GENOMIC RESOURCES FOR PHYLOGENETICS, POPULATION GENETICS AND DETECTION OF OOMYCETE PLANT PATHOGENS

By

MARIA FERNANDA PROAÑO CUENCA

Bachelor of Science in Biotechnology Engineering
Universidad de las Fuerzas Armadas ESPE
Quito, Ecuador
2014

Master of Science in Entomology and Plant Pathology
Oklahoma State University
Stillwater, Oklahoma
2016

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, 2020

GENOMIC RESOURCES FOR PHYLOGENETICS, POPULATION GENETICS AND DETECTION OF OOMYCETE PLANT PATHOGENS

Dissertation Approved:

Dr. Carla Garzon

Dissertation Adviser

Dr. Stephen Marek

Dr. Nathan Walker

Dr. Hassan Melouk

Dr. Darren Hagen

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor, Dr. Carla Garzon, for her help, encouragement, guidance, and support through the years and became my mentor and a dear friend. I would like to thank my committee members, Dr. Stephen Marek, Dr. Nathan Walker, Dr. Hasan Melouk, and Dr. Darren Hagen, for their help. I would also like to thank the staff and faculty of the Department of Entomology and Plant Pathology at Oklahoma State University. It has been a blessing to work in a friendly environment.

I thank the visiting scientists from Ecuador (Universidad de las Fuerzas Armadas – ESPE; Universidad Catolica del Ecuador) and Colombia (Universidad Francisco de Paula Santander) for the help through the years and their contribution to the improvement of my teaching, leadership, and communication skills.

I would like to express my eternal gratitude to my parents, Patricio and Doris, my sisters Gabriela and Patricia, and my lovely niece and nephew Gabita and Fer for their everlasting love, support, and inspiration. Finally, I would like to thank my partner and my family in Stillwater, Fernando.

Name: MARIA FERNANDA PROANO CUENCA

Date of Degree: DECEMBER, 2020

Title of Study: GENOMIC RESOURCES FOR PHYLOGENETICS, POPULATION GENETICS AND DETECTION OF OOMYCETE PLANT PATHOGENS

Major Field: PLANT PATHOLOGY

Abstract: Plant pathogenic oomycetes are a diverse group of fungal-like eukaryotes that cause disease in natural and agricultural ecosystems. Due to their biological properties that, in cases, require alternative technologies to culture and study, new methods have been developed. The development of those technologies has driven efforts in basic mycological techniques and genomics, diagnostics, and functional genomics since rapid and robust technologies for accurate identification and characterization are needed to develop and implement effective disease management strategies. My thesis involved integrating plant pathology, genomics, and bioinformatics to explore the evolutionary relationships, population biology, and detection alternatives of important species of the genera *Pythium* and *Globisporangium*. *Globisporangium irregulare* is a complex species that includes *G. regulare, G. irregulare* sensu stricto (s.s.), *G. cryptoirregulare*, and *G. cylindrosporum.* The phylogenomic analysis of the species within the complex conducted found a lack of genealogical concordance under the Genealogical Concordance Phylogenetic Species Recognition (GCPSR) and significant recombination evidence, which supports previous findings that *G. cryptoirregulare, G. irregulare* s.s. and *G. cryptoirregulare* are a single species. Later, the population diversity of *G. cryptoirregulare* in rhododendron from Oregon and Washington was analyzed. It was found that populations displayed a structure mostly associated with clonal lineages maintained in nurseries through the years. Understanding this plant pathogen's diversity and relationships can provide new insights of value for understanding disease epidemiology and disease management. Finally, the EDNA approach implemented through the MiFiTM platform was used to develop a validated database for oomycete diagnostics. It demonstrated its power for the effective detection of oomycete plant pathogens from metasamples. The results of these studies can improve the overall understanding of oomycete pathology and evolution and contribute with information to identify targets for effective and sensitive pathogen detection, allowing accurate and timely disease diagnosis and management.

TABLE OF CONTENTS

Chapter Page

Chapter Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

CHAPTER I

INTRODUCTION AND OBJECTIVES

Oomycetes are a diverse group of fungal-like eukaryotes that belong to the superphyla Heterokonta, in the kingdom Chromista (Cavalier-Smith 1981, 2018). They include heterotrophic pathogens, parasites, or saprophytes of a diverse range of organisms from protists, to plants, to animals. Oomycetes are considered one of the most successful groups of organisms for their global distribution and ease of transport, and along with fungi, are the most widely-dispersed and disease-causing organisms (Bebber and Gurr 2015; Derevnina 2016; Latijnhouwers et al. 2003). The two major orders of oomycetes are the Saprolegniales and the Peronosporales (Dick et al. 1999; Matari and Blair 2014). Saprolegniales include water molds that behave as saprophytes on fish, crustacea, amphibia, insects, and plants (Leclerc et al. 2000; Riethmüller et al. 1999). The Peronosporales include agriculturally important plant pathogenic genera such as *Phytophthora*, *Pythium*, and downy mildew (Dick et al. 1999; Matari and Blair 2014; Riethmüller et al. 1999; 2002).

Plant pathogenic oomycetes cause disease in natural and agricultural ecosystems (Bebber and Gurr 2015; Derevnina 2016; Latijnhouwers et al. 2003). The plant diseases

are of two types. Diseases that affect plant parts in contact with the soil and diseases that affect aboveground plant parts *(*Agrios 2005). Some of the most important plant pathogens include the genera *Albugo, Aphanomyces, Bremia, Peronospora, Phytophthora, Plasmopara, Pythium,* and *Saprolegnia,* among several others *(*Agrios 2005).

Oomycetes represent a unique eukaryotic group that, for many years, has been extensively examined. Because of their economic impact and the challenges they present as objects of research, their study has been a driving force in biotechnological developments (Kamoun 2003). Due to their biological properties that, in cases, require alternative technologies to culture and study, new methods have been developed. The development of those technologies has driven efforts in basic mycological methods, as well as genomics, diagnostics, and functional genomics since rapid and robust technologies for accurate identification and characterization are needed for the development and implementation of effective disease management strategies (Arafa and Shirasawa 2018; Grünwald et al. 2016; Kamoun et al. 2015; Lamour et al. 2007; Lévesque 2011). Nowadays, the availability of genetic information generated from genome sequencing has widely contributed to oomycete research.

In the area of systematics, although some pathogenic plant genera have been extensively studied, particularly *Phytophthora*, research on several other oomycetes is limited due to the difficulty of identifying informative loci. Single-gene and concatenated phylogenetic studies have resulted in conflicting species phylogenies, resulting in a lack of robust phylogenies within clades (Ascunce et al. 2017; Beakes et al. 2014). Resolving

closely related species is a particular challenge. Thus genome-scale phylogenomic studies are being incorporated to resolve conflicting relationships among species of multiple oomycete genera, including *Pythium* (Ascunce et al. 2017; McCarthy and Fitzpatrick 2017; Rujirawat et al. 2018).

The genus *Pythium*, which includes oomycete plant pathogens, was first described in 1858 by Pringsheim. *Pythium monospermum* Pringsheim was described as the type species (Van der Plaats-Niterink 1981). Since then, several taxonomic descriptions emerged, most of them based on the comparison of morphological characteristics of sexual and asexual structures (Martin and Kristler 1990; Middleton 1943; Paul and Masih 2000; Schroeder et al. 2013; Van der Plaats-Niterink 1981). However, due to the similarity, intraspecific variation, and overlapping morphology between closely related species, biochemical and molecular criteria have been used for species identification to supplement the morphology-based taxonomy (Ho 2018; Lévesque 2011; Martin and Kristler 1990; Middleton 1943). Most *Pythium* species are non-host selective, and the diseases they cause are influenced by environmental conditions (Agrios 2005; Daughtrey 2011). For example, *Pythium aphanidermatum* and *Pythium deliense* are aggressive pathogens that cause damping-off of seedlings and soft rots of roots and other tissues in contact with soil, the severity of the diseases they cause increases in warmer temperatures. In contrast, milder pathogens such as *Globisporangium irregulare* and *G. ultimum* (synonym with *Pythium irregulare* and *Pythium ultimum*), prefer more moderate conditions (Agrios 2005; Schroeder et al. 2013) and cause infections that can remain

latent until the host is under stress, and then aggressively invade host tissues and trigger symptom manifestation.

Globisporangium irregulare is an important soil-borne pathogen that has been proposed as a complex of species with worldwide distribution and broad host range. The complex currently includes multiple species recognized as valid in the MycoBank Database (www.mycobank.org): *G. regulare, G. cylindrosporum, G, irregulare* sensu stricto (s.s.), and *G. cryptoirregulare* (Garzon et al. 2005; Garzon et al. 2007; Spies et al. 2011). *G. irregulare* complex is a challenging group for identification because of the variable morphology characteristic of the species of this group and the high level of intraspecific genetic diversity (Garzon et al. 2005). Until now, the relationships between the species of the complex have not been entirely resolved. Hence the need to further explore this species complex with phylogenomic approaches to reconstruct their evolutionary history using alignments of large numbers of genes (Derevnina 2016; Eisen 1998).

From a different perspective, population genetics studies the effects of evolutionary processes on genomes and populations in order to answer epidemiological questions, such as factors affecting population structure, adaptation, and speciation, including migration, natural selection, genetic drift, and mutation (Grünwald and Goss 2011; Grünwald et al. 2016; Hedrick 2011; Karlin 2012; Milgroom 2015; Stinchcombe and Hoekstra 2008). Advances in genome sequencing have influenced population genetics studies, providing a framework for understanding the importance of specific genes, the genome organization of a species, and its evolution (Martin 2009; Restrepo et al. 2014;

Zody and Nusbaum 2009). Unfortunately, most of our current genomic knowledge is limited to species of *Phytophthora*, while little is known about species of *Pythium* and related genera.

The rapid advancement of high-throughput DNA sequencing technologies and computational tools has contributed to the development of novel applications. One promising approach that uses metagenomic sequencing data to detect pathogens termed E-probe Diagnostic of Nucleic acid Analysis (EDNA). EDNA is a pipeline that uses high-throughput sequencing, metagenomics, and bioinformatics to detect the presence or absence of target pathogens in metagenomes by using signature pathogen-associated sequences, named e-probes (Stobbe et al. 2013). E-probes are short species-specific markers that match the organism of interest only from all the metagenomic sequence data, reducing the computational time needed to detect pathogenic microbes (Stobbe et al. 2013, 2014). The approach was validated *in silico* for sensitivity, specificity, and limit of detection of several pathogens, including oomycete, fungal, bacterial, and viral plant pathogens (Espindola et al. 2015; Stobbe et al. 2013, 2014). Furthermore, the EDNA pipeline has been modified to detect the expression of genes associated with aflatoxin production in *Aspergillus flavus* strains, showing its applicability to detect the expression of genes associated with metabolic functions involved in pathogenicity. Consequently, it has shown its applicability and potential in the diagnosis of plant diseases and functional studies (Espindola et al. 2018).

The increased availability of genome sequences offers unique opportunities to conduct studies of plant pathogenic oomycetes, including important species of the genera *Pythium*

and *Globisporangium*. Such studies will improve the overall understanding of oomycete pathology and evolution and will contribute with information to identify targets for effective and sensitive pathogen detection, which will allow accurate and timely disease diagnosis and management.

The objectives of this research are as follows:

OBJECTIVES

1. To examine the evolutionary relationships between species of the *Globisporangium irregulare* complex and define boundaries between them through a phylogenomic approach.

2. To describe the genetic diversity and structure of populations of *Globisporangium cryptoirregulare* from *Rhododendron* nurseries in Oregon and Washington

3. To develop a detection database for oomycete plant pathogens by using E-probe Diagnostic Nucleic acid Analysis (EDNA).

CHAPTER II

LITERATURE REVIEW

Oomycetes

Oomycetes are a diverse group of fungal-like eukaryotes that belong to the superphyla Heterokonta (Stramenopiles), in the kingdom Chromista (Cavalier-Smith 1981, 2018; Margulis and Chapman 2009).The Stramenopiles (Patterson 1989) branches with Rhizaria and Alveolata within the SAR supergroup (Adl et al. 2012; Burki 2014). They include heterotrophic pathogens, parasites, or saprophytes of a diverse range of organisms from protists, to plants, to animals. Oomycetes are considered one of the most successful groups of organisms for their global distribution and ease of transport, and along with fungi, are the most widely-dispersed and disease-causing organisms (Bebber and Gurr 2015; Derevnina 2016; Latijnhouwers et al. 2003**).**

Oomycetes resemble fungi in aspects such as their filamentous growth habit, heterotrophic lifestyle, feed by absorption, and production of specialized structures for infection. Those characteristics made them be inaccurately referred to as fungi (Latijnhouwers et al. 2003; Richards et al. 2011). They are more closely related to golden

brown algae and heterokont algae based on biochemical and phylogenetic analyses of ribosomal and mitochondrial gene sequences than to true fungi (Kamoun 1999; Tyler et al. 2006). Oomycetes produce branching aseptate hyphal systems (i.e., mycelia), which contains cellulose, β-1,3 glucans, and β-1,6 glucans in their walls and minimal amounts of chitin. Oomycetes have diploid nuclei in the vegetative stage; produce biflagellate swimming zoospores in sporangia in the asexual stage and oospores in the sexual stage (Agrios 2005; Fry and Grünwald 2010; Kamoun et al. 2015, Lamour and Kamoun 2007). Their sexual and asexual reproduction, along with interspecific hybridization, also documented, had made them able to gain genetic diversity, spread rapidly, and also facilitates occasional host jumps (Derevnina 2016).

Plant pathogenic oomycetes cause devastating diseases affecting ornamental plants, crops, and trees, causing damage and economic losses (Agrios 2005; Derevnina 2016). The diseases produced are of two types. Diseases that affect plant parts in contact with the soil and diseases that affect aboveground plant parts (Agrios 2005; Kamoun et al. 2015). Oomycetes are a unique group with the ability to infect plants using biochemical and genetic mechanisms distinct from those used by true fungi (Kamoun 2003). Part of its success as a group is its ability to overcome host resistance and evolve by host jumps. Consequently, the exploration of their biology is useful to understand their evolutionary potential (Derevnina, 2016).

Genus *Pythium*

Taxonomy and representative species

The genus *Pythium* was first described in 1858 by Pringsheim. *Pythium monospermum* Pringsheim was described as the type species (Van der Plaats-Niterink 1981). Since then, several taxonomic descriptions have appeared, most of them based on the comparison of morphological characteristics of sexual and asexual structures among species (Martin and Kristler 1990; Middleton 1943; Schroeder et al. 2013; Van der Plaats-Niterink 1981), but the most recent taxonomic revisions of *Pythium* have been based on phylogenetic analyses using DNA sequence data of molecular barcodes (Lévesque and de Cock 2004; Martin 2000; Paul and Masih 2000; Uzuhashi et al. 2010).

The genus belongs to the domain Eukaryota, kingdom Chromista, phylum Oomycota, class Oomycetes, order Pythiales, and family Pythiaceae (Cavalier-Smith 1981, 1998; 2018; Ruggiero et al. 2015; Schroeder et al. 2013; Uzuhashi et al. 2010). Members of the genus have ubiquitous distribution and occupy various ecological niches, from facultative plant pathogens and soil saprophytes in terrestrial, marine, and freshwater environments (Kirk et al. 2001; Martin 2009; Moorman 2004). More than 200 host species have been described, and many more have been reported since 2000 (Feng et al. 2015; Kageyama 2014).

Most *Pythium* species are plant pathogens with broad host ranges. Notable representatives are the species *P. ultimum* (syn. *Globisporangium ultimum*) and *P. aphanidermatum*, responsible for causing significant losses in important crops

(Daughtrey et al. 1995; Kammarnjesadakul et al. 2011; Martin 2009; Schroeder et al. 2013). *Pythium ultimum* has additionally been used for the production of dietary supplements, polyunsaturated fatty acid, arachidonic acid, and eicosapentaenoic acid (Gandhi and Weete 1991). *Pythium oligandrum* and *Pythium nunn* are exceptional species reported as mycoparasites and biological control agents (Agrios 2005). *Pythium insidiosum* is an animal and human pathogen that causes phytiosis (Mendoza 1998).

Reproduction and disease cycle

Pythium spp. are routinely identified based on the morphology of asexual and sexual structures. From those, the shape of the sporangia and the ornamentation of the oogonium are key characters (Van der Plaats-Niterink 1981). Asexual reproduction occurs through sporangia, which germinate and produce hyphae or vesicles. Asexual motile spores, known as zoospores, are formed in the vesicles in the presence of free water. When they are mature, zoospores are released into a liquid environment where they can swim thanks to two unequal flagella. Zoospores encyst after several minutes of their release, either on a host surface or in water. If a suitable host is available, the cyst germinates, and a germinal tube infects host tissues (Agrios 2005). Zoospores have access to adjacent healthy plants due to their mobility and represent a means for disease spread. Species that generate few zoospores can infect hosts using mycelia harbored in the soil and plant debris as a source of vegetative inoculum (Allen et al. 2004).

In sexual reproduction, the antheridium (male) fertilizes the oogonium (female) and produces the oospore. The oogonia can be globular to limoniform, terminal or intercalary

on undifferentiated hyphae, with the oogonial wall smooth or ornamented (Dick 1991; Ho 2018). The unicellular antheridia can be sessile or intercalary on the hyphae, or terminal on a stalk. Monoclinous, if they originate from the oogonial stalk, and diclinous, if they originate from a different hypha (Dick 1991; Ho, 2018). One or more antheridia attaches to the oogonium, but only one forms a fertilization tube and penetrates the oogonium cell wall. The nucleus of the antheridium is transferred to the oogonium egg cell. The zygote formed, oospore, by karyogamy is diploid. Oospores have a dormant phase, pre- or post- karyogamy, after which they germinate and produce a germ tube. Oospores have a thin to a thick cell wall and can survive under adverse conditions, remaining viable for a few years, sometimes decades. Oospores and sporangia serve as primary inoculum in the disease cycle of *Pythium*, and their germination is determined by environmental conditions, chemical signals of the host, and temperature (Agrios 2005; Allen et al. 2004; Ho, 2018; Nzungize et al. 2012).

Heterothallic and homothallic sexual reproduction can also be used as characteristics for species identification (Dick 1991; Garrido 2014). Most *Pythium* species are homothallic and self-fertile, and only seven species have been reported as heterothallic, namely: *P. catenalatum, P. flevoense, P. heterothallicum, P. intermedium, P. macrosporum, P. splendens* and *P. sylvaticum* (Dick 1991; Van der Plaats-Niterink 1981). Data from *in vitro* crosses with *Globisporangium ultimum* and *Globisporangium irregulare* have reported outcrossing in approximately 10% of the progeny (oospores) produced when combining two distinct mycelia between isolates of the same homothallic species (Francis and St-Clair 1993; Harvey et al. 2001). Polyploidy, aneuploidy, or

heterokaryosis have also been observed from the results of field isolates on the same species; those results depict a genetic variation mechanism. (Martin, 2009).

Diseases and importance

Most *Pythium* species are plant pathogenic and cause significant crop losses. Some species can cause diverse diseases such as seed diseases of valuable crops, damping-off of seedlings, pre- and post-emergence root rots, and blackleg of cuttings. Economically, they are important pathogens of monocotyledonous, herbaceous, and woody plants. Seed and seedling diseases can also cause severe economic losses. Several *Pythium* species cause severe damage to adult plants, usually at the root level of vegetables (Al-Sa' di et al. 2008), field crops (Nzungize et al. 2012), trees (Lazreg et al. 2013; Weiland et al. 2015), lawns, floricultural and other ornamental crops, cereal crops and turf-grasses (Agrios 2005; Del Castillo-Munera 2016; Moorman 2002).

Most species are non-host selective, and the diseases they cause are influenced by environmental conditions (Agrios 2005; Daughtrey 2011). For example, *Pythium aphanidermatum* and *Pythium deliense* prefer warm and higher temperatures, contrary to *Globisporangium irregulare* and *Globisporangium ultimum*, which prefer moist and cooler conditions. Further awareness about this genus arises among growers because, under field and greenhouse conditions, they represent a problematic pathogenic group due to the killing potential to emerging or newly emerged seedlings, while also reducing crop yield and quality (Schroeder et al. 2013).

Diseases lead to a reduction in plant growth, vigor, and yield. Their broad host range, distribution, and the occasional lack of visible symptoms make it challenging to be aware of their presence until significant reductions occur (Agrios 2005). The type of diseases they cause are generally not lethal to mature plants. *Pythium* species mostly invade the meristematic tips, epidermis, and cortex of roots and fruits (Kamoun 2003; Vijayan et al. 1998); hence, there is limited information about *Pythium*, particularly its genetics (Martin 2009).

