technol.* **137**, 265–282.

Blagden, T., Schneider, W., Melcher, U., Daniels, J. and Fletcher, J. (2016) Adaptation and Validation of E-Probe Diagnostic Nucleic Acid Analysis for Detection of *Escherichia coli* O157:H7 in Metagenomic Data from Complex Food Matrices. *J. Food Prot.* **79**.

Buée, M., Reich, M., Murat, C., Morin, E., Nilsson, R.H., Uroz, S. and Martin, F. (2009) 454 Pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. *New Phytol.* **184**, 449–456.

Bulgarelli, D., Schlaeppi, K., Spaepen, S., Themaat, E.V.L. van and Schulze-Lefert, P. (2013) Structure and Functions of the Bacterial Microbiota of Plants. *Annu. Rev. Plant Biol.* **64**, 807–838.

Callicott, K. a. and Cotty, P.J. (2015) Method for monitoring deletions in the aflatoxin biosynthesis gene cluster of *Aspergillus flavus* with multiplex PCR. *Lett. Appl. Microbiol.* **60**, 60–65.

Calvo, M.A., Guarro, J., Suarez, G. and Ramirez, C. (1980) Airborne fungi in the air of barcelona, Spain. V. The yeasts. *Ann. Allergy* **45**, 115–6.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., et al. (2010) QIIME allows analysis of high- throughput community sequencing data. *Nature* **7**, 335–336.

Chang, P.K., Horn, B.W. and Dorner, J.W. (2005) Sequence breakpoints in the aflatoxin biosynthesis gene cluster and flanking regions in nonaflatoxigenic *Aspergillus flavus*

isolates. *Fungal Genet. Biol.* **42**, 914–923.

Chellemi, D.O., Mitchell, D.J., Kannwischer-Mitchell, M.E., Rayside, P. a. and Rosskopf, E.N. (2000) Pythium spp. Associated with Bell Pepper Production in Florida. *Plant Dis.* **84**, 1271–1274.

Coley-Smith, J.R., Cooke, R.C., Of, G. and Sclerotia, F. (1971) Survival and Germination of Fungal Sclerotia. *Annu. Rev. Phytopathol.* **9**, 65–92.

Columbia, I. and English, L. (1988) Pythium spp. Associated with Crown Rot of Cucumbers in British Columbia Greenhouses. *Plant Dis.* **72**, 683–687.

Cotty, P. (1988) Aflatoxin and Sclerotial Production by *Aspergillus flavus*; Influence of pH. *Growth* **78**, 1250–1253.

Cotty, P.J. (1997) Aflatoxin-producing potential of communities of Aspergillus section Flavi from cotton producing areas in the United States. *Mycol. Res.* **101**, 698–704.

Cotty, P.J. (1994) Influence of field application of an atoxigenic strains of *Aspergillus flavus* on the populations of A. flavus infection cotton balls and on the aflatoxin content of cottonseed. *Phytopathology* **84**, 1270–1277.

Cotty, P.J. (1989) Virulence and cultural characteristics of two Aspergillus Strains Pathogenic to Cotton. *Phytopathology* **79**, 808–814.

Crowley, D.E., Wang, Y.C., Reid, C.P.P. and Szaniszlo, P.J. (1991) Mechanisms of iron acquisition from siderophores by microorganisms and plants. *Plant Soil* **130**, 179–198.

Curtis, T.P., Sloan, W.T. and Scannell, J.W. (2002) Estimating prokaryotic diversity and

its limits. *Proc. Natl. Acad. Sci.* **99**, 10494–10499.

Derylo, M. and Skorupska, A. (1992) Rhizobial siderophore as an iron source for clover. *Physiol. Plant.* **85**, 549–553.

Dillard, H.R. and Grogan, R.G. (1985) Relationship between sclerotial spatial pattern and density of *Sclerotinia minor* and the incidence of lettuce drop. *Phytopathology* **75**, 90–94.

Donner, M., Atehnkeng, J., Sikora, R. a, Bandyopadhyay, R. and Cotty, P.J. (2010) Molecular characterization of atoxigenic strains for biological control of aflatoxins in Nigeria. *Food Addit. Contam. Part A. Chem. Anal. Control. Expo. Risk Assess.* **27**, 576–590.

Eid, J., Fehr, A., Gray, J., et al. (2009) Real-time DNA sequencing from single polymerase molecules. *Science.* **323**, 133–138.

Ekins, M., Aitken, E. and Goulter, K. (2002) Carpogenic germination of *Sclerotinia minor* and potential distribution in Australia. *Australas. Plant Pathol.* **31**, 259–265.

Engelmoer, D.J.P., Behm, J.E. and Toby Kiers, E. (2014) Intense competition between arbuscular mycorrhizal mutualists in an in vitro root microbiome negatively affects total fungal abundance. *Mol. Ecol.* **23**, 1584–1593.

FDA (2015) Compliance Policy Guides - CPG Sec. 575.100 Pesticide Residues in Food and Feed - Enforcement Criteria. *Food Drug Adm.*

Fukuta, S., Takahashi, R., Kuroyanagi, S., et al. (2013) Detection of *Pythium aphanidermatum* in tomato using loop-mediated isothermal amplification (LAMP)

with species-specific primers. *Eur. J. Plant Pathol.* **136**, 689–701.

Gams, W. (2007) Biodiversity of soil-inhabiting fungi. *Biodivers. Conserv.* **16**, 69–72.

Garber, R.K. and Cotty, P.J. (1997) Formation of Sclerotia and Aflatoxins in Developing Cotton Bolls Infected by the S Strain of *Aspergillus flavus* and Potential for Biocontrol with an Atoxigenic Strain. *Phytopathology* **87**, 940–945.

Geiser, D.M., Dorner, J.W., Horn, B.W. and Taylor, J.W. (2000) The phylogenetics of mycotoxin and sclerotium production in *Aspergillus flavus* and *Aspergillus oryzae*. *Fungal Genet. Biol.* **31**, 169–179.

Geoff Spencer (2001) International Human Genome Sequencing Consortium Publishes Sequence and Analysis of the Human Genome. *WASHINGTON, D.C. - Hum. Genome Proj. Int. Consort. today Announc. Publ. a Draft Seq. Initial Anal. Hum. genome - Genet. Bluepr. a Hum. being. Pap. Appear. Feb. 15 issue jour*.

Grabherr, M.G., Haas, B.J., Yassour, M., et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–52.

Grünwald, N.J., Hu, S. and Bruggen, a. H.C. Van (2000) Short-term cover crop decomposition in organic and conventional soils: Characterization of soil C, N, microbial and plant pathogen dynamics. *Eur. J. Plant Pathol.* **106**, 37–50.

Guinea, J., Peláez, T., Alcalá, L. and Bouza, E. (2005) Evaluation of Czapeck agar and Sabouraud dextrose agar for the culture of airborne Aspergillus conidia. *Diagn. Microbiol. Infect. Dis.* **53**, 333–334.

Hale, I.L., Broders, K. and Iriarte, G. (2014) A Vavilovian approach to discovering crop-associated microbes with potential to enhance plant immunity. *Front. Plant Sci.* **5**, 492.

Harvey, P. and Lawrence, L. (2008) Managing Pythium Root Disease Complexes to Improve Productivity of Crop Rotations. *Outlooks Pest Manag.* **19**, 127–129.

Hawthorne, B. (1973) Production of apothecia of *Sclerotinia minor*. *New Zeal. J. Agric. Res.* **16**, 559–560.

Hedayati, M.T., Pasqualotto, a. C., Warn, P. a., Bowyer, P. and Denning, D.W. (2007) *Aspergillus flavus*: Human pathogen, allergen and mycotoxin producer. *Microbiology* **153**, 1677–1692.

Heijden, M.G.A. Van Der, Bardgett, R.D. and Straalen, N.M. Van (2008) The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecol. Lett.* **11**, 296–310.

Hendrix, F.F. and Campbell, W.A. (1973) Pythiums as Plant Pathogens. *Annu. Rev. Phytopathol.* **11**, 77–98.

Horn, B.W. and Dorner, J.W. (1999) Regional Differences in Production of Aflatoxin B1 and Cyclopiazonic Acid by Soil Isolates of *Aspergillus flavus* along a Transect within the United States Regional Differences in Production of Aflatoxin B1 and Cyclopiazonic Acid by Soil Isolates of Aspe. **65**, 1444–1449.

Horn, B.W. and Dorner, J.W. (1998) Soil populations of Aspergillus species from section Flavi along a transect through peanut-growing regions of the United States. *Mycologia*

90, 767–776.

Horn, B.W., Moore, G.G. and Carbone, I. (2009) Sexual reproduction in *Aspergillus flavus*. *Mycologia* **101**, 423–429.

Horn, B.W. and Pitt, J.I. (1997) Yellow mold and aflatoxin. In Compendium of peanut diseases, Second Edition. (Kokalis-Burelle, N., Porter, D.M., Rodríguez-Kábana, R., Smith, D.H., and Subrahmanyam, P., eds), pp. 40–42.

Hultberg, M., Alsanius, B. and Sundin, P. (2000) In vivo and in vitro interactions between *Pseudomonas fluorescens* and *Pythium ultimum* in the suppression of damping-off in tomato seedlings. *Biol. Control* **19**, 1–8.

Imolehin, E. and Grogan, R. (1980) Effect of temperature and moisture tension on growth, sclerotial production, germination, and infection by *Sclerotinia minor*. *Phytopathology* **70**, 1153–1157.

Ishiguro, Y., Asano, T., Otsubo, K., Suga, H. and Kageyama, K. (2013) Simultaneous detection by multiplex PCR of the high-temperature-growing Pythium species: *P. aphanidermatum*, *P. helicoides* and *P. myriotylum*. *J. Gen. Plant Pathol.* **79**, 350–358.

Jones, R.H., Duncan, H.E., Payne, G. a and Leonard, H.J. (1980) Factors influencing infection by *Aspergillus flavus* in silk-inoculated corn. *Plant Dis* **64**, 859.

Klich, M. a. (2007) *Aspergillus flavus*: The major producer of aflatoxin. *Mol. Plant Pathol.* **8**, 713–722.

Larsen, J., Graham, J.H., Cubero, J. and Ravnskov, S. (2011) Biocontrol traits of plant growth suppressive arbuscular mycorrhizal fungi against root rot in tomato caused by

Pythium aphanidermatum. *Eur. J. Plant Pathol.* **133**, 361–369.

Leamon, J.H., Lee, W.L., Tartaro, K.R., Lanza, J.R., Sarkis, G.J., deWinter, A.D., Berka, J. and Lohman, K.L. (2003) A massively parallel PicoTiterPlate™ based platform for discrete picoliter-scale polymerase chain reactions. *Electrophoresis* **24**, 3769–3777.

Lee, S., Garzon, C.D. and Moorman, G.W. (2010) Genetic structure and distribution of *Pythium aphanidermatum* populations in Pennsylvania greenhouses based on analysis of AFLP and SSR markers. *Mycologia* **102**, 774–784.

Leslie, J.F. (1993) Fungal vegetative compatibility. *Annu. Rev. Phytopathol.* **31**, 127–150.

Lumsden, R.D. and Ayers, W.A. (1975) Influence of soil environment on the germinability of constitutively dormant oospores of *Pythium ultimum*. *Phytopathology* **65**, 1101–1107.

Lundberg, D.S., Lebeis, S.L., Paredes, S.H., et al. (2012) Defining the core *Arabidopsis thaliana* root microbiome. *Nature* **488**, 86–90.

Manici, L.M., Caputo, F. and Babini, V. (2004) Effect of green manure on Pythium spp. population and microbial communities in intensive cropping systems. *Plant Soil* **263**, 133–142.

Margulies, M., Egholm, M., Altman, W.E., et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380.

Marsh, P.B., Bollenbacher, K., San Antonio, J.P. and Merola, G.V. (1955) Observations on Certain Fluorescent Spots in Raw Cotton Associated with the Growth

of Microorganisms. *Text. Res. J.* **25**, 1007–1016.

Marsh, S.F. and Payne, G. a. (1984) Preharvest Infection of Corn Silks and Kernels by *Aspergillus flavus*. *Phytopathology* **74**, 1284–1289.

Martin, F.N. and Loper, J.E. (1999) Soilborne plant diseases caused by Pythium spp. ecology, epidemiology, and prospects for biological control. *CRC. Crit. Rev. Plant Sci.* **18**, 111–181.

Maxam, A.M. and Gilbert, W. (1977) A new method for sequencing DNA. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 560–564.

Mitra, S., Rupek, P., Richter, D.C., Urich, T., Gilbert, J.A., Meyer, F., Wilke, A. and Huson, D.H. (2011) Functional analysis of metagenomes and metatranscriptomes using SEED and KEGG. *BMC Bioinformatics* **12**, 1–8.

Moulin, F., Lemanceau, P. and Alabouvette, C. (1994) Pathogenicity of Pythium species on cucumber in peat-sand, rockwool and hydroponics. *Eur. J. plant Pathol.* **100**, 3–17.

Namiki, T., Hachiya, T., Tanaka, H. and Sakakibara, Y. (2012) MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Res.* **40**, e155.

Nelson, E. and Craft, C. (1989) Comparative germination of culture-produced and plant-produced sporangia of *Pythium ultimum* in response to soluble seed exudates and exudate. *Phytopathology* **79**, 1009–1013.

Olasagasti, F., Lieberman, K.R., Benner, S., Cherf, G.M., Dahl, J.M., Deamer, D.W.

and Akeson, M. (2010) Replication of individual DNA molecules under electronic control using a protein nanopore. *Nat. Nanotechnol.* **5**, 798–806.

Partridge, D. (2006) Potential for management of Sclerotinia blight of peanut (*Arachis hypogaea* L.) caused by *Sclerotinia minor* with the biological control agent *Coniothyrium minitans*. North Carolina State University.

Patterson, C.L. and Grogan, R.G. (1985) Differences in epidemiology and control of lettuce drop caused by *Sclerotinia minor* and *S. sclerotiorum. Plant Dis.* **69**, 766–770.

Paulitz, T.C., Bélanger, R.R. and Richard, R.B. (2001) Biological control in greenhouse systems. *Annu. Rev. Phytopathol.* **39**, 103–33.

Plaats-Niterink, A. van der (1981) Monograph of the genus Pythium 1st ed., Centraalbureau voor Schimmelcultures Baarn.

Postma, J., Geraats, B.P.J., Pastoor, R. and Elsas, J.D. van (2005) Characterization of the Microbial Community Involved in the Suppression of *Pythium aphanidermatum* in Cucumber Grown on Rockwool. *Phytopathology* **95**, 808–18.

Postma, J., Willemsen-de Klein, M.J. and Elsas, J.D. van (2000) Effect of the Indigenous Microflora on the Development of Root and Crown Rot Caused by *Pythium aphanidermatum* in Cucumber Grown on Rockwool. *Phytopathology* **90**, 125–33.

Samson, R.A., Visagie, C.M., Houbraken, J., et al. (2014) Phylogeny, identification and nomenclature of the genus Aspergillus. *Stud. Mycol.* **78**, 141–173.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA Sequencing with Chain-

terminating Inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467.

Schirmer, M., Ijaz, U.Z., D'Amore, R., Hall, N., Sloan, W.T. and Quince, C. (2015) Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res.* **43**, e37.

Schmale, D. and Munkvold, G. (2009) Mycotoxins in crops: A threat to human and domestic animal health. *plant Heal. Instr*.

Schroeder, H.W. and Boller, R. a (1973) Aflatoxin production of species and strains of the *Aspergillus flavus* group isolated from field crops. *Appl. Microbiol.* **25**, 885–889.

Schroeder, K.L., Okubara, P.A., Tambong, J.T., Lévesque, C.A. and Paulitz, T.C. (2006) Identification and Quantification of Pathogenic Pythium spp. from Soils in Eastern Washington Using Real-Time Polymerase Chain Reaction. *Phytopathology* **96**, 637–47.

Schulz, M.H., Zerbino, D.R., Vingron, M. and Birney, E. (2012) Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28**, 1086–1092.

Stanghellini, M. and Jr, E.N. (1972) Occurrence and survival of *Pythium aphanidermatum* under arid soil conditions in Arizona. *Plant Dis. Report.* **56**, 507–510.

Stoddart, D., Heron, A.J., Mikhailova, E., Maglia, G. and Bayley, H. (2009) Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *Proc. Natl. Acad. Sci.* **106**, 7702–7707.

Strange, R.N. and Scott, P.R. (2005) Plant disease: a threat to global food security. *Annu. Rev. Phytopathol.* **43**, 83–116.

Taubenhaus, J. (1920) A study of the black and the yellow molds of ear corn 270th ed., Texas Agricultural Experiment Station.

Torsvik, V., Øvre\r as, L. and Thingstad, T.F. (2002) Prokaryotic Diversity--Magnitude, Dynamics, and Controlling Factors. *Science.* **296**, 1064–1066.

Turner, T.R., James, E.K. and Poole, P.S. (2013) The plant microbiome. *Genome Biol.* **14**, 1–10.

Venter, J.C., Adams, M.D., Myers, E.W., et al. (2001) The Sequence of the Human Genome. *Science (80-.).* **291**, 1304–1351.

Wang, P.H., Wang, Y.T. and White, J.G. (2003) Species-specific PCR primers for Pythium developed from ribosomal ITS1 region. *Lett. Appl. Microbiol.* **37**, 127–132.

Wardle, D.A., Bardgett, R.D., Klironomos, J.N., Setälä, H., Putten, W.H. van der and Wall, D.H. (2004) Ecological Linkages Between Aboveground and Belowground Biota. *Science.* **304**, 1629–1633.

Wu, B. and Subbarao, K. (2006) Analyses of lettuce drop incidence and population structure of *Sclerotinia sclerotiorum* and *S. minor*. *Phytopathology* **96**, 1322–1329.

Wu, B.M. and Subbarao, K. V (2008) Effects of soil temperature, moisture, and burial depths on carpogenic germination of *Sclerotinia sclerotiorum* and *S. minor*. *Phytopathology* **98**, 1144–1152.

Ye, Y. and Tang, H. (2016) Utilizing de Bruijn graph of metagenome assembly for

metatranscriptome analysis. *Bioinformatics* **32**, 1001–1008.

Zhao, W., He, X., Hoadley, K.A., Parker, J.S., Hayes, D.N. and Perou, C.M. (2014) Comparison of RNA-Seq by poly(A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC Genomics* **15**, 1–11.

CHAPTER III

III GENOME SEQUENCING AND COMPARATIVE GENOMICS OF *SCLEROTINIA MINOR*, THE CAUSAL AGENT OF SCLEROTINIA BLIGHT PROVIDES INSIGHTS INTO ITS EVOLUTION AND INFECTION STRATEGY

Abstract

Sclerotinia minor is a necrotrophic plant pathogen that is closely related to *Sclerotinia sclerotiorum* and *Botrytis cinerea.* These pathogens have shown to infect a broad number of hosts and usually *S. sclerotiorum* is found co-infecting with *S. minor*. Although their infection strategy is similar, their life cycles follow different patterns. Despite extensive phenotypic information available for *S. minor*, and the availability of the *S. sclerotiorum* genome, a comprehensive genomic analysis of *S. minor* has not been performed yet. In this study we generated a first draft genome sequence of *S. minor*, yielding 33.98 Mbp distributed in 9,060 contigs. Genome annotation revealed 12,357 protein coding genes, including 3,572 orthologous genes shared with closely related ascomycetes. Protein prediction revealed the presence of a variety of plant cell wall degrading enzymes likely associated with its necrotrophic infection style and/or saprophic growth on plant debris.

Phylogenetic and phylogenomic analyses reinforced *S. minor* taxonomical placement as a sister clade of *S. sclerotiorum*. The data obtained in this study will facilitate further research on plant pathogens with necrotrophic properties. Additionally, this data will help with the development of diagnostic tools and detection techniques to discriminate *S. minor* from *S. sclerotiorum.* In particular, the use of loci that have been predicted to play important roles in the infection of this necrotrophic plant pathogen will be beneficial to the design of specific molecular-based and sequencing-based detection techniques.

Introduction

Sclerotinia minor Jagger (Ascomycota, Leotiomycetes, Sclerotiniaceae) is a soil-borne pathogen and a causal agent of Sclerotinia blight. The pathogen was first reported in 1900 in Massachusetts by R. E. Smith and named by Jagger in 1913 after the first report of lettuce drop caused by this pathogen in New York (Arthur et al., 1900; Jagger, 1913). The pathogen has a diverse hosts range including multiple economically important crops (Melzer, Smith & Boland, 1997). Among the economically important crops, losses have been reported in lettuce (*Lactuca sativa*), chicory (*Cichorium intybus*), green bean (*Phaseolus vulgaris*), peanut (*Arachis hypogaea*) and sunflower (*Helianthus annuus*). Infection results in wilting, stem collapse and finally the death of diseased plants (Melzer, Smith & Boland, 1997).

The *S. minor* disease cycle includes a sclerotial stage (resting/survival structure). Infection of the host is triggered by high humidity causing direct germination of mycelia or indirectly (less common *in vitro*) by the production of fruiting bodies (ascocarps) and eventually, ascospores (Beach, 1921; Abawi & Grogan, 1979; Adams, P. B., Ayers, 1979; Dillard & Grogan, 1985; Patterson & Grogan, 1985). Secondary inoculum has not been

reported for *S. minor*, therefore, dispersal of the inoculum is restricted to sclerotia movement by either animals, with seeds or on agricultural tools and by mycelial growth plant-to-plant (Abawi & Grogan, 1979). The incidence of *S. minor* diseases are directly correlated with the density of sclerotia in soil, as opposed to *S. sclerotiorum* Lib. disease incidence, which is directly correlated with a conducive environment for ascospore dispersal (Hawthorne, 1973; Ekins, Aitken & Goulter, 2002). *S. minor* survives in soil as sclerotia or dry mycelia on plant debris. Infection with *S. minor* to lettuce can occur at any growth stage and irrigation systems are the main dispersal mechanism for the pathogen (Wu & Subbarao, 2006).

The morphology of *S. minor* has been extensively described, however, genetic description of the pathogen is lacking. Phylogenetically close relatives — *S. sclerotiorum* Lib. and *Botrytis cinerea* — have been genetically described in literature (Carbone et al., 1993; Tudzynski & Kokkelink, 2009; Amselem et al., 2011). The secretion of cell wall degrading enzymes (CWDE) and toxins to prepare host tissues prior colonization is the most important peculiarity about *S. minor* and other necrotrophs (Oliver & Solomon, 2010). Horizontal Gene Transfer (HGT) from bacterial genomes in the Pezizomycotina has been suggested to play a role in pathogenicity of necrotrophic plant pathogens like *Botrytis* and *Sclerotinia* (Marcet-Houben & Gabaldón, 2010). Evolutionarily, necrotrophs are considered less adapted to a specific host, which explains their wide host range, as opposed of biotrophs which co-evolve with their hosts. However, it has been revealed that some necrotrophs secrete effector proteins that interact with the host (gene-for-gene) in an inverse manner compared to biotroph effectors (Oliver & Solomon, 2010).

The phylogenetically close relative *B. cinerea* has the highest number of genes with functional annotations available in literature (Tudzynski & Kokkelink, 2009). The second most studied pathogen genome in this group is *S. sclerotiorum*, however its genes have mostly structural annotations (Amselem et al., 2011). Although *S. minor*, *S. sclerotiorum*, and *B. cinerea* share physiological features and infection styles by the production of CWDE, they differ in a crucial aspect of their life cycle, their dispersal strategy. In both, *S. sclerotiorum* and *B. cinerea*, dispersal is mainly by air-borne spores. While *S. sclerotiorum* is capable of producing ascospores and not conidia, *B. cinerea* is capable of producing both conidia and ascospores, using conidia as its main method of dispersal. *Sclerotinia minor*, *S. sclerotiorum* and *B. cinerea* primary inocula are sclerotia and their dispersal is principally by irrigation and soil movement, since these resting structures are too heavy to be air-borne. Their different dispersal mechanisms has prevented the development of a single management strategy that fits all three pathogens, but distinct management strategies have been designed for each of them (Brenneman, Phipps & Stipes, 1988; Melouk, Akem & Bowen, 1992; Watson, 2007; Chitrampalam et al., 2010; Bennett, Payton & Chamberlin, 2015).

The genomes of soil-borne plant pathogens have not been studied as extensively as those of air-borne plant pathogens, probably due to the increased epidemiological impact of aerial dispersal and their potential economic impacts, but also because many soil-borne fungi are difficult to study. Currently, there exists a disequilibrium between airborne and soilborne fungi in data available for comparative studies. Here we structurally annotated the draft genome sequence of *S. minor*. Concomitantly, the genome was compared with that of its closest relative, *S. sclerotiorum.* The report of this draft genome offers new

insights into the biology of *S. minor* and a more profound genetic understanding of important organisms in the family Sclerotiniaceae. Additionally, by comparing them phylogenomically we could gain a better understanding about their true evolutionary relationship, and its association with pathogenicity factors encoded by its genes.

Experimental Procedures

Fungal culture preparation and inoculation

S. minor isolate Sm120 was provided by Dr. Hassan Melouk in an sclerotial stage. The isolate was reactivated by plating one sclerotia per Potato dextrose agar (PDA) plate and incubated for 2 days at 24 °C (until mycelia development was observed). Concomitantly, peanut seeds (*Arachis hypogaea*) provided by Dr. Hassan Melouk were germinated in petri dishes containing sterile distilled water for 2 days. The germinated seeds were planted in 16oz cups with autoclaved soil (40 minutes at 121 °C and 15 psi). Peanut plants were watered twice a week with an atomizer for four weeks.

Two inoculation categories were performed. Potato dextrose broth (PDB) was inoculated with one PDA plug containing 2-days old mycelia of *S. minor* and incubated at 24 °C for 3 days. Similarly, 4-weeks old peanut plants were inoculated (on the stem where a node was present) with one PDA plug containing 2-days old mycelia of *S. minor*. A small lesion was created near the inoculation point to facilitate infection. Inoculated plants were kept at 24 °C and 80% relative humidity for up to five days. Each inoculation category (host and media) had 5 replicates.

Genome sequence and assembly

Whole genomic DNA (gDNA) was extracted from mycelium of a 2 days old culture of *S. minor* growing on PDB media by following a protocol modified from Weising *et al.*, 2005. Approximately 1µg of gDNA was checked for quality and quantity before sequencing with a Nanodrop® spectrophotometer. Three samples from the same isolate were sequenced to achieve higher genome coverage. A single-end sequencing library was created for each of the three samples by following 454 Roche's user protocol (Roche Applied Science, Mannheim, Germany) (Sm120.1, Sm120.2 and Sm120.4) and sequenced using the Genome sequencer FLX instrument following the manufacturer's user manual (Roche Applied Science, Mannheim, Germany). For *S. minor* samples Sm120.1 and Sm120.2 the Jr sequencing method was used and for Sm120.4 the titanium sequencing method was utilized. Genome completeness was assessed by using CEGMA, a bioinformatics tool that uses a database of Core Eukaryotic Genes (CEGs) (Parra, Bradnam & Korf, 2007).

Genome annotation

The assembled genome was annotated using MAKER v2.31.8 (Cantarel et al., 2008). Large contigs and protein information from the genus *Sclerotinia* along with ESTs information from *S. sclerotiorum* were used for *ab initio* gene prediction using SNAP (Korf, 2004). SNAP was trained with ESTs and proteins from *S. sclerotiorum* before being used for the annotation of the *S. minor* genome. Illumina HiSeq 2500 was used to perform RNA sequencing in both infecting (3-days post inoculation peanut plant) and non-infecting mycelium (PDA grown mycelia). Both RNA sequencing libraries were assembled using the Trinity software (Grabherr et al., 2011) and transcripts were utilized for the genome

annotation of *S. minor* using Maker v2.31.8. Functional annotation was performed with Interproscan where Gene Ontology (GO) terms as well as Kyoto Encyclopedia of Genes and Genomes (KEGG) terms for putative biochemical pathways inference were added to the annotation files (Zdobnov & Apweiler, 2001; Kanehisa et al., 2004). Carbohydrate active enzymes were annotated by using dbCAN (Yin et al., 2012).

Phylogenetic analysis

Pectate lyase protein amino acid sequence was queried against the non-redundant database of NCBI. All hits having amino acid percent identities higher than 20% were retrieved and their protein sequence collected by using an in house perl script. Approximately 100 protein sequences were retrieved and our *S. minor* protein was added to a multi-FASTA file that was used to perform a multiple sequence alignment (MSA) with MUSCLE (Edgar, 2004). The alignment then was exported to FASTA format and a Maximum Likelihood phylogenetic analysis was performed in RaxML by using 1000 bootstrap iterations (Stamatakis, Hoover & Rougemont, 2008).

Phylogenomic analysis

Phylogenomics analysis for 10 taxa in the Leotiomycetes including *S. minor* was performed on selected ortholog genes that had protein domains identifiable through the databases of the protein collections Pfam and CAZy (Park et al., 2010). Multiple sequence alignment was performed for each of 880 selected orthologous genes with MUSCLE (Edgar, 2004), the alignments were concatenated using GBLOCKS (Castresana, 2000), and the phylogenetic inference was performed by RaxML-HPC (Stamatakis, 2006). The phylogenetic analysis used Maximum Likelihood inference with a WAG model and Gamma distributed rates among sites. Bootstrap analysis with 1000 replicates was

performed and the tree was visualized using Evolview (Zhang *et al.*, 2012). The pectate lyase enzyme alignment was used to design a graphic representation of conservation sites using WebLogo (Crooks et al., 2004).

Carbohydrate active enzymes annotation

Two Carbohydrate-active-enzymes (CAZymes) databases were used to annotate *S. minor* predicted proteins with CAZyme potential activity, the CAZyme database using CAT, as well as dbCAN (Cantarel et al., 2009; Park et al., 2010; Yin et al., 2012). The annotation was performed based on sequence similarity (CAT) and signature-based domains for every CAZyme family (dbCAN). Annotated proteins were compared with orthologous genes found in ten annotated genomes in the Leotiomycetes by using BLAST (Camacho et al., 2009). The presence of Plant Cell Wall (PCW) Degrading Enzymes in the *S. minor* genome was examined and compared with its taxonomically nearest relatives (*B. cynerea* and *S. sclerotiorum*). Additionally, an unrelated necrotroph (*Aspergillus niger*) and a biotroph (*B. graminis*) were added to the comparative genomic analysis.

Pectate lyase protein modeling

Pectate lyase protein modeling was performed using the SWISS-MODEL automated protein structure homology-modeling server (Arnold et al., 2006). Three putative models were obtained, however the one with the highest sequence identity was selected. Model coordinates were downloaded and used for further analysis in PyMol (Schrödinger, LLC, 2015). Prolines were shown as spheres and pink-colored to show the putative active-site.

Results/Discussion

Genome annotation

The estimated genome size of *S. minor* was 43.4 Mb, of which 33.98 Mb were assembled into 9,060 contigs (Figure III-1). The average coverage for the *S. minor* assembly was 6x, but a maximum coverage of 850x was achieved (Table III-1). The estimated genome size of the sequenced *S. minor* genome (43.4 Mb) is similar in size to those of phylogenetically closely related organisms like *S. sclerotiorum* and *B. cynerea* (38 Mb and 37.9 Mb respectively) (Amselem et al., 2011). GC content of the *S. minor* genome is 42.31% which is similar to the 41.78% GC content of *S. sclerotiorum,* 42% of *B. cinerea* and 41.87% of *S. borealis*, but significantly lower than those of other organisms in the Leotiomycetes, like *S. homoeocarpa* (44.61%) and *Blumeria graminis* (43.95%) (Table III-2). A total of 218 (87.9%) ultra-conserved Core Eukaryotic Genes (CEGs) are present in the *S. minor* genome. Among the 218 CEGs, 56 are the most highly conserved among eukaryotes, and 60 genes are among the less conserved in eukaryotes (group 4 and group 1 respectively) (Table III-3). A 40.37% of the CEGs that aligned with the *S. minor* genome have orthologs in other species. Completeness of the genome could also be assessed by using optical maps (Cai *et al.*, 1995), however, there are no available optical maps on literature for *S. minor* yet, although there is one available for *S. sclerotiorum* (Amselem *et al.*, 2011).

Transcriptome annotations can be divided into two types, structural and functional annotations (Hawthorne, 1976). Here we describe both structural annotations, like exons, introns, UTRs or splice forms, as well as functional annotations where RNA sequencing is used to infer potential function of proteins and transcripts. *S. minor* transcripts were

obtained from both, infecting and non-infecting mycelium, assembled RNA sequencing reads produced 22,144 and 84,071 assembled transcripts for *S. minor* growing on PDB and *S. minor* infecting peanut plants, respectively. Both assembled transcripts were utilized during the genome annotation process. Structural annotation yielded a total of 12,357 protein coding genes and among those, 2,458 were single-exon genes and 9,899 were multi-exon genes. The mean exon length was 416.98 nucleotides and the mean intron length was 95.99 nucleotides (Figure III-2). Genes having potential alternative splicing totaled 46. Functional annotations retrieved PFAM domain information for 6,750 genes and gene ontology (GO) information with GO terms for 4,227 genes. Putative protein functions were added to the annotation descriptions, 1,326 potential metabolic pathways were inferred as reactomes, and 506 KEGG inferred pathways were also predicted.

Comparative genomics might give insights into the biology of *S. minor* and into the differences between *S. minor* and other groups of the Sclerotiniaceae family, like *S. sclerotiorum* and *B. cynerea* (Figure III-3). Comparative genomics with *S. sclerotiorum* resulted in a total of 2,808,352 SNPs identified with a whole genome alignment using NUCmer (Kurtz et al., 2004). A total of 723,851 (25.78%) insertions/deletions were identified. Information about intron or exon coordinates is necessary in order to infer if such genome variations might be subject to higher/lower selective pressures based on environmental characteristics.

Orthologs

Orthology analysis of *S. minor* proteome found 3,572 genes having at least one orthologous protein sequence in other species of the Leotiomycetes (Figure III-4). The data suggests that most of its genes might have similar functions. Since not all organisms

in the Leotiomycetes group are plant pathogens or necrotrophs, most of the core orthologous genes described here might be essential genes. Further functional analyses are needed to identify target genes for gene expression analysis experiments, like qPCR or microarrays. Comparisons across genomes of Leotiomycete species permitted to identify 206 unique genes in our *S. minor* assembly. Such result contrasts with the 525, 238, 194, and 131 unique genes of *B. cinerea* and *S. boraealis, S. sclerotiorum,* and *B. graminis,* respectively (Figure III-5). *S. minor* shared 3572 genes with other four members of the Sclerotineaceae family, while 875 genes were shared with *S. sclerotiorum* and 274 with *B. cinerea* (Figure III-5). These results suggest that *S. minor* shares a high proportion of genes in the Sclerotineaceae family, suggesting that most of these genes might be housekeeping genes or genes in shared pathways, including genes related to necrotrophic metabolism.

CAZymes analysis

Like other necrotroph plant pathogens, *S. minor* is capable of producing multiple enzymes targeting plant cell wall components. Previous studies have shown that *S. sclerotiorum* and *B. cinerea* degrade complex plant carbohydrates, such as cellulose, hemicellulose and pectin in plants (Amselem et al., 2011). Databases of enzymes targeting carbohydrate degradation have been created by a variety of research groups and are continuously curated (Cantarel et al., 2009; Yin et al., 2012). Higher numbers of CAZy signature domains were identified in the *S. minor* proteome using dbCAN than with CAZYme (Cantarel et al., 2009). CAZymes classifies enzymes based on their carbohydrate degrading functions. CAZymes are currently classified as five classes that include Glycoside Hydrolases (GHs), Glycosyl Transferases (GTs), Polysaccharide Lyases (PLs), Carbohydrate Esterases (CEs) and Auxiliary Activities (AAs). In some cases, the enzymes

might contain a contiguous amino acid sequence which has carbohydrate-binding activity known as Carbohydrate-binding module (CBM) and formerly known as Cellulose-binding module (CBM). The genome of *S. minor* contains 458 proteins containing putative CAZyme modules which is similar to the 415 and 441 that *S. sclerotiorum* and *B. cinerea*, respectively, encode. Specifically, 67 AAs, 216 GHs, 89 GTs, 5 PLs and 81 CEs modules were predicted, as well as 72 CBMs which in some cases were not coupled with any CAZymes peptide domain (Figure III-6).

Further analysis of CAZYmes modules in *S. minor* permits their association with plant cell wall (PCW) degradation and in some cases fungal cell wall (FCW) degradation as suggested by Amselem et al., 2011. CAZYme modules like GH6, GH7, GH12, GH45, GH61, GH74 and GH94 have been associated with cellulose degradation, GH10, GH11, GH26, GH27, GH29, GH31, GH35, GH36, GH39, GH67, CE1, CE2, CE3, CE5, CE15 and CE16 have been associated with hemicellulose degradation. Sidechains of pectins have been shown to be part of plant cell walls and play key roles in their structural features (Hwang, Pyun & Kokini, 1993) therefore they have also been included in our search. CAZyme modules involved in the degradation sidechains of pectins are GH43, GH51, GH53, GH54, GH62, GH93 and CE12. Pectin degrading enzymes are part of the following modules GH28, GH78, GH88, GH95, GH105, GH115, PL1, PL3, PL4, PL9, PL11 and CE8. Necrotrophic plant pathogens take advantage of a variety of enzymes to colonize plant tissues. PCW degrading enzyme analysis (Figure III-7) revealed that pectinases were present in the *S. minor* genome in higher number than other enzymes, like hemicellulases and cellulases. Among the enzymes that degrade hemicellulose, 13 and 8 enzymes belonging to the CAZyme modules CE5 and CE1 were found, respectively. The CE5

module is known for enzymatic activities like cutinases and acetyl xylan esterases. Cutin is a polymer found as part of the main components of the plant cuticule, most importantly covering the aerial portions of the plant. The CE1 module is formed by enzymes that catalyze reactions of breaking up xylan which is a group of hemicelluloses predominantly formed of β-D-xylose. Interestingly, not many enzymes belonging to the CE1 module were found in *S. sclerotiorum* or *B. cynerea* (Figure III-7). Another interesting finding is the reclassification of the GH61 module to AA9 which are enzymes associated with the cleavage of cellulose chains by oxidizing various carbons. Similar numbers of enzymes (nine) belonging to the AA9 module are found in both *S. minor, S. sclerotiorum* and *B. cynerea* when compared with *A. niger* which has seven. Another group that stands out in *S. minor* is GH74 with three detected cellulases, when compared with zero found in *B. cynerea* and *S. sclerotiorum.*

An enzyme that is likely to be found in organisms that are capable of degrading host cell walls is Pectate lyase. This enzyme degrades plant tissues by eliminative cleavage of the (1→4)-α-D-galacturonan. The predicted *S. minor* sequence (Figure III-8 and Figure III-10) was queried against the Pezizomycotina non-redundant database and found multiple orthologs that were analyzed phylogeneticaly (Figure III-9). As expected, the pectate lyase tree grouped *S. minor* with *S. sclerotiorum*, and *B. cinerea*.

Oxalic acid

Oxalic acid biosynthesis and its physiological roles have been thoroughly documented for *B. cynerea* and *S. sclerotiorum* (Cessna, 2000; Favaron, Sella & D'Ovidio, 2004). Roles like niche acidification (suitable pH for endo-polygalacturonase activity), suppression of oxidative burst of the host plant and deregulation of guard cells during infection have been

suggested (Godoy et al., 1990; Cessna, 2000; Guimarães, Stotz & Guimara, 2004; Favaron, Sella & D'Ovidio, 2004). Its synthesis can follow either the TCA cycle, the glyoxalate cycle (KEGG:00630) or possibly both (Amselem et al., 2011). Genes associated to the production of oxalic acid in *S. sclerotiorum* and *B. cynerea* have been identified in previous studies (Amselem et al., 2011). However, literature suggests that an oxaloacetate acetyl hydrolase (OAH) is a major responsible for catalyzing an hydrolytic cleavage of oxaloacetate, being a crucial enzyme (Genbank: EDN92355.1 | KEGG: R00338) in the production of oxalate in *S. sclerotiorum* (Han et al., 2007) and *B. cynerea* (Genbank: XP_001557891.1). By analyzing the *S. minor* proteome we were able to identify its corresponding orthologous protein/enzyme (Study code: SMIN_00012238-RA) which was also shared with *Pseudogymnoascus destructans* (A bat pathogen, Genbank: ELR04445.1), and *Sclerotinia borealis* (A pathogen of Barley, rye and wheat, Genbank: ESZ97744.1). Additionally we retrieved 26 enzymes putatively associated with the glyoxylate pathway (KEGG: 00630) which could be potentially used for the development of molecular-based pathogen detection tests as well as gene expression analyses (Table III-4).

Oxalic acid interact with polygalacturonases by enhancing their activity (Favaron, Sella & D'Ovidio, 2004). We have been able to identify 16 polygalacturonases (GH28) in the *S. minor* proteome which is similar to the 17 and 18 polygalacturonases found in the *S. sclerotiorum* and *B. cynerea* genome.

Phylogenomics of the Leotiomycetes

S. sclerotiorum, *S. homoeocarpa*, *B. cynerea* (3 strains), *B. paeoniae* and *S. borealis* are as of today the only fully sequenced species within the Sclerotineaceae family (Amselem et al., 2011; Staats, Kan & van Kan, 2012; Hulvey et al., 2012; Blanco-Ulate &

Allen, 2013; Mardanov et al., 2014). However, within the Leotiomycetes group, 16 new genomes have been published and the only complete genome available is from *B. cinerea* B0510 (Table III-2). The lack of genome sequences to infer phylogenies has led to low resolution clades in the Sclerotineaceae family. Various phylogenies have been suggested in the literature, one of the oldest studies using the ribosomal region Internal Transcribed Spacer (ITS) suggested to place *S. minor* in the same clade with *S. sclerotiorum* being a sister clade of *S. trifoliorum* (Carbone et al., 1993). However, in the most recent phylogeny studies using protein coding genes, *S. minor* is considered a sister group of *S. trifoliorum*, and it is placed in a sister clade of *S. sclerotiorum*, where *Sclerotinia* species is a sister group of *S. sclerotiorum* (Amselem et al., 2011). The mentioned studies used both single gene phylogeny and multiple gene phylogeny, although in a whole genome phylogeny inference the Leotiomycetes group is underrepresented due to the lack of genome sequences available (Wang et al., 2009).

The phylogenomic analysis of 10 whole genomes confirmed the placement of *S. minor* in the same clade where *S. sclerotiorum* and *S. borealis* (Amselem et al., 2011). Since phylogenomic analysis resolution is directly related to the number of available annotated genomes and available proteomes in taxonomically related species this analysis is preliminary (Figure III-9). A better representation of the Pezizomycotina genomes is needed to conduct a comprehensive phylogenomic study of this subdivision.

Phylogenetic analysis

Cell Wall degrading enzymes (CWDEs) play a key role in the infection of necrotrophic plant pathogens. Specifically pectate lyase (Figure III-10) is an enzyme widely known for its involvement in the maceration and soft rotting of plant tissue. The

pectate lyase gene was chosen for phylogenetic analysis of *S. minor* alongside with other species of the Pezizomycotina subdivision since most of its members had an orthologous gene with at least 25% identity with the *S. minor* pectate lyase gene. The pectate lyase sequence diversity is high in the Pezizomycotina subdivision, suggesting that positive selective pressure has been shaping this gene to adapt to different ecological conditions. In some instances, pectate lyase may act as an intracellular enzyme, however, in other instances it may act as an extracellular enzyme, as is the case in some members of the Sclerotineaceae plant pathogenic group (Riou *et al.*, 1992).

In conclusion, *S. minor* is genetically similar to *S. sclerotiorum* and *B. cinerea.* The number of genes associated with plant cell wall degradation and their classification are very similar, in some cases differing only by one enzyme. Often times, *S. minor* is found co-infecting with *S. sclerotiorum* and their genetic similarity might suggest a relatively recent speciation event given their phylogenomic & phylogenetic close association, their high number of orthologous genes and their infection style. The information obtained from sequencing the *S. minor* genome could be used to better understand how these pathogens co-evolved with their hosts as well as to help in the development of new diagnostics techniques that would help to discriminate among these necrotrophic plant pathogens.

Acknowledgements

This work was supported by the USDA-CSREES Plant Biosecurity Program, grant number 2010-85605-20542. The computing for this project was performed at the OSU High Performance Computing Center at Oklahoma State University supported in part through the National Science Foundation grant OCI-1126330.

Literature cited

Abawi, G.S. and Grogan, R.G. (1979) Epidemiology of Diseases Caused by Sclerotinia Species. *Phytopathology* **69**, 899–904.

Adams, P. B., Ayers, W. a. (1979) Ecology of Sclerotinia Species. *Phytopathology* **69**, 896–899.

Amselem, J., Cuomo, C.A., Kan, J.A.L. van, et al. (2011) Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genet.* **7**, e1002230.

Arnold, K., Bordoli, L., Kopp, J. and Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195–201.

Arthur, J.C., Barnes, C.R., Coulter, J.M., Coulter, M.S. and Chicago., U. of (1900) Botrytis and Sclerotinia: their relation to certain plant diseases and to each other. *Bot. Gaz.* **29**, 1–427.

Beach, W. (1921) The lettuce“ drop” due to Sclerotinia minor 165th ed., Pennsylvania State College Agricultural Experiment Station.

Bennett, R., Payton, M. and Chamberlin, K. (2015) Response to oxalic acid as a resistance assay for *Sclerotinia minor* in peanut. *Peanut Sci.*

Blanco-Ulate, B. and Allen, G. (2013) Draft genome sequence of *Botrytis cinerea* BcDW1, inoculum for noble rot of grape berries. *Genome Announc. ASM* **1**, e00252-13.

Brenneman, T., Phipps, P. and Stipes, R. (1988) A Rapid Method for Evaluating Genotype Resistance, Fungicide Activity, and isolate pathogenicity of *Sclerotinia minor* in Peanut. *Peanut Sci.* **15**, 104–107.

Cai, W., Aburatani, H., Stanton, V.P., Housman, D.E., Wang, Y.K. and Schwartz, D.C. (1995) Ordered restriction endonuclease maps of yeast artificial chromosomes created by optical mapping on surfaces. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 5164–5168.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T.L. (2009) BLAST plus : architecture and applications. *BMC Bioinformatics* **10**, 421.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V. and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.* **37**, D233–D238.

Cantarel, B.L., Korf, I., Robb, S.M.C., Parra, G., Ross, E., Moore, B., Holt, C., Sánchez Alvarado, A. and Yandell, M. (2008) MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**, 188–96.

Carbone, I., Kohn, L.L.M.L.M., Url, S. and Kohn, L.L.M.L.M. (1993) Ribosomal DNA Sequence Divergence within Internal Transcribed Spacer 1 of the Sclerotiniaceae. *Mycol. Soc. Am.* **85**, 415–427.

Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552.

Cessna, S.G. (2000) Oxalic Acid, a Pathogenicity Factor for *Sclerotinia sclerotiorum*,

Suppresses the Oxidative Burst of the Host Plant. *PLANT CELL ONLINE* **12**, 2191–2200.

Chitrampalam, P., Cox, C., Turini, T. and Pryor, B. (2010) Efficacy of *Coniothyrium minitans* on lettuce drop caused by *Sclerotinia minor* in desert agroecosystem. *Biol. Control* **55**, 92–96.

Crooks, G.E., Hon, G., Chandonia, J.-M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–90.

Dillard, H.R. and Grogan, R.G. (1985) Relationship between sclerotial spatial pattern and density of *Sclerotinia minor* and the incidence of lettuce drop. *Phytopathology* **75**, 90–94.

Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113.

Ekins, M., Aitken, E. and Goulter, K. (2002) Carpogenic germination of *Sclerotinia minor* and potential distribution in Australia. *Australas. Plant Pathol.* **31**, 259–265.

Favaron, F., Sella, L. and D'Ovidio, R. (2004) Relationships Among Endo-Polygalacturonase, Oxalate, pH, and Plant Polygalacturonase-Inhibiting Protein (PGIP) in the Interaction Between *Sclerotinia sclerotiorum* and Soybean. *Mol. Plant-Microbe Interact.* **17**, 1402–1409.

Godoy, G., Steadman, J.R., Dickman, M.B. and Dam, R. (1990) Use of mutants to demonstrate the role of oxalic acid in pathogenicity of *Sclerotinia sclerotiorum* on Phaseolus vulgaris. *Physiol. Mol. Plant Pathol.* **37**, 179–191.

Grabherr, M.G., Haas, B.J., Yassour, M., et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–52.

Guimarães, R.L., Stotz, H.U. and Guimara, R.L. (2004) Oxalate Production by *Sclerotinia sclerotiorum* Deregulates Guard Cells during Infection. *Plant Physiol.* **136**, 3703–3711.

Han, Y., Joosten, H.-J., Niu, W., Zhao, Z., Mariano, P.S., McCalman, M., Kan, J. van, Schaap, P.J. and Dunaway-Mariano, D. (2007) Oxaloacetate hydrolase, the C-C bond lyase of oxalate secreting fungi. *J. Biol. Chem.* **282**, 9581–90.

Hawthorne, B. (1976) Observations on the development of apothecia of *Sclerotinia minor* Jagg. in the field. *New Zeal. J. Agric. Res.* **19**, 383–386.

Hawthorne, B. (1973) Production of apothecia of *Sclerotinia minor*. *New Zeal. J. Agric. Res.* **16**, 559–560.

Hulvey, J., Popko, J.T., Sang, H., Berg, A. and Jung, G. (2012) Overexpression of ShCYP51B and ShatrD in *Sclerotinia homoeocarpa* isolates exhibiting practical field resistance to a demethylation inhibitor fungicide. *Appl. Environ. Microbiol.* **78**, 6674–82.

Hwang, J., Pyun, Y.R. and Kokini, J.L. (1993) Sidechains of pectins: some thoughts on their role in plant cell walls and foods. *Food Hydrocoll.* **7**, 39–53.

Jagger, I. (1913) *Sclerotinia minor*, n. sp., the cause of a decay of lettuce, celery, and other crops. *J. Agric. Res* **XX**, 331–334.

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. and Hattori, M. (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **32**, D277–D280.

Korf, I. (2004) Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59.

Kurtz, S., Phillippy, A., Delcher, A., Smoot, M., Shumway, M., Antonescu, C. and Salzberg, S. (2004) Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12.

Marcet-Houben, M. and Gabaldón, T. (2010) Acquisition of prokaryotic genes by fungal genomes. *Trends Genet.* **26**, 5–8.

Mardanov, A.V.A., Beletsky, A.V.A., Kadnikov, V. V, Ignatov, A.N. and Ravin, N. V (2014) Draft Genome Sequence of *Sclerotinia borealis*, a Psychrophilic Plant pathogenic fungus. *Genome Announc.* **2**, 1–2.

Melouk, H., Akem, C. and Bowen, C. (1992) A Detached Shoot Technique To Evaluate the Reaction of Peanut Genotypes to *Sclerotinia minor*. *Peanut Sci.* **19**, 58–62.

Melzer, M.S., Smith, E. a. and Boland, G.J. (1997) Index of plant hosts of *Sclerotinia minor*. *Can. J. Plant Pathol.* **19**, 272–280.

Oliver, R. and Solomon, P. (2010) New developments in pathogenicity and virulence of necrotrophs. *Curr. Opin. Plant Biol.* **13**, 415–419.

Park, B.H., Karpinets, T. V., Syed, M.H., Leuze, M.R. and Uberbacher, E.C. (2010) CAZymes Analysis Toolkit (cat): Web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. *Glycobiology* **20**, 1574–1584.

Parra, G., Bradnam, K. and Korf, I. (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**, 1061–7.

Patterson, C.L. and Grogan, R.G. (1985) Differences in epidemiology and control of lettuce drop caused by *Sclerotinia minor* and *S. sclerotiorum*. *Plant Dis.* **69**, 766–770.

Riou, C., Freyssinet, G. and Fevre, M. (1992) Purification and Characterization of Extracellular Pectinolytic Enzymes Produced by *Sclerotinia sclerotiorum*. *Appl. Environ. Microbiol.* **58**, 578–583.

Schrödinger, L. (2015) The PyMOL Molecular Graphics System, Version 1.8. *Unpubl. Work.*

Staats, M., Kan, J. van and Kan, J. a L. van (2012) Genome update of *Botrytis cinerea* strains B05.10 and T4. *Eukaryot. Cell* **11**, 1413–1414.

Stamatakis, A. (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690.

Stamatakis, A., Hoover, P. and Rougemont, J. (2008) A Rapid Bootstrap Algorithm for the RAxML Web Servers. *Syst. Biol.* **57**, 758–771.

Tudzynski, P. and Kokkelink, L. (2009) *Botrytis cinerea* : Molecular Aspects of a Necrotrophic Life Style. *Style DeKalb IL* **5**, 29–50.

Wang, H., Xu, Z., Gao, L. and Hao, B. (2009) A fungal phylogeny based on 82 complete genomes using the composition vector method. *BMC Evol. Biol.* **9**, 195.

Watson, A. (2007) *Sclerotinia minor*—Biocontrol target or agent? In Novel Biotechnologies for Biocontrol Agent Enhacement and Management. (Vurro, M. and

Gressel, J., eds), pp. 205–211. Quebec, Canada.

Weising, K., Nybom, H., Pfenninger, M., Wolff, K. and Kahl, G. (2005) DNA Fingerprinting in Plants: Principles, Methods, and Applications, Second Edition 2nd ed., Boca Raton, FL: CRC Press.

Wu, B. and Subbarao, K. (2006) Analyses of lettuce drop incidence and population structure of *Sclerotinia sclerotiorum* and *S. minor*. *Phytopathology* **96**, 1322–1329.

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F. and Xu, Y. (2012) dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **40**, W445–W451.

Zdobnov, E.M. and Apweiler, R. (2001) InterProScan – an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847–848.

Zhang, H., Gao, S., Lercher, M.J., Hu, S. and Chen, W.-H. (2012) EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. *Nucleic Acids Res.* **40**, W569–W572.

Figures

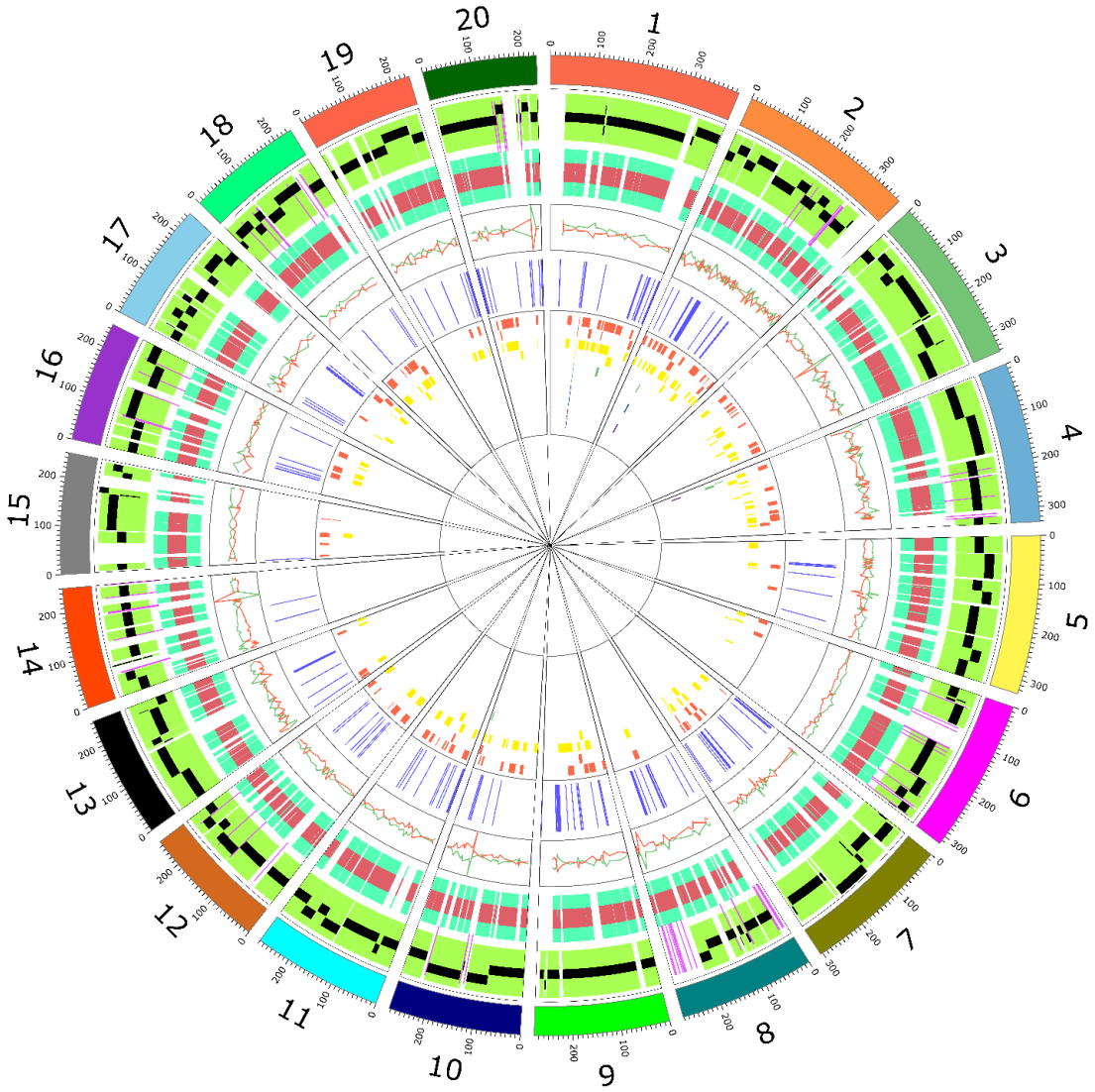

Figure III-1. Partial Genome visualization of *Sclerotinia minor* represented by the twenty largest contigs. The figure has ten shells. The outer shell shows the contigs delimited by different colors. Second shell (green highlight with black stacked bars) shows syntenic regions in *S. minor* and *Sclerotinia sclerotiorum* genomes. The stacked black bars represent the alignment lengths with *S. sclerotiorum* genome. Third shell (green highlight with centered red highlight) shows structural annotated genes, the red centered highlight shows CDS. Fourth shell shows GC content variation (green line) and AT content variation (red line). The fifth shell (blue highlight) shows repetitive sequences in both exons and introns. The sixth shell includes comparative genomic visualizations of closely related organisms by using stacked bars representing alignment lengths. Red = *Botrytis cynerea*; yellow = *Sclerotinia borealis*; green = *Sclerotinia homoeocarpa*; blue = *Pseudogymnoascus destructans*; purple = *Geomyces pannorum*. A purple highlight in the second shell shows coordinates where e-probes have been generated.