Disease symptoms and disease management

Pythium diseases can be diagnosed based on symptoms followed by pathogen isolation and identification. Symptoms of *Pythium* diseases are usually similar in most host plants and include wilting, stunting, chlorosis, decayed roots, poor seed germination, and emergence, resulting in the reduction of crop quality (Daughtrey and Miller 2009). On mature plants, diagnosis is difficult because plants often remain asymptomatic until symptoms start to manifest after a period of stress (Schroeder et al. 2013). These diseases affect all plant parts in contact with the soil and are described as seed rots, seedling damping-off, root rots, black leg of cuttings, and fruit rots (Agrios 2005).

Management strategies for *Pythium* diseases combine prevention, sanitation, and eradication measures. Prevention starts with the use of disease-free plant material or surface-sterilized propagative materials. Sanitation includes discarding potential sources of inoculum and infected plant material. Eradication includes scouting, disposal of symptomatic plants, and application of fungicide treatments (Agrios 2005; Al-Sa'di et al.

2008b; Garzón et al. 2011). In regards to chemical control, fungicide resistance to specific chemistries has been reported in several *Pythium* and *Globisporangium* species, particularly to metalaxyl, mefenoxam (Fungicide Resistance Action Committee [FRAC] group 4) and propamocarb (FRAC 28) (FRAC 2018; Moorman and Kim 2004), which has also been reported to affect the genetic diversity of some species (Lee et al. 2010). Metalaxyl and metalaxyl-M (mefenoxam) resistance have been reported in *P. aphanidermatum, P. cylindrosporum, P. dissotocum, P. heterotallicum, P. irregulare, P. splendes, P. ultimum,* and *Pythium spp.* (FRAC 2018; Moorman et al. 2002).

Molecular phylogeny

The basis for the morphological classification of *Pythium* species was developed by Van der Plaats-Niterink (1981). His work primarily focused on the differences in structural features such as the globose or filamentous sporangia, oogonia size, ornamentation, paragynous or hypogynous antheridia, and development of the zoospores, among others, which are used as primary diagnosis characters. However, due to morphological similarity, intraspecific variation, and overlapping between closely related species, biochemical and molecular criteria have been used for species identification to supplement the morphology-based taxonomy (Ho 2018; Lévesque 2011; Martin and Kristler 1990; Middleton 1943).

The DNA technology exploration for *Pythium* species increased in the 1990s accompanied by the development of universal ribosomal DNA (rDNA) primers for fungi that worked for oomycetes (White et al. 1990). Criteria used for taxonomic comparisons

included endonuclease restriction maps (Klassen et al. 1987), isozyme patterns and polymorphisms (Chen et al. 1991; Chen et al. 1992; Lee 1993), restriction endonuclease of mitochondrial DNA and restriction length fragment polymorphisms (RFLP) analysis of DNA (Lévesque et al. 1993; Martin and Kristler 1990). Briard et al. (1995) analyzed the ribosomal large subunit (LSU) of RNA of several *Pythium* species and found a correlation with the sporangia morphology, either globose or filamentous. Many other reports showed the importance of representing the morphological variation in the genus for studies using molecular data (Dick et al. 1999; Cooke et al. 2000; Martin 2000; Villa et al. 2006).

In the 2000's studies centered on the use of the non-coding internal transcribed spacer (ITS1 and ITS2) regions of rDNA, for the separation of taxa at the species and genus level (Cooke et al. 2000; Matsumoto et al. 1999; Matsumoto et al. 2000; Moorman et al. 2002). Many other reports conclude that the genus is polyphyletic with several monophyletic groups that cluster together based on the sporangia morphology (Bedard et al. 2006; Lévesque and de Cock 2004; Robideau et al. 2011; Schroeder et al. 2013; Uzuhashi et al. 2010). Additionally, other independent genes are used to compare phylogenies and reflect the evolution of the entire genome (Villa et al. 2006). Martin (2000), using the cytochrome oxidase 2 (*cox*II) gene, observed the formation of three phylogenetic groups. Species with filamentous sporangia formed one clade, and species with globose sporangia formed the other two groups. Villa et al. (2006) found similar results using the ITS region, *cox* II gene, and β-tubulin gene, this work, and others strengthened the importance of the sporangial morphology in *Pythium* systematics and

the use of DNA barcoding for identification of species (Belbahri et al. 2008; Choi et al. 2015; Moralejo et al. 2008).

Lévesque and de Cock (2004), using the sequencing of the ITS region and sequences of the LSU rDNA, separated the *Pythium* sensu lato species into 11 clades (A-K). Their study supported the morphological clustering of two monophyletic groups, one group composed of clades A-D, characterized by their filamentous sporangia, and a second group formed by clades E-J, characterized by their globose sporangia. Clade K (described later as *Phytopythium*) was phylogenetically distant from the other ten clades and was distinctive because of its *Pythium*-like zoospores and *Phytophthora*-like sporangia (Bala et al. 2010). The K clade forms a morphologically and phylogenetically intermediate clade between *Pythium* and *Phytophthora* (Bala et al. 2010; De Cock et al. 2015).

Several attempts have been made to split the genus and classify it based on the sporangia shape. For example, *Aquaperonospora* was described for species producing rigid, erect, and branched sporangiophores forming sporangia on branchlet tips. The genus was characterized by their morphological similarity to *Pythium* (=*Phytopythium*) *helicoides* (Ko et al. 2010). Bala et al. (2010) proposed the genus *Phytopythium* for those species with globose to ovoid sporangia. Uzuhashi et al. (2010) proposed dividing *Pythium* sensu lato into five genera based on sporangia morphology. They delimited the genus *Pythium* sensu stricto to those species with inflated or non-inflated filamentous sporangia (Lévesque and de Cock clades A-D) and created four new genera for species with non-filamentous sporangia. *Globisporangium* for species with globose and sometimes proliferating sporangia (Lévesque and de Cock, 2004 clades: E, F, G, I and J),

Elongiosporangium for species with clavate to elongate sporangia (Lévesque and de Cock clade H), *Pilasporangium*, for species with globose, non-proliferating sporangia, and *Ovatisporangium* for species with ovoid, pyriform, sometimes irregular shaped sporangia (Lévesque and de Cock clade K).

Phytopythium has been treated as synonymous with *Ovatisporangium* (De Cock et al. 2015), and the name *Phytopythium* takes precedence over *Ovatisporangium* because it was published earlier in 2010 (Schroeder et al. 2013). The genus *Phythopythium* is classified into three monophyletic clades (Baten et al. 2015; de Jesus et al. 2016), had been reclassified as a genus sister to *Phytophthora*, which evolutionary status has not been adequately resolved (Bala et al. 2010; De Cock et al. 2015).

***Pythium irregulare* complex**

The examination of multiple isolates from individual species helps to identify intraspecific variation in physiological and molecular traits. Bar et al. (1997) separated *Pythium irregulare* Buisman (syn. *Globisporangium irregulare*) isolates into two groups based on 11 isozyme loci. Matsumoto et al. (1999, 2000) performed Random Amplified Polymorphic DNA (RAPD) and phylogenetic analyses using sequences of the ITS (Internal Transcribed Spacer) region and found the formation of four groups (I, II, III, IV). Groups III and IV were not significant plant pathogens and were genetically distinct from groups I and II, clustering closer to *P. sylvaticum*. Similar results were observed by Lévesque and de Cock (2004) and by Garzon et al. (2005a) using ITS data. *P. irregulare* is part of clade F (Lévesque and de Cock, 2004), has great genetic diversity, and a genetic

structure that includes four clusters with significant bootstrap support. Based on its genetic structure, *P. irregulare* has been proposed as a complex of species with worldwide distribution and broad host range (Garzon et al. 2007; Spies et al. 2011).

Garzon et al. (2005a, 2007), using Amplified Fragment Length Polymorphism (AFLP) fingerprinting and sequencing of the ITS region and the *cox* I-II mitochondrial region, reported that *P.irregulare* had undergone a speciation process, supporting the separation of groups I and II sensu Matsumoto 1999. Garzon et al. (2007) identified groups I and II as *Pythium irregulare* (syn. *Globisporangium irregulare*) sensu stricto (s.s.) and *Pythium cryptoirregulare* (syn. *Globisporangium cryptoirregulare*) respectively, but groups III and IV were not included in the study due to lack of access to samples. *G. irregulare* and *G. cryptoirregulare* form a complex of cryptic and closely related species with *G. regulare,* and *G. cylindrosporum*, all of which have very similar ITS sequences but have limited gene flow between them (Garzon et al. 2007). The ITS cluster that includes the sequences of the type specimens of *G. cryptoirregulare, G. regulare* and *G. cylindrosporum* has significant genetic diversity, and the species limits cannot be resolved based on ITS and *cox* I-II sequences alone or combined, even though these species have morphological differences.

Garrido (2014), with a multigene phylogeny, found results similar to those in the Matsumoto (1999, 2000) and Garzon et al. (2005a, 2007) studies, supporting the formation of four clades that represent at least two of the sister species, *P. irregulare* s.s. and *P. cryptoirregulare*, separated with significant statistical support.

The existence of hybrids among *P. irregulare* s.s. and *P. cryptoirregulare* has been suggested (Garzón et al. 2007; Lee and Moorman 2008). Although Spies et al. (2011) found a similar genetic structure using ITS sequences, their analysis found evidence of aneuploidy and putative hybrids, using additional DNA sequence analyses of β-tubulin, mitochondrial regions (*cox* I and *cox* II), and isozymes. Based on which it was suggested that the complex should be maintained as one single and variable species *P. irregulare*. Nonetheless, the multilocus phylogeny reported by Hyde et al. (2014) analyzed the sequences of four phylogenetically informative loci and provided strong support for the separation of *P. irregulare* and *P. cryptoirregulare.*

The *P.irregulare* species complex is a challenging group for identification because of the limited morphological variation between three of the described species, the high levels of intraspecific genetic diversity (Garzón et al. 2005a, 2005b, 2007), and the lack of informative loci to resolve closely related oomycete species. Until now, the relationships among the species within the complex have not been completely resolved.

Diagnostics in the management of oomycete plant pathogens

Plant pathogenic oomycetes cause some of the most destructive plant diseases. Therefore, rapid and robust detection technologies are required for their accurate identification and characterization, and for timely management of crop diseases (Agrios 2005; Arafa and Shirasawa 2018). Currently, many methods used for isolation and diagnosis are relatively slow and are usually applied after symptoms of diseases are already evident.

Baiting, isolation, and direct culturing of pathogens, followed by microscopic examination of pure cultures, are part of the conventional protocols used for identification based on morphological characteristics (Pettitt et al. 2002). One of the principal constraints of these techniques is the time needed to generate information, one to two weeks depending on the species, which slows disease management decisions (Schroeder et al. 2013; Wakeham and Pettitt 2017).

In the last three decades, many advances have been made on molecular-based techniques for *Pythium* diagnostics, which fall into two categories: immunologically-based assays and nucleotide-based assays (Wakeham and Pettitt 2017). Immunoassays use polyclonal antibodies and monoclonal antibodies, which can identify molecules present on particular structures of the pathogen, providing sensitive and specific tests. Enzyme-linked immunosorbent assay (ELISA) commercial assays and lateral flow devices (LFDs) are available *Pythium* and *Phytophthora* and detection (Kyuchukova et al. 2006; Lyons and White 1992; Rafin et al. 1994). Serological tests are beneficial because of their ease and cost (Ali-Shtayeh et al. 1991; Avila et al. 1995; Murillo-Williams and Pedersen 2008; Yuen et al. 1998). However, they are limited in their ability to discriminate oomycete pathogens at the species level (MacDonald et al. 1990).

Nucleotide-base detection

To assist detection and overcome limitations of immunological assays, better extraction systems, and sample pre-treatments have been developed (Wakeham and Pettitt 2017). Recently, work has been done in the improvement of LFDs using nucleic

acid-based techniques. Loop-mediated isothermal amplification (LAMP) has been used successfully to detect *Pythium* and *Phytophthora* species in infected plant tissues (Feng et al. 2019; Fukuta et al. 2013, 2014; Shen et al. 2017; Takahashi et al. 2014). LAMP method amplifies DNA under isothermal conditions and has the advantage of only require a heating block instead of a thermal cycler than a conventional PCR (Polymerase Chain Reaction), or real-time PCR assay requires (Wakeham and Pettitt 2017).

Most molecular diagnostics methods are based on PCR and are widely and routinely used in plant disease diagnostic laboratories (Wakeham and Pettitt 2017). DNA markers provide stable results and constitute the basis of a range of techniques in basic and applied research (Collard et al. 2005). The ITS region of the rDNA is recommended as the universal barcode sequence for oomycetes (Lévesque 2011) since it is a region composed of conserved nucleotide sequences between species with interspersed variable sections and multiple copies along the genome, which increases assay sensitivity (Binagwa et al. 2016; Cooke et al. 2000; Klemsdal et al. 2008; Lévesque and de Cock 2004; Lévesque 2011; Matsumoto et al. 1999; Moorman et al. 2002; Nzungize et al. 2011; Wang et al. 2003). Other regions of the genome are also sequenced for identification and diagnostic purposes (Gomez-Alpizar et al. 2010; Schroeder et al. 2013). These include *cox* I and *cox* II spacer regions, as well as the beta-tubulin gene (Blair et al. 2008; Kroon et al. 2004; Lévesque and de Cock 2004; Martin 2000; Moralejo et al. 2008; Robideau et al. 2011; Villa et al. 2006). Unfortunately, these loci are not variable enough to consistently differentiate closely related species in simple or multiplex assays.

DNA markers are used for various purposes: detect polymorphisms between individuals in populations, determine genetic biodiversity among genotypes, investigate plant-pathogen interactions, analyze population genetic structure, and identify species identification (Patwardhan et al. 2014). DNA fragment polymorphisms can be used to characterize and differentiate oomycete species (Garzon et al. 2005b), some of those methods include restriction fragment length polymorphism (RFLP)-PCR (Gomez-Alpizar et al. 2010; Kageyama et al. 2005; Matsumoto et al. 2000; Rafin et al. 1995; Wang and White 1997), random amplified polymorphic DNA (RAPD) (Herrero and Klemsdal 1998) and amplified fragment length polymorphism (AFLP) (Garzon et al. 2005a, 2005b; Perneel et al. 2006). However, intraspecific variation makes their use challenging (Schroeder et al. 2013). Subsequently, microsatellites (SSRs) are also used and have been characterized for many species. SSRs showed potential in population genetics and diagnostics (Al-Sa'di et al. 2008; Garzon et al. 2005b; Lee and Moorman 2008; Lee et al. 2010).

Progress in molecular diagnostics has coupled PCR with multiplex PCR (Asano et al. 2010; Ishiguro et al. 2013), multiplex arrays, and real-time PCR. Tambong et al. (2006) described a DNA macroarray to detect more than 100 *Pythium* species. This study, along with others, used array-based detection to identify of all species in a sample (Le Floch et al. 2007; Lévesque et al. 1997; Lievens et al. 2005; Njambere et al. 2011). In the case of real-time PCR, it has also been used to quantify multiple pathogenic species (Kernagan et al. 2008; Le Floch et al. 2007; Li et al. 2010, 2014; Lievens et al. 2006; Okubara et al. 2005; Schroeder et al. 2006). As a result, further studies using various molecular biology

techniques, targeting novel loci, are needed to study the phylogeny of isolates and to develop rapid tools for the identification of unknown isolates, especially when some critical morphological characters are lacking (Ho 2018).

DNA sequencing and Next generation sequencing

In the 1970s, Sanger dideoxy sequencing of DNA started a revolution by allowing to decipher complete genomes (Sanger et al. 1997). The automation of the Sanger's chain-termination sequencing let it becoming the dominant sequencing method for three decades and was the foundation for several techniques developed to overcome some of its limitations. The Sanger method's limitations include the amount of data produced, the speed, the sequencing quality, time required, and the applications it has (Ari and Arikan 2016). However, the Sanger method remains a useful tool that can to produce about 250 kb data (1000 bp DNA sequences) at a low average cost (Verma et al. 2017). The cost of sequencing has decreased as new advances and applications have emerged (Ari and Arikan 2016; Verma et al. 2017). Also, useful databases have been developed to allow species identification by comparisons of DNA sequence similarity using the Basic Local Alignment Search Tool (BLAST). However, caution is needed since many have not been carefully annotated to species and curated, and mistakes are not uncommon (Ho 2018; Wakeham and Pettitt 2017).

The advent of high-throughput DNA sequencing technologies and their applications has profoundly impacted the advancement of molecular biology and molecular genetics research (Mardis 2008). The sequencing technologies developed after the Sanger method

are known as Next-generation sequencing (NGS). NGS technologies started a second revolution by making genome sequencing cheaper and faster. NGS is a term used for technologies that have enabled the massively parallel analysis of DNA sequences. The essential steps involved in NGS sequencing are generically applicable to all current technologies available (Schuster 2008).

The most significant advantage of NGS is the ability to produce an immense volume of data and reach millions or billions of reads in just one run. This is an enormous advantage compared to Sanger's sequencing, which only reaches up to hundreds of reads, but with a length of up to approximately 1000 base pairs (bp). Therefore, with NGS, genome coverage is considerably increased. However, NGS technologies produce shorter sequences (e.g., those produced utilizing technology and equipment from Illumina), except for those produced utilizing technology and equipment from PacBio or Oxford Nanopore platforms, which represent a computational challenge for assembly, due to the length and the huge number of sequences (Pop 2009).

The sequencing platforms available through Illumina uses the principle of sequencing by synthesizing DNA strands one nucleotide at a time in parallel (Momand and McCurdy 2017). It generates short sequences with high error rates. Reads are between 100-350 bp and reach depths of at least 30X per library (Alkan et al. 2011). Short-read sequencing can reveal details about the genomic content, but it is often insufficient for assessing the genomic structure. Short reads cannot span repetitive structures in the genome because when a read comes from a region that includes a short repetitive motif, it is not known with certainty which copy of the repetition was obtained. Thus, during assembly, false

joints can be created in the genome. Due to the repetitive regions, it is often difficult to assemble all the fragments, so it is not easy to reconstruct an entire genome sequence (De Maio et al. 2019; Pop 2009).

The sequencing platforms available through PacBio and Oxford Nanopore uses single-molecule, long-read sequencing that can resolve complex genomes when combined with short-read data. They generate reads of several kilobases, which can exceed the length of repeats in a genome, making a possible complete assembly. Although this is a key advantage, the accuracy of the results varies (Wick et al. 2017). The sequencing error rates are higher than those produced by Illumina, although this issue has improved in recent years (De Maio et al. 2019; Pop, 2009).

High-throughput sequencing applications for oomycete plant pathogens

Whole-genome sequencing analysis of many plant pathogens has contributed to the understanding of pathogenicity, host preferences, secreted effector proteins, and fungicide resistances (Arafa and Shirasawa 2018; Grünwald et al. 2016). As a diagnostic tool, it does not require previous knowledge of genome sequences of a specific host or pathogen but requires specific procedures in the lab and computing skills for the analysis of large amounts of sequence data (Schuster 2008). These technologies provide the capacity of developing novel applications such as analysis of complex environmental samples without culturing, and the identification of known, unknown and new pathogens within host samples, a field known as metagenomics (Adams et al. 2009; Mardis 2008). Advancement of NGS technology enables the analysis of the genetic variations of

pathogens and crops at the genome level. The information generated provides knowledge in the evolution of oomycete pathogens and their interactions with their hosts (Yin et al. 2017). Moreover, NGS-based genotyping techniques can be used for diagnosis and monitoring epidemics, new diseases, and pathogen populations (Grünwald et al. 2016).

Phylogenetic relationships among oomycete through phylogenomic approaches

In oomycetes, molecular studies have used the sequences of nuclear-encoded and mitochondrial-encoded genes, individually or combined, to resolve the phylogenetic relationships of species. Multilocus phylogenies have been successfully used to determine the phylogeny of oomycetes, combined with significant sampling for a better understanding of relationships between taxa (Beakes and Sekimoto 2009; Choi et al. 2015; Robideau et al. 2011). Nevertheless, research on oomycete systematics is limited due to conflicting phylogenies and the lack of robust phylogenies within clades. Single-gene and concatenated phylogenetic studies have resulted in conflicting species phylogenies. Thus genome-scale phylogenomic studies have been used to resolve relationships and to answer questions about genome evolution (Ascunce at al 2017).

Phylogenomics uses genome data to reconstruct evolutionary histories by using large alignments of genes (Eisen 1998; Eisen and Fraser 2003). Phylogenomic methods are recommended when the history of gene families is complicated, and accuracy is an objective. Consequently, in the last years, phylogenomic approaches have been incorporated to study the evolutionary relationships in *Pythium* species and other

oomycete plant pathogens (Ascunce et al. 2017; McCarthy and Fitzpatrick 2017; Rujirawat et al. 2018).

McCarthy and Fitzpatrick (2017) used supertree methods to perform a phylogenomic reconstruction of 37 oomycete genomes. Supertree methods rely on the analysis of several trees when individual phylogenies cannot resolve evolutionary relationships. The study was done with data from 8,355 gene families, which was able to resolve oomycete classes and individual clades, particularly within the genus *Phytophthora* and *Pythium.* *Pythium* clades A-D (*Pythium aphanidermatum, Pythium arrhenomanes, Pythium insidiosum,* and *Pythium oligandrum* respectively) clustered with 100% bootstrap support (BP), *Pythium* clades F, G, and I (*Pythium irregulare, Pythium iwayami, and Pythium ultimum* respectively) clustered with 100% BP, and *Phythopythium vexans* placed at the base of the Peronosporales order. The results were in agreement with previous reports and represent an important step in the analysis of oomycete systematics, generating information useful for further studies.

Ascunce et al. (2017) used a phylogenomic approach to clarify the phylogeny of *Pythium*. The results were in agreement with previous reports: *Pythium* sensu lato (s.l.) formed two paraphyletic groups (Lévesque and de Cock 2004; Uzuhashi et al. 2010). One group was composed of clades A, B and C, and the other was composed of clades F, G and I. The group formed by clades F, G and I was placed as sister to the *Phytophthora-Hyaloperonospora-Phytopythium* clade, supporting clade K (*Pythopythium*) as a sister clade of *Phytophthora* and *Hyaloperonospora.* Thus, this study also confirmed *Pythium*

as a polyphyletic group in agreement with previous phylogenetic studies (Hyde et al. 2014; Robideau et al. 2014; Villa et al. 2006).

The most recent phylogenomic analysis of *Pythium* species was a study on the animal pathogen *Pythium insidiosum*. This study investigated the evolutionary relatedness and differences between 20 oomycetes using a set of 2,073 core genes. The results were able to separate all oomycetes according to their genera. *Phytopythium vexans* was separated from *Pythium* species and were proximal to *Phytophthora* species, and *Hyloperonospora arabidopsis* (Rujirawat et al., 2018), which is congruent with previous analyses (Ascunce et al. 2017; McCarthy and Fitzpatrick 2017; Rujirawat et al. 2018).

Population genetics of oomycete plant pathogens

Genome sequencing and genomics have also impacted population genetic studies providing a framework for understanding the importance of specific genes, genome organization of species, and genome evolution (Martin 2009; Zody and Nusbaum 2009). Population genetics study processes on genomes and populations to answer three types of questions: evolutionary and demographic processes affecting population structure; evolutionary processes affecting speciation; and locus-specific effects acting on genes or chromosomes that affect adaptation (Grünwald and Goss 2011; Grünwald et al. 2016; Hedrick 2011; Karlin 2012; Milgroom 2015).

Many studies have demonstrated the potential of population genetics to identify intraspecific variations within oomycete species (Calerson et al. 2019; Del Castillo Munera et al. 2019; Huzar-Novakowiski and Dorrance 2018; Kamvar et al. 2015; Lee

and Moorman 2008; Weiland et al. 2015). In plant pathology, population biology uses population genetics approaches to address epidemiological questions (Milgroom and Peever 2003). Studies like these provide evidence of the many applications in plant pathology, which can help improve our understanding of pathogen biology, disease management, and disease resistance (Grünwald et al. 2016).

With the increase of genomic information available, traditional population genetics has been extended to explain the same and more detailed type of questions with population genomics but using a large number of markers (Grünwald et al. 2016; Stinchcombe and Hoekstra 2008). Population genomics embraces detailed genetic analyses to discover the mechanisms underlying phenotypes associated with adaptive traits (Grünwald et al. 2016). The advances in oomycete genetics and genomics are possible because of the availability of complete genome sequences, which provide access to large numbers of intraspecific single-nucleotide polymorphisms (SNPs) and microsatellite or simple sequence repeat (SSR) loci that can be used as genetic markers for population genetics studies (Grünwald et al. 2016; Lévesque 2011).

Reduced representation methods are alternatives to whole-genome sequencing, which mitigate the amounts of sequencing data needed to identify polymorphism in populations and are useful for an organism with large genome (Hass et al. 2009; van Orsouw et al. 2007). Genotyping by sequencing (GBS) is one method that can be used for genotyping and for developing novel molecular markers (Elshire et al. 2011; He et al. 2014; Milgroom et al. 2014; Sonah et al. 2013).

Some applications of population genetics and genomics approaches for fungal and oomycete pathogens include comparative genomic analyses of *Phytophthora infestans* and closely related species (Goos et al. 2011; Haas et al. 2009; Ioos et al. 2006; Raffaele et al. 2010; Tyler et al. 2006); demographic processes and genome-wide effects on *Fusarium graminerarum* (Talas et al. 2015; Talas et al. 2016) and *Verticillum dahliae* (Milgroom et al. 2014; Short et al. 2014); identification of genes under selection, or quantitative trait locus mapping (QTL mapping) to explain variation in *Zymoseptoria tritici* (Lendenmann et al. 2014; Poppe et al. 2015; Stewart et al. 2016), among others.

E-probe Diagnostic of Nucleic acid Analysis (EDNA)

Plant pathogen detection requires an early, rapid, and accurate assessment. Although several protocols exist for detection and identification of oomycete plant pathogens, there are several limitations in their use, such as the detection of one or a few pathogens using single target or multiplex assays, their level of sensitivity is variable, and cannot detect uncharacterized organisms (Adams et al. 2009). In contrast, NGS had been applied on environmental data, allowing the development of metagenomic based diagnostics, for simultaneous detection of all organisms present within a sample. Metagenomics studies the composition of microbial populations present in a sample by analyzing the nucleotide sequences obtained from genome analysis of the entire sample (metasample) (Adam et al. 2009; Escobar-Zepeda et al. 2015). In diagnostics, NGS is followed by sequence filtering and trimming, assembly, and BLAST comparisons (Huson et al. 2011). Next-generation sequencing technologies are producing an exponential increase in environmental data in

public databases. Therefore, it is required specialized solutions and methods for dealing with large and complex metagenome data sets (Ritcher et al. 2008).

E-probe Diagnostic of Nucleic acid Analysis (EDNA) is a pipeline that uses high throughput sequencing, metagenomics, and bioinformatics to detect the presence or absence of sequences of the target pathogen in metagenomes by using pathogen-specific sequences, named e-probes (Stobbe et al. 2013), as reference for identification. This approach can identify one or multiple target organisms present in the metasample. E-probes are short species-specific markers that allow targeting the pathogens of interest only, reducing the computational time needed to detect and identify a pathogen (Stobbe et al. 2014). E-probes are identified using a modified version of the microarray probe software tool for oligonucleotide fingerprinting identification (TOFI) (Vijaya Satya et al. 2008). EDNA approach applies four steps: e-probe design, comparison of target pathogen genomes with near neighbors sequences, construction of mock sequencing databases (MSDs), the metagenomes, and detection of the pathogen from databases using the e-probes as queries (Stobbe et al. 2013).

Stobbe et al. (2013) reported for the first time the use of EDNA for the detection of multiple pathogens including two viruses (Bean golden mosaic virus and Plum pox virus), five bacteria (*Spiroplasma citri*, *Xyllela fastidiosa*, *Xanthomonas oryzae*, *Candidatus Leberibacter asiaticus* and *Ralstonia solanacearum*), two fungi (*Puccinia graminis* and *Phakopsora pachyrhizi*) and two oomycetes (*Pythium ultimun* and *Phytophthora ramorum*). This work highlighted the potential of the EDNA approach to address diagnostic needs with application in a wide range of fields. It offers advantages

over other diagnostic protocols, such as reducing the computational work needed for conventional metagenomic-based diagnostics, and its flexibility allows it to be adapted for many pathogens, hosts, and any NGS platform (Stobbe et al. 2013).

The EDNA pipeline has been validated for sensitivity, specificity, and limit of detection *in silico* in several pathogens (Blagden et al. 2016; Espindola et al. 2015; Stobbe et al. 2013; Stobbe et al. 2014). EDNA was applied to metagenomic data sets obtained from virus-infected plant material to differentiate closely related strains of the Plum pox virus (PPV) (Stobbe et al. 2014). It also compared results when using the traditional metagenomic approach for NGS data and the EDNA approach. The study showed that EDNA is a powerful tool for diagnostics because of its ability to reduce the size of the known sequence databases and its ability to detect samples infected with the viral strains used. Espindola et al. (2015) reported the application of EDNA for diagnostics of fungi and oomycetes using NGS output databases. The target genome information used included *Pythium ultimum*, *Puccinia graminis* f. sp. *tritici*, and *Phytophthora ramorum*. The study used EDNA for eukaryotic plant pathogens and demonstrated its reliability as a diagnostic tool, based on the high sensitivity and specificity values found during detection, attributed to the species-specific e-probes designed.

The EDNA approach was also evaluated for the detection of foodborne pathogens, and it was validated using NGS data sets from alfalfa sprouts inoculated with *Escherichia coli* O157: H7 (Blagden et al. 2016). The authors recommended EDNA as a rapid detection

methodology for foodborne outbreaks and for creating microbial profiles of complex food samples.

Moreover, EDNA capacities were extended to detect specific gene targets present in transcriptomic databases by Espindola et al. (2018), using the approach to investigate the presence of the aflatoxin producing genes in different substrates colonized by *Aspergillus flavus* strains. In this study, the original EDNA pipeline was modified, and EDNA-transcriptomics was developed, which incorporated functional genome annotations. EDNA-transcriptomics detected aflatoxin production in a corn-based substrate without assembly or mapping to reference genomes and was the first attempt to identify actively metabolizing plant pathogens by using sequencing technologies. The authors recommended this novel approach for studies in the detection of metabolic functions associated with pathogenicity in other pathosystems.

CHAPTER III

PHYLOGENOMIC ANALYSIS OF *GLOBISPORANGIUM IRREGULARE* COMPLEX

Abstract

Globisporangium irregulare (syn. *Pythium irregulare*) is an oomycete plant pathogen responsible for diverse soilborne and waterborne diseases in a broad host range. *G. irregulare* has been proposed as a complex of species based on morphology and multilocus phylogenetic analysis. The species described in the complex are *G. regulare, G, irregulare* s.s., *G. cryptoirregulare*, and *G. cylindrosporum*, the last one having very distinct morphology even though its *cox* I-II and ITS region sequences overlap with those of *G. cryptoirregulare*. The phylogenetic relationships of an international collection of 34 isolates of the *G. irregulare* complex was analyzed using a phylogenomic approach and applying the Genealogical Concordance Phylogenetic Species Recognition (GCPSR). Based on the well-supported branch and concordance factors in the species trees generated by concatenation and coalescence approaches using BUSCO genes, the isolates were grouped in a monophyletic clade representing one phylogenetic species, *G. irregulare*. Although two well-supported groups (I and II) were found within the

G. irregulare clade, the lack of genealogical concordance under the GCPSR and significant recombination evidence supports previous findings that *G. cryptoirregulare*, *G. irregulare* s.s. and *G. cryptoirregulare* are a single species. Further research is required to explore the role of recombination as the source of discordance

1. Introduction

Globisporangium Uzuhashi, Tojo & Kakish (formerly *Pythium* Pringsh.) species are important phytopathogens that establish rapid infections in crops and wild plants, with severe impacts on the economy and ecosystems worldwide. *Globisporangium irregulare* (Buisman) Uzuhashi, Tojo & Kakish (syn. *Pythium irregulare* Buisman.) is an oomycete responsible for diverse soilborne and waterborne diseases in a broad host range (Levesque and De Cock 2004; Van der Plaats-Niterink 1981). Significant variation within the genus in terms of virulence, global distribution, and host range contributes to considerable economic loss caused by pre-emergence and post-emergence damping-off and root rot of seedlings in ornamentals and vegetable crops (Adhikari et al. 2013; Garzon et al. 2005b; Hendrix and Campbell 1973; Martin 2009; Martin and Loper 1999).

Globisporangium irregulare has been proposed to be a species complex that exhibits high morphological and genetic diversity based on morphology and multilocus phylogenetic analysis (Harvey et al. 2001; Levesque and De Cock 2004; Spies et al. 2011). The species within the complex include *G. irregulare sensu stricto* (s.s.), *G. cryptoirregulare*, *G. cylindrosporum*, and *G. regulare*. Efforts to resolve the relationships of this species complex have been made using isozymes (Bar et al. 1997), random

amplified polymorphic DNA (RAPD) (Matsumoto et al. 1999, 2000), amplified fragment length polymorphisms (Garzon et al. 2005a, 2005b) and nuclear and mitochondrial DNA sequence data including the ITS region, the *cox* I-II region, β-tubulin region and the heat shock protein 90 (*Hsp*90) (Garrido 2014; Garzon et al. 2007; Levesque and De Cock 2004; Spies et al. 2011). However, until now, the relationships of the species have not been entirely resolved. Differences in mefenoxam sensitivity in isolates of these species suggest that disease management may benefit from discriminating between the species (Garzon et al. 2007; Moorman et al. 2002).

Molecular phylogenetics in oomycetes commonly uses the sequences of nuclear-encoded and mitochondrial-encoded markers for single-gene or multigene approaches (Beakes and Sekimoto 2009; Choi et al. 2015; Robideau et al. 2011). Phylogenetic studies have contributed to defining new species by using the phylogenetic species concept and its extension, the genealogical concordance phylogenetic species recognition (GCPSR) (Cai et al. 2011; Taylor et al. 2000). GCPSR uses multigene genealogies to determine the concordance of branches supporting the presence of separate species (Taylor et al. 2000). Accurate taxonomical classification is required to describe new species, particularly with cryptic species where phylogenetic and morphological concepts do not draw clear distinctions between them (Huzar-Novakowiski and Dorrance 2018; Milgroom 2015; Eggertson 2012). Such is the case with *G. irregulare* complex where *G. cylindrosporum* and *G. regulare* are morphologically distinct species but not distinct phylogenetic species. The inverse occurs with *G. irregula*re s.s. and *G. cryptoirregulare*,

which are phylogenetically distinct species but morphologically indistinguishable (Garzon et al. 2007; Spies et al. 2011).

Research on oomycete systematics is limited due to conflicting phylogenies and the lack of robust phylogenies within clades. Thus, genome-scale phylogenomic studies have been incorporated to resolve incongruent relationships, understand population dynamics, and genome evolution (Ascunce et al. 2017; Rokas et al. 2003). Phylogenomics, which intersects genomics and evolution, uses large datasets to infer evolutionary relationships and reconstruct species phylogenies (Eisen 1998; Eisen and Fraser 2003). Therefore, these approaches have been included to study the evolutionary relationships in various species of oomycete plant pathogens, including *Pythium* (Ascunce et al. 2017; McCarthy and Fitzpatrick 2017; Rujirawat et al. 2018).

The objective of this study was to define the species boundaries of an international collection of isolates within the *G. irregulare* complex through a phylogenomic approach. Resolving the genomic differences between the clades can contribute to understanding the biology and evolution of these species.

2. Materials and Methods

2.1.Isolates information

An international collection of thirty-four *G. irregulare* isolates, two *Pythium 'vipa'* isolates, and one outgroup *Pythium mamillatum* (syn. *Globisporangium mamillatum*) were obtained from diverse sources (Table 3-1). The Ottawa Research and Development Center, Agriculture and Agri-Food in Canada provided 21 genomes (LEV isolates;

Nguyen et al. 2018); the 16 DNA samples were isolates selected to represent lineages identified in previous studies (Garrido 2014; Garzon et al. 2007) and maintained in Dr. Garzon's DNA collection (OOM isolates).

2.2.Isolates identification

Samples with good DNA concentration, quality, and integrity were selected from Dr. Garzon's *Pythium* DNA collection. DNA quality was measured by a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific, Inc., Waltham, MA, USA). In order to confirm the species identity of each sample (OOM isolates), PCR amplification of the ITS (internal transcribed spacer) region of the ribosomal DNA was conducted using primers ITS5 (sense) and ITS4 (antisense) (White et al. 1990), and a fragment of the *cox* I and cox II genes using primers FM35 (sense) and FM52 (antisense) (Martin 2000), following previously reported PCR conditions. Sequencing reaction products were run on an ABI 3100 DNA sequencer (Applied Biosystems) at the Recombinant DNA and Protein Core Facility at Oklahoma State University.

2.3.Whole genome sequencing with Illumina technology

Total double-stranded DNA (dsDNA) concentration was assessed with the Qubit 3.0 fluorimeter (Thermo Fisher Scientific , Inc., Waltham, MA, USA) using the Qubit ® dsDNA BR (Broad Range) assay kit. The integrity of the gDNA was verified by electrophoresis in a 0.8% agarose gel. Isolates selected to represent the genetic diversity of each of the species in the *G. irregulare* species complex (OOM samples) were sent to the Ottawa Research and Development Center, Agriculture and Agri-Food in Canada for

subsequent whole-genome sequencing. Briefly, the gDNA was normalized to 300 ng and sheared to 350 bp fragments using the Covaris M220 focused-ultrasonicator (Covaris, Inc., Massachusetts, USA). The obtained fragments were used as templates to construct PCR free libraries with NxSeq AmpFREE Low DNA Library kit (Lucigen) and TruSeq CD dual indices (Illumina, San Diego, CA, USA) following Lucigen's library protocol. Indexed libraries were pooled and sequenced on a NextSeq 500/550 sequencing system (Illumina, San Diego, CA, USA) using the 2 x 150-bp NextSeq High Output Reagent Kit (Illumina) according to the manufacturer recommendations to obtain the paired-end reads. Additionally, whole-genome sequences (LEV isolates) were provided by the Ottawa Research and Development Center, Agriculture and Agri-Food in Canada. Illumina MiSeq (300 bp) or NextSeq (150 bp) sequencing system were used to obtain paired-end reads for these samples (Table 3-1).

2.4.Genome assembly and genome statistics

Read quality on post-sequencing raw data was assessed with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adaptor sequences and trimming of poor quality data was performed with BBDuk (https://manpages.debian.org/testing/bbmap/bbduk.sh.1.en.html) from the Joint Genome Institute BBTools (https://jgi.doe.gov/data-and-tools/bbtools/). FastQC was used to check the impact of quality control on trimmed reads, and multi-QC reports were created for all genomes in R version 4.0 (R core team 2020) using the package fastqcr (Kassambara 2019) to inspect its quality.

De-novo genome assembly was performed on filtered reads with SPAdes v3.13.0 with error correction, and mismatch correction enabled (Bankevich et al. 2012). Basic genome quality and statistics were determined with QUAST v4.6.3 (Gurevich et al. 2013). To assess each assembly genome completeness, we used the Benchmarking Universal Single Copy Orthologs (BUSCO) tool v3.0.2 (Simão et al. 2015), which for genome annotation used Augustus (Keller et al. 2011), a program for the prediction of genes in eukaryotic genome data. BUSCO was run with the stramenopiles database (stramenopiles_odb10 database) (Table 3-2) part of the OrthoDB v10 database. OrthoDB database has a catalog of orthologous protein-coding genes across many organisms (Kriventseva et al. 2018).

2.5. Phylogenomic analysis

2.5.1. Alignments

BUSCO genes, single-copy orthologs, were used as phylogenetic markers (Waterhouse et al. 2018). The shared, complete BUSCO genes common in all genome assemblies were extracted into fasta files. Each BUSCO gene was individually aligned using the multiple alignment program MAFFT v7.245 (Katoh and Standley 2013) and trimmed using the tool for automated alignment trimming trimAl v1.2rev59 (Capella-Guitierrez et al. 2009) with the parameter "–automated1". We constructed nucleotide (NT), and amino acid (AA) versions comprised of the aligned and trimmed BUSCO genes.

2.5.2. Individual gene trees

Individual phylogenetic gene trees were constructed for each BUSCO alignment by maximum likelihood (ML) using RAxML v8.2.8 (Stamatakis et al., 2014). The best substitution (evolutionary) model for every single gene was estimated with the PROTGAMMAAUTO option for the amino acid data matrix and the GTRGAMMA for the nucleotide data matrix. The evolutionary models were ranked based on the Bayesian information criterion (BIC) (Schwarz 1978). PROTGAMMAAUTO option determined the best protein substitution model for the dataset. GTRGAMMA is the general reversible model with gamma-distributed rates across sites (Stamatakis et al. 2014). We assessed node support by bootstrapping (100 replicates) using the nearest-neighbor interchange (NNI) branch swapping method.

To compared the agreement or disagreement amongst the BUSCO gene trees and evaluate topologies using the newick utilities command-line tool was used (Junier and Zdobnov 2010). Gene trees were concatenated into a single file, then rooted, sorted, and each topology occurrence was counted. All trees were summarized all trees with a majority-rule consensus tree using the *consense* program part of the phylip package (Falstein 2005). The trees were visualized and annotated using the Interactive Tree of Life (iTOL) (Letunic and Bork 2007).

2.5.3. Species tree

Phylogenetic species trees were constructed separately for the AA and NT alignments under two approaches: concatenation (Huelsenbeck et al. 1996; Philippe et al. 2005;

Rokas et al. 2003) and coalescence (Edwards 2009).