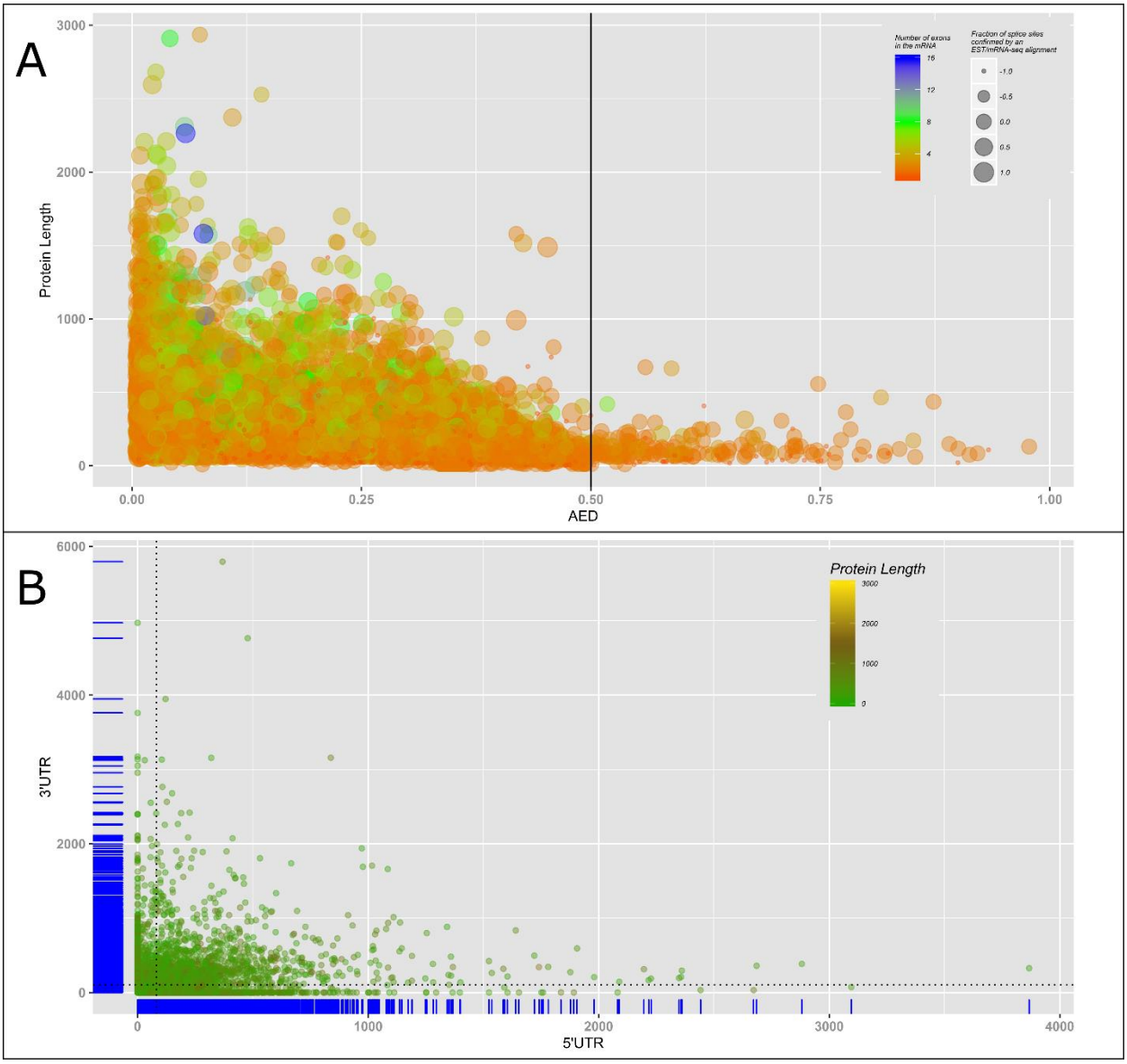

Figure III-2. *Sclerotinia minor* transcriptome and genome annotation metrics. A. Annotation Edit Distance (AED) values vs. protein length, number of exons per mRNA and fraction of splice sites confirmed by mRNA-seq alignments. B. Frequency of predicted 5' UTR regions and 3'UTR regions with marginal density plots and its association with protein length.

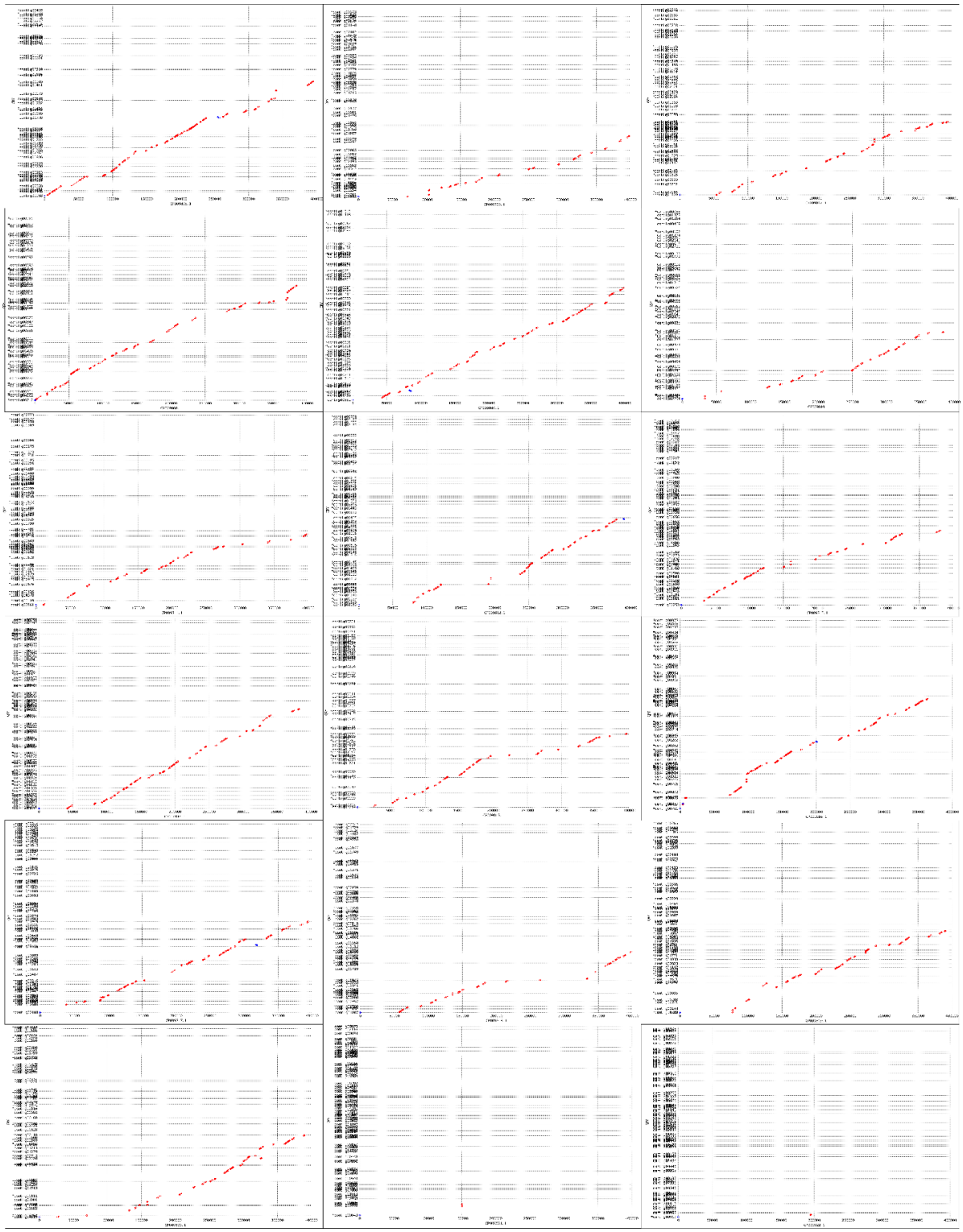

Figure III-3. Whole genome partially filtered comparative genomics of *Sclerotinia minor* vs. *Botrytis cinerea*.

Figure III-4. Phylogenomic tree showing the taxonomic relationship between *Sclerotinia minor* and 10 other fungi with sequenced genomes in the Leotiomycetes.

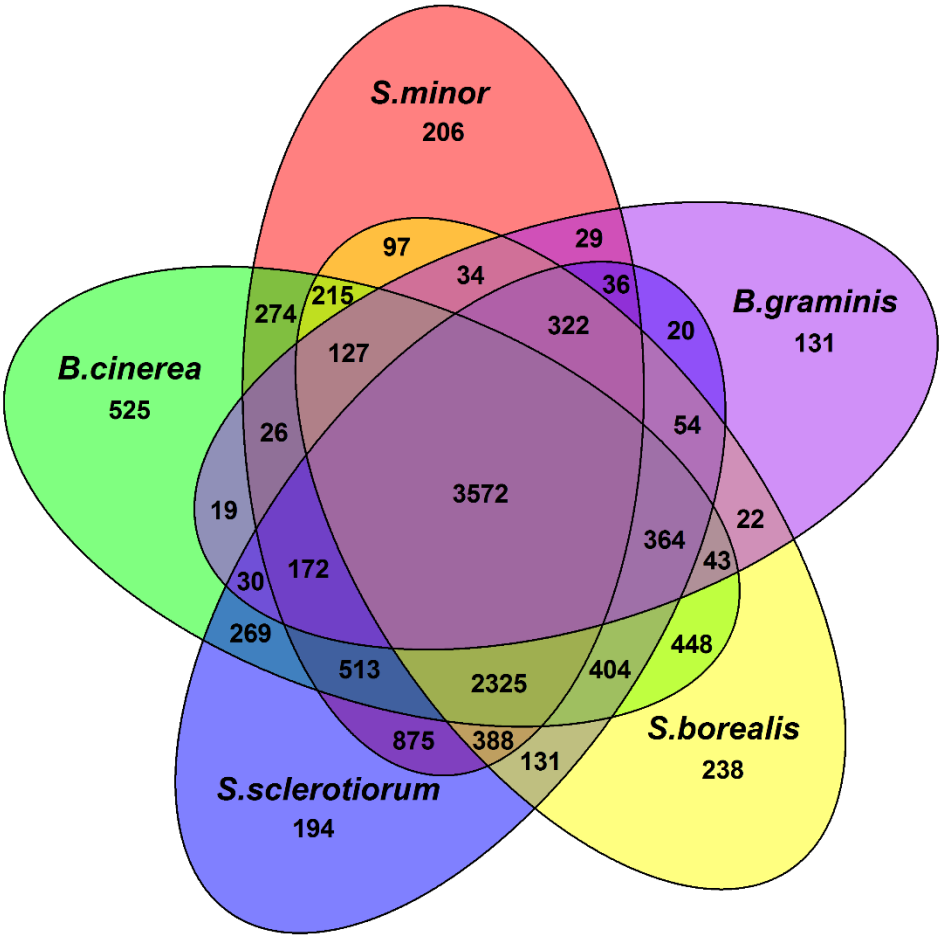

Figure III-5. Venn diagram depicting number of unique and orthologous genes for 5 species in the Order Eurotiales.

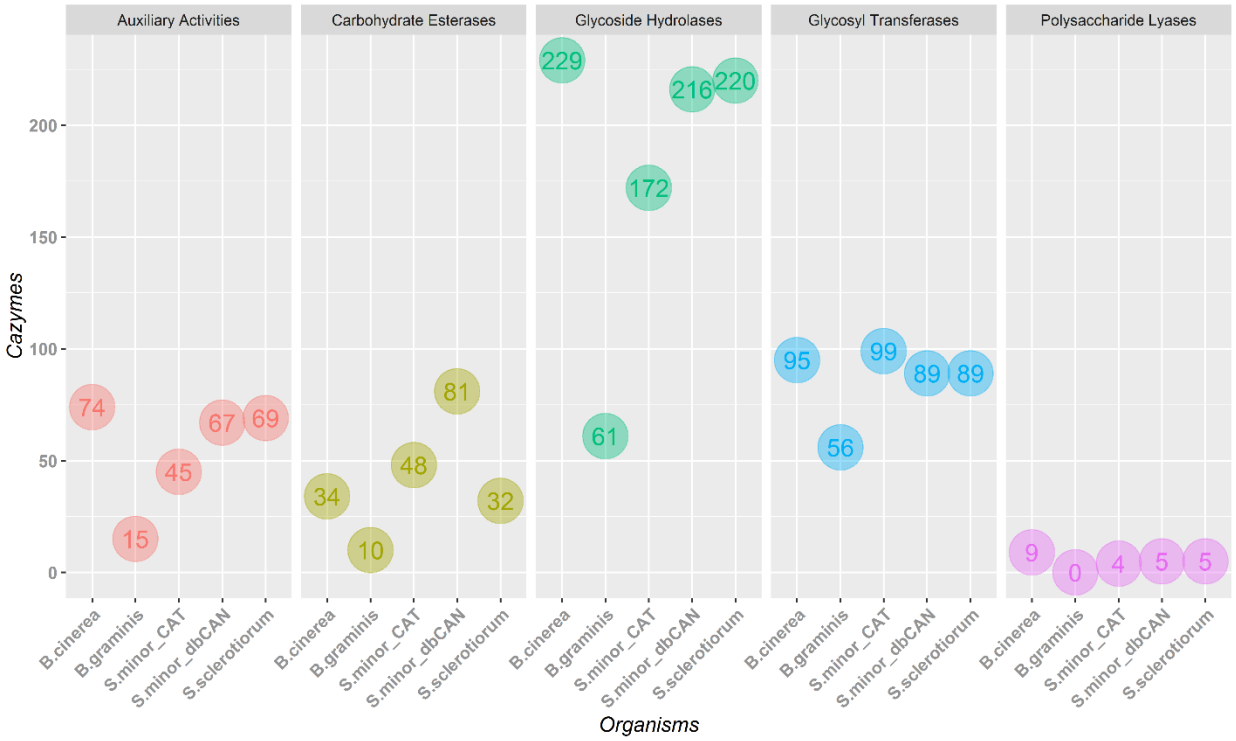

Figure III-6. Carbohydrate Active Enzymes annotated in the *Sclerotinia minor* genome and other species in the Sclerotiniaceae.

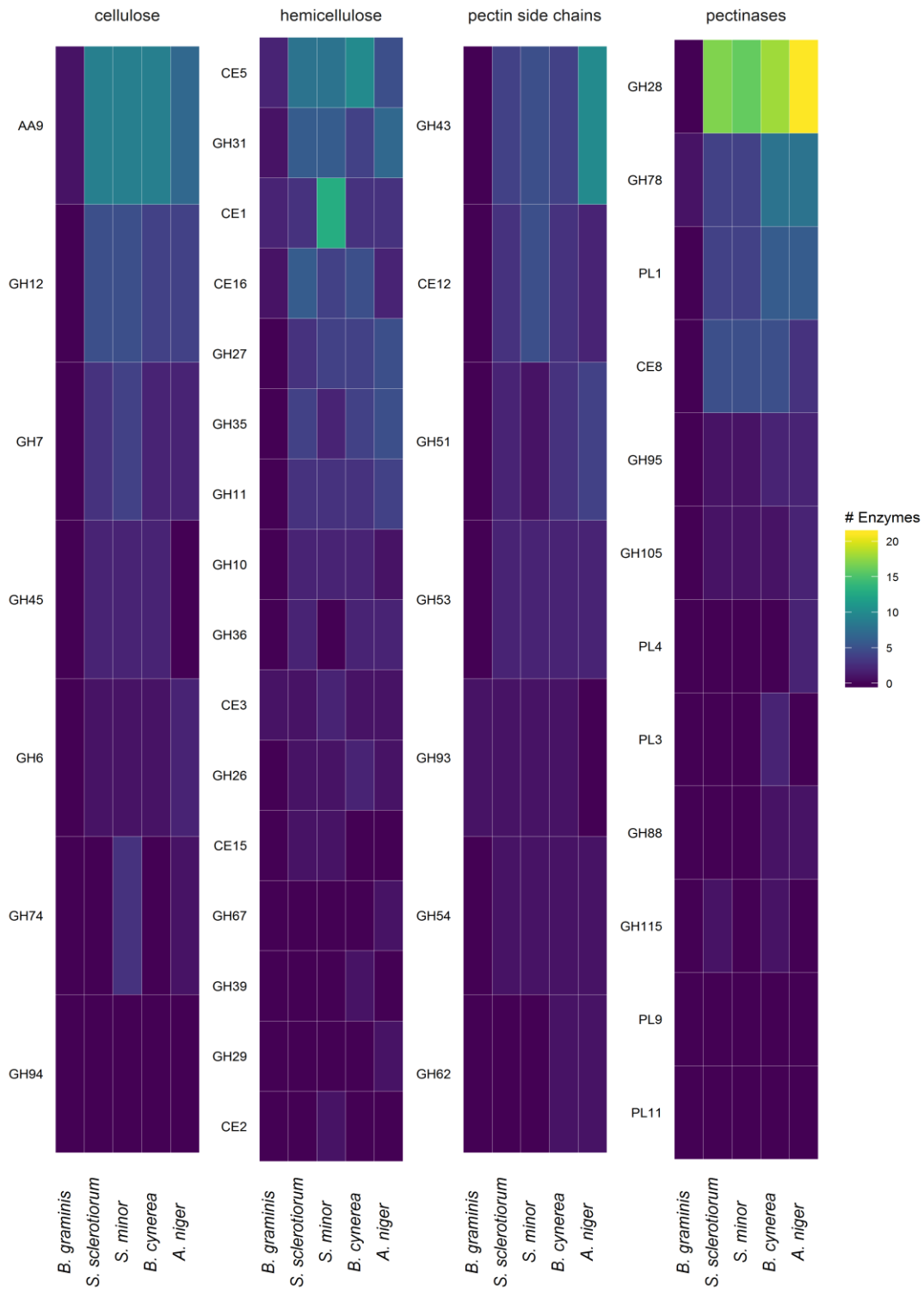

Figure III-7. Comparison of plant cell wall (PCW) degrading enzymes between *Sclerotinia minor*, *Sclerotinia sclerotiorum* and other ascomycetes.

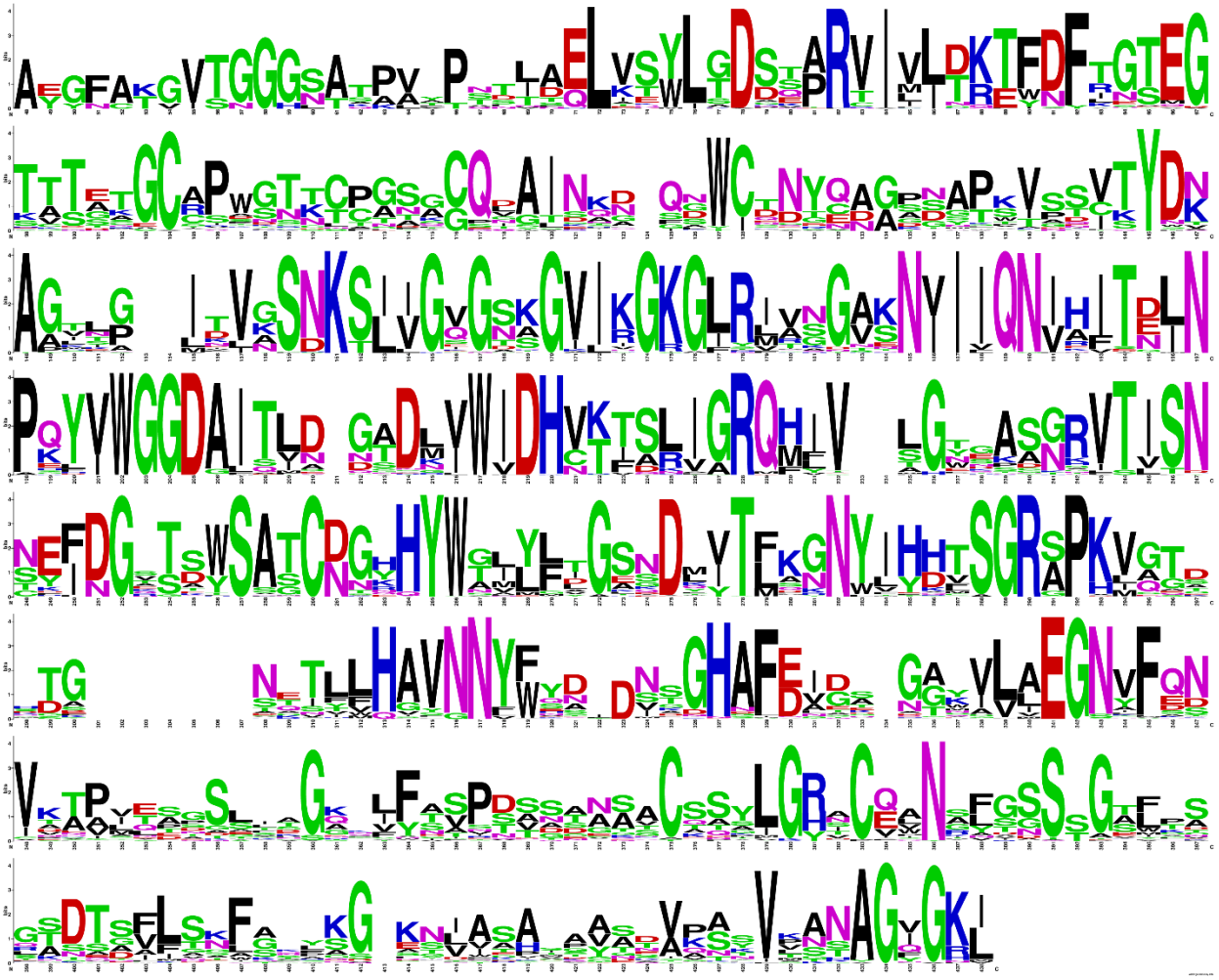

Figure III-8. *Sclerotinia minor* pectate lyase Sequence Logo obtained from a position-specific scoring matrix (PSSM) and multiple sequence alignment (MSA).

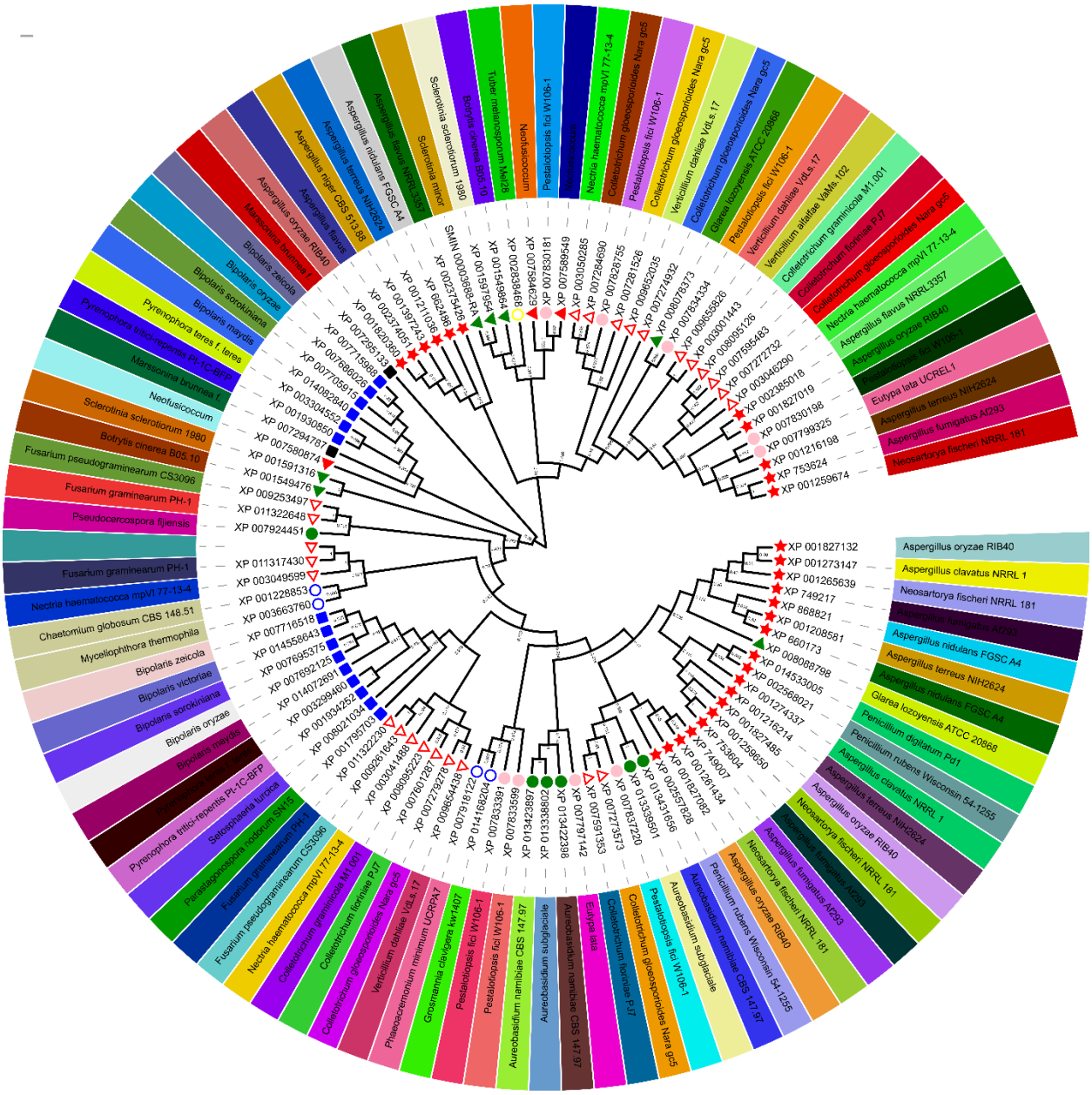

Figure III-9. Phylogeny of the Pezizomycotina using pectate lyase orthologous genes of *Sclerotinia minor*.

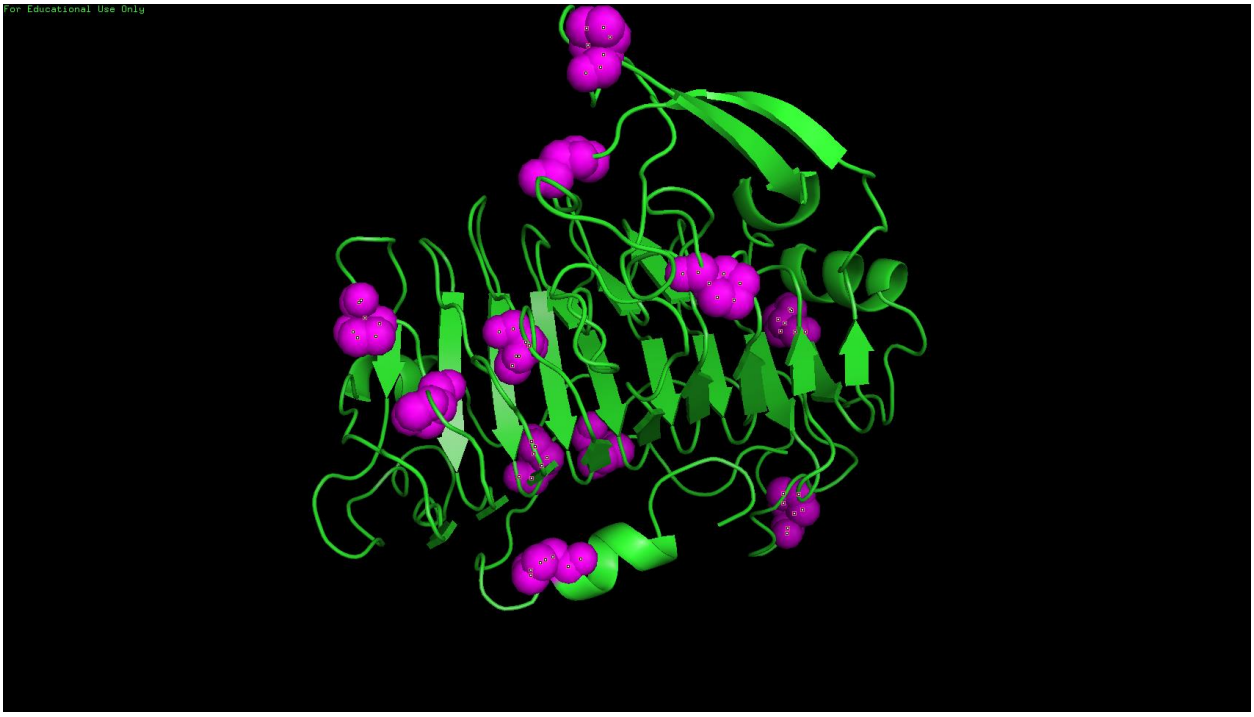

Figure III-10. *Sclerotinia minor* pectate lyase protein structure modeling.

Tables

Table III-1.*Sclerotinia minor* genome assembly statistics and metrics

Genome metrics	
Largest Contig size	39,554
Avg. Contig size	4,245
N50	6,348
Aligned reads	525,545
Estimated Genome Size	43.4 Mb
Sequencing Depth	6.0x
Maximum Depth	850x

Table III-2. Genome information of Leotimycete genomes for whole genome phylogenomics.

Organism	Contigs/Chromosomes	Scaffolds	Size (Mb)	GC content	Proteins
Ascocoryne sarcoides NRRL 50072	219	13	34.17	46.38	0
Blumeria graminis f. sp. *hordei* DH14	15056	6843	87.91	43.96	6495
Botrytis cinerea B05.10	18	-	42.63	42.00	16581
Botrytis paeoniae	11700	-	44.24	41.11	0
Erysiphe necator	5935	-	52.51	38.74	6484
Erysiphe pisi	35300	-	69.26	39.16	0
Glarea lozoyensis ATCC 20868	239	22	39.17	45.82	13083
Marssonina brunnea f. sp. *'multigermtubi'*	2415	89	51.72	42.92	10027
Oidiodendron maius Zn	433	100	46.24	47.11	16702
Poculum sydowianum	11777	-	51.99	43.07	0
Pseudogymnoascus destructans 20631-21	3580	1848	28.36	50.12	9153
Pseudogymnoascus pannorum VKM F-3557	3339	-	27.65	50.21	9482
Rutstroemia echinophila	7345	-	40.25	43.11	0
Sclerotinia borealis F-4157	1741	1241	39.24	41.87	10166
Sclerotinia homoeocarpa LT30	31623	-	29.73	44.61	0
Sclerotinia minor	9060	-	33.98	42.31	12357
Sclerotinia sclerotiorum 1980 UF-70	682	39	38.20	41.79	14446

Table III-3. Core eukaryotic genes (CEG) mapped to the *Sclerotinia minor* genome.

	Prots	% Completeness	Total	Average	% Orthologs
Complete	218	87.9	350	1.61	40.37
Group 1	60	90.91	99	1.65	38.33
Group 2	48	85.71	75	1.56	41.67
Group 3	54	88.52	89	1.65	40.74
Group 4	56	86.15	87	1.55	41.07
Partial	239	96.37	402	1.68	43.93
Group 1	60	90.91	104	1.73	40
Group 2	55	98.21	94	1.71	49.09
Group 3	61	100	102	1.67	40.98
Group 4	63	96.92	102	1.62	46.03

Table III-4. *Sclerotinia minor* proteins potentially involved with the glyoxylate pathway, one of the potential precursors of oxalic acid.

Protein code	Protein length	PFAM Code	Putative function	Gene Ontology term	Potential KEGG, Reactome
SMIN_00010183-RA	756	PF00199	Catalase	GO:0004096|GO:0020037|GO:0055114	KEGG: 00380+1.11.1.6|KEGG: 00630+1.11.1.6|Reactome: R-HSA-3299685|Reactome: R-HSA-74259
SMIN_00010183-RA	756	PF06628	Catalase-related immune-responsive		KEGG: 00380+1.11.1.6|KEGG: 00630+1.11.1.6|Reactome: R-HSA-3299685|Reactome: R-HSA-74259
SMIN_00003907-RA	410	PF00120	Glutamine synthetase, catalytic domain	GO:0004356|GO:0006807	KEGG: 00220+6.3.1.2|KEGG: 00250+6.3.1.2|KEGG: 00630+6.3.1.2|KEGG: 00910+6.3.1.2|MetaCyc: PWY-381|MetaCyc: PWY-5675|MetaCyc: PWY-6549|MetaCyc: PWY-6963|MetaCyc: PWY-6964
SMIN_00006768-RA	300	PF00199	Catalase	GO:0004096|GO:0020037|GO:0055114	KEGG: 00380+1.11.1.6|KEGG: 00630+1.11.1.6|Reactome: R-HSA-3299685|Reactome: R-HSA-74259
SMIN_00008757-RA	647	PF00199	Catalase	GO:0004096|GO:0020037|GO:0055114	KEGG: 00380+1.11.1.6|KEGG: 00630+1.11.1.6|Reactome: R-HSA-3299685|Reactome: R-HSA-74259
SMIN_00008757-RA	647	PF06628	Catalase-related immune-responsive		KEGG: 00380+1.11.1.6|KEGG: 00630+1.11.1.6|Reactome: R-HSA-3299685|Reactome: R-HSA-74259
SMIN_00006203-RA	477	PF00464	Serine hydroxymethyltransferase	GO:0016740	KEGG: 00260+2.1.2.1|KEGG: 00460+2.1.2.1|KEGG: 00630+2.1.2.1|KEGG: 00670+2.1.2.1|KEGG: 00680+2.1.2.1|MetaCyc: PWY-1622|MetaCyc: PWY-181|MetaCyc: PWY-2161|MetaCyc: PWY-2201|MetaCyc: PWY-3661|MetaCyc: PWY-3661-1|MetaCyc: PWY-3841|MetaCyc: PWY-5497|Reactome: R-HSA-196757
SMIN_00005491-RA	541	PF00199	Catalase	GO:0004096|GO:0020037|GO:0055114	KEGG: 00380+1.11.1.6|KEGG: 00630+1.11.1.6|Reactome: R-HSA-3299685|Reactome: R-HSA-74259
SMIN_00005491-RA	541	PF06628	Catalase-related immune-responsive		KEGG: 00380+1.11.1.6|KEGG: 00630+1.11.1.6|Reactome: R-HSA-3299685|Reactome: R-HSA-74259
SMIN_00014211-RA	443	PF06628	Catalase-related immune-responsive		KEGG: 00380+1.11.1.6|KEGG: 00630+1.11.1.6|Reactome: R-HSA-3299685|Reactome: R-HSA-74259
SMIN_00014211-RA	443	PF00199	Catalase	GO:0004096|GO:0020037|GO:0055114	KEGG: 00380+1.11.1.6|KEGG: 00630+1.11.1.6|Reactome:

					R-HSA-3299685|Reactome: R-HSA-74259
SMIN_00014211-RA	443	PF00199	Catalase	GO:0004096|GO:0020037|GO:0055114	KEGG: 00380+1.11.1.6|KEGG: 00630+1.11.1.6|Reactome: R-HSA-3299685|Reactome: R-HSA-74259
SMIN_00014652-RA	126	PF00199	Catalase	GO:0004096|GO:0020037|GO:0055114	KEGG: 00380+1.11.1.6|KEGG: 00630+1.11.1.6|Reactome: R-HSA-3299685|Reactome: R-HSA-74259
SMIN_00012487-RA	90	PF00463	Isocitrate lyase family	GO:0004451|GO:0019752	KEGG: 00630+4.1.3.1|MetaCyc: PWY-6969
SMIN_00005373-RA	322	PF04199	Putative cyclase	GO:0004061|GO:0019441	KEGG: 00380+3.5.1.9|KEGG: 00630+3.5.1.9|MetaCyc: PWY-5651|MetaCyc: PWY-6309|MetaCyc: PWY-7717|MetaCyc: PWY-7733|MetaCyc: PWY-7734
SMIN_00007143-RA	612	PF00463	Isocitrate lyase family	GO:0004451|GO:0019752	KEGG: 00630+4.1.3.1|MetaCyc: PWY-6969
SMIN_00006955-RA	364	PF03951	Glutamine synthetase, beta-Grasp domain	GO:0004356|GO:0006542|GO:0006807	KEGG: 00220+6.3.1.2|KEGG: 00250+6.3.1.2|KEGG: 00630+6.3.1.2|KEGG: 00910+6.3.1.2|MetaCyc: PWY-381|MetaCyc: PWY-5675|MetaCyc: PWY-6549|MetaCyc: PWY-6963|MetaCyc: PWY-6964
SMIN_00006955-RA	364	PF00120	Glutamine synthetase, catalytic domain	GO:0004356|GO:0006807	KEGG: 00220+6.3.1.2|KEGG: 00250+6.3.1.2|KEGG: 00630+6.3.1.2|KEGG: 00910+6.3.1.2|MetaCyc: PWY-381|MetaCyc: PWY-5675|MetaCyc: PWY-6549|MetaCyc: PWY-6963|MetaCyc: PWY-6964
SMIN_00008114-RA	366	PF04199	Putative cyclase	GO:0004061|GO:0019441	KEGG: 00380+3.5.1.9|KEGG: 00630+3.5.1.9|MetaCyc: PWY-5651|MetaCyc: PWY-6309|MetaCyc: PWY-7717|MetaCyc: PWY-7733|MetaCyc: PWY-7734
SMIN_00009289-RA	502	PF06628	Catalase-related immune-responsive		KEGG: 00380+1.11.1.6|KEGG: 00630+1.11.1.6|Reactome: R-HSA-3299685|Reactome: R-HSA-74259
SMIN_00009289-RA	502	PF00199	Catalase	GO:0004096|GO:0020037|GO:0055114	KEGG: 00380+1.11.1.6|KEGG: 00630+1.11.1.6|Reactome: R-HSA-3299685|Reactome: R-HSA-74259
SMIN_00012486-RA	231	PF00463	Isocitrate lyase family	GO:0004451|GO:0019752	KEGG: 00630+4.1.3.1|MetaCyc: PWY-6969
SMIN_00007649-RA	521	PF00464	Serine hydroxymethyltransferase	GO:0016740	KEGG: 00260+2.1.2.1|KEGG: 00460+2.1.2.1|KEGG: 00630+2.1.2.1|KEGG: 00670+2.1.2.1|KEGG: 00680+2.1.2.1|MetaCyc: PWY-1622|MetaCyc: PWY-181|MetaCyc: PWY-2161|MetaCyc: PWY-2201|MetaCyc: PWY-3661|MetaCyc: PWY-3661-

					1|MetaCyc: PWY-3841|MetaCyc: PWY-5497|Reactome: R-HSA-196757
SMIN_00007315-RA	495	PF00120	Glutamine synthetase, catalytic domain	GO:0004356|GO:0006807	KEGG: 00220+6.3.1.2|KEGG: 00250+6.3.1.2|KEGG: 00630+6.3.1.2|KEGG: 00910+6.3.1.2|MetaCyc: PWY-381|MetaCyc: PWY-5675|MetaCyc: PWY-6549|MetaCyc: PWY-6963|MetaCyc: PWY-6964
SMIN_00014346-RA	325	PF01274	Malate synthase	GO:0004474|GO:0006097	KEGG: 00620+2.3.3.9|KEGG: 00630+2.3.3.9|MetaCyc: PWY-6728|MetaCyc: PWY-6969|MetaCyc: PWY-7118|MetaCyc: PWY-7294|MetaCyc: PWY-7295
SMIN_00008086-RA	134	PF01274	Malate synthase	GO:0004474|GO:0006097	KEGG: 00620+2.3.3.9|KEGG: 00630+2.3.3.9|MetaCyc: PWY-6728|MetaCyc: PWY-6969|MetaCyc: PWY-7118|MetaCyc: PWY-7294|MetaCyc: PWY-7295

CHAPTER IV

IV EVALUATING EDNA-TRANSCRIPTOMICS: A VARIATION OF EDNA FOR THE DETECTION OF PLANT PATHOGENS USING RNA SEQUENCING

Abstract

Traditional plant disease diagnostics use signs and/or symptoms for initial assessment of the presence of a pathogen. In cases where more sensitive and specific analysis is required, both molecular and immunological assays are utilized. However, current molecular-based plant pathogen detection techniques use genetic information of the pathogen to determine its presence or absence in the lack of symptoms or to confirm presumptive diagnostics based on signs and symptoms. Yet, the presence of a pathogenic organism in a plant sample, does not reflect the status of an infectious process, since DNA of the resting pathogen propagules may be detected even if the pathogen is already dead or in latent state. Common protocols to determine pathogen viability include the isolation and culture of the organism, which can be time consuming if the sample is taken from a microorganism rich environment. E-probe Diagnostics Nucleic acid Analysis (EDNA), a bioinformatic tool that performs accurate and sensitive metagenomic-based plant pathogen detection, was developed to offer a diagnostics alternative for simultaneous detection of diverse microorganisms from infected plant samples that do not require isolation of pure cultures or assembly of genomic data.

Plant pathogen viability assessment can assist in the prediction of potential disease outbreaks. Although EDNA was originally developed to detect plant pathogens at the metagenomic level, here we present a revised version of EDNA — named EDNAtran because it uses transcriptomic analysis — designed to detect plant pathogens that could potentially be found actively developing in asymptomatic and symptomatic plants. EDNAtran was validated *in vitro* using two nutritional substrate sources (Potato Dextrose Broth (PDB) and living peanut plant). EDNAtran successfully detected actively transcribed genes of *S. minor* growing on PDB, however, genes were not detected in *S. minor* growing on peanut plants possibly due to sample collection timing. No further validation was done with other sample collection times from *S. minor* growing on peanut plants due to sequencing budget limitations. Further analysis with *S. minor* and other organisms is necessary to successfully validate this promising tool.

Introduction

E-probe Diagnostics Nucleic acid Analysis (EDNA) is a tool originally developed as a theoretical approach aiming to detect all/most plant pathogens in a given plant sample by using Next Generation Sequencing (NGS) and bioinformatic pipelines (Stobbe et al., 2013). Rapidly decreasing sequencing prices (Wetterstrand, 2013) have lately permitted the access to affordable genome sequencing, which allows the deployment of EDNA for real case diagnostics (Espindola et al., 2015). EDNA is a phytobiome targeted diagnostics tool. Its target plant pathogen detection scenarios include agricultural and/or non-agricultural environments where samples are retrieved from the rhizosphere, phyllosphere, soil or water, and are subsequently processed for nucleic acid purification and NGS to generate unassembled sequence data for all its genetic constituents. EDNA uses then unique or signature sequences derived from genomes of target organisms (e-probes) to detect their presence in the metagenomic data (Espindola et al., 2015).

Sequencing environmental samples without previous molecular amplification (amplicon sequencing) is termed metagenomic sequencing or "shotgun sequencing" (Eisen, 2007). Metagenomic sequencing allows retrieving genetic information about most organisms found in a sample and permits the study of its biodiversity. Originally presented as an alternative to study unculturable microorganisms (Chen & Pachter, 2005), metagenomic sequencing has become widely utilized since the advent of NGS. Nonetheless, metagenomic studies have limitations related to the laborious and time consuming data analysis (Pop & Salzberg, 2008). Consequently, a variety of tools were rapidly developed aiming to analyze metagenomic data, but most relied on non-curated databases and lacked efficiency due to the unavoidable pairwise alignments with public databases (Huson et al., 2007).

Certain bioinformatic tools, such as MetaPhlan, reduce analysis time for profiling metagenomic microbial communities by using clade specific marker genes to identify microbial clades in the microbiomes of model systems (Truong et al., 2015). Currently, EDNA is the only available bioinformatic tool that uses species-specific e-probes designed to detect target plant pathogens in metagenomic samples (Stobbe et al., 2013; Truong et al., 2015; Espindola et al., 2015). Although EDNA is faster than other tools that use non-curated databases, a considerable amount of time is spent designing e-probes (Espindola et al., 2015). Although genomic based techniques are capable of providing information about the presence/absence of the pathogen, transcriptome based detection can be used as reflection of biological activity of the target organism.

Plant pathogen DNA residues may be present in plant samples due to accidental contamination. Propagule viability assessment can be helpful to decision makers for risk assessment and consideration of eradication measures in cases of contamination with pathogen

DNA or asymptomatic infections. Many plant pathogens are capable to remain viable but dormant for long periods of time until they encounter the appropriate conditions to germinate, colonize and infect a new host (Agrios, 2005). Preventive measures are used to avoid the potential infection of plant propagative materials traded worldwide to reduce the risk of introduction of exotic pathogens to new areas. Soilborne pathogens in particular may require specific treatments to destroy their resting structures (Boehm & Hoitink, 1992; Pascual et al., 2000; Swain et al., 2006; Pane et al., 2011) as they can be highly tolerant to environmental changes and withstand high and/or low temperatures, sudden humidity changes among other extreme weather conditions (Koike et al., 2003).

The soil-borne ascomycete *Sclerotinia minor* is the causal agent of Sclerotinia blight of peanut and has caused severe losses in peanut producing states, including Oklahoma, Texas, Virginia and North Carolina (Wadsworth, 1979; Goldman et al., 1995). The pathogen has a wide host range, including several economically important crops, such as peanut (*Arachis hypogaea* L), and many species of weeds which are considered alternative hosts of the pathogen and play an important role by increasing the prevalence of *S. minor* in the soils (Melzer, Smith & Boland, 1997; Cousens & Croft, 2000). Sclerotinia blight of peanut can cause approximately 50% yield losses in severely affected fields (Butzler, Bailey & Beute, 1998). Lettuce drop is another economically important disease caused by *S. minor* on lettuce (*Lactuca sativa* L) where yield losses can reach up to 75% (Melzer, Smith & Boland, 1997). *S. minor* was chosen as a model organism to take advantage of the fully sequenced and annotated genomes of their taxonomically nearest neighbors *S. sclerotiorum* and *Botrytis cinerea* for comparison purposes (Amselem et al., 2011). Additionally, *S. minor* is a necrotrophic pathogen that has been extensively described in the literature, capable of degrading its host tissue to access nutrients, colonize the host internally, and

produce resting structures (sclerotia). The arsenal of this fungus includes a variety of cell-wall degrading enzymes (CWDE) like carbohydrate active enzymes (CAZ) and oxalic acid (OA), among other pathogenicity factors. The *S. minor* genome sequencing and annotation were addressed in Chapter 3.

The purpose of this study was to create a variation of EDNA that allows detecting physiologically active and/or actively growing plant pathogens based on RNA sequencing. Here we describe EDNA transcriptomics (EDNAtran) which takes advantage of annotated eukaryotic genomes, and RNA sequencing to detect actively growing or infecting plant pathogens in transcriptomic data sets.

Experimental Procedures

RNA sequencing

S. minor isolate Sm120 was provided by Dr. Hassan Melouk in an sclerotial stage. The isolate was reactivated by plating one sclerotia per potato dextrose agar (PDA) plate and incubated for 2 days at 24 °C (until mycelia development was observed). Concomitantly, peanut seeds (*Arachis hypogaea*) provided by Dr. Hassan Melouk were germinated in petri dishes containing sterile distilled water for 2 days. The germinated seeds were planted in 16oz cups with autoclaved soil (40 minutes at 121 °C and 15 psi). Peanut plants were watered twice a week with an atomizer for four weeks.

Two inoculation categories were performed. Potato dextrose broth (PDB) was inoculated with one PDA plug containing 2-days old mycelia of *S. minor* and incubated at 24 °C for 3 days. Similarly, 4-weeks old peanut plants were inoculated (on the stem where a node was present) with one PDA plug containing 2-days old mycelia of *S. minor*. A small lesion was created near the

inoculation point to facilitate infection. Inoculated plants were kept at 24 °C and 80% relative humidity for up to five days. Each inoculation category (host and media) had 5 replicates.

Three days old mycelia growing on PDB was filtered using whatman grade 1 filter paper and fast-frozen with liquid nitrogen until RNA extraction. Concomitantly (3 days post inoculation), a 2 inch piece of peanut infected tissue (stem) was aseptically collected on 15 mL glass tubes and fast-frozen on liquid nitrogen until RNA extraction. RNA from both frozen samples was extracted using the RNeasy Plant Mini Kit from Qiagen®. RNA quality was checked for integrity using the assay "eukaryote total RNA nano series II" in a bioanalyzer 2100 from Agilent Technologies™ housed at Oklahoma State University. Library preparation used a Poly(A) enrichment methodology and RNA sequencing in both infecting (3 days post inoculation Peanut plant) and non-infecting mycelium (PDA growing mycelia) was done using an Ilumina HiSeq 2500 housed at the core facility of the University of Illinois at Urbana-Champaign.

EDNA modification

The modification of the original version of EDNA to produce EDNAtran occurs mainly at the e-probe design level. E-probe databases were made in Fasta format and the header of each e-probe had an annotation format as observed in Figure IV-2. The Fasta headers were space delimited lines containing annotations for each e-probe which were retrieved by an extra blastn step added during stage 3 of the e-probe design pipeline (Figure IV-3). The pairwise alignment retrieved the coordinates of the e-probe in the target genome and compared it with the genome annotation coordinates found in a gff3, gff or gtf files (annotation files). Once the coordinates were retrieved, e-probes generated only on exons of the genome were kept in a separate file. Further analysis was done to verify if the selected e-probes were found in genes of interest. In this case the genes of interest included genes encoding for CWDE and genes associated with the production of oxalic

acid (OA). Gene ontology terms were retrieved from the *S. minor* genome annotation and used for the selection of the e-probes. However, being this analysis a proof of concept, all the analyses were done using all the exonic regions of the *S. minor* genome.

EDNAtran in metatranscriptomic data

The e-probe database containing the annotated exonic e-probes were aligned to the target transcriptomic databases as described in Stobbe *et al.,* 2013 and Espindola et al., 2015 for metagenomes. The alignment output was parsed based on e-probe hit frequencies and HQMs. Statistical assessment using ANOVA, a Tukey HSD, and pairwise T-test was incorporated to compare hits of near neighbor's e-probes (internal negative controls) with the hits of *S. minor* exonic e-probes. Transcriptomic databases in this study included *S. minor* growing on PDB and *S. minor* infecting a peanut plant. Therefore, e-probes specific to *S. minor* as well as e-probes specific for two near neighbors (*S. sclerotiorum* and *B. cinerea*) were queried against the transcriptome obtained from *S. minor* growing on PDB and the transcriptome of *S. minor* infecting a peanut plant. It was expected that the frequency of hits in *S. minor* infecting peanut was higher than the frequency of hits in *S. minor* growing on PDB. Conversely, negative control e-probes were expected to have very low hit frequency (zero or close to zero). The output was displayed in the terminal screen as well as in tab-delimited tables created in the linux working directory.

Results/Discussion

EDNAtran is a modified version of EDNA that works in most linux environments having installed the appropriate dependencies. Until now it has been successfully tested in personal computers, although it can be scalable for High Performance Computing (HPC) infrastructures. We successfully analyzed the sequenced transcriptomes of *S. minor* growing on different substrate

elements and determined EDNAtran's usefulness for the detection of actively growing plant pathogens.

EDNA transcriptomics

RNA sequencing libraries yielded 21.91 million reads for *S. minor* infecting peanut plant and 22.38 million reads for the *S. minor* growing on PDA. The original EDNA tool relied on the presence of the target gDNA in the metagenomic sample to successfully detect the pathogen of interest. Although RNA was part of the metagenomics databases obtained due to whole genome and whole transcriptome amplification of the sample, RNA sequencing was not originally analyzed as a single variable in the detection process using EDNA on eukaryotes yet. EDNAtran was created as a new approach to consider RNA sequencing as an alternative option that provides insights into the presence of physiologically active target organisms. It was expected that all reads obtained during the sequencing of *S. minor* infecting peanut plants metatranscriptome

EDNAtran requires fully sequenced and annotated eukaryotic genomes for e-probe design. Additionally, EDNAtran relies on upregulated genes that will likely be reflected in their frequency's presence in RNA sequencing databases. It was expected that genes associated with the necrotrophic behavior of the pathogen were activated and will serve as the main source of detection in RNA sequencing databases. However, selecting the appropriate time when these genes are activated is a crucial task that can only be completed by running multiple gene expression analyses with RNA sequencing at different points in the infection time. The selected model organism, *S. minor,* presents various stages during its life cycle. Primarily, it produces resting structures (sclerotia) a few days after infection and can be found in the soil usually as sclerotia in dormant stage (Abawi & Grogan, 1979) once the infection has completely decomposed plant tissues. Under optimal conditions, with 95 - 100% humidity and temperatures ranging between 18

and 25 °C, *S. minor* activates and starts an eruptive germination directly from the sclerotia (Dow, Porter & Powell, 1988; Wu & Subbarao, 2008). Theoretically, a variety of transcriptional factors may be activated during saprophytic and necrotrophic activity to produce CWDE, including beta-1,3-glucanases, cellulases, xylanases, cutinases and glyoxydases, as well as OA. We previously identified more than 400 Carbohydrate active enzymes (CAZymes) present throughout the *S. minor* genome with potential important roles in plant tissue maceration and saprophytic activity. However, when the production of these enzymes starts is still unknown. Yet, we were able to use a variety of exonic regions to accurately detect actively growing *S. minor*.

E-probe generation

E-probe generation followed four systematic steps to build e-probes that are specific for the detection of exonic regions in RNA sequencing databases (Figure IV-1). E-probes were tagged with coordinates (Figure IV-2) and protein coding gene information. The first step used mummer and nucmer to compare whole genomes. The software identified regions that are unique in the *S. minor* genome when compared to the *S. sclerotiorum* and *B. cinerea* genomes. Once the unique regions were identified, they were size selected to a desired length. As expected from previously reported data, e-probe optimal sizes ranged from 60 nt to 80 nt (Espindola et al., 2015). Pairwise alignments of the *S. minor* e-probes against the target genome were done to retrieve their coordinates. Annotation of each designed e-probe was based on functional annotations of *S. minor* (Chapter 3). During stage 3, the fasta file containing the e-probes was modified. The header of each e-probe was annotated with gene identification as well as coordinate information about their localization in the genome including intron/exon information. E-probes were also tagged based on potential gene functions, for example CWDE e-probes. Specifically for *S. minor*, e-probes were designed in genes that encode for carbohydrate-active enzymes (Figure IV-2).

E-probe curation steps are crucial to be able to accurately identify the pathogen. Therefore, two layers of curation were added in EDNAtran since multiple eukaryotic organisms having similar orthologous proteins might also be holding the same genetic code for those proteins. Therefore, a fourth stage used pairwise alignments against their near neighbor genomes, in this case *B. cinerea* and *S. sclerotiorum* to further eliminate e-probes that produce false positives when running EDNAtran. The output e-probes were also curated by blastn (Camacho et al., 2009) against the nt database of NCBI. E-probes hitting on species other than *S. minor* are eliminated.

E-probes of two different lengths were created (60-mer and 80-mer). A total of 14,191 e-probes were generated for the 60-mer length and a total of 2,947 e-probes were generated for the 80-mer length. For comparative purposes, e-probes were also generated for *S. sclerotiorum* and *B. cinerea* which were used as negative controls. Number of e-probes generated for *S. sclerotiorum* ranged from 13,366 for 60-mer e-probe length to 4,190 for 80-mer e-probe length. Similarly, numbers of e-probes generated for *B. cinerea* ranged from 1,409 for 60-mer e-probes to 4,552 for 80-mer e-probes. As a rule of thumb, generating shorter e-probes yields higher number of e-probes, however, *B. cinerea's* 60-mer e-probes were less numerous than 80-mer e-probes.

EDNAtran assessment

EDNAtran aims to identify physiologically active plant pathogens by using RNA sequencing. E-probes were generated in exons potentially up-regulated during *S. minor* necrotrophic and saprophytic activity but potentially not during sclerotial development. Initial and basic assessment of detection was performed by counting High Quality Matches (HQMs). A higher number of high quality matches gave a rough estimate that the *S. minor* transcriptomic sample was positive for the *S. minor* e-probes tested (Table IV-1 & Figure IV-4).

However, EDNAtran contains a semi-quantitative component that uses e-probe hit frequencies to infer potential upregulating genes based on the phenotypic behavior of the pathogen. Traditional EDNA's statistical analysis relied on T-student test that compared e-probe hit frequencies of an unknown sample vs. decoy e-probe hit frequencies. The decoy e-probes are sets of reversed e-probes (used as a negative control) that are generated from the original set of e-probes which are pathogen specific. Shuffled e-probes have been traditionally used for eukaryotic plant pathogens. Although this has been still used as an internal negative control, an experimental negative control was added to the EDNAtran's pipeline. The experimental negative control can consist of 1). metagenomic/metatranscriptomic databases that do not to contain sequences from the pathogen of interests. If there is not a real metagenomic/metatranscriptomic negative control available, 2). an *in silico* negative control can be generated by using the host genome information. Yet, if there is not any possibility to create an *in silico* simulated negative control; 3). e-probes generated on exonic regions from the *S. minor* taxonomically nearest neighbor can be utilized. The latter was utilized for this analysis due to budget limitations.

EDNAtran validation used two real RNA sequencing databases as "unknown" samples. Ideally, higher number of transcripts associated to CWDE will be found in an actively infecting *S. minor*. Therefore, it was assumed that there will be a higher e-probe hit frequency in *S. minor* infecting peanut than in *S. minor* growing on PDB. Thus, a comparison between the two RNA sequencing databases was performed, and e-probes generated for *B. cinerea* and *S. sclerotiorum* were used as negative controls where e-probe hit frequencies were expected to be zero or close to zero.