Species tree by concatenation method

Alignments were concatenated to construct a supermatrix of nucleotide and amino acid residues. Alignments were then trimmed using trimAl with the option "-automated1" to keep only the best-aligned regions. The trimmed and concatenated supermatrix was used to estimate the ML species tree using IQ-TREE v2.0 (Nguyen et al. 2015). The best substitution model was estimated using ModelFinder (Kalyaanamoorthy et al. 2017), ranked based on the Bayesian information criterion (Schwarz 1978). Branch support was calculated on 1000 bootstrap replicates with the ultrafast bootstrap approximation (UFBoot) (Hoang et al. 2017), an accurate and faster alternative to the classic bootstrap approach. Brach support was also measured with the SH-like approximate likelihood ratio test (SH-alRT) (Guindon et al. 2010) and the approximate Bayes test (aBayes) (Anisimova et al. 2011). The phylogenies were visualized and annotated using the Interactive Tree of Life (iTOL) (Letunic and Bork 2007).

Species tree by the coalescent method

The newick single-gene trees were used as input in ASTRAL-III, version 5.7.4 (Chao et al. 2018), with default parameters. The Accurate Species Tree Algorithm (ASTRAL) enables coalescent-based analyses of datasets. This method estimates species phylogeny from single-gene trees under the multispecies coalescent (Edwards 2009). To assess node support, we used the local posterior probabilities (LPP) (Sayyari and Mirarab 2016). LPP is the probability that a branch is a true branch given the set of gene trees, computed

based on the quartet score, and assuming incomplete lineage sorting. The quartet score is the proportion of the quartets in the gene trees that support the branch (Sayyari and Mirarab 2016). The Interactive Tree of Life (iTOL) (Letunic and Bork 2007) software was used to visualize and annotate the phylogeny.

2.6. Concordance factors

The genealogical concordance was measured in the species tree (Astral) with the gene concordance factor (gCF) and the site concordance factor (sCF) as implemented in IQ-TREE (Nguyen et al. 2015). These values indicate how well each gene, or each site of the alignments for a given node in the species tree, are concordant with that node. Both values range from 0% (high discordance of gene trees and informative sites) to 100% (high concordance of gene trees and informative sites).

2.7. Detection of recombination with the pairwise homoplasy index test

The clades within the *G. irregulare* complex defined in the coalescence species tree were analyzed using the Genealogical Species Recognition model (GCPSR) (Taylor et al. 2000) by performing a pairwise homoplasy index (Φw) test. The test determines whether there is a recombinant signal present within the sequences (Bruen et al. 2006). The pairwise homoplasy index test (PHI) using the GCPSR concept was performed in SplitsTree v4 (Huson 1998; Huson and Bryant 2006) to determine recombination levels within the *G. irregulare* clades using the concatenated sequence alignments obtained with the 75 BUSCO genes. The BUSCO stramenopiles odb10 set has 100 genes, only 75 genes were used because they were present in all of the assemblies. The values of Φw

below the 0.05 threshold indicate significant recombination. The relationships were visualized by the construction of split graphs. A network representation of the evolutionary relationships between the isolates in this study was constructed on the concatenated sequences using the neighbor-net algorithm available in SplitsTree v4.

3. Results

3.1.Assembly characteristics of *Globisporangium* and *Pythium* genomes

The sequences of thirty-seven *Pythium* and *Globisporangium* genomes were generated using the Illumina sequencing system (Table 3-3). The assembled genomes differed in size ranging from 43 Mb to50 Mb. *Pythium mamillatum* (LEV1570) and '*Pythium vipa' '*(OOM1116 and OOM1117) isolates have bigger genomes than the rest of the isolates that are part of the *G. irregulare* complex (Table 3-2). All genomes shared a similar G+C content (53%), yielded 4,664 to 22,897 contigs with an N_{50} contig length ranging from 2.7 to 28.8 Kb (Table 3-3).

3.2.Genome completeness and ortologs selection

The completeness of the thirty-seven genomes was assessed using the BUSCO set of 100 single-copy, conserved genes among 26 stramenopiles genomes. The complete BUSCO genes among the genomes ranged from 88% to 100% of the 100 stramenopiles single-copy BUSCO genes. Genomes showed low numbers of duplicated genes, little fragmentation, and low levels of missing genes (Figure 3-3). After removing genes that were not present in all 37 assemblies, BUSCO analysis revealed 75 BUSCO genes that

were present and single-copy in the genome assemblies (Table 3-4). These BUSCO sequences were retained for the phylogenomic analysis.

3.3.Gene trees

The gene trees were compared by counting each topology's occurrence, and many different topologies on the NT gene trees and the AA gene trees were obtained. Later, the majority-rule consensus tree summarized the most frequent tree among the multiple gene trees. For the nucleotide and amino acid gene trees, the number of total trees was 7500 since, for every BUSCO gene, a tree with 100 replicates was generated. The number on each indicates how often the topology occurred, showing higher values in the internodes and lower values on the external branches (Figure 3-2). Those numbers indicated that those relationships did not occur in all gene trees. Thus there are disagreements between individual genes trees. However, the tree topology with the NT and AA alignments were similar. One clade within the *G. irregulare* isolates was subdivided into two subclades; the reference isolates clustered in one of each clade (Figure 3-2 A and B). The outgroup (*P. mamillatum*) and the isolates corresponding to P. vipa did not cluster with the rest of the isolates (Figure 3-2).

3.4.Species trees

Concatenation

Two maximum likelihood species trees were generated using the concatenated amino acid sequence alignments (39,352 sites) and nucleotide alignments (117,472 sites). The concatenated species tree based on the nucleotide alignments were constructed under the

TIM3+F+I+G4 substitution model. TIM3 is a transition model that assumes AC=CG, AT=GT, and unequal base frequencies. I + G accounts for a proportion of invariable sites and the discrete Gamma model (Yang 1994) with four rate categories. F is used by default because the model has unequal base frequencies. The tree clustered all *G. irregulare* complex in one well supported clade that could be divided into two subclades. The values on the main branch were the following: SH-alrt (100%), aBayes support (1), and the ultrafast bootstrap support (100%) (Figure 3-3). The subclades had well supported values. The first clade (colored orange) clustered isolates that were either *G. irregulare* or *G. cryptoirregulare*. The second clade, which clustered the remain *G. irregulare* isolates, could be subdivided into two subclades. The first subclade (colored pink) had high support (100/1/100) and contained 12 isolates (*G. irregulare* and *G. cryptoirregulare)* and the *G. cylindrosporum* (LEV4744) and *G. cryptoirregulare (*LEV4534) ex-type specimens. The second subclade (colored green) had high support as well (100/1/100) and contained 13 isolates (*G. irregulare* and *G. cryptoirregulare)* and the *G. irregulare* (LEV1481) type. *Pythium vipa* isolates formed a separated clade with high support (100/1/100) (Figure 3-3).

The concatenated species tree based on the amino acid alignments were constructed under the JTT+F+I+G4 substitution model. JTT is the general matrix model (Jones et al. 1992). The tree clustered the *G. irregulare* isolates in one well-supported clade (100/1/100) that could be divided into two subclades (Figure 3-4). The first subclade (colored pink) had moderate support (71.6/0.941/38) and contained ten isolates (*G. irregulare* and *G. cryptoirregulare)* and the *G. cylindrosporum* (LEV4744) and *G.*

cryptoirregulare (LEV4534) ex-type specimens. The second subclade (colored green) had low support (26.3/0.941/37) and contained 21 isolates (*G. irregulare and G. cryptoirregulare)* and the *G. irregulare* (LEV14810) type. *Pythium vipa* isolates formed a separated clade with high support (100/1/100) (Figure 3-4).

Overall, both trees clustered isolates differently; nevertheless, they helped to identify clustering (two main clades) that may prove to be two phylogenetic species with the coalescence approach.

Coalescence

The two coalescent-based species trees recovered a similar overall topology. The main well supported monophyletic *G. irregulare* clade was divided into two subclades (Figure 3-5A and 3-5B). In the nucleotide-based species tree, the first subclade (colored green) had high support (LPP=1) and contained 12 isolates (*G.irregulare* and *G. cryptoirregulare)* and the *G. irregulare* (LEV1481) type specimen. The second subclade (colored pink) had high support (LPP=1) and contained 19 isolates (*G. irregulare* and *G. cryptoirregulare)* and the *G. cryptoirregulare* (LEV1481) and the *G. cylindrosporum* (LEV4744) ex-type specimens (Figure 3-5A). *Pythium vipa* isolates formed a separated clade with high support (LPP =1).

In the amino acid-based species tree, the first subclade (colored green) had high support (LPP=1) and contained 13 isolates (*G. irregulare* and *G. cryptoirregulare)* and the *G. irregulare* (LEV1481) type specimen. The second subclade (colored pink) had high support (LPP=1) and contained 18 isolates (*G. irregulare* and *G. cryptoirregulare)*

and the *G. cryptoirregulare* (LEV1481) and the *G. cylindrosporum* (LEV4744) ex-type specimens (Figure 3-5B). *Pythium vipa* isolates formed a separated clade with high support (LPP =1).

The only disagreement in topology between both trees was the isolate LEV3046 that clustered in different subclades on each tree. Overall, high support values (LPP=1) separated the *G. irregulare* clade into the two subclades. However, both subclades had representatives of each species: *G. irregulare, G cryptoirregulare,* and *G. cylindrosporum.*

3.5.Concordance factors

In order to identify instances of significant discordance between gene trees and the consensus species tree within the *G. irregulare* complex, the gene concordance factor (gCF) and the site concordance factor (sCF) were measured (Figure 3-6 and Figure 3-7). The split, which corresponded to the monophyletic *G. irregulare* clade, had a LPP value of 1, a gCF of 94.7%, and a sCF of 90% in the nucleotide species tree (Figure 3-6). High LPP and concordance values were observed on the amino acid species tree (1/65.3/90.6) at the same split (Figure 3-7) as well. Those results mean that most single-locus trees contained this grouping, while most of the sites informative for this branch supported it. Similar results were observed in the branch that clustered the *P. vipa* isolates, a different species, and not part of the complex.

In the rest of the branches, the LLP and concordance factors were very different, particularly in the subclades within the complex. The clade that clustered *G. irregulare*

type (colored green), on the nucleotide-based species tree, had a gCF of 24% and a sCF of 51.3% (Figure 3-6). This means that only 18 (gCF_N = 18) trees were concordant with the branch and over half of the sites informative for this branch support it. It was also observed that three other single-locus trees supported a second resolution of that branch (gDF1 = 4%), and three trees supported a third resolution (gDF2 = 4%) (Table 3-5). On the other hand, the clade that clustered *G. cryptoirregulare* and *G. cylindrosporum* ex-types (colored green) had a gCF of 9.33% and a sCF of 39.37% (Figure 3-6). This means that only 7(gCF_N = 7) trees were concordant with the branch, and over a third of the sites informative for this branch support it. It was also found that 12 other single-locus trees supported a third resolution of that branch (gD2 = 16%) (Table 3-5).

In the case of these two subclades within *G. iregulare* complex in the amino acid-based species tree, the concordance values were lower. The clade that clustered *G. irregulare* type (colored green) had a gCF of 4% and a sCF of 47.44% (Figure 3-7). This means that only 3 (gCF_N = 3) trees were concordant with the branch and over almost a half of the sites informative for this branch support it. It was also observed that three other single locus trees supported a second resolution of that branch (gDF1 = 4%), and two supported a third resolution (gDF2 = 2.67%) (Table 3-6). On the other hand, the clade that clustered *G. cryptoirregulare* and *G. cylindrosporum* ex-types (colored green) had a gCF of 2.67% and a sCF of 45.12% (Figure 3-6). This means that only 2 (gCF_N = 2) trees were concordant with the branch and over almost half of the sites informative for this branch support it. It was also observed that one single locus tree supported a second

resolution of that branch (gD2 = 1.33%), and two trees supported a third resolution (gDF2 = 2.67%) (Table 3-6).

Overall, these results mean that the remaining trees, out of the 75 gene trees, had a topology that was not any one of the three possible arrangements around the two subclades (pink and green), which is a signal of noisy single-locus trees. Regardless of the high LPP values (LPP=1), the sCF and gCF values indicates that the data contain discordance around these two subclades. In conclusion, this data indicates that the species tree's resolution represents the relationships of the isolates analyzed but that there is a conflicting signal in the gene trees.

3.6. Recombination test

The PHI tests revealed significant evidence of recombination in the main *G. irregulare* clade (Φw= 0.0; Figure 3-7a) and the underlining subclades. Clade I, which contained the *G. irregulare* type, had a Φw = 0.0. (Figure 3-7b). Clade II-A, had a Φw = 2.165e-12. (Figure 3-7c). Clade II-B, which contained the *G. cryptoirregulare* and *G. cylindrosporum* ex-types, had a Φw = 0.0. (Figure 3-7c). Thus, genealogical exclusivity was not confirmed with the PHI test and the GCPSR model, which supports the hypothesis that the *G. irregulare, G. cryptoirregulare,* and *G. cylindrosporum* isolates analyzed represent one phylogenetic species. A phylogenetic network was used to observe the evolutionary relationships of the isolates under study since there is evidence of recombination and gene discordance (Figure 3-7d).

4. Discussion

Reconstructing the evolutionary relationship of species is a significant challenge in biology. Currently, few phylogenetic and genetic resources exist that enable the identification and characterization of *Pythium* and *Globisporangium* species (Schroeder et al. 2013), which are required for accurate species identification and are part of plant disease management. Many studies have been carried out to understand the relationships of *G. irregulare* using combinations of methods. Bar et al. (1997) separated *Pythium irregulare* Buisman (syn. *Globisporangium irregulare*) isolates into two groups based on 11 isozyme loci. Matsumoto et al. (1999, 2000) performed Random Amplified Polymorphic DNA (RAPD) and phylogenetic analyses using sequences of the ITS region and found the formation of four groups (I, II, III, IV). Similar results were observed by Lévesque and de Cock (2004) and by Garzon et al. (2005a, 2005b) using the ITS region and the cox I-II mitochondrial region.

G. irregulare is part of *Pythium* clade F (Lévesque and de Cock, 2004), has remarkable genetic diversity, and a genetic structure that forms clusters with significant bootstrap support. Based on its genetic structure, *G. irregulare* was proposed as a complex with worldwide distribution and broad host range (Garzon et al. 2007). The differentiation of two clades corresponding to Matsumoto et al. (2000) groups I and II lead to the description of *G. cryptoirregulare* (Garzon et al. 2007). *G. cylindrosporum* and *G. regulare* were also included in the complex. All of them represent a challenging group for identification because of its variable morphology and high levels of intraspecific genetic diversity (Garzon et al. 2005a, Spies et al. 2011).

In this study, the phylogenetic relationships of an international collection of 34 isolates part of the *G. irregulare* complex was analyzed using a phylogenomic approach and applying the Genealogical Concordance Phylogenetic Species Recognition (GCPSR). Based on the high support values and high values on the concordance factors in the species trees generated by concatenation and coalescence using 75 BUSCO genes, the isolates were grouped in a monophyletic clade representing one phylogenetic species (*G. irregulare*). Despite two well-supported groups (I and II) and other well supported branches were found within the *G. irregulare* clade, the lack of genealogical concordance under the GCPSR, and significant recombination evidence did not support the hypothesis of a complex of species.

Criteria to identify plant pathogenic oomycetes include morphological, biological, ecological, and phylogenetic species concepts (Schardl and Craven 2003). A phylogenetic species is defined as a population sharing evolutionary relevant characters derived from a common ancestor (Schardl and Craven 2003). Delineation of cryptic species such as the putative species in the *G. irregulare* complex is challenging since they have significant genetic diversity, and their species limits cannot be resolved based on ITS and *cox* I-II sequences alone or combined. Thus, our study made use of 75 BUSCO genes as markers for species delimitation. BUSCO genes are commonly used to quantify genome completeness (Waterhouse et al. 2013) and have been extended to other applications, including identifying markers for phylogenomics studies (Waterhouse et al. 2018).

This analysis used two approaches to infer the relationships within the species under study. The concatenation method concatenates multiple genes to estimate the species tree (Xi et al. 2014; Jian et al. 2014). A high support was found for the clade that clustered all the *G. irregulare* species and well-supported subclades under this approach. However, a comparison of the trees generated with the nucleotide and amino acid data matrices clustered isolates differently. The reference isolates clustered in two different groups, *G. irregulare* in group I and *G. cryptoirregulare + G cylindrosporum* in group II, and the rest of the isolates clustered randomly on each group. The approach of concatenating sequence data from multiple genes can lead to poor species discrimination (Kubatko and Degnan 2007) since it can mask discordance between individual gene trees and speciation events that can be misidentified. These is evidence of discordance between gene trees. Thus, to test the hypothesis that distinct clades were different species, a coalescent method was used to compare results.

Coalescent-based species delimitation methods incorporate the process of lineage sorting and the presence of incongruent genomic regions into phylogenetic estimations (Cartstens and Knowles 2007). To look for discrepancies in the species trees produced by concatenation, Astral-III (Chao et al. 2018) was used as the coalescent method. Astral infers the species tree by searching for the topology that agrees with the largest number of gene trees (Mirarab et al. 2014, 2015). Then, concordance factors were incorporated to estimate how well each gene, and each site of the alignments, agreed with the topology. The Astral species trees were congruent with the concatenated analysis, a high supported clade was observed, the *G.irregulare* clade (LPP=1) that was divided into two well-

supported subclades (I and II). Even with high support values on the branches in these two subclades, a high amount of discordance was confirmed based on the two concordance factors (gCF and sCF) analyzed. Those values reflected a disagreement between the species trees and the gene trees on the branches separating the clades. Also, significant recombination was detected among the subclades when applying the PHI test with the GCPSR model.

GCPSR compares more than one gene genealogy, and conflict among genealogies is likely due to recombination (Liu et al. 2016; Taylor et al. 2000). Boundaries in closely related taxa are difficult to resolve since genes can differ in their evolutionary histories (Liu et al. 2016; Steward et al. 2014). Many factors can affect the reconstruction of gene trees, including the amount of phylogenetic signal, recombination (Lanier and Knowles 2012), and biological reasons including incomplete lineage sorting (ILS) and introgression (Edwards 2009; Mallet et al. 2016). Thus, many processes can make gene trees and species trees discordant, masking true evolutionary relationships (Degnan and Rosemberg 2009).

When the grouping, support values, and concordance factors in the *P. 'vipa'* isolates was analyzed, it was found that most genes and sites support the identification of this clade as a single species. Globisporangium (*Pythium) 'vipa'* is an informally *Pythium* species associated with cavity spot on carrot in Norway (Hermansen et al. 2006). This species has a unique morphology, ITS sequence (Klemsdal et al. 2008), and was found to be similar to *G. irregulare* group IV sensu Matsumoto et al. (2000) (Weiland 2011). The species is part of clade F (Levesque and de Cock 2004), along with *G. irregulare* and

Pythium mamillatum used to root our trees. On the other hand, the node that clustered all *G irregulare* isolates had high LPP support and concordance factors, meaning that most gene trees and the alignment sites support the monophyly of *G. irregulare*. Comparisons on the geographic origin, mefenoxam test, and host of the isolates under study did not reveal any pattern that distinguished them.

This study did not support the presence of the three phylogenetic species (*G. irregulare* s.s. *G. cylindrosporum* and *G. cryptoirregulare*) previously described in the *G. irregulare* species complex in agreement with the findings of Spies et al. (2011) and Nguyen et al. (2018), and suggest instead that *G. irregulare* may be a highly diverse monophyletic species with significant intraspecific structure. The clades that could represent separate species revealed genetic recombination, indicating they do not have evolutionary independence. Data from *in vitro* crosses and with field *Globisporangium* populations indicates that outcrosses can occur in nature (Harvey et al. 2001), even in homothallic species, such as *G. irregulare*. This implies that the species has a mechanism for genetic variation (Martin 2009) and that can explain the incongruence between nuclear and mitochondrial gene regions (Spies et al. 2011). Although this study provides evidence of recombination, its role as a discordance source will require further research.

The methodology presented here has shown to be useful for delimiting species with closely related taxa and morphologically indistinguishable. However, apart from molecular phylogenetics, morphology, ecology, and pathogenicity should also be assessed for *Pythium* and *Globisporangium* species. The inclusion of representatives of *G. irregulare* III and IV groups and larger samples is recommended, including the known

clades as well as isolates that form independent and well supported branches, to investigate further the relevance of intraspecific variation of this important plant pathogenic species.

5. Acknowledgments

The assistance by Dr. Hai Nguyen from the Ottawa Research and Development Center, Agriculture and Agri-Food in Canada and his team is greatly appreciated, for providing funding and some of the genomes used in this study. Some of the computational analysis done in this project was performed at the Oklahoma State University High-Performance Computing Center, supported in part through the National Science Foundation grant OAC-1126330.

Tables

Table 3-1. Information of the isolates included in this study, their host, year of isolation, mefenoxam sensitivity, Illumina technology used for sequencing and source.