The statistical analysis had one independent variable which is the e-probe set (*S. minor, S. sclerotiorum* or *B. cynerea*); and one dependent variable which is e-probe hit frequency. Analysis

of variance (ANOVA) was selected as the statistical model for EDNAtran, although further enhancements will include other analysis like factorial-ANOVA and MANOVA. Thus, 6 treatments per e-probe length were analyzed as a group. Theoretically, EDNAtran statistical analysis is designed to use only one e-probe length. Yet, both e-probe lengths were utilized in this statistical analysis yet, they are not statistically comparable in this study. A total of 12 averaged hit frequencies (treatments) where compared and analyzed, the resulting ANOVA showed that at least one of the participating frequencies (treatments) was different than the others. TukeyHSD (Figure IV-6) and a pairwise T-test (Table IV-2) were performed to identify which EDNAtran result differs from the others. As expected, e-probes from *S. minor* had significantly higher frequencies of hits than e-probes from *S. sclerotiorum* and *B. cinerea* in the transcriptomic databases. Confirming that EDNAtran approach can successfully detect *S. minor* using an RNA sequencing data set (Figure IV-4 & Figure IV-5A, 5B).

No hits were observed on the infected peanut metatranscriptomic database, possibly because sclerotia formation was in progress and little or no mycelial growth and/or saprophytic activity was underway (Table IV-1 & Figure IV-5C, 5D). We could hypothesize that during sclerotia development, CWDE and other proteins are not produced in detectable amounts. Future studies should include metatranscriptomic replicates at different times post inoculation that will allow to clearly define the time at which transcripts are activated and e-probes targeting coding regions associated with sclerotial formation, or targeting housekeeping genes could be better designed.

In conclusion, EDNAtran effectively detected *S. minor* using RNA sequencing databases generated from mycelium growing on PDB using e-probes designed on exonic regions. Theoretically, CWDE should account for most of the transcripts present in both *S. minor* in PDB

and *S. minor* infecting a peanut plant sequencing databases. However, the highest e-probe hit frequencies were found in databases originating from *S. minor* growing on PDB. Yet, no e-probe hit frequencies were recorded for *S. minor* infecting a peanut plant. Suggesting that the transcripts from *S. minor* were not being produced actively when the sample was collected, or that the plant immunity system could have degraded most of the pathogenic transcripts. Further analysis should include more replicates and different sample collection times from *S. minor* infecting peanut plants.

Literature cited

Abawi, G.S. and Grogan, R.G. (1979) Epidemiology of Diseases Caused by Sclerotinia Species. *Phytopathology* **69**, 899–904.

Agrios, G. (2005) Plant Pathology Fifth, ed, Elsevier.

Amselem, J., Cuomo, C.A., Kan, J.A.L. van, et al. (2011) Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genet.* **7**, e1002230.

Boehm, M.J. and Hoitink, H.A.J. (1992) Sustenance of microbial activity in potting mixes and its impact on severity of pythium root-rot of poinsettia. *Phytopathology* **82**, 259–264.

Butzler, T.M., Bailey, J. and Beute, M.K. (1998) Integrated Management of Sclerotinia Blight in Peanut: Utilizing Canopy Morphology, Mechanical Pruning, and Fungicide Timing. *Plant Dis.* **82**, 1312–1318.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421.

Chen, K. and Pachter, L. (2005) Bioinformatics for Whole-Genome Shotgun Sequencing of

Microbial Communities. *PLoS Comput Biol* **1**, e24.

Cousens and Croft (2000) Weed populations and pathogens. *Weed Res.* **40**, 63–82.

Dow, R., Porter, D. and Powell, N. (1988) Effect of environmental factors on *Sclerotinia minor* and Sclerotinia blight of peanut. *Phytopathology* **78**, 672–676.

Eisen, J.A. (2007) Environmental Shotgun Sequencing: Its Potential and Challenges for Studying the Hidden World of Microbes. *PLoS Biol* **5**, e82.

Espindola, A.S., Garzon, C.D., Schneider, W.L., Hoyt, P.R., Marek, S.M. and Garzon, C.D. (2015) A new approach for detecting Fungal and Oomycete plant pathogens in Next Generation Sequencing metagenome data utilizing Electronic Probes. *Int. J. Data Min. Bioinfomatics* **12**, 115–128.

Goldman, J.J., Smith, O.D., Simpson, C.E. and Melouk, H.A. (1995) Progress in Breeding Sclerotinia Blight-Resistant Runner-Type Peanut. *Peanut Sci.* **22**, 109–113.

Huson, D.H., Auch, A.F., Qi, J. and Schuster, S.C. (2007) MEGAN analysis of metagenomic data. *Genome Res.* **17**, 377–386.

Koike, S., Subbarao, K., Davis, R.M. and Turini, T. (2003) Vegetable diseases caused by soilborne pathogens 8099th ed., UCANR Publications.

Melzer, M.S., Smith, E. a. and Boland, G.J. (1997) Index of plant hosts of *Sclerotinia minor*. *Can. J. Plant Pathol.* **19**, 272–280.

Pane, C., Spaccini, R., Piccolo, A., Scala, F. and Bonanomi, G. (2011) Compost amendments enhance peat suppressiveness to *Pythium ultimum*, Rhizoctonia solani and *Sclerotinia minor*. *Biol. Control* **56**, 115–124.

Pascual, J.A., Hernandez, T., Garcia, C., Leij, F. De and Lynch, J.M. (2000) Long-term suppression of *Pythium ultimum* in arid soil using fresh and composted municipal wastes. *Biol. Fertil. Soils* **30**, 478–484.

Pop, M. and Salzberg, S.L. (2008) Bioinformatics challenges of new sequencing technology. *Trends Genet.* **24**, 142–149.

Stobbe, A.H., Daniels, J., Espindola, A.S., Verma, R., Melcher, U., Ochoa-Corona, F., Garzon, C., Fletcher, J. and Schneider, W. (2013) E-probe Diagnostic Nucleic acid Analysis (EDNA): A theoretical approach for handling of next generation sequencing data for diagnostics. *Microbiol. Methods* **94**, 356–366.

Swain, S., Harnik, T., Mejia-Chang, M., Hayden, K., Bakx, W., Creque, J. and Garbelotto, M. (2006) Composting is an effective treatment option for sanitization of *Phytophthora ramorum*-infected plant material. *J. Appl. Microbiol.* **101**, 815–827.

Truong, D.T., Franzosa, E. a, Tickle, T.L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C. and Segata, N. (2015) MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* **12**, 902–903.

Wadsworth, D.F. (1979) Sclerotinia Blight of Peanuts in Oklahoma and Occurrence of the Sexual Stage of the Pathogen. *Peanut Sci.* **6**, 77–79.

Wetterstrand, K.A. (2013) DNA sequencing costs: data from the NHGRI Genome Sequencing Program (GSP). *Natl. Hum. Genome Res. Inst.*

Wu, B.M. and Subbarao, K. V (2008) Effects of soil temperature, moisture, and burial depths on carpogenic germination of *Sclerotinia sclerotiorum* and *S. minor*. *Phytopathology* **98**, 1144–

1152.

Figures

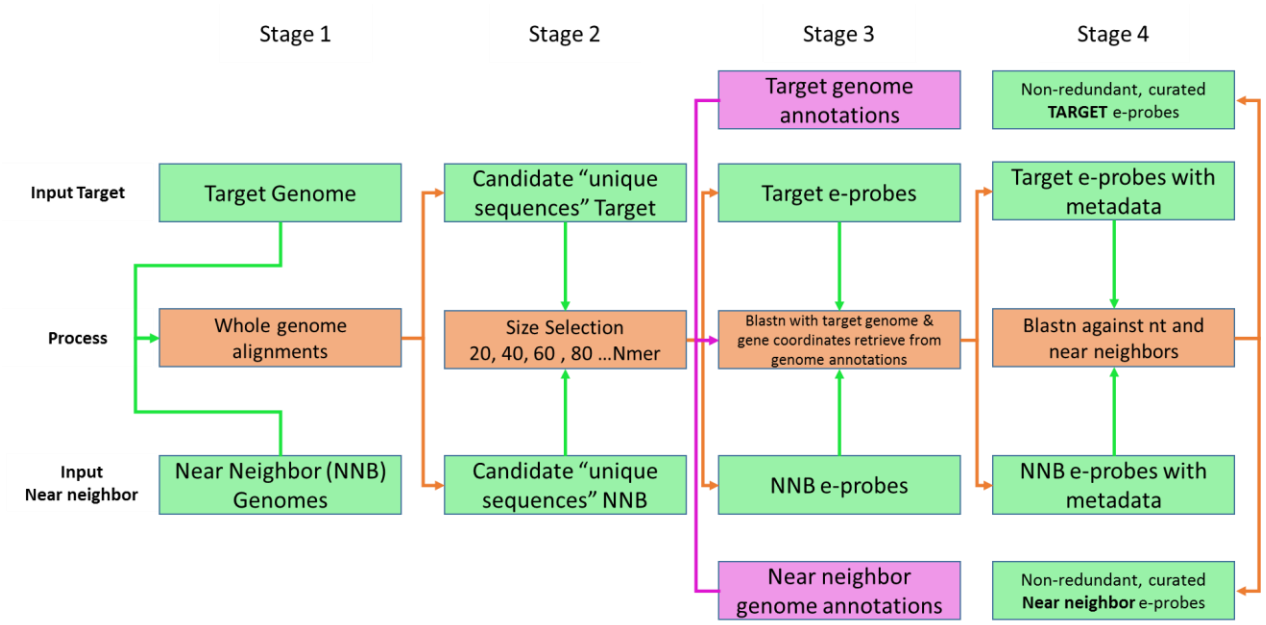

Figure IV-1. Pipeline of EDNA Transcriptomics (EDNAtran). The figure depicts the complete pipeline that was designed specifically to create curated e-probes that will be able to identify actively infecting eukaryotic plant pathogens. The pipeline takes advantage of full genomes with their corresponding annotation files (gff3). It contains four stages: Stage 1. Refers to the whole genome comparisons that are performed to the target and the nearest neighbor. Stage 2. Candidate sequences are size selected. Stage 3. E-probe files are annotated with functional genome annotation; Stage 4. Where e-probe databases are blasted against their near neighbor as well as the nt database to eliminate potential false positives.

```
>minor-80-33 contig00016 Exon HitStart: 389 HitEnd: 468
CAAATCAATCGTCCGATACCACTCAAATTTTCAATTCTAGTCAGCCATCCAATAGGCTTCGGACATTCGACTCCATTTCA
>minor-80-34 contig00016 Intron HitStart: 7562 HitEnd: 7641
ATTCTGCGGGGTTTACCCCTGGTTACCCTGAGTATATACCCTAGGTACCTTTTGCCCAACCCCGATTCCTACTTTAATGC
>minor-80-35 contig00017 Intron HitStart: 2779 HitEnd: 2858
GTAGAGAGTAATAAGCACTGTACAGTGTAGTTACCGGTACGGATAACCTAACCCTAGTAACCTGAGAGTTTACCCCGCCA
>minor-80-36 contig00017 Intron HitStart: 14176 HitEnd: 14255
TCTAAACTATCAGCAACAACCTAAATGTGTGGTCACTTGAGATAAAATGGAGTGAGCGGAGTGTTAATCTTGGTTGATAA
>minor-80-37 contig00017 Intron HitStart: 16386 HitEnd: 16465
TGATTGTAATCGCGTCACTTTCATGTCCCGCAACACTGAGGACGGAAGGAATGCGTCTTCGTGAGATAAGTTGAAACGGG
>minor-80-38 contig00018 Intron HitStart: 19796 HitEnd: 19875
ATTGCAAGCTAAACCTCTTCACTTGAAAAGCACTACCCGAAACTCGTGTGTGCCACTACAGACAGTGAAAAATTCTCCAC
>minor-80-39 contig00019 Intron HitStart: 13303 HitEnd: 13382
TACACCTCCCCTCCCATATACTCTCACGAGAAGCATCTGAACCCGTACTAACACGACATCAGGTGACGCCATGTACCCAT
>minor-80-40 contig00020 Intron HitStart: 15563 HitEnd: 15642
AATACTTACTCTCTACGTACTCTGTTACTCGGTTGTTGACATATCTGGTGATCCAACAAGTATGACTGCATGCCATGCAT
>minor-80-42 contig00021 Exon HitStart: 3509 HitEnd: 3588
TAGAATAAGGTCGAGTTATGACTGTGTATCCATGCGTAACAGAGTGCCTGCTACTTTTAGCATAGCAATCCCGCCGAGAA
>minor-80-43 contig00022 Intron HitStart: 2414 HitEnd: 2493
AGTAGGAAGGAGCAAATTATAGCAGACTTGTTGTTATGGATGCTAGATTTAATAGCCCGCAGGCAAAATTCAACAATGAA
>minor-80-44 contig00022 Intron HitStart: 2513 HitEnd: 2592
TGATGTGAGAAGTTGAAAATCAGCTGTTCATTCCTAACTTCCTAGCTTCTCGACCATTAAAAGTTAGGATCTGGAATAGC
>minor-80-45 contig00022 Intron HitStart: 22274 HitEnd: 22353
TTAGTACATAGATGGCGCGTAATCTGGAGTTGAATAACTGCGAGAAGAATGTCTCGGCCAGTGTGAGAGAGCGATTTCCA
>minor-80-46 contig00023 Intron HitStart: 3496 HitEnd: 3575
CTTATACCATCCATCTTGAAACGTTCGAGTACTAAAGCAGTGCTCATCTGCCCCATCATGACTTACGTTCGGAAATCTTC
>minor-80-47 contig00024 Intron HitStart: 21572 HitEnd: 21651
CTGAATAAATTTTGAGGTCTATGGAACTGGGTAGCTTGGAGAAGCTGCTTATTAAGGTGAGAAGTAGGGTTCAATAGATT
>minor-80-48 contig00025 Intron HitStart: 15917 HitEnd: 15996
ACCGACAGGATCCCCAGCTTTTGGCCTATTTCACTTATAGCTCCAGATCTTTCCTACTAGATGTTCTGTGAATCTGGCTG
>minor-80-49 contig00025 Exon HitStart: 18491 HitEnd: 18570
TCAAAGGCAACCGGCAGATGCCGCAATATTCGTCAGGTGGGATATCTGCGAACAAAACCGGTGCAAGGATGTCGGGATTG
```

Figure IV-2. Example of the e-probe tagging step output (e-probes + metadata) while designing e-probes for EDNA-transcriptomics. In this specific figure which represents a fasta file containing 80-mer e-probes for *Sclerotinia minor*, the headers of each sequence are tagged with four informational tags. The tags are tab-delimited in the header of each sequence and contain information about: 1. e-probe unique identifier, 2. Genome identifier where the sequence is found, 3. Genome annotation information (intron/exon). 4. E-probe coordinates in the *S. minor* genomic sequence.

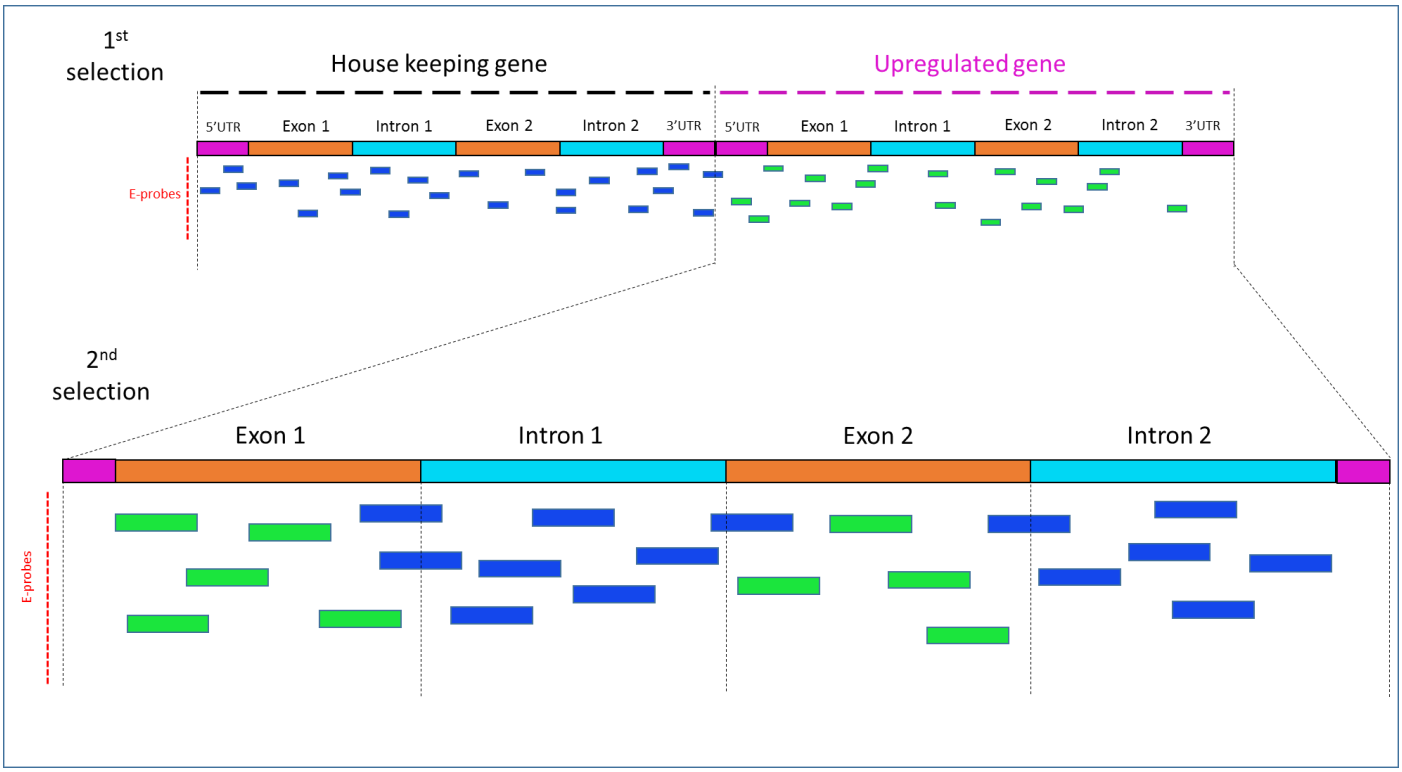

Figure IV-3. E-probe selection process during e-probe design stage 3. Annotated genomes are required to design e-probes for EDNA transcriptomics. The figure compares the selection process by using an example of two contiguous genes (for demonstration purposes, one has been named housekeeping gene). During the e-probe generation process regions that are known to be upregulated are selected (green) and regions suspected to be downregulated or non-transcribed are eliminated (blue).

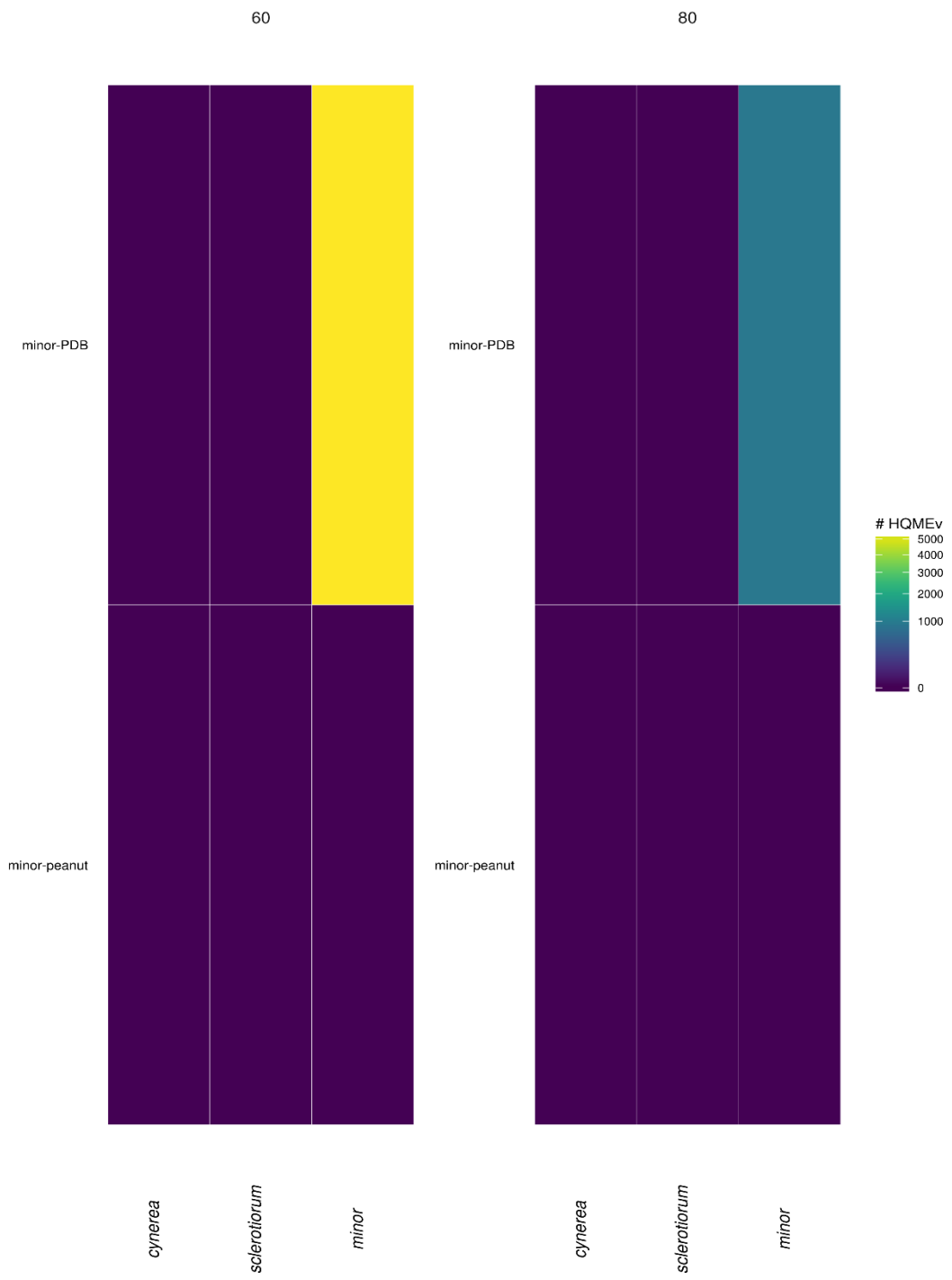

Figure IV-4. Heat map hierarchy clustered by High Quality Matches (HQMs) that include E-value into its diagnostics calculation. The graphic shows positive detection of *Sclerotinia minor* from RNA sequencing databases using 60-mer and 80-mer e-probes.

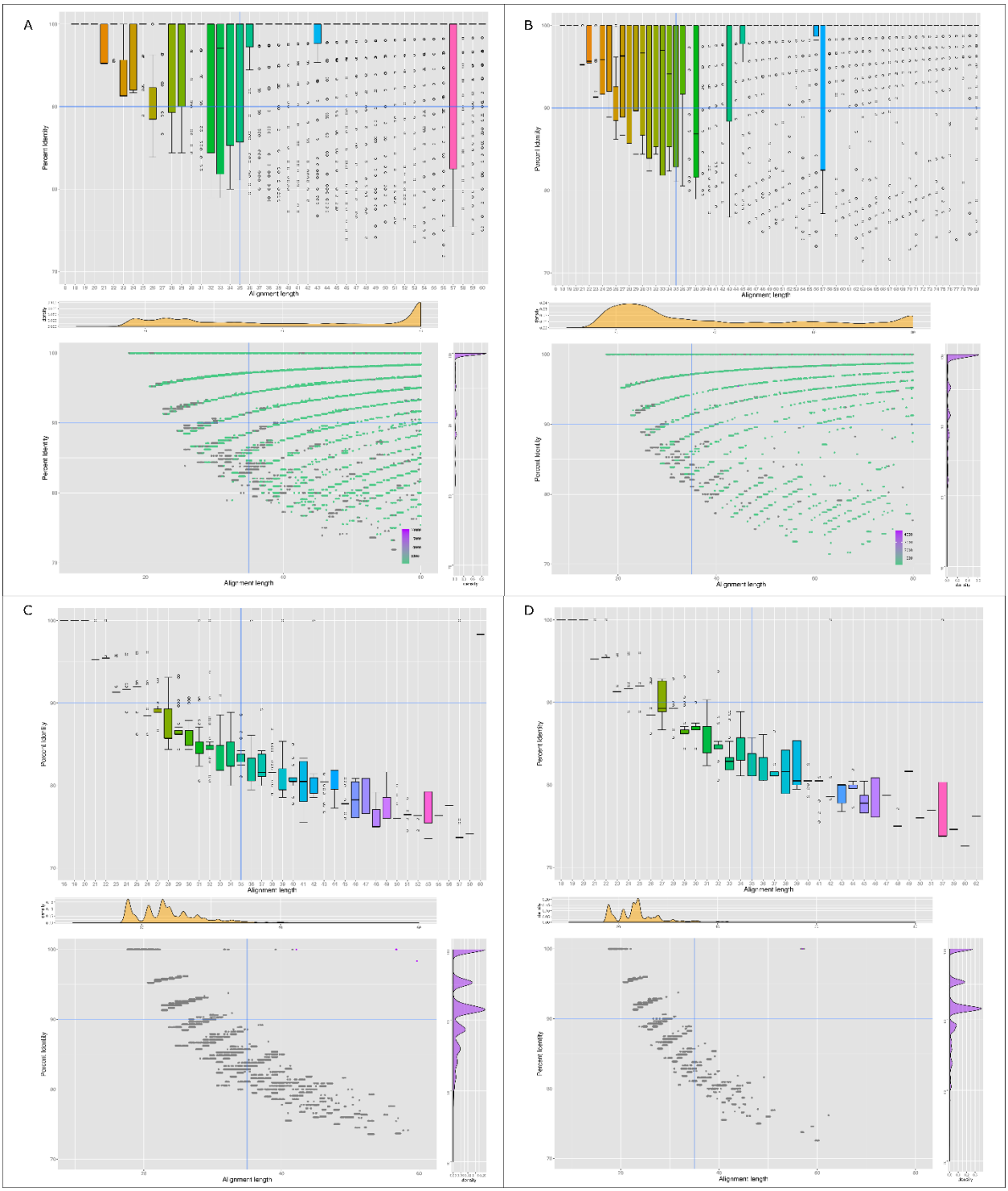

Figure IV-5. EDNA transcriptomics hits distribution and frequencies based on alignment length and percent identity of *Sclerotinia minor* exonic e-probes. (A,B) RNA sequencing of *Sclerotinia minor* growing on PDB identified with 60-mer and 80-mer EDNA transcriptomics e-probes. (C,D). RNA sequencing of *S. minor* infecting 4-weeks old peanut plant. The sample was taken at 3-days post inoculation.

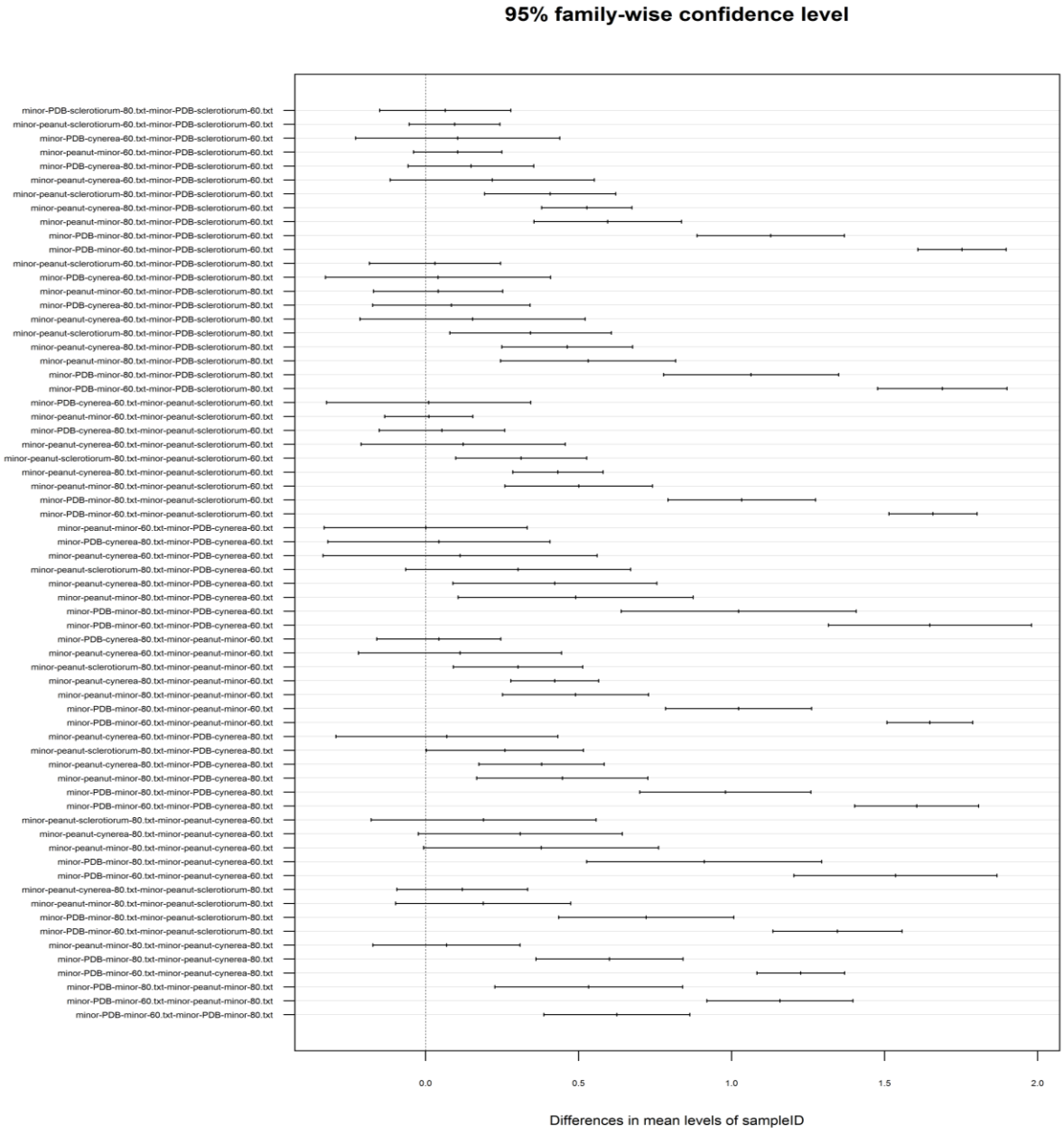

Figure IV-6. Post-hoc analysis of variance (ANOVA) using Tukey Honest Significant Difference (HSD) with 95% of confidence for the EDNA transcriptomic analysis of *Sclerotinia minor*. Lines close to zero are interactions that had no difference in their hit frequency mean while lines closer to 2 are interactions that had different hit frequency means between each other.