Study code	Strain code	Status[d]	Genus	Species	Host	Year	Mefenoxam[e]	Location	Sequencing technology	Source
OOM1101[a]	73028-97	-	*Globisporangium*	*irregulare* s.s.	*Euphorbia pulcherrima*	1997	S	PA,USA	Nextseq	Kim
OOM1102[a]	13-30	-	*Globisporangium*	*irregulare*	*Pelargonium*		S	OK, USA	Nextseq	Gary Moorman/Elizabeth Hudgins
OOM1103[a]	C20	-	*Globisporangium*	*irregulare*	*Geranium*	2009	-	NY, USA	Nextseq	Margery Daughtrey
OOM1101[a]	C216	-	*Globisporangium*	*irregulare*	*Zonal geranium*	2011	-	NY, USA	Nextseq	Margery Daughtrey
OOM1105[a]	C246	-	*Globisporangium*	*irregulare*	*Mixed gerbera*	2011	-	NY, USA	Nextseq	Margery Daughtrey
OOM1106[a]	81774-98	-	*Globisporangium*	*cryptoirregulare*	*Euphorbia pulcherrima*	1998	S	PA, USA	Nextseq	Kim
OOM1107[a]	C205	-	*Globisporangium*	*cryptoirregulare*	*Zonal geranium*	2011	-	NY, USA	Nextseq	Margery Daughtrey
OOM1108[a]	C206	-	*Globisporangium*	*cryptoirregulare*	*Zonal geranium*	2011	-	NY, USA	Nextseq	Margery Daughtrey
OOM1109[a]	13-22	-	*Globisporangium*	*cryptoirregulare*	*Pelargonium*	-	R	CA, USA	Nextseq	Gary Moorman/Daryl Thomas
OOM1110[a]	Pl40	-	*Globisporangium*	*cryptoirregulare*	*Euphorbia pulcherrima*	1999	-	PA, USA	Nextseq	Moorman
OOM1111[a]	C228	-	*Globisporangium*	*cryptoirregulare*	*Gerbera*	2011	R	NY, USA	Nextseq	Margery Daughtrey
OOM1112[a]	C223	-	*Globisporangium*	*cryptoirregulare*	*Zonal geranium*	2011	R	NY, USA	Nextseq	Margery Daughtrey
OOM1113[a]	42143-99	-	*Globisporangium*	*cryptoirregulare*	*Pelargonium*	1999	S	PA, USA	Nextseq	Kim
OOM1115[a]	P50/99-1880	-	*Globisporangium*	*cryptoirregulare*	*Euphorbia pulcherrima*	1999	S	PA, USA	Nextseq	Gary Moorman
OOM1116[a]	Pl20	-	*Pythium*	*vipa*	*Soil*	2008	-	OR, USA	Nextseq	Jerry Weiland
OOM1117[a]	Pl21	-	*Pythium*	*vipa*	*Soil*	2008	-	OR, USA	Nextseq	Jerry Weiland
LEV4910d[b]	CBS 461.48, DAOMC BR969	-	*Globisporangium*	*irregulare*	-	-	-	South Australia	Nextseq	Hai Nguyen
LEV4840a[b]	DAOMC BR946	-	*Globisporangium*	*irregulare*	*Alfalfa feeder roots*	-	-	CA, USA	Nextseq	Hai Nguyen
LEV3105d[b]	NBRC 100110, DAOMC 239172	-	*Globisporangium*	*irregulare*	*Kidney bean*	1974	-	Hokkaido, Japan	Nextseq	Hai Nguyen
LEV4761a[b]	DAOMC BR722	-	*Globisporangium*	*irregulare*	Soil	1992	-	British Columbia, Canada	Nextseq	Hai Nguyen

Study code	Strain code	Status[d]	Genus	Species	Host	Year	Mefenoxam[e]	Location	Sequencing technology	Source
LEV4845b[b]	DAOMC BR958	-	*Globisporangium*	*irregulare*	*Cyperus sp.* (sedge), *root*	1970	-	Auckland, New Zealand,	Nextseq	Hai Nguyen
LEV4863a[b]	IMI 344526, DAOMC BR1025	-	*Globisporangium*	*irregulare*	*Lactuca sativa*	1991	-	Essex, United Kingdom,	Nextseq	Hai Nguyen
LEV4866a[b]	DAOMC BR1051	-	*Globisporangium*	*irregulare*	*Cucumber*	-	-	British Columbia, Canada	Nextseq	Hai Nguyen
LEV4874c[b]	DAOMC BR426	-	*Globisporangium*	*irregulare*	*Rhododendron sp., roots*	1982	-	Alberta, Canada	Nextseq	Hai Nguyen
LEV4917a[b]	DAOMC BR995	-	*Globisporangium*	*irregulare*	*Aspalanthus linearis*	-	-	Clan William, South Africa	Nextseq	Hai Nguyen
LEV 1861[b]	DAOMC BR1001	-	*Globisporangium*	*irregulare*	*Lucerne sp.*	-	-	George, South Africa	MiSeq	Hai Nguyen
LEV 1862[b]	DAOMC BR1013	-	*Globisporangium*	*irregulare*	*Wheat*	-	-	Australia	MiSeq	Hai Nguyen
LEV 3046[b]	DAOMC 232336	-	*Globisporangium*	*irregulare*	*Unknown*	-	-	Quebec, Canada	MiSeq	Hai Nguyen
LEV 4775[b]	DAOMC BR818	-	*Globisporangium*	*irregulare*	*Wheat roots*	-	-	Transvaal, South Africa	MiSeq	Hai Nguyen
LEV 4859[b]	DAOMC BR1019	-	*Globisporangium*	*irregulare*	*Lupin*	-	-	New South Wales, Australia	MiSeq	Hai Nguyen
LEV 1481[b]	CBS 250.28	PN	*Globisporangium*	*irregulare*	*Phaseolus vulgaris*	-	-	Netherlands	MiSeq	Hai Nguyen
LEV 4534[b]	CBS 118731	T	*Globisporangium*	*cryptoirregulare*	*Euphorbia pulcherima*	1999	-	PA, USA	Nextseq	Hai Nguyen
LEV3045c[b]	DAOMC 232335	-	*Globisporangium*	*cylindrosporum*	*Basil*	-	-	Quebec, Canada	Nextseq	Hai Nguyen
LEV4852a[b]	IMI 358071, DAOMC BR1011	-	*Globisporangium*	*cylindrosporum*	*Cucumis sativus*	1994	-	Norway	Nextseq	Hai Nguyen
LEV 4867[b]	DAOMC BR1052	-	*Globisporangium*	*cylindrosporum*	*Cucumis sativus*	-	-	British Columbia, Canada,	Nextseq	Hai Nguyen
LEV 4744[b]	CBS 218.94	T	*Globisporangium*	*cylindrosporum*	Soil	-	-	Germany	Nextseq	Hai Nguyen
LEV 1570[b,c]	CBS 251.28, DAOMC BR648	PN	*Pythium*	*mamillatum*	*Beta vulgaris*	-	-	Netherlands	Nextseq	Hai Nguyen

[a] = isolates from Dr. Garzon's DNA collection.

[b] = isolates provided by Dr. Hai Nguyen from the Levesque (LEV) legacy collection part of the Canadian Collection of Fungal Cultures (CCFC, DAOMC).

[c] = Outgroup.

[d] = Status: T, ex-type strain; PN, strain used for description in the monograph of van der Plaats-Niterink (1981).

[e] = Mefenoxam test: R: resistant; S: sensitive.

Table 3-2. OrthoDB v10 database relevant to oomycete genomics used in this study.

Database	Number of genes	Default Augustus training set	Number of reference genomes	Stramenopiles reference genomes
Stramenopiles	100	*Toxoplasma gondii*	26	*Phytophthora parasitica* P1569, *Thraustotheca clavata, Blastocystis hominis, Albugo candida, Thalassiosira oceanica* CCMP1005, *Achlya hypogyna, Ectocarpus siliculosus, Blastocystis sp.* ATCC 50177/Nand II, *Aphanomyces astaci, Thalassiosira pseudonana CCMP1335, Phytophthora sojae, Plasmopara halstedii, Phytophthora palmivora var. palmivora, Phytophthora nicotianae, Blastocystis sp. subtype 4, Nannochloropsis gaditana CCMP526, Phaeodactylum tricornutum CCAP 1055/1, Aureococcus anophagefferens, Saprolegnia diclina VS20, Fistulifera solaris, Phytophthora infestans T30-4, Fragilariopsis cylindrus CCMP1102, Phytophthora megakarya, Aphanomyces invadans, Nannochloropsis gaditana, Saprolegnia parasitica* CBS 223.65, *Pythium insidiosum*

Table 3-3. Genome assembly statistics of *Globisporangium and Pythium* isolates. Description of each genome includes species, genome size, number of contigs, G+C content, largest contig, N_{50}, L_{50}.

Study code	Genus	Species	Estimated genome size (bp)	Number contigs	G+C content (%)	Largest contig (bp)	N_{50}[a]	L_{50}[b]
OOM1101	*Globisporangium*	*irregulare*	*45,862,580*	6,505	53.56	194,719	25,635	465
OOM1102	*Globisporangium*	*irregulare*	*43,987,540*	*7,333*	53.68	150,871	15,801	736
OOM1103	*Globisporangium*	*irregulare*	*44,245,923*	*5420*	53.66	260,167	27,502	420
OOM1104	*Globisporangium*	*irregulare*	*43,821,643*	*19,512*	53.66	123,721	3,450	3195
OMM1105	*Globisporangium*	*irregulare*	*43,459,689*	*7,106*	53.75	97,873	15,734	776
OOM1106	*Globisporangium*	*cryptoirregulare*	*45,031,153*	10,784	53.83	145,412	10,199	1,198
OOM1107	*Globisporangium*	*cryptoirregulare*	*43,888,350*	9,880	53.7	121,804	9,799	1,209
OOM1108	*Globisporangium*	*cryptoirregulare*	*43,799,775*	10,460	53.7	121,804	8,807	1,327
OOM1109	*Globisporangium*	*cryptoirregulare*	*43,316,896*	19,390	53.75	161,612	3,504	3,123
OOM1110	*Globisporangium*	*cryptoirregulare*	*43,912,875*	14,878	53.72	85,767	5,409	2,132
OOM1111	*Globisporangium*	*cryptoirregulare*	*43,346,792*	22,057	53.7	116,553	2,897	3,844
OOM1112	*Globisporangium*	*cryptoirregulare*	*43,737,747*	19,852	53.66	116,513	3,406	3,309
OOM1113	*Globisporangium*	*cryptoirregulare*	*46,719,806*	22,580	53.59	187,569	3,000	3,573
OOM1115	*Globisporangium*	*cryptoirregulare*	*43,849,277*	17,892	53.7	164,089	4,018	2,621
OOM1116	*Pythium*	*vipa*	*50,909,982*	7,607	52.85	162,708	20,424	692
OOM1117	*Pythium*	*vipa*	*48,642,797*	6,529	52.9	162,442	20,641	664
LEV4910d	*Globisporangium*	*irregulare*	*46,752,289*	9,360	53.62	146,239	15,415	713
LEV4840a	*Globisporangium*	*irregulare*	*45,902,764*	6,819	53.61	227,266	26,200	444
LEV3105d	*Globisporangium*	*irregulare*	*43,924,420*	5,533	53.7	258,104	28,383	407
LEV4761a	*Globisporangium*	*irregulare*	*43,497,442*	19,700	53.72	117,541	3,425	3,126
LEV4845b	*Globisporangium*	*irregulare*	*43,919,000*	13,388	53.71	293,519	7,809	888
LEV4863a	*Globisporangium*	*irregulare*	*43,568,599*	21,651	53.7	117,509	2,963	3,635
LEV4866a	*Globisporangium*	*irregulare*	*43,629,689*	19,904	53.71	119,499	3,350	3,002

Study code	Genus	Species	Estimated genome size (bp)	Number contigs	G+C content (%)	Largest contig (bp)	N_{50}[a]	L_{50}[b]
LEV4874c	*Globisporangium*	*irregulare*	*43,262,060*	22,897	53.7	101,932	2,727	4,161
LEV4917a	*Globisporangium*	*irregulare*	*43,964,472*	5,236	53.71	257,377	28,857	387
LEV1861	*Globisporangium*	*irregulare*	*46,569,560*	7,473	53.61	150,889	20,315	614
LEV1862	*Globisporangium*	*irregulare*	*47,599,756*	7,291	53.58	192,602	23,120	522
LEV3046	*Globisporangium*	*irregulare*	*45,931,194*	5,156	53.61	182,340	25,334	469
LEV4775	*Globisporangium*	*irregulare*	*46,624,308*	7,675	53.6	141,932	19,343	632
LEV4859	*Globisporangium*	*irregulare*	*47,883,205*	7,420	53.56	192,180	22,435	543
LEV 1481	*Globisporangium*	*irregulare*	*44,840,884*	4,817	53.69	138,122	20,301	584
LEV 4534	*Globisporangium*	*cryptoirregulare*	*43,038,241*	19,037	53.72	193,149	3,579	2,837
LEV3045c	*Globisporangium*	*cylindrosporum*	*43,323,513*	21,162	53.71	162,643	3,065	3,406
LEV4852a	*Globisporangium*	*cylindrosporum*	*43,439,690*	20,281	53.72	153,261	3,287	3,061
LEV4867	*Globisporangium*	*cylindrosporum*	*45,135,118*	4,664	53.68	219,870	28,447	419
LEV 4744	*Globisporangium*	*cylindrosporum*	*45,387,642*	14,774	53.68	117,183	5,873	1,841
LEV 1570*	*Pythium*	*mamillatum*	*46,234,300*	6,905	53.69	250,550	25,693	481

[a] = length such that sequence contigs of this length or longer include half the bases of the assembly.

[b] = number of sequence contigs that are longer than, or equal to, the N50 length.

Table 3-4. List of the 75 BUSCOS genes found in the thirty-seven assemblies used in this study. Gene information was obtained from the OrthoDB database.

ID	GENE
26307at33634	Deoxyuridine 5'-triphosphate nucleotidohydrolase
3405at33634	Pre-mRNA-processing factor 6
6989at33634	Pescadillo homolog
15351at33634	Proteasome subunit alpha type
7485at33634	RNA helicase
3892at33634	SDA1 family protein
5960at33634	tRNA uridine 5-carboxymethylaminomethyl modification enzyme GidA
28151at33634	YbaK/aminoacyl-tRNA synthetase-associated domain
26258at33634	Ribonucleoprotein-associated protein
723at33634	valyl-tRNA synthetase
1102at33634	DNA-directed RNA polymerase subunit
9286at33634	Mitochondrial-processing peptidase subunit beta
20769at33634	Adenylate kinase isoenzyme 6 homolog
12684at33634	pap-associated domain-containing protein 5-like
21097at33634	Signal recognition particle receptor subunit alpha
31838at33634	predicted protein
17057at33634	Pumilio homology domain family member 6
25187at33634	H/ACA ribonucleoprotein complex subunit 2
17622at33634	Histone acetyltransferase
8760at33634	MutY A/G-specific DNA repair glycosylase
1342at33634	Coatomer subunit beta
971at33634	Isoleucine-tRNA ligase
1326at33634	Regulator of nonsense transcripts 2
13102at33634	Large subunit GTPase 1
23030at33634	Ribosomal RNA-processing protein
12562at33634	WD40-repeat-containing domain
4148at33634	Eukaryotic translation initiation factor 5B
13832at33634	protein phosphatase
6981at33634	Exportin-1
302at33634	rRNA biogenesis protein rrp5
5144at33634	Signal recognition particle 54 kDa protein
20044at33634	DNA mismatch repair protein
21875at33634	Deoxyribonuclease TATDN1-like protein
12069at33634	Cytosolic Fe-S cluster assembly factor NUBP1 homolog
13819at33634	ubiquitin-activating enzyme E1 C
22080at33634	60S ribosome subunit biogenesis protein NIP7
6684at33634	T-complex protein 1 subunit alpha
13476at33634	Proteasome subunit beta type
22950at33634	Proteasome subunit alpha type
4406at33634	WD40-repeat-containing domain

ID	GENE
13537at33634	Methyltransferase WBSCR22
18432at33634	regulatory proteasome non-atpase subunit 10
4166at33634	EF-hand domain
26441at33634	LSM domain, eukaryotic/archaea-type
5145at33634	pre-mRNA-processing factor 17
15242at33634	Proteasome subunit beta type
5074at33634	Glycerol kinase
25189at33634	Mitochondrial import inner membrane translocase subunit Tim17, putative
19836at33634	Cactin, central domain
8842at33634	26S Proteasome regulatory subunit 6A
27268at33634	BUD32 protein kinase
693at33634	DNA-directed RNA polymerase, subunit 2, hybrid-binding domain
10425at33634	H/ACA ribonucleoprotein complex subunit 4
5578at33634	Eukaryotic translation initiation factor 2 subunit 3
15520at33634	Flap endonuclease 1
27410at33634	Protein RER1
2880at33634	AP complex subunit beta
27411at33634	GrpE nucleotide exchange factor
27409at33634	Protein tyrosine phosphatase type IVA protein 1
15177at33634	Eukaryotic translation initiation factor 3 subunit I
18429at33634	Cyclin
12626at33634	MFS transporter, PAT family, solute carrier family 33 (acetyl-CoA transportor), member 1
12363at33634	Coatomer subunit epsilon
28445at33634	nuclear transport factor 2
12249at33634	Methylthiotransferase, N-terminal
9413at33634	DNA primase
16248at33634	Impact, N-terminal
24919at33634	DNA-directed RNA polymerase subunit
9731at33634	AP-3 complex subunit delta
15216at33634	Pantoate-beta-alanine ligase
10123at33634	Gdp-man
6366at33634	Inosine-5'-monophosphate dehydrogenase
15969at33634	Calreticulin
9939at33634	Glycerol-3-phosphate dehydrogenase
18615at33634	Rab proteins geranylgeranyltransferase component A

Table 3-5. Concordance and discordance factors for the coalescent-based species tree using the nucleotide data matrix.