Tables

Table IV-1. Output table produced by the EDNA transcriptomics pipeline for the *Sclerotinia minor* analysis. The table contains 11 columns. A). the metagenome codes B). the metagenome read number C). metagenome average read length, D). the metagenome maximum read length, E). the metagenome minimum read length, F). the e-probe identification code, G). the e-probe length, H). the number of e-probes I) the high quality match (HQM) numbers calculated with e-values, J). the HQMs calculated without e-value, K). high scoring general matches (HSGM) that uses only percent identity higher than 90% as a stringency parameter.

A	B	C	D	E	F	G	H	I	J	K
minor-PDB	22378513	98.95291269	100	35	minor	60	14191	5582	5645	5657
minor-PDB	22378513	98.95291269	100	35	minor	80	2947	890	905	906
minor-peanut	21910920	98.97474287	100	35	minor	60	14191	0	0	0
minor-peanut	21910920	98.97474287	100	35	minor	80	2947	0	0	0
minor-peanut	21910920	98.97474287	100	35	cinerea	60	1409	0	0	0
minor-peanut	21910920	98.97474287	100	35	cinerea	80	13395	0	0	0
minor-PDB	22378513	98.95291269	100	35	cinerea	60	1409	0	1	3
minor-peanut	21910920	98.97474287	100	35	sclerotiorum	60	13366	0	0	0
minor-PDB	22378513	98.95291269	100	35	cinerea	80	4552	0	5	51
minor-PDB	22378513	98.95291269	100	35	sclerotiorum	60	13366	0	18	213
minor-PDB	22378513	98.95291269	100	35	sclerotiorum	80	4190	0	15	145
minor-peanut	21910920	98.97474287	100	35	sclerotiorum	80	4190	0	0	0

Table IV-2.Pairwise T-test analysis showing multiple comparisons of e-probe hit frequencies in EDNAtran of *Sclerotinia minor*. P-values lower than 0.05 are highlighted with red.

	minor-PDB-cinerea-60.txt	minor-PDB-cinerea-80.txt	minor-PDB-minor-60.txt	minor-PDB-minor-80.txt	minor-PDB-sclerotiorum-60.txt	minor-PDB-sclerotiorum-80.txt	minor-peanut-cinerea-60.txt	minor-peanut-cinerea-80.txt	minor-peanut-minor-60.txt	minor-peanut-minor-80.txt	minor-peanut-sclerotiorum-60.txt
minor-PDB-cinerea-80.txt	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
minor-PDB-minor-60.txt	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA
minor-PDB-minor-80.txt	1.27E-66	1.27E-66	NA	NA	NA	NA	NA	NA	NA	NA	NA
minor-PDB-sclerotiorum-60.txt	1	1	0	1.27E-66	NA	NA	NA	NA	NA	NA	NA
minor-PDB-sclerotiorum-80.txt	1	1	0	1.27E-66	1	NA	NA	NA	NA	NA	NA
minor-peanut-cinerea-60.txt	1	1	0	1.27E-66	1	1	NA	NA	NA	NA	NA
minor-peanut-cinerea-80.txt	1	1	0	1.27E-66	1	1	1	NA	NA	NA	NA
minor-peanut-minor-60.txt	1	1	0	2.04E-66	1	1	1	1	NA	NA	NA
minor-peanut-minor-80.txt	1	1	0	1.45E-66	1	1	1	1	1	NA	NA
minor-peanut-sclerotiorum-60.txt	1	1	0	1.27E-66	1	1	1	1	1	1	NA
minor-peanut-sclerotiorum-80.txt	1	1	0	1.27E-66	1	1	1	1	1	1	1

CHAPTER V

V METATRANSCRIPTOMICS FOR MONITORING TOXIN PRODUCING *ASPERGILLUS FLAVUS*

Abstract

E-probe Diagnostic Nucleic acid Analysis (EDNA) is a bioinformatic tool originally developed to detect plant pathogens in metagenomic databases. However, enhancements made to EDNA permitted to increase its capacity to conduct hypothesis directed detection of specific gene targets present in metatranscriptomic databases. To target specific pathogenicity factors used by the pathogen to infect its host or other targets of interest, e-probes need to be developed in transcripts related to that function. In this study, EDNAtran was used to semi quantitatively detect the expression of genes related to aflatoxin production at the metatranscriptomic level in *A. flavus*. E-probes were designed from up-regulated genes during *A. flavus* aflatoxin production. EDNAtran successfully inferred aflatoxin production using e-probes that targeted the aflatoxin gene cluster metabolic pathway. Treatments included *A. flavus* using corn as nutrient source and *A. flavus* using Potato Dextrose Broth (PDB) as a nutrient source.

Introduction

Maize (Marsh & Payne, 1984), peanuts (Hill et al., 1983), tree nuts, dried spices (Llewellyn et al., 1992) and cottonseed (Cotty, 1997) are crops that can be infected during the pre-harvest, post-harvest and/or storage period with *Aspergillus flavus* Link. This fungus produces polyketide secondary metabolites named aflatoxins. Among the four known aflatoxins (B_1, B_2, G_1, G_2), B_1 has been of special interest to food biosecurity due to its toxicity and potent carcinogenic properties (Squire, 1981). *A. flavus* is an ubiquitous, saprophyte ascomycete fungus grouped in the *Aspergillus* section Flavi. Species containing aflatoxin-producing strains include *A. flavus, A. parasiticus* and *A. nomius* (Kurtzman, Horn & Hesseltine, 1987; Klich & Pitt, 1988).

Aflatoxin contamination in food is highly regulated in multiple countries, consequently increasing management costs and final product price (Robens & Cardwell, 2003; Van Egmond, Schothorst & Jonker, 2007; Wu, Liu & Bhatnagar, 2008). In the United States alone, the maximum allowed concentration of aflatoxin in food for human consumption is 20 ppb, as dictated by the U.S. Food and Drug Administration (FDA). Appropriate and accurate aflatoxin testing is necessary to opportunely control *A. flavus* infected crops. Among the most used techniques for aflatoxin detection and quantification are TLC, LC, ELISA and fluorometry (Van Egmond, Schothorst & Jonker, 2007). Industry costs for testing crops for aflatoxins in the United States alone range from $30 to $50 million per year at approximately at $10 to $20 per sample tested (Robens & Cardwell, 2003).

Aflatoxin is produced through a polyketide metabolic pathway with the interaction of approximately 25 genes encoded by the aflatoxin gene cluster (Trail et al., 1995; Yu et al., 1995; Minto & Townsend, 1997). *aflR* and *aflS* (*aflJ*) are regulatory genes where *aflR* specifically encodes for a transcriptional factor of the type $Zn(II)_2Cys_6$ which can bind promoter regions of

many aflatoxin genes (Chang et al., 1993; Woloshuk et al., 1994; Yu et al., 1996; Fernandes, Keller & Adams, 1998). In contrast aflS (*aflJ*) regulates aflatoxin production trough binding and activating aflR (Chang, 2004).

Although host resistance to *A. flavus* has not been developed, promising studies have found genetic factors suitable to confer resistance in maize (Chen et al., 2011). Another promising management strategy for aflatoxin reduction includes biocontrol using atoxigenic strains of *A. flavus* (Aflaguard® & AF36) (Garber & Cotty, 1997; Probst et al., 2011). For appropriate management using atoxigenic strains, the use of indigenous atoxigenic strains is recommended to avoid potential adaptation problems (Mehl et al., 2012). The indigenous isolated atoxigenic strains can be multiplied *in vitro* and later inoculated on crop fields with the purposes of excluding toxigenic strains through competition, since they occupy the same niche (Cotty, 1994). Screening after inoculation of the atoxigenic strain is performed regularly for both the production of aflatoxin and in some cases strain viability ratio atoxigenic:toxigenic. The viability of the atoxigenic strain can be inferred by testing for the presence of aflatoxin. If higher concentrations of toxin are found, inoculation of the field with additional atoxigenic is recommended (Mauro, Battilani & Cotty, 2015). Both, the identification of newly-infecting toxigenic *A. flavus* strains and biocontrol screening require sensitive testing for their identification.

While there have been multiple attempts to use genetic features to discriminate toxigenic/atoxigenic strains, the rapidness of some quantitative kits like ELISA makes them more practical than nucleic acid based methods in spite of the power of such tools. However, aflatoxin using immunoassay detection might be useful for *A. flavus* strain screening only when a substantial growth of the fungus and relevant amounts of the toxin are present. This limits their sensitivity for potential detection of asymptomatic infections with toxin levels under the detection limit of the

assays. Hence, their use is not feasible as preventative approaches, but aims to serve as an assessment approach followed by therapeutic methods. On the contrary, genetic-based tools have the flexibility of being used as early infection detection tool for most organisms, due to their sensitivity and specificity. Nevertheless, the large amount of genes that need to be targeted for proper discrimination of aflatoxin producing strains has limited the development of such tools. Currently, RT-PCR tests targeting coding regions in the aflatoxin gene cluster have been suggested to replace microbiological and chemical methods for the identification of aflatoxin-producing strains of *A. flavus* (Niessen, 2008). Furthermore, newly developed monitoring techniques focus mainly on genetic characterization of the aflatoxin gene cluster (Chang, Horn & Dorner, 2005; Donner et al., 2010). The most recent is a DNA based monitoring technique with 32 markers amplified in four multiplex PCR. The protocol relies on deletions occurring in the aflatoxin gene cluster of atoxigenic strains (Callicott & Cotty, 2015).

A fast and tentatively less expensive screening tool for toxigenic *A. flavus* strains might be sequencing the whole metagenome of the pathogen niche and determining the presence of potential toxigenic inoculum. Second generation sequencing or next generation sequencing (NGS) techniques can sequence billions of nucleotides in one single run. Sequencing costs based on the U.S. National Human Genome Research Institute calculates that the cost for sequencing one million bases is $0.015 per Mb as of July of 2015. Such decreasing costs have allowed to sequence approximately 3 billion reads (one human genome) for $1,363 (Geoff Spencer, 2001). Furthermore, sequencing costs will continue to drop with the advent of third generation sequencing techniques which use single molecule sequencing. Consequently, metagenome taxa assignment (MTA) is becoming a challenging task. Multiple tools have been developed to assign taxa for metagenomic outputs. Many of them are designed for clade specificity while others have a more

general approach (Huson et al., 2011; Truong et al., 2015). One of them addresses specifically MTA for plant pathogens precisely for viruses, bacteria, fungi & oomycete by using species specific markers or e-probes (Stobbe et al., 2013; Espindola et al., 2015). However, none of them have addressed the detection of genes involved in mycotoxin production, an application with great potential for mycotoxin assessment in food biosecurity.

In the present study, *A. flavus* was used as a model system for the development of e-probes that target genes in the aflatoxin gene cluster. This EDNAtran protocol aimed to discriminate metagenomics databases from samples infected by either toxigenic or atoxigenic *A. flavus* strains, which will permit the screening of fields that could be potentially infected with toxigenic *A. flavus.*

Discrimination with e-probes can be performed at both, genome and transcriptome level. A previous study focused on *Sclerotinia minor* used e-probes targeting exonic regions of CAZyme genes to detect physiologically active mycelium (Chapter 4).

EDNAtran has not been tested for detection of expression of gene clusters, smaller target than the whole exome of a fungus. The aim of this study was to detect *A. flavus* in metatranscriptomic samples by using e-probes designed in coding regions of the aflatoxin gene cluster, and compare the differential transcription of the aflatoxin gene cluster without the hassle of assembling the metagenome.

Experimental Procedures

Fungal isolates and culture methods

Aspergillus flavus strains were obtained as freeze dried (AF36; ATCC 96045; atoxigenic) or frozen (AF70; ATCC MYA-384; toxigenic) cultures from ATCC (Manassas, VA). AF36 was reactivated by rehydration, adding autoclaved distilled water inside the vial and then the contents

were plated on Malt extract agar Blakeslee's formula (MEAbl) and incubated at 31°C until mycelium was developed (72 hours), according to ATCC instructions. AF70 was thawed for 5 minutes and directly plated onto Malt extract agar and incubated at 25 °C until mycelium was developed (72 hours), according to ATCC instructions. Plugs with actively growing mycelia were used for plating onto PDA plates. *A. flavus* was grown on these plates at their optimal temperatures in the dark until extensive conidial development (5 days) was observed.

Both strains (AF70 and AF36) were inoculated and grown in ground corn and PDB. Dry corn kernels (*Zea mays*) were weighted (20g) and ground until obtaining pieces with the approximate texture of coarse sand (0.5-1mm in diameter). The coarse grains were autoclaved dry (dry cycle) for 20 minutes in polycarbonate containers (Magenta GA-7, Plantmedia, US) and its humidity was adjusted to hold between 25 – 33% w/v (Modified from (Woloshuk, Cavaletto & Cleveland, 1997). Simultaneously, yeast-extract sucrose (YES) media was prepared with 2% yeast extract and 15% sucrose (Probst & Cotty, 2012).

Corn and Potato Dextrose Broth (PDB) media were inoculated with conidial suspensions obtained by washing *A. flavus* MEAbl plates with 2 mL of sterile distilled water. Conidia collected were then added to a single vial containing 4mL of distilled water for a final dilution of 3:1 v/v. Six mL of spore suspension was used for each replicate (20 g of ground grains and PDB). The ground grain was inoculated with the *A. flavus* suspension in polycarbonate containers and homogeneously mixed by rolling the containers to allow uniform distribution of the conidia. On the other hand, PDB plates were inoculated with *A. flavus* spore suspension. Finally, the containers and PDB plates were incubated at 31°C until mycelial growth and conidia development which took 10 days. All environmental factors that affect aflatoxin production, like temperature, pH, metals/trace elements, nitrogen source, lipids contents and even light color, were the same during

culture of the two experimental strains (Georgianna & Payne, 2009), to achieve differentiation of the strains from transcriptomic data that reflected their genetic differences alone.

RNA extraction and sequencing

Mycelia of the two strains grown on PDB and on corn were used for RNA extraction by using the RNeasy Plant Mini Kit from Qiagen®. After quality control RNA was submitted to be sequenced using the Illumina HiSeq 2500 sequencer at the Core Facility of the University of Illinois at Urbana-Champaign, IL. The mRNA sequencing library was performed with poly(A) capture method per manufacturer's protocol and the metatranscriptome was sequenced as single-end.

Gene expression analysis

RNA sequencing reads were mapped to the *A. flavus* AF70 genome with STAR (Dobin et al., 2013) and bam binary files were created with SAMtools (http:// samtools.sourceforge.net). Gene expression analysis was performed by using DeSeq2 in R (Anders & Huber, 2010). Upregulated genes were retrieved by an in house linux bash script.

Transcriptomic discrimination

Aflatoxin detection by using transcriptomic approaches was achieved by selecting appropriate genetic signatures (e-probes) targeting genes that are up-regulated when aflatoxin is produced in AF70. The identification of up-regulated genes was achieved by challenging toxigenic *A. flavus* strains with environmental conditions that favor the production of aflatoxin and comparing them with environmental conditions that are not conducive for the production of aflatoxin. Up-regulated genes were retrieved and e-probes were generated targeting loci containing single nucleotide polymorphisms (SNPs) by comparative genetics using a local alignment search

against the transcriptome of the atoxigenic *A. flavus* strain on both conducive and non-conducive growing substrates. E-probes specificity and sensitivity was assessed comparing metatranscriptomic databases of *A. flavus* strains subjected to a variety of environmental conditions.

EDNAtran discrimination of toxin producing vs. non-toxin producing strains of A. flavus

The genomes from *A. flavus* AF70 (Accession: JZDT00000000.1) and NRRL3357 (Accession: AAIH00000000.2) (Nierman et al., 2015) were obtained from Genbank. In addition, the sequences for the aflatoxin gene cluster of AF70 (AY510453) and AF36 (AY510455) were also retrieved from GenBank (Ehrlich, Yu & Cotty, 2005).

E-probes were generated in lengths from 20 to 80 nt long using the e-probe generation pipeline for EDNA (Espindola *et al.*, 2015; Stobbe *et al.*, 2013). The target sequence for e-probe design was the aflatoxin gene cluster of both AF70 and AF36 *A. flavus* strains. Their specificity was verified by aligning the developed e-probes with the intended target and non-target sequences using a stringency of 100% identity and 100% query coverage. However, for purposes of this chapter, only AF70 e-probes were utilized for the analysis because the main objective is to discriminate between aflatoxin active producer and non-aflatoxin producers. It was expected that AF70 in PDB will not produce aflatoxin when compared with AF-70 Corn. Therefore, hit frequencies in AF-70-Corn should be higher than hit frequencies in AF70-PDB, AF-36-Corn and AF-36-PDB. Differences in hit frequencies was assesed using central tendency statistics. When only two samples were used a T-test was utilized, however, whenever more than 2 samples were analyzed an analysis of variance (ANOVA) was used to determine any differences on hit frequencies (p-value < 0.05). In the event of detectable differences, a post-hoc analysis using a

Tukey honest significance difference (HSD) test was performed to identify which samples are different from positive and negative controls.

Results and Discussion

Assessing appropriate growing conditions for the production of aflatoxin

The isolates of *A. flavus* AF70 (toxigenic) and AF36 (atoxigenic) showed different growing patterns and morphology on the different growing media tested (YES media, ground corn and ground rice). AF36 and AF70 grown on MEAbl produced conidia and sclerotia after 5 days, respectively. Therefore, for the inoculation of the three substrates, a spore suspension was used for AF36 and a sclerotia suspension was used for AF70. Conidia production was observed on both strains growing on ground rice and ground corn, however, AF36 produced mycelium only on YES media. Aflatoxin production is directly correlated with concentrations of free saccharides (Mateles & Adye, 1965; Davis, Diener & Eldridge, 1966; Mellon, Dowd & Cotty, 2005; Probst & Cotty, 2012); therefore AF70 and AF36 were grown in a toxin-inducing substrate (ground corn) and non-toxin inducing media (PDB). Increased sclerotia production was observed in AF70 compared with AF36, which produced mostly conidia in all media. However, AF70 produced conidia 10-days post inoculation in corn.

RNA sequencing and Gene expression analysis

RNA extracted from AF36 and AF70 strains growing on PDB and ground corn yielded from 20 to 24,9 million reads per sequencing run (Table V-1). The sequenced reads then were mapped to the *A. flavus* AF70 strain genome to retrieve information about potential up-regulation and down-regulation of genes by using STAR (Dobin et al., 2013) and DESeq2. A total of 129 genes were identified as up-regulated and 44 genes were down-regulated in the whole *A. flavus* genome. Not all the upregulated genes were found to be part of the aflatoxin gene cluster. Gene

regulation fold changes ranging from two to six were observed and have been plotted in a hierarchical clustering heat map as well as in a mean average plot (MA Plot) (Figure V-1, Figure V-2).

E-probe generation for aflatoxin detection

In total, 231 highly specific e-probes were generated to detect the production of aflatoxin specifically for AF70. AF36 genome wide e-probes were not generated because there is not a genome sequence available yet for that specific strain.

Multiple genes are involved in specific metabolic pathways in living organisms. Such genes tend to be found in gene clusters since they co-evolve at the same rate (Brown, Brown-Jenco & Payne, 1999; Ehrlich et al., 2004). Therefore, selecting up-regulated genes is crucial in EDNAtran since the likelihood of detecting the pathogen is higher than when e-probes are designed randomly throughout the genome. Previous RNA sequencing analysis needs to be performed to detect the up-regulated genes of interest (*A. flavus*), yet the literature can also be used as source of information to design the e-probes in up-regulated genomic regions when RNA sequencing does not provide enough information (Chapter 4). EDNAtran takes advantage of e-probes designed in key genes that are up-regulated during particular metabolic stages of the pathogen. EDNAtran capacity to detect plant pathogens in RNA sequencing databases was previously assessed using the fungal plant pathogen *S. minor* as a model system (Chapter 4).

Detecting aflatoxin production using EDNAtran in A. flavus

As expected, 231 e-probes had hits creating High Quality Matches (HQMs) in AF70-corn transcriptome data sets, meanwhile AF70-PDB had only 39 HQMs (Figure V-3 & Figure V-4). AF36-corn had only 2 HQMs and AF36-PDB had 12 HQMs (Table V-1). EDNAtran was able to

discriminate between the transcriptomic databases with abundant aflatoxin production and the non-toxigenic transcriptomes based on EDNA eukaryotic metrics (Table V-1). However, to indirectly assess the presence of aflatoxin we use frequencies of hits as a measuring value. In this case, the number of times a read was mapped to an e-probe was recorded and counted without any limits. An easy way of visualizing e-probe hit frequencies is by plotting a dot plot of alignment length vs. percent identity with marginal hit frequencies. Specifically for *A. flavus* AF70 in corn it was observable that the hit frequencies were very high — around 9,000 hits per e-probe — when the alignments are above 90% identity and the alignment length was approaching to the total length of the e-probe (Figure V-4A). Conversely, for AF70 in PDB and AF36 the marginal plots show a low frequency of hits when alignment lengths and percent identities were above the threshold of 35nt and 90% respectively (Figure V-4B-4D).

Averaged frequencies of hits were square root converted and statistically analyzed with ANOVA. The ANOVA in the *A. flavus* experiment had a p-value lower than 0.05 (Figure V-5) which rejects the null hypotheses that all treatments were similar, therefore, a post-hoc analysis was automatically performed using the Tukey HSD function in R. The post-hoc analysis and T-test for *A. flavus* showed that e-probes hitting on RNA sequencing databases obtained from *A. flavus* AF70 growing on ground corn were different from those of AF70 growing on PDB, and AF36 on corn and PDB (Figure V-6 & Table V-2). In conclusion, EDNAtran was able to find statistically significant differences between the transcriptomic data set of the highly toxigenic sample, from the non-toxigenic samples, using 231 e-probes generated in this study and was able to transcriptomically assess the production of aflatoxin by solely using EDNAtran.

Future studies need to include multiple blind samples to assess the usefulness of the new EDNAtran protocol to identify aflatoxin producing *A. flavus* strains. In this study we have showed

that in a known positive transcriptomic database, EDNAtran is capable of discriminating between production and no-production of aflatoxin. However, blind samples will provide a more realistic assessment of the newly developed tool.

Literature cited

Anders, S. and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biol.* **11**, 1–12.

Brown, M.P., Brown-Jenco, C.S. and Payne, G.A. (1999) Genetic and molecular analysis of aflatoxin biosynthesis. *Fungal Genet. Biol.* **26**, 81–98.

Callicott, K. a. and Cotty, P.J. (2015) Method for monitoring deletions in the aflatoxin biosynthesis gene cluster of *Aspergillus flavus* with multiplex PCR. *Lett. Appl. Microbiol.* **60**, 60–65.

Chang, P.K. (2004) Lack of interaction between AFLR and AFLJ contributes to nonaflatoxigenicity of *Aspergillus sojae*. *J. Biotechnol.* **107**, 245–253.

Chang, P.K., Cary, J.W., Bhatnagar, D., Cleveland, T.E., Bennett, J.W., Linz, J.E., Woloshuk, C.P. and Payne, G. a. (1993) Cloning of the *Aspergillus parasiticus* apa-2 gene associated with the regulation of aflatoxin biosynthesis. *Appl. Environ. Microbiol.* **59**, 3273–3279.

Chang, P.K., Horn, B.W. and Dorner, J.W. (2005) Sequence breakpoints in the aflatoxin biosynthesis gene cluster and flanking regions in nonaflatoxigenic *Aspergillus flavus* isolates. *Fungal Genet. Biol.* **42**, 914–923.

Chen, Z.-Y., Brown, R.L., Menkir, A. and Cleveland, T.E. (2011) Identification of resistance-associated proteins in closely-related maize lines varying in aflatoxin accumulation. *Mol.*

Breed. **30**, 53–68.

Cotty, P.J. (1997) Aflatoxin-producing potential of communities of Aspergillus section Flavi from cotton producing areas in the United States. *Mycol. Res.* **101**, 698–704.

Cotty, P.J. (1994) Influence of field application of an atoxigenic strains of *Aspergillus flavus* on the populations of *A. flavus* infection cotton balls and on the aflatoxin content of cottonseed. *Phytopathology* **84**, 1270–1277.

Davis, N.D., Diener, U.L. and Eldridge, D.W. (1966) Production of aflatoxins B1 and G1 by *Aspergillus flavus* in a semisynthetic medium. *Appl. Microbiol.* **14**, 378–380.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21.

Donner, M., Atehnkeng, J., Sikora, R. a, Bandyopadhyay, R. and Cotty, P.J. (2010) Molecular characterization of atoxigenic strains for biological control of aflatoxins in Nigeria. *Food Addit. Contam. Part A. Chem. Anal. Control. Expo. Risk Assess.* **27**, 576–590.

Egmond, H.P. Van, Schothorst, R.C. and Jonker, M. a. (2007) Regulations relating to mycotoxins in food : Perspectives in a global and European context. *Anal. Bioanal. Chem.* **389**, 147–157.

Ehrlich, K.C., Chang, P., Yu, J. and Cotty, P.J. (2004) Aflatoxin Biosynthesis Cluster Gene cypA Is Required for G Aflatoxin Formation. *Appl. Environ. Microbiol.* **70**, 6518–6524.

Ehrlich, K.C., Yu, J. and Cotty, P.J. (2005) Aflatoxin biosynthesis gene clusters and flanking regions. *J. Appl. Microbiol.* **99**, 518–527.

Espindola, A.S., Garzon, C.D., Schneider, W.L., Hoyt, P.R., Marek, S.M. and Garzon, C.D. (2015) A new approach for detecting Fungal and Oomycete plant pathogens in Next Generation Sequencing metagenome data utilizing Electronic Probes. *Int. J. Data Min. Bioinfomatics* **12**, 115–128.

Fernandes, M., Keller, N.P. and Adams, T.H. (1998) Sequence-specific binding by *Aspergillus nidulans* AflR, a C6 zinc cluster protein regulating mycotoxin biosynthesis. *Mol. Microbiol.* **28**, 1355–1365.

Garber, R.K. and Cotty, P.J. (1997) Formation of Sclerotia and Aflatoxins in Developing Cotton Bolls Infected by the S Strain of *Aspergillus flavus* and Potential for Biocontrol with an Atoxigenic Strain. *Phytopathology* **87**, 940–945.

Geoff Spencer (2001) International Human Genome Sequencing Consortium Publishes Sequence and Analysis of the Human Genome. *WASHINGTON, D.C. - Hum. Genome Proj. Int. Consort. today Announc. Publ. a Draft Seq. Initial Anal. Hum. genome - Genet. Bluepr. a Hum. being. Pap. Appear. Feb. 15 issue jour*.

Georgianna, D.R. and Payne, G.A. (2009) Genetic regulation of aflatoxin biosynthesis: from gene to genome. *Fungal Genet. Biol.* **46**, 113–25.

Hill, R.A., Blankenship, P.D., Cole, R.J. and Sanders, T.H. (1983) Effects of soil moisture and temperature on preharvest invasion of peanuts by the *Aspergillus flavus* group and subsequent aflatoxin development. *Appl. Environ. Microbiol.* **45**, 628–33.

Huson, D.H., Mitra, S., Ruscheweyh, H.-J., Weber, N. and Schuster, S.C. (2011) Integrative analysis of environmental sequences using MEGAN4. *Genome Res.* **21**, 1552–1560.

Klich, M. a. and Pitt, J.I. (1988) Differentiation of *Aspergillus flavus* from *A. parasiticus* and

other closely related species. *Trans. Br. Mycol. Soc.* **91**, 99–108.

Kurtzman, C.P., Horn, B.W. and Hesseltine, C.W. (1987) *Aspergillus nomius*, a new aflatoxin-producing species related to *Aspergillus flavus* and *Aspergillus tamarii*. *Antonie Van Leeuwenhoek* **53**, 147–158.

Llewellyn, G.C., Mooney, R.L., Cheatle, T.F. and Flannigan, B. (1992) Mycotoxin contamination of spices : An update. *Int. Biodeterior. Biodegradation* **29**, 111–121.

Marsh, S.F. and Payne, G. a. (1984) Preharvest Infection of Corn Silks and Kernels by *Aspergillus flavus*. *Phytopathology* **74**, 1284–1289.

Mateles, R.I. and Adye, J.C. (1965) Production of Aflatoxins in Submerged Culture. *Appl. Microbiol.* **13**, 208–211.

Mauro, A., Battilani, P. and Cotty, P.J. (2015) Atoxigenic *Aspergillus flavus* endemic to Italy for biocontrol of aflatoxins in maize. *BioControl* **60**, 125–134.

Mehl, H.L., Jaime, R., Callicott, K. a., Probst, C., Garber, N.P., Ortega-Beltran, A., Grubisha, L.C. and Cotty, P.J. (2012) *Aspergillus flavus* diversity on crops and in the environment can be exploited to reduce aflatoxin exposure and improve health. *Ann. N. Y. Acad. Sci.* **1273**, 7–17.

Mellon, J.E., Dowd, M.K. and Cotty, P.J. (2005) Substrate utilization by *Aspergillus flavus* in inoculated whole corn kernels and isolated tissues. *J. Agric. Food Chem.* **53**, 2351–2357.

Minto, R.E. and Townsend, C.A. (1997) Enzymology and Molecular Biology of Aflatoxin Biosynthesis. *Chem. Rev.* **97**, 2537–2556.

Nierman, W.C., Yu, J., Fedorova-Abrams, N.D., Losada, L., Cleveland, T.E., Bhatnagar, D.,

Bennett, J.W., Dean, R. and Payne, G.A. (2015) Genome Sequence of *Aspergillus flavus* NRRL 3357, a Strain That Causes Aflatoxin Contamination of Food and Feed. *Genome Announc.* **3**, e00168-15.

Niessen, L. (2008) PCR Based Diagnosis and Quantification of Mycotoxin Producing Fungi. *Adv. Food Nutr. Res.* **54**, 81–138.

Probst, C., Bandyopadhyay, R., Price, L.E. and Cotty, P.J. (2011) Identification of Atoxigenic *Aspergillus flavus* Isolates to Reduce Aflatoxin Contamination of maize in Kenya. *Plant Dis.* **95**, 212–218.

Probst, C. and Cotty, P.J. (2012) Relationships between in vivo and in vitro aflatoxin production: Reliable prediction of fungal ability to contaminate maize with aflatoxins. *Fungal Biol.* **116**, 503–510.

Robens, J. and Cardwell, K. (2003) The Costs of Mycotoxin Management to the USA: Management of Aflatoxins in the United States. *Toxin Rev.* **22**, 139–152.

Squire, R.A. (1981) Ranking animal carcinogens: a proposed regulatory approach. *Sci.* **214**, 877–880.

Stobbe, A.H., Daniels, J., Espindola, A.S., Verma, R., Melcher, U., Ochoa-Corona, F., Garzon, C., Fletcher, J. and Schneider, W. (2013) E-probe Diagnostic Nucleic acid Analysis (EDNA): A theoretical approach for handling of next generation sequencing data for diagnostics. *Microbiol. Methods* **94**, 356–366.

Trail, F., Mahanti, N., Rarick, M., Mehigh, R., Liang, S.H., Zhou, R. and Linz, J.E. (1995) Physical and transcriptional map of an aflatoxin gene cluster in *Aspergillus parasiticus* and functional disruption of a gene involved early in the aflatoxin pathway. *Appl. Environ.*

Microbiol. **61**, 2665–2673.

Truong, D.T., Franzosa, E. a, Tickle, T.L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C. and Segata, N. (2015) MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* **12**, 902–903.

Woloshuk, C.P., Cavaletto, J.R. and Cleveland, T.E. (1997) Inducers of aflatoxin biosynthesis from colonized maize kernels are generated by an amylase activity from *Aspergillus flavus*. *Phytopathology* **87**, 164–9.

Woloshuk, C.P., Foutz, K.R., Brewer, J.F., Bhatnagar, D., Cleveland, T.E. and Payne, G.A. (1994) Molecular characterization of aflR, a regulatory locus for aflatoxin biosynthesis. *Appl Env. Microbiol* **60**, 2408–2414.

Wu, F., Liu, Y. and Bhatnagar, D. (2008) Cost-Effectiveness of aflatoxin control methods: economic incentives. **9543**, 203–225.

Yu, J.H., Butchko, R. a E., Fernandes, M., Keller, N.P., Leonard, T.J. and Adams, T.H. (1996) Conservation of structure and function of the aflatoxin regulatory gene aflR from *Aspergillus nidulans* and *A. flavus*. *Curr. Genet.* **29**, 549–555.

Yu, J.J., Chang, P.K., Cary, J.W., Wright, M., Bhatnagar, D., Cleveland, T.E., Payne, G.A. and Linz, J.E. (1995) Comparative Mapping of Aflatoxin Pathway Gene Clusters in *Aspergillus parasiticus* and *Aspergillus flavus*. *Appl. Environ. Microbiol.* **61**, 2365–2371.

Figures

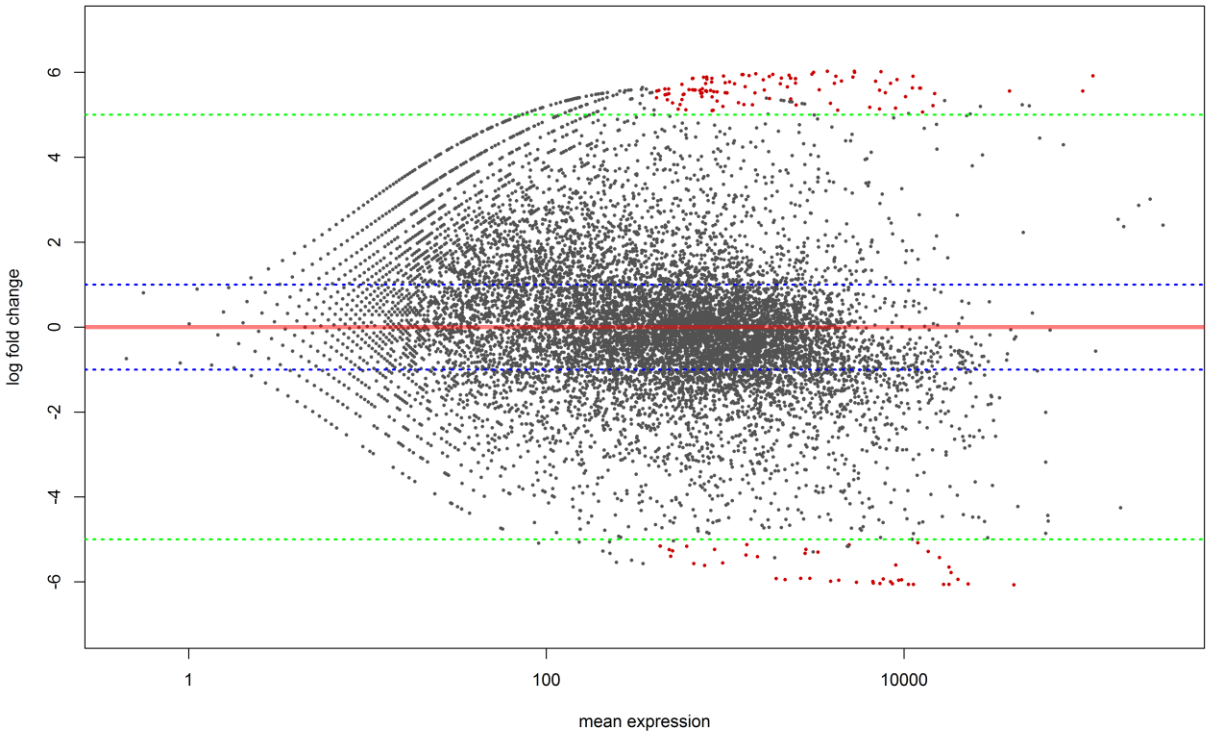

Figure V-1. Gene expression analysis Mean Average (MA) Plot for *Aspergillus flavus* AF70. Red line shows zero change in gene expression. Blue dashed lines show 1-fold change in gene expression and green dashed lines show a 5-fold change in gene expression. Red dots are genes that have been either up-regulated or down-regulated in *A. flavus* AF70 growing on ground corn. Gray dots depict genes that have not had enough statistical evidence to be assigned a gene expression fold change.

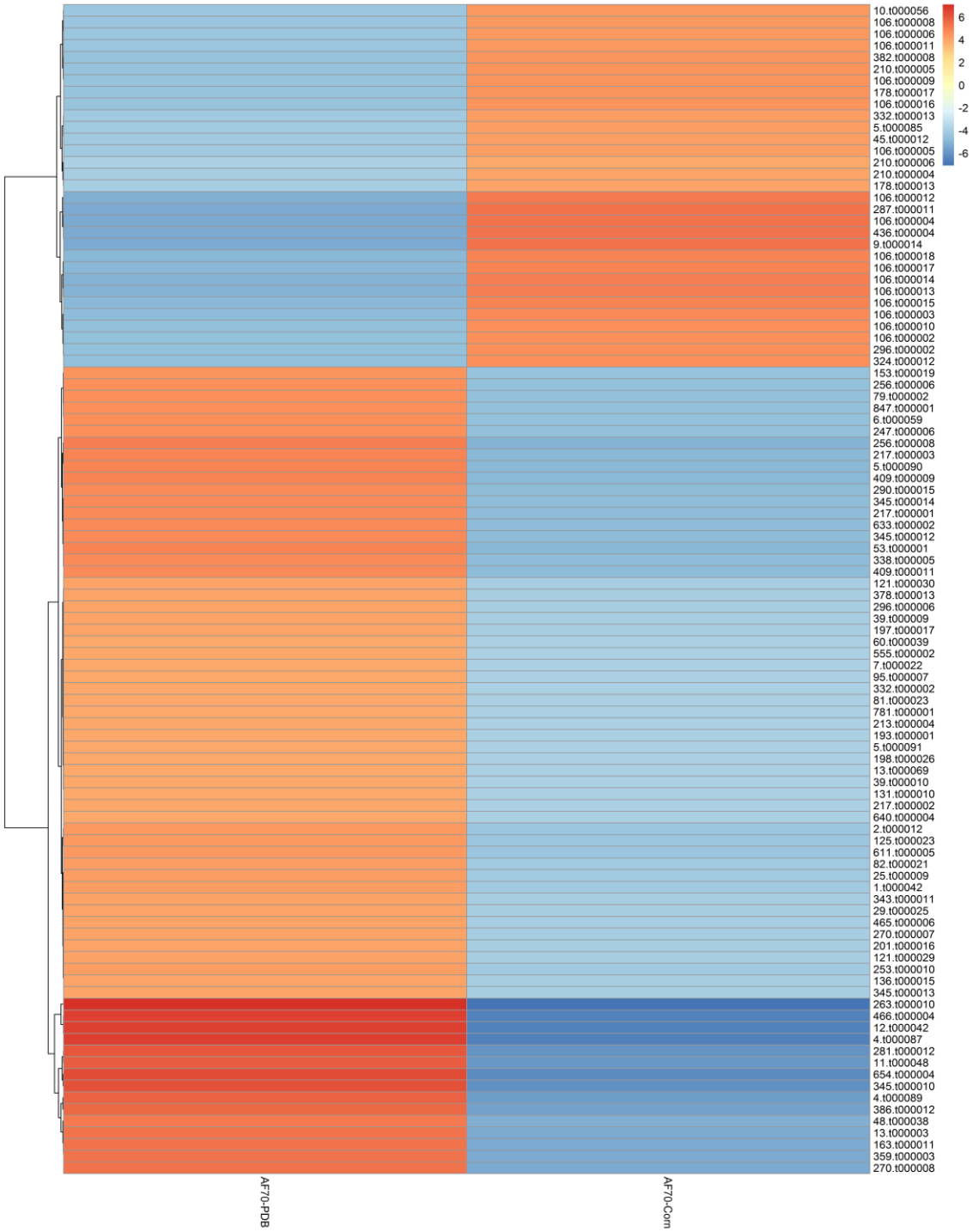

Figure V-2. Hierarchical clustering map depicting gene expression of *Aspergillus flavus* AF70 growing on potato dextrose agar (PDA) and ground corn. Gene expression fold change is differentiated by a color palette ranging from red (most up-regulated genes have plus 6-fold changes) to blue (most downregulated genes have minus 6-fold change). Genes are clustered based on their gene expression fold change to facilitate gene co-expression analysis.

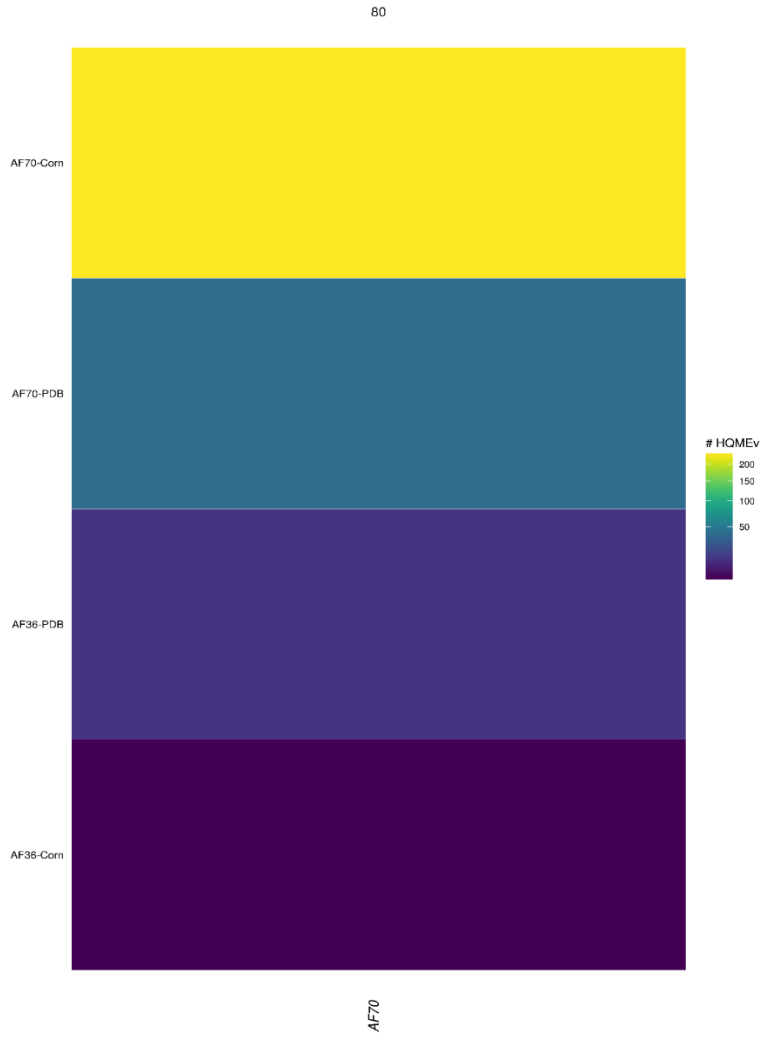

Figure V-3. Hierarchal-clustered heat map depicting the number of High Quality Matches (HQMs) of e-probes designed on the aflatoxin gene cluster hitting on RNA sequencing runs containing *Aspergillus flavus* AF70 (toxigenic) and AF36 (atoxigenic) growing on Potato Dextrose Agar (PDB) and ground corn. Higher number of HQMs are colored yellow and lower number of HQMs are colored blue.

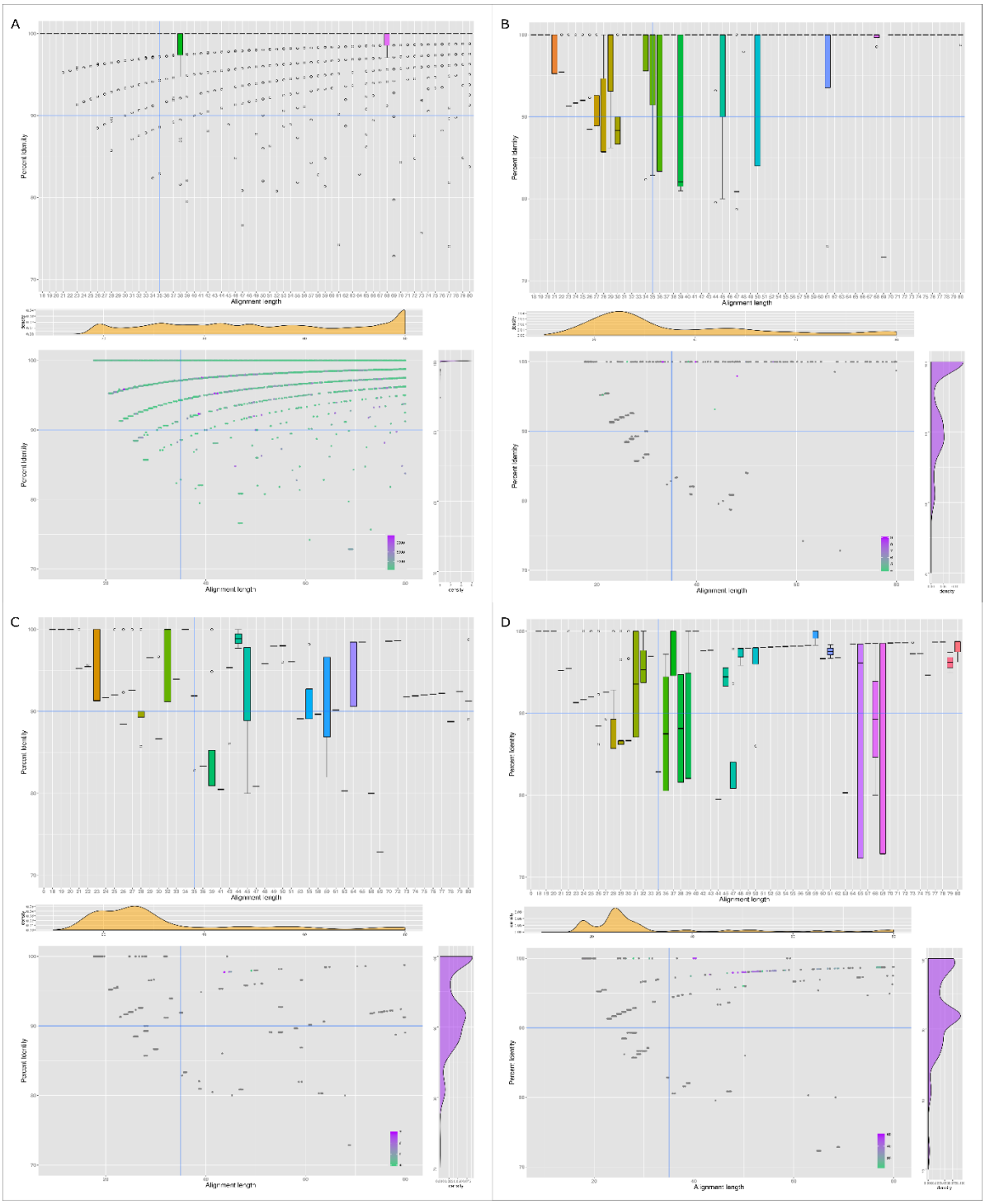

Figure V-4. EDNA transcriptomics hits distribution and frequencies based on alignment length and percent identity of *Aspergillus flavus* AF70 e-probes for aflatoxin-related gene detection. (A,C) RNA sequencing of *A. flavus* AF70 and AF36 respectively growing on corn identified with 80-mer AF70 aflatoxin-specific e-probes. (B,D). RNA sequencing of *A. flavus* AF70 and AF36 respectively growing on PDB.

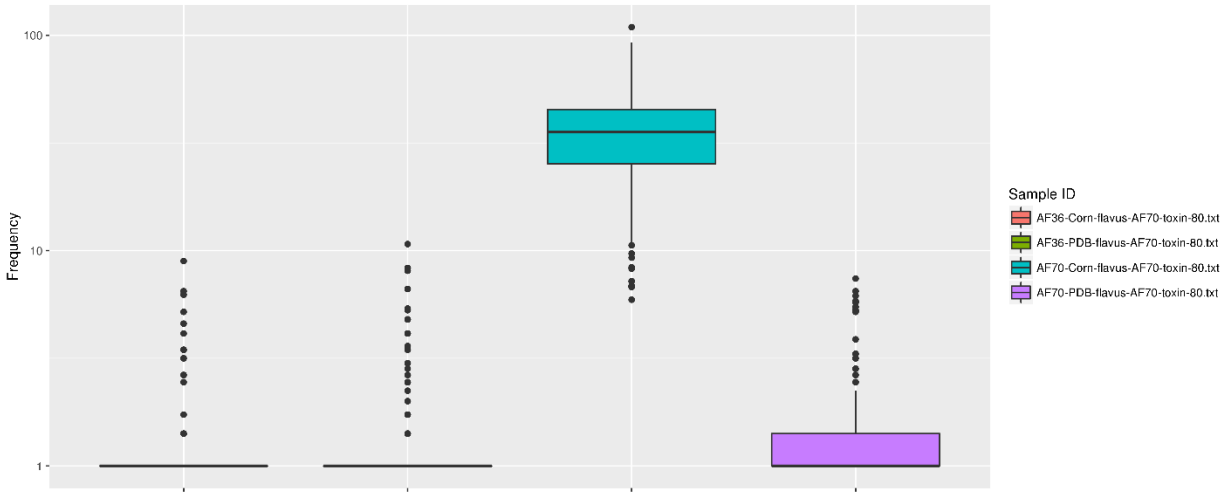

Figure V-5. Hit frequencies of *Aspergillus flavus* AF70 e-probes in RNA sequencing runs of toxigenic (AF70) and atoxigenic (AF36) *A. flavus* strains

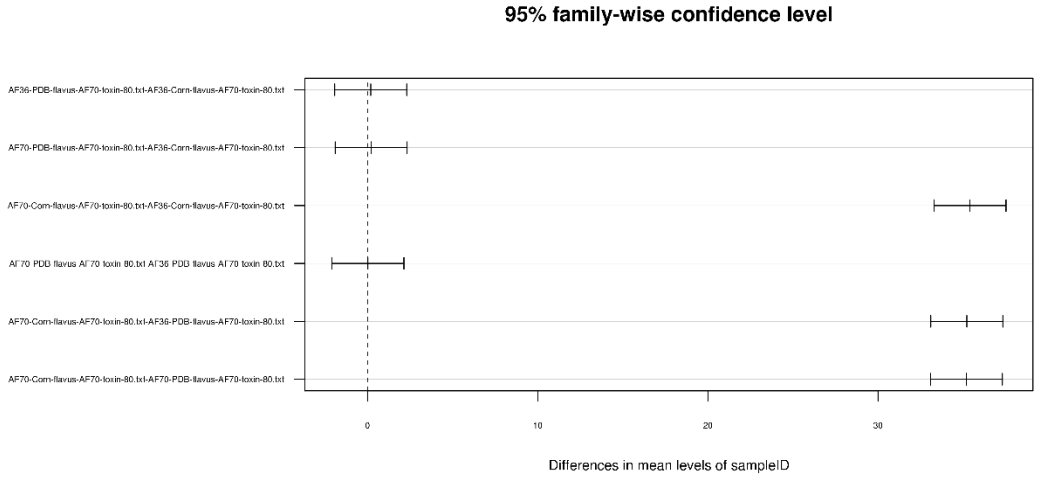

Figure V-6. Post-hoc analysis of variance (ANOVA) using Tukey honest significant difference (HSD) with 95% of confidence for the EDNA metatranscriptomic detection of aflatoxin-related genes. Lines close to zero are interactions that had no difference in their hit frequency means while lines closer to 30 are interactions that had different hit frequency means.

Tables

Table V-1. Output table produced by the EDNA transcriptomics pipeline for the *Aspergillus flavus* aflatoxin detection analysis. The table contains 11 columns. A). the metagenome codes B). the metagenome read number C). metagenome average read length, D). the metagenome maximum read length, E). the metagenome minimum read length, F). the e-probe identification code, G). the e-probe length, H). the number of e-probes I) the high quality match (HQM) numbers calculated with e-values, J). the HQMs calculated without e-value, K). high scoring general matches (HSGM) that uses only percent identity higher than 90% as a stringency parameter.

A	B	C	D	E	flavus-AF70-toxin	G	H	I	J	K
AF70-PDB	20657024	98.902862	100	35	flavus-AF70-toxin	80	231	39	40	39
AF70-PDB	20657024	98.902862	100	35	flavus-AF70-toxin	80	231	39	40	39
AF36-Corn	22495368	99.03803299	100	35	flavus-AF70-toxin	80	231	2	2	2
AF36-PDB	24134226	98.72027899	100	35	flavus-AF70-toxin	80	231	12	12	12
AF70-Corn	24902500	98.90925321	100	35	flavus-AF70-toxin	80	231	231	231	231

Table V-2. Pairwise T-test analysis showing multiple comparisons of e-probe hit frequencies for *Aspergillus flavus* toxin detection analysis. P-values lower than 0.05 are highlighted with red.

	AF36-Corn-AF70-80.txt	AF36-PDB-AF70-80.txt	AF70-Corn-AF70-80.txt
AF36-PDB-AF70-80.txt	1	NA	NA
AF70-Corn-AF70-80.txt	3.72E-222	8.38E-221	NA
AF70-PDB-AF70-80.txt	1	1	1.04E-220

CHAPTER VI

VI USING HIGH-RESOLUTION GENOMIC SIGNATURES FOR THE DISCRIMINATION OF OOMYCETE STRAINS IN PHYTOBIOMES

Abstract

Timely and accurate plant pathogen detection is crucial to implement effective crop management strategies. Molecular and serological plant pathogen detection tools are available for most plant pathogens; however, detecting or discriminating strains is difficult due to low genetic diversity within a species and/or phenotypic similarities. Low genetic diversity impacts the development of new detection techniques since most of them use protein coding regions to design molecular-based detection tools like Polymerase Chain Reaction (PCR). The bioinformatic tool E-probe Diagnostics Nucleic acid Analysis (EDNA) —originally designed to detect plant pathogens at the species level — was adapted to detect and discriminate among eukaryotic plant pathogenic strains. High-resolution genomic signatures (e-probes) capable of discriminating *Pythium aphanidermatum* strains were generated. *P. aphanidermatum*, an oomycete soilborne plant pathogen, was selected as model system because of its well characterized limited genetic variability.

EDNA was able to discriminate among two very similar strains of *P. aphanidermatum* (P16 and BR444) and simultaneously discriminate them from the sister species, *Pythium deliense*, and *Globisporangium irregulare* (formerly *P. irregulare*). The present study gives new insights into the potential uses and scopes of EDNA in the plant pathology diagnostics field due to its scalability and detection accuracy.

Introduction

In Plant Pathology, the necessity to discriminate among plant pathogen strains or isolates is becoming progressively important due to the emergence of new strains with significantly increased pathogenicity levels (Singh et al., 2011). Detecting pathogenic strains/isolates in a timely manner allows the application of proper management and control strategies on time. Typical plant pathogen characterization and detection is performed initially by visually determining signs, symptoms or host specificity (Martin & Loper, 1999; Sutton et al., 2006). However, once pathogens have been phenotypically characterized, molecular markers are often preferred over traditional detection methods due to their rapidness and potential parallelization (Ishiguro et al., 2013). The preferred molecular markers for plant pathogen detection include the serological methods and nucleic acid based methods (Wang et al., 2002; Ishiguro et al., 2013). Among the nucleic-acid based detection methods, amplicon Sanger sequencing of genetic barcodes is one of the most sensitive and specific when targeting loci that have been described as genetically different among strains or isolates (Robideau et al., 2011). However, mutations might decrease the sensitivity of single target molecular methods, potentially leading to false negative results (van der Sande et al., 1992). Therefore, new or revised PCR tools are often needed for detecting rapidly changing pathogenic strains, which traditionally includes de novo primer design and the standardization of wet lab protocols for primer amplification.

New sequencing technologies have permitted higher sequencing coverage and yield, consequently increasing the likelihood of detecting almost any organism when sequencing metagenomes (Daniel, 2005; Stobbe et al., 2013; Bragg & Tyson, 2014; Espindola et al., 2015). EDNA is a bioinformatics tool that takes advantage of next generation sequencing (NGS) technologies to detect plant pathogens in metagenomics samples (Stobbe et al., 2013; Espindola et al., 2015). EDNA detects plant pathogens at the species level in metagenomes and metatranscriptomes (Stobbe et al., 2013; Espindola et al., 2015; Chaper 4; Chapter 5). Yet, strain or isolate identification in eukaryotic plant pathogens is a field that has not been extensively studied due to the complexity of eukaryotic genomes. In fact, strain discrimination research with NGS has been limited only to fungal organisms genetically well-characterized at the strain level (i.e. *Aspergillus flavus*) (Callicott & Cotty, 2015; Chapter 5). EDNA's flexibility to adapt to different eukaryotic detection scenarios suggests that its adaptation to detect strains or isolates in eukaryotic plant pathogens is theoretically feasible.