ID	Clade	gCF	gCF_N	gDF1	gDF1_N	gDF2	gDF2_N	gDFP	gDFP_N	gN	sCF	sCF_N	sDF1	sDF1_N	sDF2	sDF2_N	sN	Label	Length
39		NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA		0.0000
40		29.33	22	6.67	5	0	0	64	48	75	65.43	198.52	23.63	76.59	10.94	37.14	312.25	1	0.4977
41		9.33	7	2.67	2	6.67	5	81.33	61	75	46.66	93.4	18.96	31.56	34.37	71.55	196.51	1	0.2004
42		17.33	13	1.33	1	1.33	1	80	60	75	70.97	204.66	8.33	23.43	20.7	83.89	311.98	1	0.6612
43		70.67	53	0	0	1.33	1	28	21	75	56.83	115.54	12.97	29.62	30.2	110.89	256.05	1	0.8941
44		16	12	14.67	11	8	6	61.33	46	75	38.97	97.62	40.94	73.1	20.09	32.34	203.06	1	0.1464
45		18.67	14	6.67	5	0	0	74.67	56	75	49.53	167.47	38.32	122.29	12.16	36.99	326.75	1	0.1660
46		81.33	61	0	0	0	0	18.67	14	75	100	248.01	0	0	0	0	248.01	1	1.6454
47		16	12	6.67	5	8	6	69.33	52	75	29.06	78.87	36.32	94.31	34.62	100.35	273.53	1	0.0420
48		25.33	19	10.67	8	13.33	10	50.67	38	75	54.18	150.99	29.05	88.35	16.78	47.8	287.14	1	0.2126
49		22.67	17	14.67	11	2.67	2	60	45	75	78.62	182.75	17.01	34.9	4.38	9.25	226.9	1	0.3916
50		80	60	5.33	4	4	3	10.67	8	75	100	118.75	0	0	0	0	118.75	1	1.0664
51	*G. iregulare* clade	24	18	4	3	4	3	68	51	75	51.29	206.32	35.56	140.98	13.14	61.14	408.44	1	0.2386
52	Main *G. irregulare* complex clade	94.67	71	0	0	0	0	5.33	4	75	89.99	1532.16	4.37	74.35	5.63	95.63	1702.14	1	1.6127
53		100	75	0	0	0	0	0	0	75	99.73	5448.86	0.09	4.66	0.18	10.01	5463.53	1	8.5173
54	*G. cyptoirregulare* clade	9.33	7	0	0	16	12	74.67	56	75	39.37	175.97	17.48	70.1	43.15	181.51	427.58	1	0.1401
55		9.33	7	0	0	1.33	1	89.33	67	75	50.54	269.72	7.16	43.06	42.31	219.34	532.12	1	0.2837
56		5.33	4	0	0	0	0	94.67	71	75	74.91	183.67	17.17	40.26	7.93	18.09	242.02	1	0.5032
57		12	9	0	0	4	3	84	63	75	44.8	69.89	34.58	54.99	20.62	31.74	156.62	1	0.1486
58		21.33	16	6.67	5	6.67	5	65.33	49	75	55.66	89.96	22.13	35.2	22.21	35.45	160.61	1	0.2086
59		17.33	13	0	0	13.33	10	69.33	52	75	58.03	79.28	13.03	15.91	28.94	38.59	133.78	1	0.2588
60		30.67	23	26.67	20	1.33	1	41.33	31	75	56.53	63.71	41.45	46.69	2.01	2.21	112.61	1	0.2266
61		1.33	1	5.33	4	0	0	93.33	70	75	35.11	124.14	36.25	140.04	28.64	115.76	379.94	1	0.0692
62		2.67	2	9.33	7	0	0	88	66	75	47.66	229.45	30.75	139.08	21.58	95.75	464.28	1	0.0396
63		1.33	1	0	0	0	0	98.67	74	75	33.86	157.86	28.51	135.09	37.63	177.91	470.86	1	0.0642
64		5.33	4	0	0	0	0	94.67	71	75	33.89	144.64	37.24	155.12	28.87	119.46	419.22	1	0.0562
65		0	0	0	0	0	0	100	75	75	50.74	293.99	24.94	124.56	24.31	124.3	542.85	0.99	0.0279
66		2.67	2	0	0	0	0	97.33	73	75	27.1	145.13	49.5	285.92	23.39	126.75	557.8	1	0.0838
67		0	0	0	0	0	0	100	75	75	33.24	168.63	26.31	129.05	40.45	220.01	517.69	0.77	0.0156
68		8	6	0	0	0	0	92	69	75	19.97	117.5	62.96	404.38	17.08	101.46	623.34	1	0.0869
69		0	0	1.33	1	0	0	98.67	74	75	47.2	262.32	26.59	129.48	26.21	119.27	511.07	0.99	0.0292
70		0	0	0	0	0	0	100	75	75	30.11	121.12	43.22	202.07	26.67	104.89	428.08	0.93	0.0265
71		2.67	2	0	0	0	0	97.33	73	75	43.9	204.6	27.65	108.48	28.45	110.29	423.37	1	0.0335
72		1.33	1	0	0	0	0	98.67	74	75	21.53	125.6	19.61	114.93	58.86	373.05	613.58	0.82	0.0123

ID	Clade	gCF	gCF_N	gDF1	gDF1_N	gDF2	gDF2_N	gDFP	gDFP_N	gN	sCF	sCF_N	sDF1	sDF1_N	sDF2	sDF2_N	sN	Label	Length
73		5.33	4	2.67	2	4	3	88	66	75	61.15	379.9	17.97	102.43	20.88	117.82	600.15	1	0.0779

ID: Branch ID - clade
gCF: Gene concordance factor (=gCF_N/gN %)
gCF_N: Number of trees concordant with the branch
gDF1: Gene discordance factor for NNI-1 branch (=gDF1_N/gN %)
gDF1_N: Number of trees concordant with NNI-1 branch
gDF2: Gene discordance factor for NNI-2 branch (=gDF2_N/gN %)
gDF2_N: Number of trees concordant with NNI-2 branch
gDFP: Gene discordance factor due to polyphyly (=gDFP_N/gN %)
gDFP_N: Number of trees decisive but discordant due to polyphyly
gN: Number of trees decisive for the branch
sCF: Site concordance factor averaged over 100 quartets (=sCF_N/sN %)
sCF_N: sCF in absolute number of sites
sDF1: Site discordance factor for alternative quartet 1 (=sDF1_N/sN %)
sDF1_N: sDF1 in absolute number of sites
sDF2: Site discordance factor for alternative quartet 2 (=sDF2_N/sN %)
sDF2_N: sDF2 in absolute number of sites
sN: Number of informative sites averaged over 100 quartets
Label: Existing branch label -LPP
Length: Branch length

Table 3-6. Concordance and discordance factors for the coalescent-based species tree using the amino acid data matrix.

ID	Clade	gCF	gCF_N	gDF1	gDF1_N	gDF2	gDF2_N	gDFP	gDFP_N	gN	sCF	sCF_N	sDF1	sDF1_N	sDF2	sDF2_N	sN	Label	Length
39		NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA		0
40		16	12	2.67	2	1.33	1	80	60	75	67.8	19.68	21.18	6.38	11.03	3.38	29.44	1	0.352306
41		1.33	1	4	3	0	0	94.67	71	75	39.4	10.02	38.91	9.75	21.69	5.39	25.16	1	0.05126
42		1.33	1	0	0	0	0	98.67	74	75	69.99	29.98	7.37	2.55	22.64	8.59	41.12	1	0.145223
43		2.67	2	0	0	0	0	97.33	73	75	45.03	27.77	27.81	23.23	27.16	17.01	68.01	1	0.05747
44		2.67	2	1.33	1	0	0	96	72	75	32.92	20.62	45.41	36.62	21.67	12.39	69.63	1	0.080365
45		0	0	0	0	0	0	100	75	75	39.75	36.05	33.37	27.82	26.88	18.09	81.96	1	0.05421
46		0	0	0	0	0	0	100	75	75	36.71	27.22	28.28	19.04	35.02	25.47	71.73	1	0.045582
47		0	0	0	0	6.67	5	93.33	70	75	34.49	21.14	28.46	17.27	37.05	25.38	63.79	1	0.035425
48		0	0	0	0	0	0	100	75	75	59.39	100.19	15.84	17.82	24.76	23.37	141.38	0.97	0.019808
49		1.33	1	0	0	0	0	98.67	74	75	30.14	26.7	52.05	78.28	17.81	13.38	118.36	1	0.048884
50		5.33	4	2.67	2	1.33	1	90.67	68	75	15.54	15.87	8.46	8.56	76	133.76	158.19	0.79	0.011593
51		0	0	0	0	0	0	100	75	75	44.31	61.61	29.46	31.09	26.23	18.95	111.65	0.76	0.010378
52		0	0	1.33	1	0	0	98.67	74	75	26.92	20.35	26.11	19.46	46.97	60.87	100.68	1	0.055954
53		1.33	1	0	0	0	0	98.67	74	75	17.18	25.45	10.44	15.6	72.37	124.24	165.29	1	0.066663
54		0	0	1.33	1	0	0	98.67	74	75	53.27	75.33	23.46	18.48	23.27	19.2	113.01	0.49	0.00971
55		8	6	0	0	1.33	1	90.67	68	75	55.41	27.38	19.57	9.68	25.03	12.36	49.42	1	0.045759
56		2.67	2	0	0	0	0	97.33	73	75	51.96	25.06	22.3	10.28	25.74	18.69	54.03	1	0.101731
57	Main *G. irregulare* complex clade	65.33	49	0	0	0	0	34.67	26	75	90.63	199.48	4.52	9.88	4.85	10.66	220.02	1	0.842846
58		90.67	68	1.33	1	0	0	8	6	75	99.69	805.39	0	0	0.31	2.5	807.89	1	2.26571
59	*G. iregulare* clade	4	3	4	3	2.67	2	89.33	67	75	47.44	24.3	14.4	8.53	38.16	19.61	52.44	1	0.103679
60		5.33	4	2.67	2	9.33	7	82.67	62	75	38.53	24.63	12.03	7.11	49.44	31.94	63.68	1	0.113458
61		6.67	5	1.33	1	1.33	1	90.67	68	75	67.76	38.66	5.47	3.94	26.77	16.01	58.61	1	0.131087
62	*G. cyptoirregulare* clade	2.67	2	1.33	1	2.67	2	93.33	70	75	45.12	13.45	15.3	3.55	39.58	19.15	36.15	1	0.058115
63		6.67	5	0	0	1.33	1	92	69	75	66.19	31.69	22.96	20.12	10.85	3.83	55.64	1	0.171902
64		1.33	1	9.33	7	6.67	5	82.67	62	75	29.86	7.87	32.91	22.28	37.24	11.23	41.38	0.59	0.017751
65		24	18	1.33	1	6.67	5	68	51	75	45.52	15.51	0.99	0.27	53.49	43.98	59.76	1	0.186456
66		9.33	7	2.67	2	0	0	88	66	75	49.94	31.79	31.8	15.92	18.26	9.21	56.92	1	0.073217
67		14.67	11	6.67	5	4	3	74.67	56	75	48.4	24.8	25.71	23.79	25.89	13.35	61.94	1	0.087717
68		4	3	1.33	1	4	3	90.67	68	75	41.18	17.08	26.83	10.5	31.99	13.27	40.85	0.95	0.017982
69		42.67	32	0	0	0	0	57.33	43	75	100	38.05	0	0	0	0	38.05	1	0.397147
70		12	9	4	3	1.33	1	82.67	62	75	77.53	23.14	8.56	2.45	13.92	4	29.59	1	0.201838
71		40	30	4	3	5.33	4	50.67	38	75	100	31.43	0	0	0	0	31.43	1	0.364146
72		5.33	4	1.33	1	0	0	93.33	70	75	72.92	15.72	18.34	4.1	8.74	1.76	21.58	1	0.137744

ID	Clade	gCF	gCF_N	gDF1	gDF1_N	gDF2	gDF2_N	gDFP	gDFP_N	gN	sCF	sCF_N	sDF1	sDF1_N	sDF2	sDF2_N	sN	Label	Length
73		8	6	1.33	1	9.33	7	81.33	61	75	40.66	6.5	2.7	0.46	56.64	8.84	15.8	1	0.154883

ID: Branch ID - clade
gCF: Gene concordance factor (=gCF_N/gN %)
gCF_N: Number of trees concordant with the branch
gDF1: Gene discordance factor for NNI-1 branch (=gDF1_N/gN %)
gDF1_N: Number of trees concordant with NNI-1 branch
gDF2: Gene discordance factor for NNI-2 branch (=gDF2_N/gN %)
gDF2_N: Number of trees concordant with NNI-2 branch
gDFP: Gene discordance factor due to polyphyly (=gDFP_N/gN %)
gDFP_N: Number of trees decisive but discordant due to polyphyly
gN: Number of trees decisive for the branch
sCF: Site concordance factor averaged over 100 quartets (=sCF_N/sN %)
sCF_N: sCF in absolute number of sites
sDF1: Site discordance factor for alternative quartet 1 (=sDF1_N/sN %)
sDF1_N: sDF1 in absolute number of sites
sDF2: Site discordance factor for alternative quartet 2 (=sDF2_N/sN %)
sDF2_N: sDF2 in absolute number of sites
sN: Number of informative sites averaged over 100 quartets
Label: Existing branch label -LPP
Length: Branch length

Figures

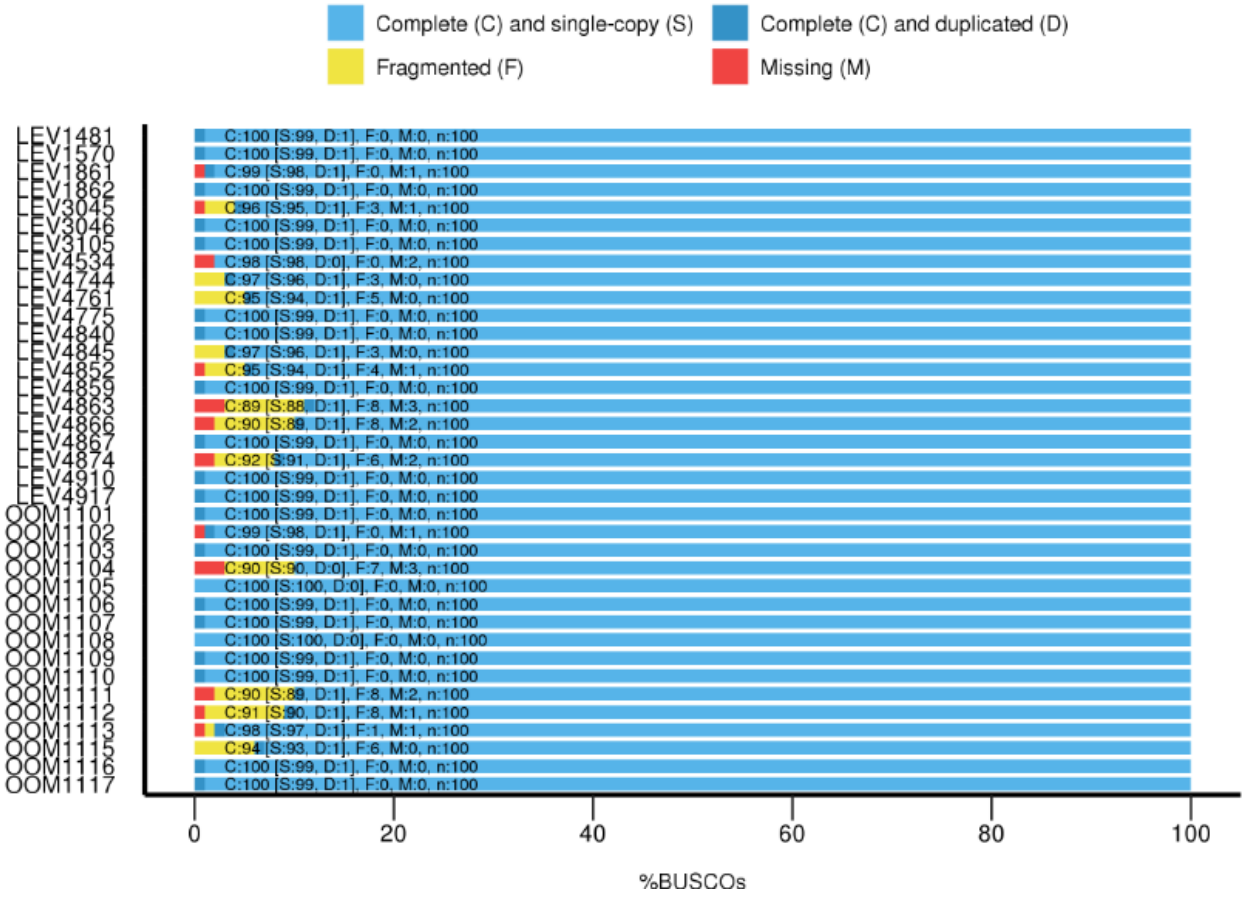


Figure 3-1. Genomic quality assesment of the thirty-seven *Globisporangium* and *Pythium* genomes using the stramenopiles_odb10 dataset. The bar plot next to each species indicates the fractions of BUSCO genes that are present or missing in each genome. “Complete”: fraction of single-copy; “Duplicated”: fraction of multiple-copy; “Fragmented”: fraction of genes with a partial sequence; “Missing”: fraction of genes not found in the genome.

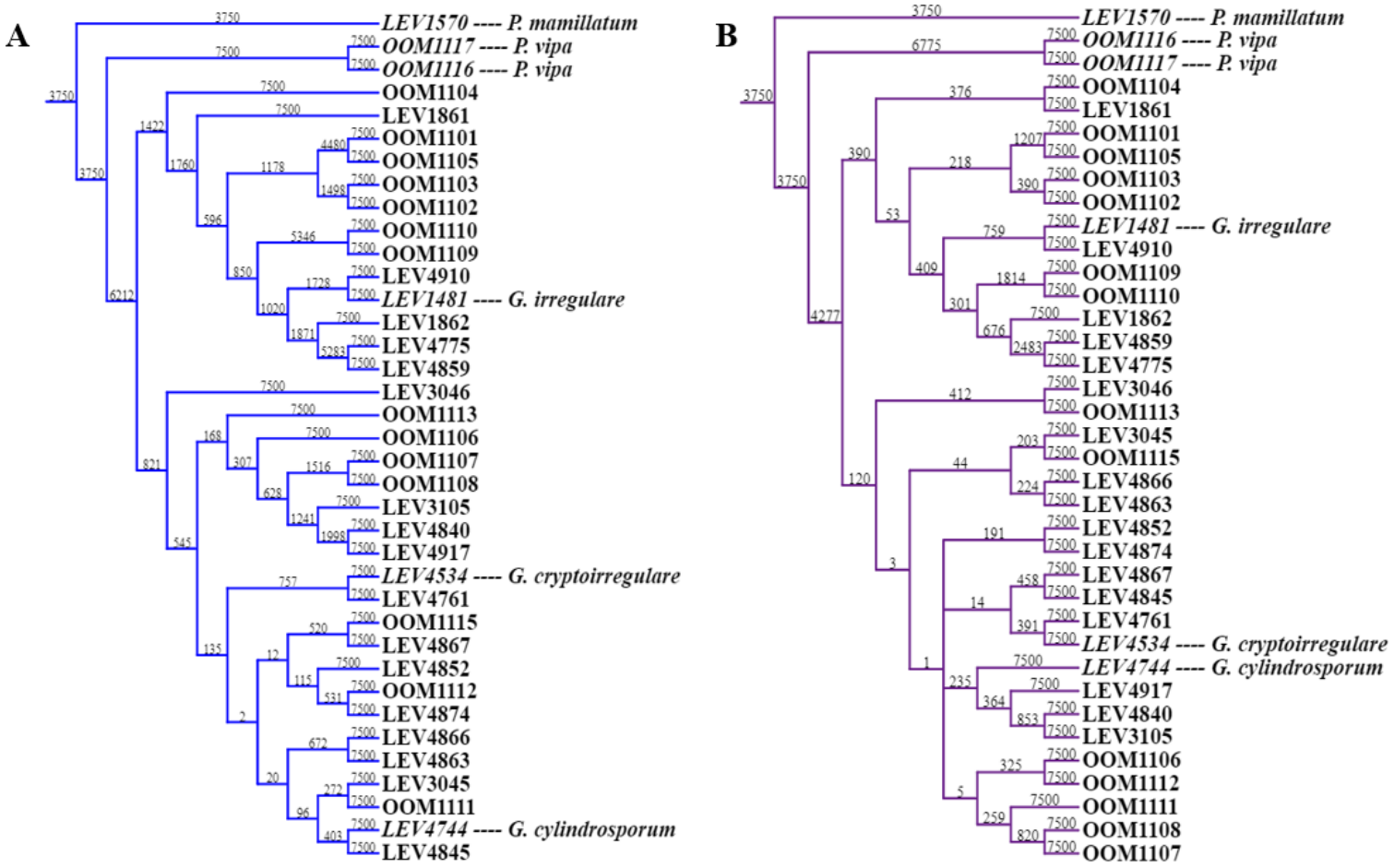


Figure 3-2. Majority-rule consensus gene trees: **A.** Majority-rule consensus gene tree obtained with the 75 single-copy BUSCO gene nucleotide (NT) data matrix. **B**. Majority-rule consensus gene tree obtained with the 75 single-copy BUSCO gene amino acid (AA) data matrix. The numbers on the branches indicate the number of times the partition of the species into the two sets which are separated by that branch occurred among the trees, out of 7500 trees. Reference specimens and the outgroup (LEV1570) are labeled.

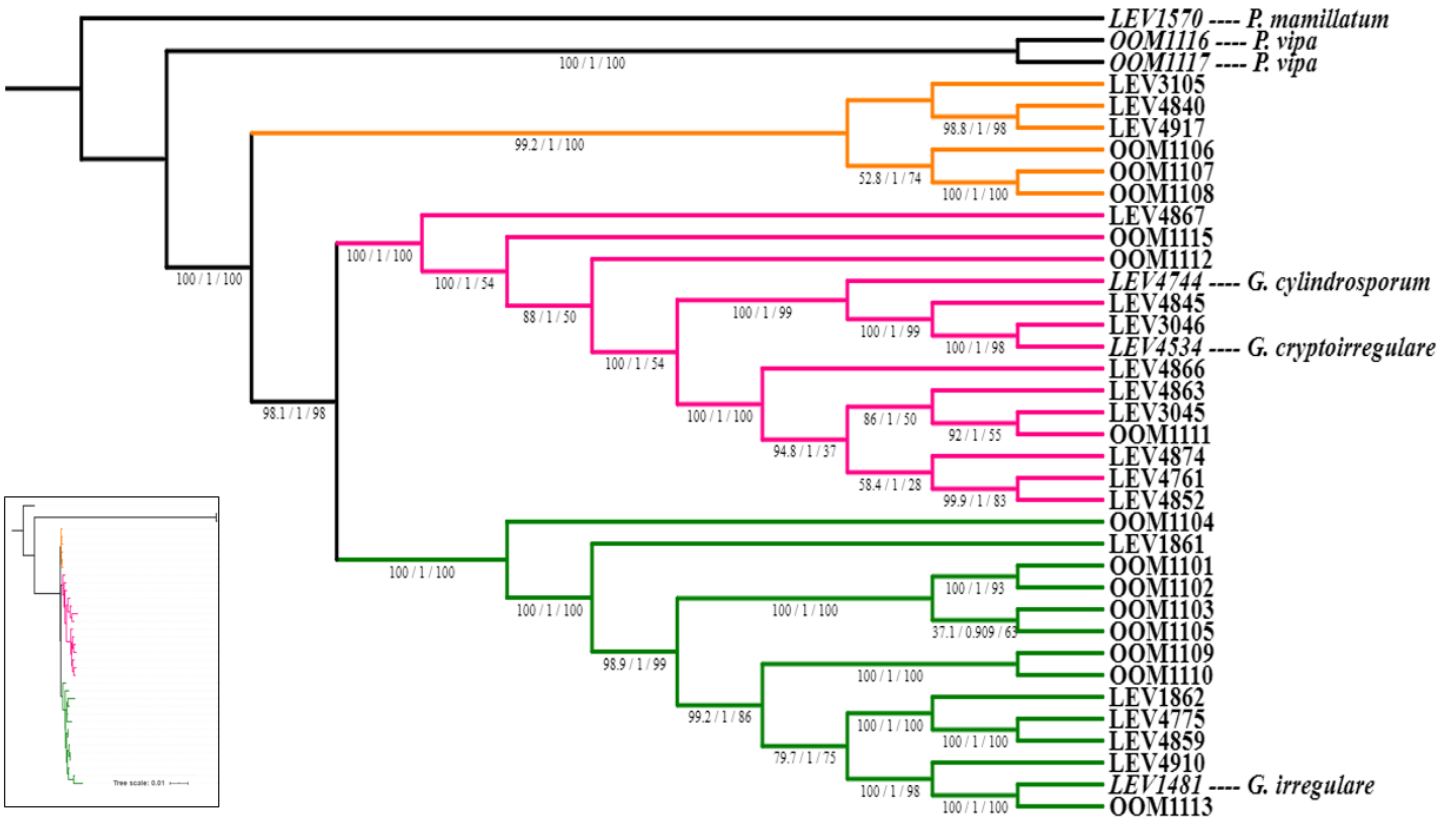


Figure 3-3. Maximum Likelihood (ML) species tree inferred from the concatenation-based analysis of a 75 single-copy BUSCO gene nucleotide (NT) data matrix. The ML phylogeny was reconstructed based on the concatenation nucleotide matrix (117,472 sites) under the TIM3+F+I+G4 substitution model. Branch support values near internodes are indicated as SH-aLRT support value, aBayes support value and ultra-fast bootstrap support value, respectively. Reference specimens and the outgroup (LEV1570) are labeled. Note, branch lengths on the ML tree are given in the inset at the bottom left.

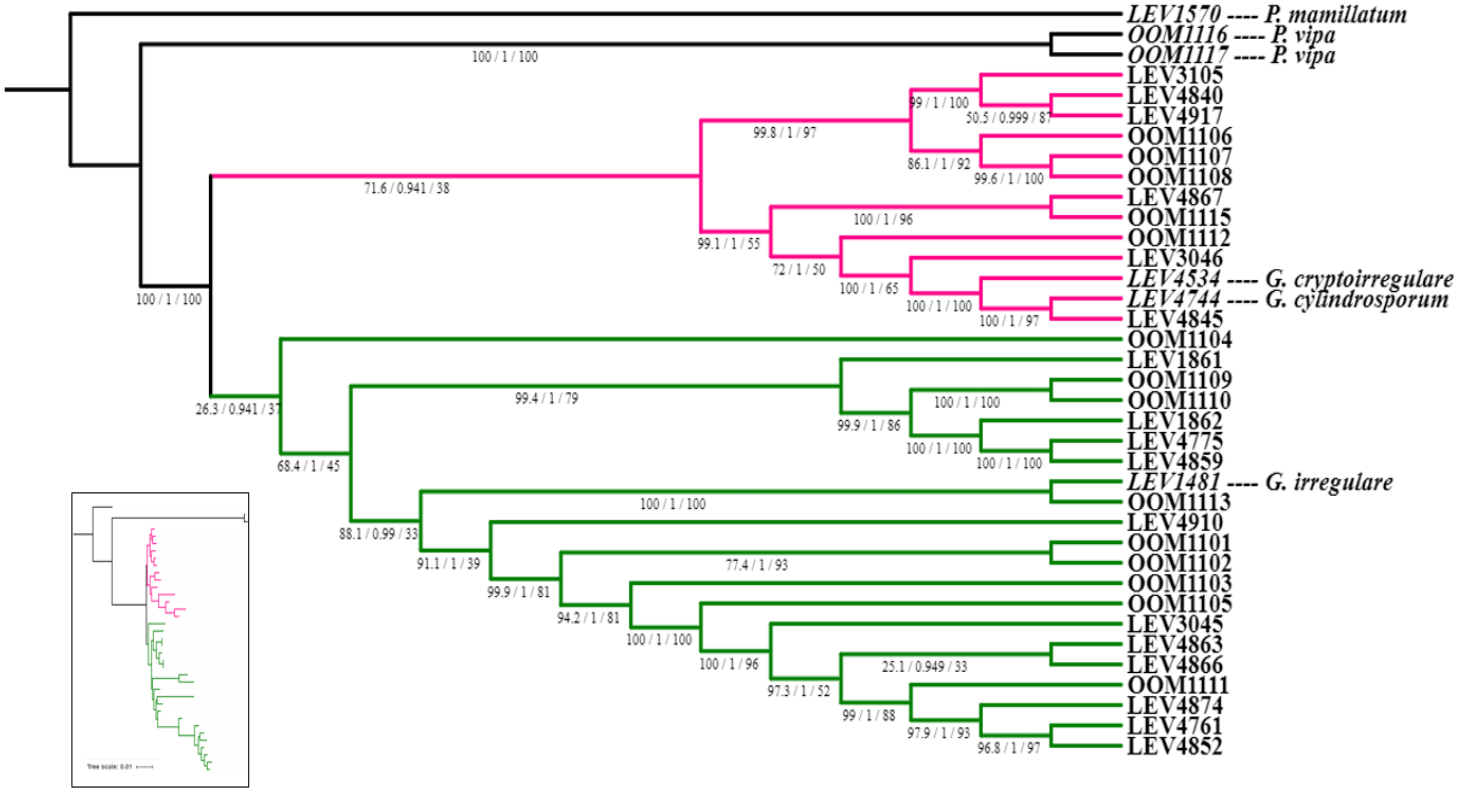


Figure 3-4. Maximum Likelihood (ML) species tree inferred from the concatenation-based analysis of a 75 single-copy BUSCO amino acid (AA) data matrix. The phylogeny was reconstructed based on the amino acid matrix (39,352 sites) under the JTT+F+I+G4 substitution model. Branch support values near internodes are indicated as SH-aLRT support value, aBayes support value and ultra-fast bootstrap support value, respectively. Reference specimens and the outgroup (LEV1570) are labeled. The tree was rooted using *P. mamillatum* (LEV1570). Note, branch lengths on the ML tree are given in the inset at the bottom left.

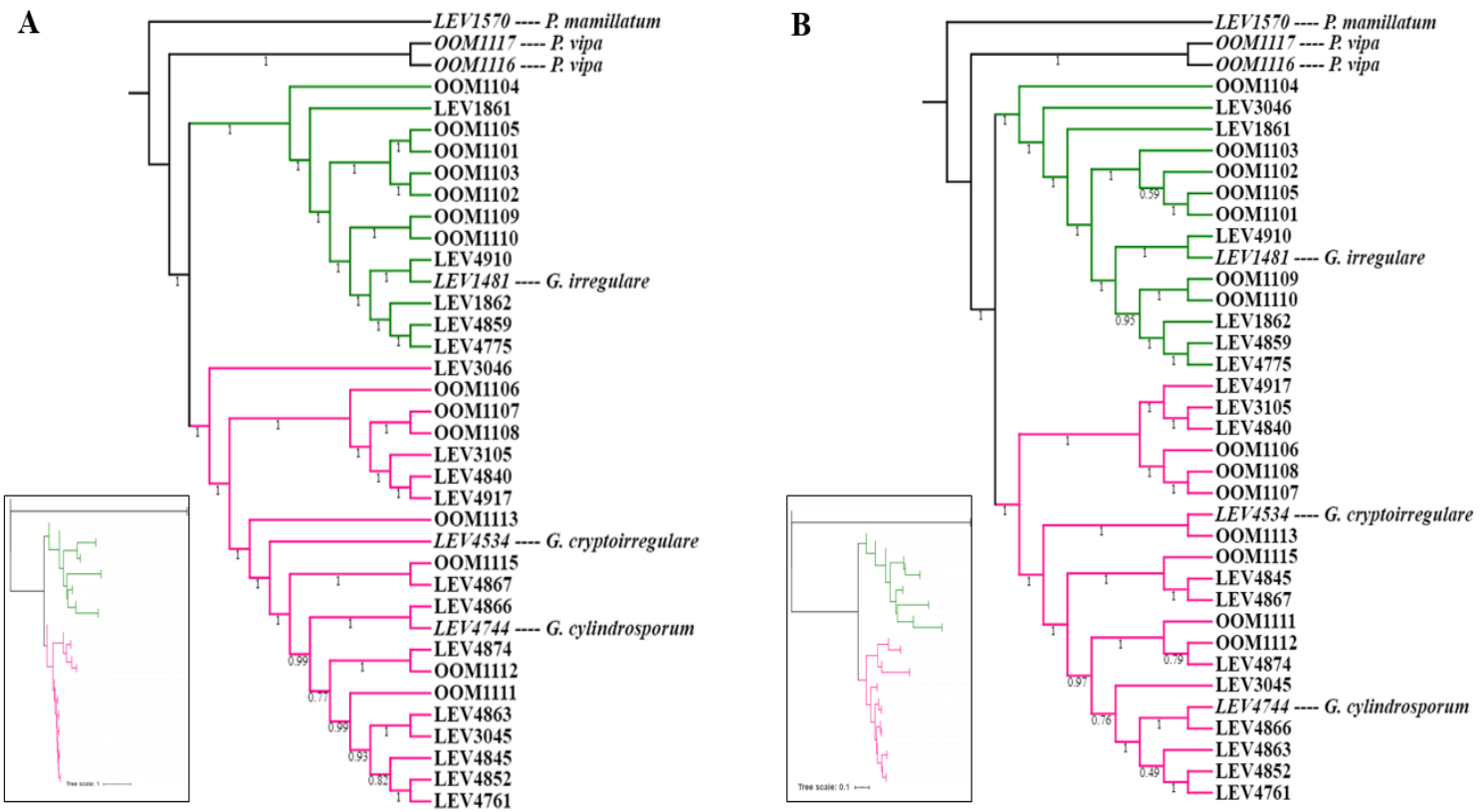


Figure 3-5. Species tree inferred from coalescence-based analysis of 75 single-copy BUSCO genes: **A.** Coalescence-based species tree estimated with the BUSCO gene nucleotide (NT) data matrix, **B**. Coalescence-based species tree estimated with the BUSCO gene amino acid (AA) data matrix. Local posterior probabilities (LPP) are displayed at the nodes. Reference specimens and the outgroup (LEV1570) are labeled. The tree was rooted using *P. mamillatum* (LEV1570). Note, branch lengths on the tree are given in the inset at the bottom left.

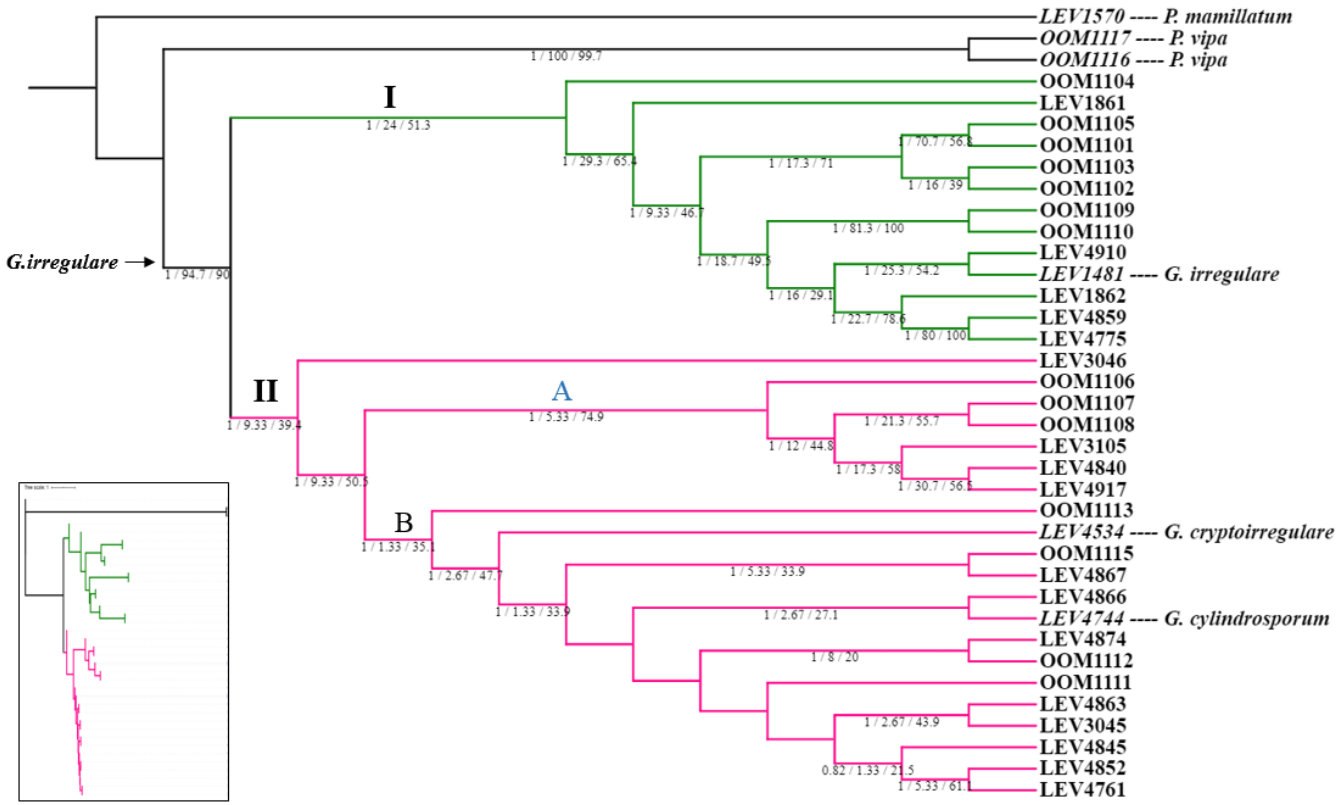


Figure 3-6. Species tree inferred from coalescence-based analysis of 75 single-copy BUSCO genes using the nucleotide (NT) data matrix. Local posterior probabilities (LPP) are displayed at the nodes, followed by the gene concordance factor (gCF) and the site concordance factor (sCF). Reference specimens and the outgroup (LEV1570) are labeled. The tree was rooted using *P. mamillatum* (LEV1570). Note, branch lengths on the tree are given in the inset at the bottom left.

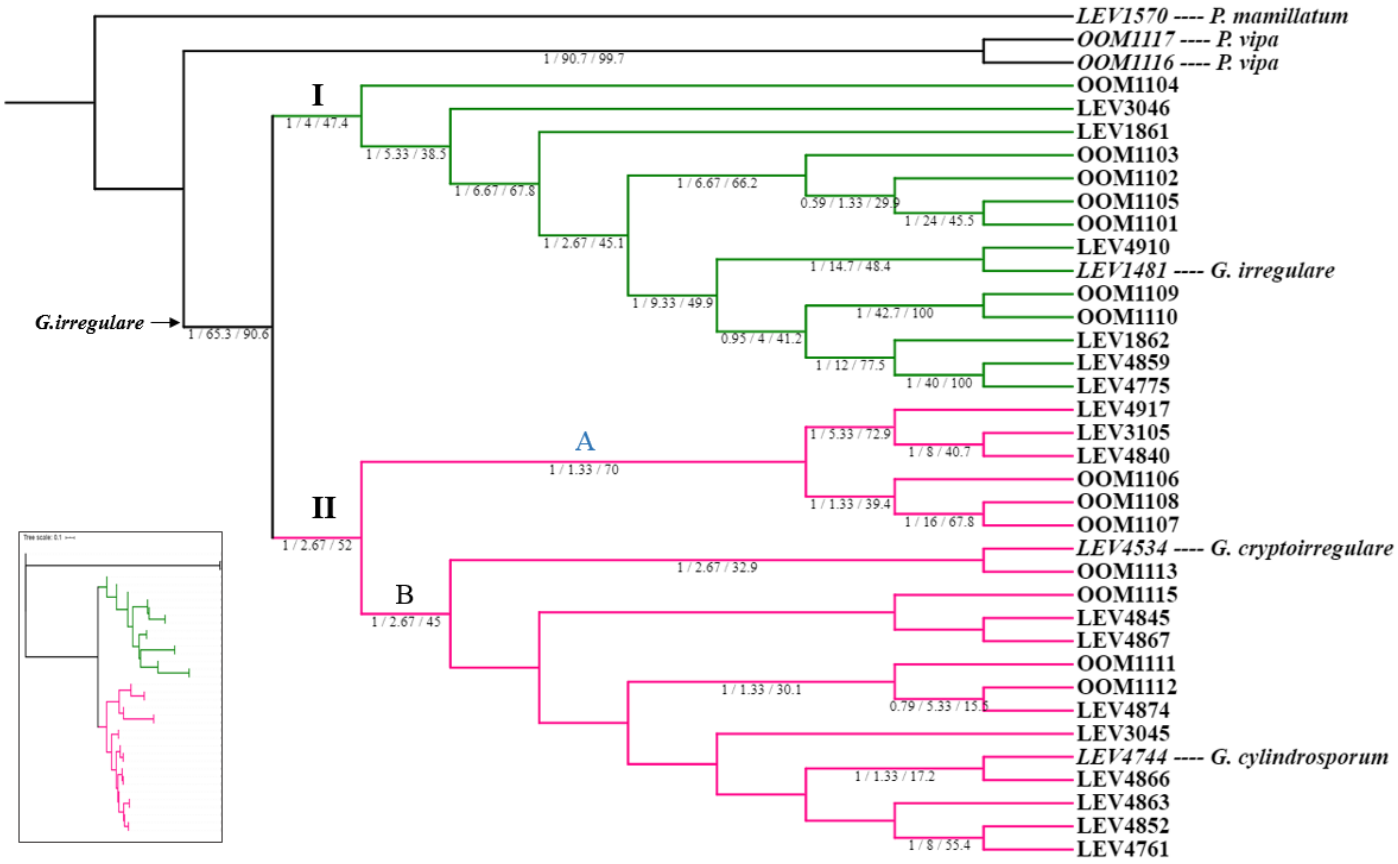
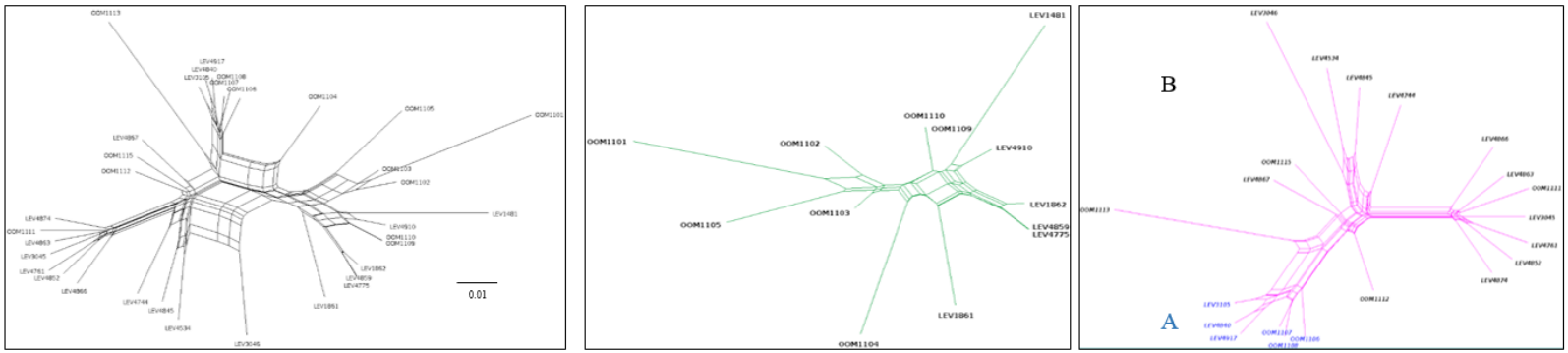


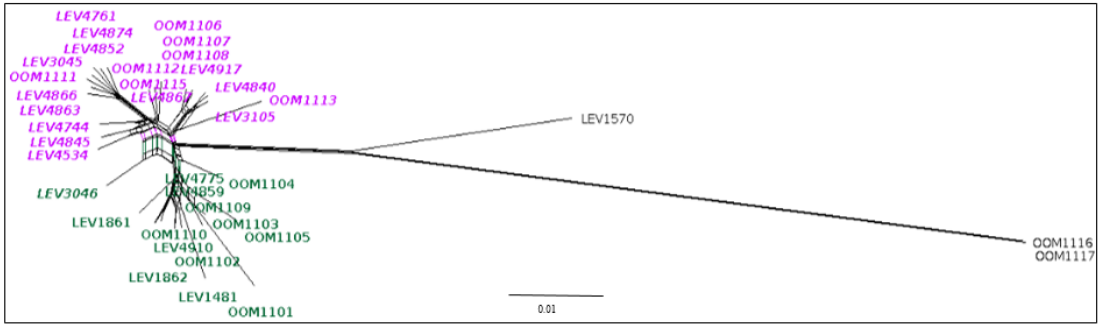
Figure 3-7. Species tree inferred from coalescence-based analysis of 75 single-copy BUSCO genes using the amino acid (AA) data matrix. Local posterior probabilities (LPP) are displayed at the nodes, followed by the gene concordance factor (gCF) and the site concordance factor (sCF). Reference specimens and the outgroup (LEV1570) are labeled. The tree was rooted using *P. mamillatum* (LEV1570). Note, branch lengths on the tree are given in the inset at the bottom left.



a) ***G. irregulare*** Φw= 0.0

b) I Φw= 0.0

c) II A. Φw = $2.165e^{-12}$
B. Φw= 0.0



d)

Figure 3-8. Split graphs showing the results of the pairwise homoplasy index (PHI) test. PHI test results (Φw) < 0.05 indicate significant recombination within the dataset. **a.** Representatives of all *G.irregulare* isolates in this study; **b**. Isolates in clade I; **c**. Isolates in clade II; **d.** Network of the evolutionary relationships of the 37 isolates used in this study.