The aim of this study was to adapt EDNA as a bioinformatic tool to discriminate among plant-pathogenic eukaryotic strains or isolates. *P. aphanidermatum* was selected as a model system due to the low genetic diversity observed in genetic barcodes at the isolate level in population genetic studies (Lee, Garzon & Moorman, 2010). Genetic barcodes like the Internal Transcribed Spacer (ITS) are highly conserved at the species level and have been widely utilized for detecting *P. aphanidermatum*, however, isolate/strain discrimination was achieved only with multilocus approaches, such as Amplified fragment length polymorphism (AFLP) or Simple Sequence Repeat (SSRs) markers (Garzon et al., 2005; Al-Sa'di et al., 2008; Lee, Garzon & Moorman, 2010). We developed genome-wide e-probes generated from an unreleased draft genome of *P.*

aphanidermatum as well as an isolate genome available in public databases (BR444) to create a robust *P. aphanidermatum* strain/isolate discrimination pipeline.

Experimental Procedures

Genome sequencing of P. aphanidermatum strains

Pythium and *Globisporangium* isolates used in this study were provided by Dr. Gary Moorman (Pennsylvania State University) in water agar (WA). Cultures were grown on potato dextrose agar (PDA), then transferred to potato dextrose broth (PDB) and incubated at 22 °C for 5 days. DNA extractions were performed using Weising et al., 2005 DNA extraction protocol to produce high quality DNA samples for sequencing.

The genomes of *P. aphanidermatum* strain (P16), *G. irregulare* (P18) as well as *P. deliense* (P154) were sequenced using the Illumina Hiseq 2500® platform at the Core Faxility of The University of Illinois at Urbana-Champaign. Partial genome assemblies were performed using Velvet with different kmer sizes ranging from 21 to 31. The best kmer assembly was selected based on the N50 and the longest assembly (Zerbino & Birney, 2008; Luo et al., 2012).

E-probe design

E-probes were designed by using one strain as target and the combined databases of the other strains as a near neighbor, by creating a multifasta file with all the near neighbors' sequences concatenated. The coordinates of their location in the genome file were included in the header of each e-probe. Strain discrimination could also be used on annotated genomes, however, in this case we did not use annotated genomes and consequently, the intron/exon information is shown only as intron. However, if well-annotated genomes were utilized (Chapter 5), the gff3 file can also be provided in the control file and take advantage of EDNA transcriptomics. The EDNA

eukaryotic strain specific e-probe design required an extra curation step which aligned the generated e-probes with their near neighbors and detected e-probes producing hits in the near neighbors' isolates/strains which were subsequently eliminated.

EDNA for P. aphanidermaturm strains and Pythium spp. discrimination

EDNA eukaryotic was used for the discrimination of *P. aphanidermatum* isolates and *Pythium* spp. with standard stringency metrics (minimum alignment length of 35 nt and minimum percent identity of 90%). The Illumina sequencing runs utilized to assemble the genomes were utilized as positive controls for each of the strains. Therefore, it is expected that most of the e-probes generated, hit in the sequencing database. No modifications were performed in the original EDNA eukaryotic script.

Results and Discussion

Although EDNA has been successfully utilized in a variety of plant pathogen detection scenarios. EDNA has not been used for detecting and discriminating isolates at the strain level. E-probe databases at various hierarchical taxonomical levels can be utilized simultaneously when detecting plant pathogens. In the present study oomycete pathogens that cause Pythium diseases (in the genera *Pythium* and *Globisporangium*) were selected as model systems to challenge the EDNA protocol to discriminate strains at three different taxonomic hierarchical levels, genus, species and strain, simultaneously. Additionally, *Pythium* spp. (and *Globisporangium* species) are often found in disease complexes (Martin, 2000), infecting mainly seedling roots in the soil (DeVay, Garber & Matheron, 1982) and the development of this pipeline creates a new application of EDNA for future phytobiome studies. The ability to discriminate among strains/isolates of soilborne pathogens, while also discriminating among other related and unrelated species might

help to understand better the complex ecological interactions among organisms in the rhizosphere microbiome and help in the development of better management strategies.

No genetic variation has been found in genetic barcodes and limited genetic variation was found in *P. aphanidermatum* using neutral genetic markers like AFLPs and RAPD (Herrero & Klemsdal, 1998; Garzon et al., 2005), in spite of using whole genome multilocus targets. In fact, when comparing AFLP profiles of *P. aphanidermatum* against *G. irregulare* and *G. ultimum*, a lower degree of genetic diversity was found in *P. aphanidermatum* (Lee & Moorman, 2008).

Genome sequencing

The genomes of *P. aphanidermatum* isolate (P16), *P. deliense* (P154) and *G. irregulare* (P18) were sequenced and partially assembled, to better identify regions of genetic diversity where e-probes could be generated. Illumina HiSeq2500 yielded 21.96 million reads for P16, 23.81 million reads for P18 and 20.08 million reads for P154. Their genomes were assembled using both Velvet and SOAPdenovo using kmers ranging from 19 to 39 (Zerbino & Birney, 2008; Luo et al., 2012). Selection of the best assembly was assessed primarily based on the largest N50, lower number of misassemblies, minimum number of contigs and total length (Table VI-2). The kmer having the best assembly parameters was 31 and the genomes assembled with that kmer were kept for downstream analyses.

E-probe design

E-probes designed for each of the strains were carefully selected based on specificity with the intended target. E-probe design for strain typing carried two extra curation steps. A basic local alignment of the e-probes originally designed using Mummer against their specific genomes to retrieve crucial information for the analysis like genomic coordinates. Extra information was

added to the header of each e-probe verifying its specificity. The e-probe headers presented as fasta files contain important e-probe parameters to be used mainly when further analysis of the hit is necessary. Among the parameters included in the headers of the e-probes were the e-probe ID number, the reference contig or sequence number from which the e-probe was originally designed in the strain genome, and the coordinates of the e-probe in that specific contig. The information about the presence of either introns or exons in the genomic area where the e-probe was generated were not determined for any of the organisms in this study since they have not been annotated yet.

The second curation step uses BLASTn to eliminate e-probes that potentially could create false positives. The procedure evaluates each e-probe for non-specific hits in the near neighbor strain/isolate or species genomes. E-probes aligning with the near neighbors having more than 98% identity and 100% query coverage were eliminated from the e-probe database. Consequently, a total of 78 and 19 e-probes were generated for P16 with 60-mer and 80-mer length, respectively. Similarly, the strain BR444 totaled 71 e-probes 60-mer long and 27 e-probes 80-mer long. P18 (*G. irregulare*) had 35,901 and 13,791 e-probes with 60-mer and 80-mer lengths respectively. Finally, P154 (*P. deliense*) generated 15,001 e-probes for the 60-mer length and 4,857 for the 80-mer length.

The pipeline utilized for the development of e-probes found few sequence differences when comparing *P. aphanidermatum* genomes since a relatively low number of e-probes were generated. As expected, when genome comparisons were done at the species and genus level, much higher numbers of e-probes were generated, and those numbers reflected genetic distances, with more e-probes generated for *G. irregulare* than for *P. deliense* (Table VI-1).

EDNA for P. aphanidermatum isolate/strain discrimination

EDNA was performed using species level parameters, since a mixture of taxonomic hierarchies were analyzed in this study. The EDNA parameters for species level include a percent identity higher than 90% and minimum alignment length of 35 nt. E-values can be defined by the user, however, for this study, the $1x10^{-9}$ e-value parameter was maintained. Expectedly, EDNA was able to detect and discriminate among P16, BR444, P154 and P18 with high confidence (Figure VI-1 & Figure VI-2). When using P16 specific e-probes we can clearly observe that High Quality Matches (HQMs) are only calculated only for the P16-genome. A total of 78 HQMs for the 60 nt e-probes and 19 HQMs for the 80 nt e-probes. Similarly, when using BR444 e-probes, HQMs are observed only on BRR444-real with 71 and 27 HQMS for 60 and 80 nt length e-probes respectively (Table VI-1, Figure VI-1 and Figure VI-2). Therefore, we can conclude that EDNA is able discriminate among eukaryotic strains that have low genetic diversity, in this case *P. aphanidermatum.* On the other hand, within the same experiment and the same EDNA run, we were able to also discriminate *P. aphanidermatum* strains from their close relative species. Specifically *P. deliense* e-probes (P154) had 14,997 and 4,857 HQMs for P154-real metagenome (Table VI-1 and Figure VI-2). Similarly, *G. irregulare* e-probes (P18) had only HQMs on their intended target (P18-real) with 35,896 and 13,791 for 60 nt and 80 nt e-probes respectively (Table VI-1 and Figure VI-2).

In conclusion, these results confirm that EDNA is, in fact, a tool that could serve to discriminate among eukaryotic plant pathogen strains. Additionally, these results suggest that EDNA's scope could be extended beyond the plant pathology limit. Microbial forensics is a field that requires extremely sensitive techniques to discriminate among very genetically similar microorganisms (Cummings & Relman, 2002). Therefore, EDNA could also be applied to track

strains that have been intentionally released. Ideally, e-probe databases for all the microorganisms that represent a threat to agriculture and human beings could be generated. Such a database should be permanently curated by adding e-probe databases for new strains or species.

Literature cited

Al-Sa'di, A.M., Drenth, A., Deadman, M.L. and Aitken, E.A.B. (2008) Genetic diversity, aggressiveness and metalaxyl sensitivity of Pythium aphanidermatum populations infecting cucumber in Oman. *Plant Pathol.* **57**, 45–56.

Bragg, L. and Tyson, G.W. (2014) Metagenomics using next-generation sequencing. *Methods Mol. Biol.* **1096**, 183–201.

Callicott, K. a. and Cotty, P.J. (2015) Method for monitoring deletions in the aflatoxin biosynthesis gene cluster of Aspergillus flavus with multiplex PCR. *Lett. Appl. Microbiol.* **60**, 60–65.

Cummings, C.A. and Relman, D.A. (2002) Microbial Forensics--"Cross-Examining Pathogens." *Science (80-.).* **296**, 1976–1979.

Daniel, R. (2005) The metagenomics of soil. *Nat Rev Micro* **3**, 470–478.

DeVay, J.E., Garber, R.H. and Matheron, D. (1982) Role of Pythium species in the seedling disease complex of cotton in California. *Plant Dis.* **66**, 151–154.

Espindola, A.S., Garzon, C.D., Schneider, W.L., Hoyt, P.R., Marek, S.M. and Garzon, C.D. (2015) A new approach for detecting Fungal and Oomycete plant pathogens in Next Generation Sequencing metagenome data utilizing Electronic Probes. *Int. J. Data Min. Bioinfomatics* **12**, 115–128.

Garzon, C.D., Geiser, D.M., Moorman, G.W., Geiser, D.M. and Moorman, G.W. (2005) Diagnosis and Population Analysis of Pythium Species Using AFLP Fingerprinting. *Plant Dis.* **89**, 81–89.

Herrero, M.L. and Klemsdal, S.S. (1998) Identification of Pythium aphanidermatum using the RAPD technique. *Mycol. Res.* **102**, 136–140.

Ishiguro, Y., Asano, T., Otsubo, K., Suga, H. and Kageyama, K. (2013) Simultaneous detection by multiplex PCR of the high-temperature-growing Pythium species: P. aphanidermatum, P. helicoides and P. myriotylum. *J. Gen. Plant Pathol.* **79**, 350–358.

Lee, S., Garzon, C.D. and Moorman, G.W. (2010) Genetic structure and distribution of Pythium aphanidermatum populations in Pennsylvania greenhouses based on analysis of AFLP and SSR markers. *Mycologia* **102**, 774–784.

Lee, S. and Moorman, G.W. (2008) Identification and characterization of simple sequence repeat markers for Pythium aphanidermatum, P-cryptoirregulare, and P-irregulare and the potential use in Pythium population genetics. *Curr. Genet.* **53**, 81–93.

Luo, R., Liu, B., Xie, Y., et al. (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 18.

Martin, F.F.N. (2000) Phylogenetic Relationships among Some Pythium Species Inferred from Sequence Analysis of Phylogenetic relationships among some Pythium species inferred from sequence analysis of the mitochondrially encoded cytochrome oxidase II gene. *Mycologia* **92**, 711–727.

Martin, F.N. and Loper, J.E. (1999) Soilborne plant diseases caused by Pythium spp. ecology,

epidemiology, and prospects for biological control. *CRC. Crit. Rev. Plant Sci.* **18**, 111–181.

Robideau, G.P., Cock, A.W.A.M. De, Coffey, M.D., et al. (2011) DNA barcoding of oomycetes with cytochrome c oxidase subunit I and internal transcribed spacer. *Mol. Ecol. Resour.* **11**, 1002–11.

Sande, C.A.F.M. van der, Kwa, M., Nues, R.W. van, Heerikhuizen, H. van, Raué, H.A. and Planta, R.J. (1992) Functional analysis of internal transcribed spacer 2 of Saccharomyces cerevisiae ribosomal DNA. *J. Mol. Biol.* **223**, 899–910.

Singh, R.P., Hodson, D.P., Huerta-Espino, J., et al. (2011) The Emergence of Ug99 Races of the Stem Rust Fungus is a Threat to World Wheat Production. In Annual Review of Phytopathology, Vol 49. (VanAlfen, N.K., Bruening, G., and Leach, J.E., eds), pp. 465–481. Palo Alto: Annual Reviews.

Stobbe, A.H., Daniels, J., Espindola, A.S., Verma, R., Melcher, U., Ochoa-Corona, F., Garzon, C., Fletcher, J. and Schneider, W. (2013) E-probe Diagnostic Nucleic acid Analysis (EDNA): A theoretical approach for handling of next generation sequencing data for diagnostics. *Microbiol. Methods* **94**, 356–366.

Sutton, J.C., Sopher, C.R., Owen-Going, T.N., Liu, W., Grodzinski, B., Hall, J.C. and Benchimol, R.L. (2006) Etiology and epidemiology of Pythium root rot in hydroponic crops: current knowledge and perspectives. *Summa Phytopathol.* **32**, 307–321.

Wang, P., Boo, L., Lin, Y. and Yeh, Y. (2002) Specific detection of<i>Pythium aphanidermatum from hydroponic nutrient solution by booster PCR with DNA primers developed from mitochondrial DNA. *Phytoparasitica* **30**, 473–485.

Weising, K., Nybom, H., Pfenninger, M., Wolff, K. and Kahl, G. (2005) DNA Fingerprinting in Plants: Principles, Methods, and Applications, Second Edition 2nd ed., Boca Raton, FL: CRC Press.

Zerbino, D.R. and Birney, E. (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–9.

Figures

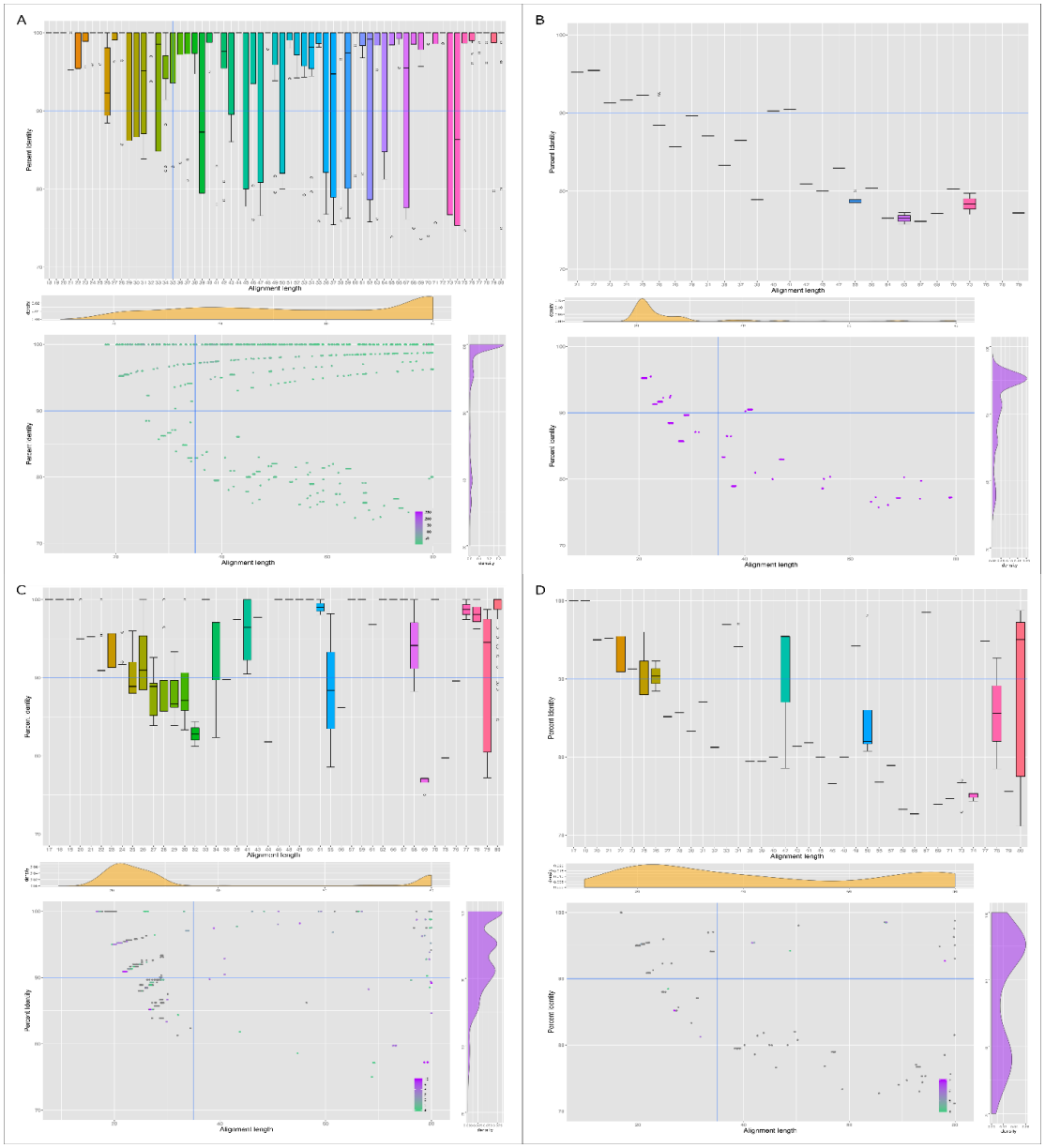

Figure VI-1. EDNA hit distribution and frequencies based on alignment length and percent identity of *Pythium aphanidermatum* strain specific e-probes. (A-B) Representations of hit percent identity distributions and its relationship to alignment length for 80-mer *P. aphanidermatum* (P16) specific e-probes hitting the metagenomic database of *P. aphanidermatum* (P16) and BR444 respectively. Similarly, (C-D) depicts 80-mer BR444 specific e-probes hitting the metagenome of *P. aphanidermatum* (BR444) and P16 metagenome respectively. (A-D) Color-coded dotplots depict e-probe alignment depths (green=lowest depth, purple=highest depth) related to percent identities and aligment lenghts in 80-mer e-probes. Additionally, marginal hit frequency plots are presented, it is expected that in positive samples, hits with alignment lengths higher than 35nt and percent identities higher than 90% have a higher frequency than other hits.

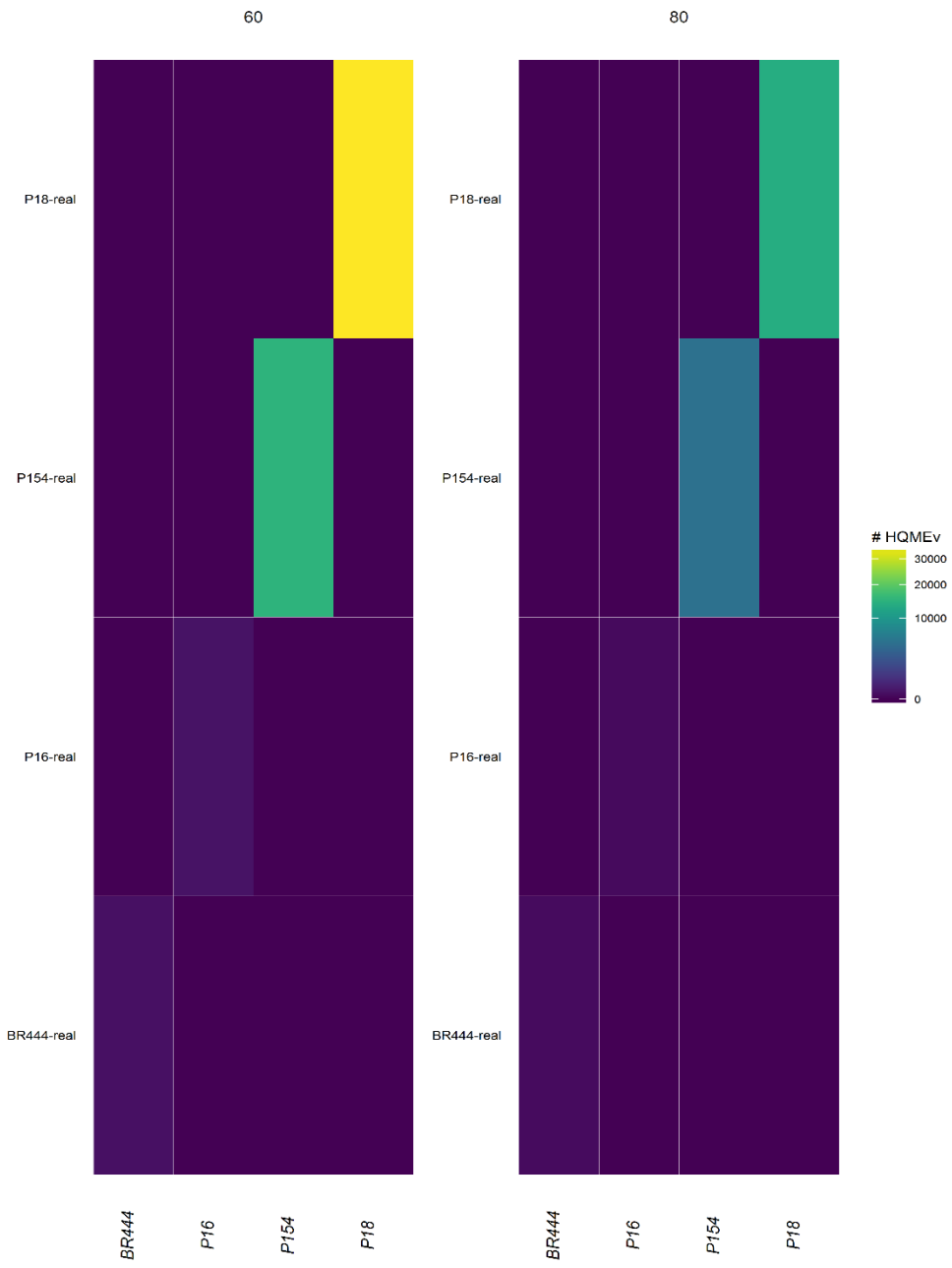

Figure VI-2. Hierarchal-clustered heat map depicting the number of High Quality Matches (HQMs) of e-probes designed for *Pythium aphanidermatum* (P16) strain detection.

Tables

Table VI-1. EDNA eukaryotic detection metrics for *Pythium aphanidermatum* strain discrimination in metagenomes. The table contains 11 columns. A). the metagenome codes B). the metagenome read number C). metagenome average read length, D). the metagenome maximum read length, E). the metagenome minimum read length, F). the e-probe identification code, G). the e-probe length, H). the number of e-probes I) the high quality match (HQM) numbers calculated with e-values, J). the HQMs calculated without e-value, K). high scoring general matches (HSGM) that uses only percent identity higher than 90% as a stringency parameter.

A	B	C	D	E	F	G	H	I	J	K
BR444-real	1299108	523.9967139	2049	51	BR444	60	71	55	55	63
BR444-real	1299108	523.9967139	2049	51	BR444	80	27	24	24	25
P16-real	21963597	99.51529597	100	35	BR444	60	71	0	3	4
P16-real	21963597	99.51529597	100	35	BR444	80	27	0	0	0
P154-real	20082961	99.5355537	100	35	BR444	60	71	0	0	0
P154-real	20082961	99.5355537	100	35	BR444	80	27	0	0	0
P18-real	23819955	99.41442379	100	35	BR444	60	71	0	0	0
P18-real	23819955	99.41442379	100	35	BR444	80	27	0	0	0
P154-real	20082961	99.5355537	100	35	P154	60	15001	14997	14997	15001
P154-real	20082961	99.5355537	100	35	P154	80	4857	4857	4857	4857
BR444-real	1299108	523.9967139	2049	51	P154	60	15001	0	6	63
BR444-real	1299108	523.9967139	2049	51	P154	80	4857	0	7	122
P16-real	21963597	99.51529597	100	35	P154	60	15001	0	32	58
P16-real	21963597	99.51529597	100	35	P154	80	4857	0	16	86
P18-real	23819955	99.41442379	100	35	P154	60	15001	0	1	0
P18-real	23819955	99.41442379	100	35	P154	80	4857	0	2	1
P16-real	21963597	99.51529597	100	35	P16	60	78	78	78	78
P16-real	21963597	99.51529597	100	35	P16	80	19	19	19	19
BR444-real	1299108	523.9967139	2049	51	P16	60	78	0	0	20
BR444-real	1299108	523.9967139	2049	51	P16	80	19	0	0	6
P154-real	20082961	99.5355537	100	35	P16	60	78	0	0	0
P154-real	20082961	99.5355537	100	35	P16	80	19	0	0	0
P18-real	23819955	99.41442379	100	35	P16	60	78	0	0	0
P18-real	23819955	99.41442379	100	35	P16	80	19	0	0	0
P18-real	23819955	99.41442379	100	35	P18	60	35901	35896	35897	35901
P18-real	23819955	99.41442379	100	35	P18	80	13791	13791	13791	13791
BR444-real	1299108	523.9967139	2049	51	P18	60	35901	0	1	2
BR444-real	1299108	523.9967139	2049	51	P18	80	13791	0	2	17
P154-real	20082961	99.5355537	100	35	P18	60	35901	0	3	2
P16-real	21963597	99.51529597	100	35	P18	60	35901	0	1	1
P154-real	20082961	99.5355537	100	35	P18	80	13791	0	2	9
P16-real	21963597	99.51529597	100	35	P18	80	13791	0	4	11

Table VI-2. Genome assembly metrics for Pythium spp. Information is distributed in 11 columns. A). shows the assembler used on each of the sequencing samples (velvet or soapdenovo), B). number of contigs assembled that are longer than 0 bp. C). number of contigs assembled that are longer than 1000 bp., D). total length of the assembly with contigs longer than zero bp., E). total length of the assembly with contigs longer than 1000 bp., F). total number of contigs for the assembly, G). largest contig length, H). total length of the assembly I). GC content ratio, J). N50 and K). N75

A	B	C	D	E	F	G	H	I	J	K
soap-P154-31	53413	8463	38070101	28774550	12863	62869	31910683	57.1	3942	1923
velvet-P16-31	39261	6933	37540482	31605254	9771	78984	33619099	53.9	6370	3023
soap-P18-31	151809	7811	49157842	19959887	19117	41001	27846374	54.52	1740	926
velvet-P154-31	23429	4212	36998829	33461932	5851	125404	34628074	56.97	15239	6563
soap-P16-31	94597	9914	40065878	22319003	18045	60866	28107999	54.12	1965	1122
velvet-P18-31	69417	10021	46274745	35545645	15114	89715	39194438	53.9	4155	1999

VITA

Andrés Sebastián Espíndola Camacho

Candidate for the Degree of

Doctor of Philosophy

Thesis: EUKARYOTIC PLANT PATHOGEN DETECTION THROUGH HIGH THROUGHPUT DNA/RNA SEQUENCING DATA ANALYSIS

Major Field: Plant Pathology

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Plant Pathology at Oklahoma State University, Stillwater, Oklahoma in December, 2016.

Completed the requirements for the Master of Science in Entomology and Plant Pathology at Oklahoma State University, Stillwater, Oklahoma in December, July, 2013

Completed the requirements for the Bachelor of Science in Biotechnology Engineering at Escuela Politécnica del Ejército, Sangolquí, Ecuador in 2009.

Experience:

Graduate Research Assistant and Teaching Assistant, Department of Entolology and Plant Pathology, Oklahoma State University, Stillwater, Oklahoma, from January 2011 to December 2016.

Instructor, Microbiology, Escuela Politécnica del Ejército, Sangolquí, Ecuador from January 2010 to June 2010.

Research Assistant, Microbiology Laboratory, Escuela Politécnica del Ejército, Sangolquí, Ecuador from January 2008 to July 2009.

Professional Memberships:

American Phytopathological Society (APS)