CHAPTER IV

GENETIC DIVERSITY AND STRUCTURE OF POPULATIONS OF *GLOBISPORANGIUM CRYPTOIRREGULARE* FROM RHODODENDRON NURSERIES IN OREGON AND WASHINGTON

Abstract

Rhododendron is an important ornamental nursery crop in the United States which production is often compromised by several *Pythium* and *Globisporangium* species. Several surveys in Oregon and Washington rhododendron nurseries have commonly found that *Globisporangium cryptoirregulare* is a prevalent pathogen in rhododendron rots, causing mild levels of disease. *G. cryptoirregulare* exhibits high morphological and genetic diversity, and its population structure and dynamics as pathogens of rhododendrons have not been explored. This study did not find significant structure based on location, nursery, the production type, and mefenoxam test. Populations displayed a structure mostly associated with the presence of certain clonal lineages. These findings also suggest the movement of the inoculum between facilities and maintenance of lineages through the years. Moreover, variability in sensitivity to mefenoxam,

particularly among isolates and nurseries, was found. Understanding this plant pathogen's diversity can provide new insights of value for understanding disease epidemiology and disease management.

1. Introduction

Rhododendron is an important ornamental nursery crop in the United States. Its production is often compromised by several species of *Phytophthora, Pythium,* and *Globisporangium*. *Pythium* and *Globisporangium* are two sister genera of plant pathogenic oomycetes with ubiquitous distribution and responsible for infecting diverse crops (Levesque and De Cock 2004; Van der Plaats-Niterink 1981), including woody ornamentals (Kirk et al. 2001; Moorman 2004).

Several surveys in Oregon and Washington rhododendron nurseries have commonly found that *Globisporangium cryptoirregulare* Garzón, Yañéz, Moorman (syn. *Pythium cryptoirregulare*) is a prevalent pathogen in rhododendron rots, causing mild levels of disease (Weiland et al. 2018, 2020). *Globisporangium cryptoirregulare* is the causal agent of root rot and damping-off of a broad range of crops (Garzon et al. 2007) and has been described as a species that forms part of the *G. irregulare* complex (Garzon et al. 2017). The complex exhibits high morphological and genetic diversity (Harvey et al. 2001; Levesque and De Cock 2004; Spies et al. 2011), and includes *G. irregulare* sensu stricto (s.s.), *G. cryptoirregulare, G. cylindrosporum*, and *G. regulare*.

Although *Phytophthora* spp. cause the most serious root rot diseases; severe losses can be found if combined with other soilborne pathogens or environmental stress (Weiland et

al. 2018). The occurrence of *G. irregulare* and *G. cryptoirregulare* in forest and rhododendron nurseries had been reported (Weiland et al. 2015, 2018), but its population structure and dynamics have not been assessed. Population genetics study processes on genomes and populations with many applications in plant pathology. These studies can provide insight into the evolution of plant pathogens to understand the epidemiology of plant diseases and explore management strategies (Milgroom 2015).

The objective of this study was to characterize the *G. cryptoirregulare* population structure, understand its genetic diversity, and determine population differentiation within or between nurseries, production types, years and sensitivity to mefenoxam in Oregon and Washington, using microsatellites (SSRs). Later a subset of samples was analyzed using single nucleotide polymorphism (SNPs) markers generated by GBS (Genotyping by sequencing) to examine structure. Understanding this plant pathogen's diversity can provide new insights of value for understanding disease epidemiology and disease management.

2. Materials and Methods

2.1. *Globisporangium cryptoirregulare* isolates and identification

Isolates of *G. cryptoirregulare* isolates were obtained from six rhododendron nurseries (labeled A, B, C, D, E, and G). Rhododendrons were sampled from propagation, container, and fields from 2013 to 2017 (Table 4-1). The sampling of *G. cryptoirregulare* was done according to previous methods (Weiland et al. 2020).

Isolation, morphological, and molecular identification was performed at the USDA ARS Horticultural Crops Research laboratory at Oregon State University. Isolates were identified to the species level using the internal transcribed spacer (ITS) region, as described by Weiland et al. (2018, 2020). Isolates plated in PARP (piramicin 0.4 ml/L, ampicillin 0.25 g/L, rifampicin 1 ml/L, pentachloronitrobenzene 5ml/L and corn meal agar 17 g/L), a semi selective media, were sent to the Soilborne Plant Pathogens Laboratory at Oklahoma State University for genotyping.

2.2.DNA extraction

Mycelia from PARP media were transferred to PDA (potato dextrose agar 39 g/L), to water agar (WA; agar 15 g/L), and back to PDB (potato dextrose agar 24 g/L) for four days. Mycelia mats were harvested from PDB, placed in 1.5 ml microtubes, frozen at - 20 ºC, and lyophilized. DNA extraction was performed using the CTAB (cetyl trimethylammonium bromide) (Doyle and Doyle 1987) method and the Qiagen DNeasy Plant Mini kit (Germantown, MD) following the manufacturer's protocol. DNA was quantified using a Nanodrop spectrophotometer ND 1000 (Thermo Fisher Scientific, Waltham, MA, USA) and stored at -20 °C.

2.3.Microsatellite genotyping

A total of eight markers described by Lee and Moorman (2008) were selected for microsatellite analysis of *G. cryptoirregulare* isolates. Fluorescently labeled primers and fluorescently labeled M13 primers (Table 4-2) were used. Forward primers were labeled with either fluorescent 6FAM, NED, or PET dyes (Applied Biosystems® Waltham,

Massachusetts, USA). The M13 primers were used as described by Barkley et al. (2007) with minor modifications (Schuelke 2000) for two markers. Briefly, each forward primer was modified by adding a 21-bp M13 tail (5'-CGTTGTAAAACGACGGCCAGT-3') to the 5' end. The reverse primer was not modified, and a third primer consisted of the same M13 sequence labeled with the HEX dye (Thermo Fisher Scientific, Waltham, MA, USA).

For all primer combinations, the reaction conditions were the same, 20 µl reactions: 1 µl of each 5µM primer, 2 µL of 35 ng μl^{-1} DNA, 10 µL of 2X GoTaq Green Master Mix (Promega Corporation, Madison, WI, USA) and 6 µl of nuclease-free water. The PCR amplification with the labeled primers consisted of an initial denaturing step of 94ºC for 2 min, followed by 35 cycles of 94 ºC for 30 s, annealing temperature 55 ºC - 60 ºC (depending on the specific primer set) for 30 s, 72 ºC for 1 min, a final extension of 72 ºC for 10 min and 4 ºC hold. The PCR amplification with the M13 primers consisted of the first reaction mentioned above and a second reaction, 20 µl: 1.6 µl M13 forward primer at 5µM, 3.2 µl of 5 µM reverse primer, 3.2 µl of 5 µM HEX-M13, 1 µl of PCR product, 9 µl of 2X GoTaq Green Master Mix (Promega Corporation, Madison, Wisconsin, USA) and 2 µl of nuclease-free water. The program for the second reaction consisted of one cycle at 94 ºC for 5 min for initial denaturing, 20 cycles of 94 ºC for 30 s, 56 ºC for 45 s and 72 ºC for 45 s, five cycles of 94 ºC for 30 s, 53 ºC for 45 s and 72 ºC for 45 s, one cycle of 72 ºC for 10 min for final extension and a 4 ºC hold.

The PCR products were diluted and pooled into a multiplex set of 3 SSRs according to optimize genotyping cost (Table 4-2). PCR products were resolved by multiplexed

capillary electrophoresis on an ABI 3730 DNA Analyzer (Applied Biosystems, Waltham, Massachusetts, USA) at the Recombinant DNA and Protein Core Facility at Oklahoma State University by loading 1 µl of the diluted PCR product, 9 µl Hi-Di™ formamide (Applied Biosystems, Waltham, Massachusetts, USA), and 0.4 µl of GeneScan 500 LIZ® size standard (Applied Biosystems, Waltham, Massachusetts, USA). Results after fragment analysis were inspected and alleles were called using Peak scanner v 1.0 (Thermo Fisher Scientific, Waltham, Massachusetts, USA) software. A genotypic matrix was constructed and analyzed using R 4.0 (R Core Team, 2020).

2.4.Statistical and population genetic analysis

The data was arranged into three hierarchical levels: year, nursery, production type, and mefenoxam test. Dr. Jerry Weiland (Oregon State University) provided information about fungicide sensitivities. Population genetic analyses were conducted in R 4.0 using various packages (R Core Team, 2020). The R package poppr (Kamvar et al. 2014, 2015) was used to inspect the data's basic statistical properties and to assess genotypic diversity and multilocus genotypes (MLGs). Based on their power and utility (Grünwald et al. 2003), the following genotypic diversity parameters were measured: Shannon-Weiner diversity (H), the Stoddart and Taylor's index (G) (Shannon 1949; Stoddart and Taylor 1988), Simpson's index lambda, evenness E5 (Grünwald et al. 2003; Ludwig and Reynolds 1988; Pielou 1975), and gene diversity (H_{exp}). G and H measure richness and adjust for evenness, increasing as the number of genotypes increases. Lambda estimates the probability that two genotypes randomly sampled are different. E5 measures the

distribution of genotype abundance within the populations. H_{exp} is the expected heterozygosity.

The standardized index of association, $\bar{r}_d$ (Milgroom 1996), was calculated on clone-corrected data to infer the population's reproduction mode. P values were calculated with a null distribution created from 999 permutations of the data. Clone-corrected data has only one individual per multilocus genotype (MLG).

Analysis of molecular variance (AMOVA) was performed to detect and compare population differentiation within and between the population levels. A randomization test with 1000 permutations was conducted to assess the statistical significance in the AMOVA statistics. To calculate *Fst* (fixation indices between isolates), the R package *hierfstat* (Goudet 2004) was used.

For assessing population structure, two methods were used: minimum spanning networks (MSN) using Bruvo's genetic distance for microsatellite loci (Bruvo et al. 2004; Kamvar et al. 2014) and Principal Component Analysis (PCA). Bruvo's genetic distance is based on a stepwise mutation model for microsatellites. Discriminant Analysis of Principal Components (DAPC), from the R package adegenet (Jombart 2008; Jombart and Ahmed 2011), was conducted to optimize the principal components' numbers and conduct the principal component analysis (PCA). The PCA was visualized using the R package *factoextra* (Kassambara and Mundt 2017).

2.5.Clustering of *G. cryptoirregulare* populations

A subset of samples was genotyped by Genotyping by Sequencing (GBS) at the Center for Genome Research and Biocomputing (CGRB) at Oregon State University (Table 4-1). Neighbor-Joining (NJ) trees were constructed in R with the package ape 5.0 (Paradis and Schliep 2019) to compare the clustering of samples when using SSRs and SNPs. *G. irregulare* isolates were included to see the level of structure (Table 4-1). Trees were constructed using matrices of genetic distances. The Bruvo's distance and the bitwise distance were used for SSRs and SNPs, respectively. The bitwise distance calculates dissimilarity or Euclidean distance for SNP data. Branch support was calculated with the bootstrap method using 1000 replicates.

2.5.1. Genotyping by sequencing, read processing and SNP calling

Genomic DNA samples, high molecular weight, and no degradation were digested with the combination of the *PstI* (CTGCA/G) and *MspI* (C/CGG) restriction enzymes (New England Biolabs, Ipswich, MA). Samples were barcoded with Illumina adapters. The library was sequenced on an Illumina HiSeq 3000 (San Diego, CA) system with 150-bp paired-end reads in the Center for Genome Research Biocomputing (CGRB) at Oregon State University.

SNP calling was made with the TASSEL3 GBS UNEAK pipeline (Lu et al. 2012) by the CGRB at Oregon State University. The pipeline work without a reference genome and generates a large number of useful information. The resulting variant caller format (VCF) file was quality filtered in R using the *vcfR* package (R Core Team 2020; Knaus

and Grünwald 2017). Variants were removed based on the percentage of missing data and based on a minimum read DP of 4. Reads were also filtered by mapping quality (MQ=44) (Tabima et al. 2018)

3. Results

3.1. Genetic diversity and mode of reproduction

A total of 148 isolates were obtained from six nurseries over five years and were analyzed using eight polymorphic microsatellites. Genotypic diversity measured by the Shannon-Wiener index (H), Stoddart and Taylor's index (G), and Simpson's index (lambda) were low, regardless of the hierarchy level analyzed. However, lower diversity is observed in populations with small sizes. Thus, the lambda* index provided a better estimation of the probability that two randomly selected genotypes were different. The values close to 1 indicated that most genotypes are different; thus, populations showed great diversity. The evenness (average) was 0.39, indicating the genotypes are not in equal abundance and that populations are not dominated by single genotypes (Table 4-3).

The standardized index of association (*rd*) was calculated for clone corrected data to infer the reproduction mode. Values were significantly different to 0, indicating selfing and clonal reproduction (P = 0.001) (Table 4-3). These results were contradictory when analyzing the data based on their mefenoxam test result. However, the measure was made considering the samples that were not tested, which can produce bias.

The number of observed MLGs across all populations was 40, indicating that most individuals belonged to clonal lineages (Table 4-3). One clonal genotype (MLG20) was

prevalent in all the nurseries (except G), years (except 2017), and production types (55 isolates). This genotype has representatives of resistant and sensitive isolates (Figure 4-1 and Figure 4-2). The second most common clonal genotype (MLG 12, n =20) was prevalent in 2016, at nurseries C, D, and E, and in propagation and container systems (Figure 4-1 and Figure 4-2).

3.2. Genetic differentiation among populations

Fst index calculated to estimate the differentiation among populations indicated low to moderate genetic differentiation at all hierarchy levels (Table 4-5 Table 4-6, Table 4-7, and Table 4-8). The individuals from nursery G sampled in 2017 were the most differentiated. The AMOVA analysis supported these results; populations were not significantly differentiated among years, nurseries, production type, or sensitivity to mefenoxam. Approximately 90% of the variation was explained due to variation within populations (year, nursery, production type. Mefenoxam test) than between populations (Table 4-4). The gene flow analysis (Nm > 1) suggested genetic exchange among the populations (Table 4-4).

The number of observed MLGs across all populations was 40, indicating that most individuals belonged to clonal lineages (Table 4-3). One clonal genotype (MLG20) was prevalent in all the nurseries (except G), years (except 2017), and production types (55 isolates). This genotype has representatives of resistant and sensitive isolates (Figure 4-1 and Figure 4-2). The second most common clonal genotype (MLG 12, n =20) was

prevalent in 2016, at nurseries C, D, and E, and in propagation and container systems (Figure 4-1 and Figure 4-2).

3.3.Population structure

Minimum spanning networks (MSNs) and Principal component analysis (PCA) suggested low levels of genetic differentiation across years, nurseries, and production types (Figure 4-3- Figure 4-8). The structure is more associated with the multilocus genotypes and the presence of clonal lineages. More than half percent of the isolates belonged to multilocus genotypes, labeled in the MSNs (Figure 4-3 and Figure 4-4). The majority of individuals clustered in the PCAs regardless of the hierarchy level analyzed (Figure 4-5-Figure 4-8). However, the individuals from nursery G sampled in 2017 are the most differentiated and did not cluster close to the main cluster.

MSNs showed that all nurseries sampled over five years and from different production systems have isolates belonging to the dominating clonal lineage (MLG20). This clonal lineage also has isolates that were characterized as resistant or sensitive to mefenoxam. Resistant isolates were more frequent in the MLG20, while a similar number of isolates, either resistant or sensitive, were found in the MLG12 (second most common lineage). Interestingly, a relation between mefenoxam sensitivity and nurseries were found. Nurseries A, E, and G had only resistant isolates, while nursery B had only sensitive isolates. Nursery C and D have the presence of resistant and sensitive isolates (Figure 4-9).

Overall, results suggest low diversity between populations (hierarchy levels), no structure, and a likelihood of gene flow. Analyses suggest that clonal lineages are getting established once introduced every year.

3.4.Clustering of *G. cryptoirregulare* isolates with *G. irregulare*

Seventy-seven samples were sequenced by GBS, which resulted in 42,530 raw SNPs and 76.43% of missing data. After filtering, the dataset went down to 633 SNPs in 64 samples, including 5 *G. irregulare* isolates used as a reference. NJ trees gave similar topologies, *G. irregulare* isolates genotyped either with SSRs or SNPs grouped with the rest of isolates (Figure 4-10 and Figure 4-11). The *G. cryptoirregulare* isolates analyzed formed two well-supported branches, bootstrap support > 50%. One of the branches has isolates that gave ambiguous results; they were identified as *G. irregulare* with ITS sequencing, or they were suspect to be hybrids based on ITS cloning.

Overall, these results indicated no structure based on the species reported as part of the *G. irregulare* complex, suggesting that the *G. cryptoirregulare* is a large and diverse group in Oregon and Washington nurseries.

4. Discussion

SSRs markers have been used to determine the population structure and distribution of *Pythium* species: *G. irregulare* s.s, *G. cryptoirregulare*, *P. aphanidermatum*, and *P. helicoides* (Lee and Moorman 2008; Zhou et al. 2009); some of them transferable to additional species. Genetic diversity and structure have been explored mostly in *G. irregulare* based on culture collections and isolates from different hosts and locations

(Garrido et al. 2014; Huzar-Novakowiski and Dorrance 2018; Proaño et al. 2016; Weiland et al. 2015). Hence, little is known about the genetic population structure and diversity of *G. cryptoirregulare*. Weiland et al. (2020) surveyed the soilborne and root infecting pathogens of rhododendrons and found that *G. cryptoirregulare* is the most common *Pythium* species isolated from roots. Weiland et al. (2018) conducted a pathogenicity experiment on *G. cryptoirregulare* isolates and found they tend to cause mild disease levels. Based on those results, the need for research to determine the impact of this pathogen on rhododendron (Weiland et al. 2018).

The population structure and diversity of *G. cryptoirregulare* in rhododendron Oregon and Washington nurseries revealed not significant structure based on location, nursery, production type, and mefenoxam test. The populations displayed a structure mostly associated with certain clonal lineages. More than half of the isolates belonged to multilocus genotypes. The most prevalent, MLG20, had 55 isolates from different nurseries, years, and production types. Interestingly, this clonal lineage has resistant and sensitive isolates to mefenoxam and was from two different geographical areas, Oregon and Washington. These findings suggest the movement of the inoculum between facilities and maintenance of lineages through the years. Clonal lineages were shared among nurseries suggesting potential common sources of inoculum. Clonal lineages were also found with *Phytophthora plurivora* in rhododendron from nurseries A, B, and C (Carleson et al. 2019). *Phytophthora plurivora* was recently found as one of the most common *Phytophthora* species causing root rot of rhododendron (Weiland et al. 2018).

Analysis of genotypic diversity is an important part of the genetic structure analysis of populations calculated from genotype frequencies (Grünwald et al. 2003). Overall, populations were clonal with moderate levels of genetic diversity and levels of genetic exchange. These results showed that there is diversity in *G. cryptoirregulare* populations that has been established over time. Genetic diversity is a source of variability of pathogenic strains in crop hosts and could be used to identify alleles controlling key pathogenicity traits (Bennett and Stone 2016; Huzar-Novakowiski and Dorrance 2018).

Fungicide resistance to certain chemistries has been reported in *Pythium* and *Globisporangium* species, particularly to mefenoxam and propamocarb (Moorman and Kim 2004; Weiland et al. 2014). In this study, most isolates reported as resistant to mefenoxam were isolated from nurseries A, E, and G. One nursery, B, has only sensitive isolates, and the remaining nurseries (C and D) had resistant and sensitive isolates; most of them grouped in the most prevalent MLG. These results suggest variability in sensitivity, particularly among isolates and nurseries. Some studies have shown that isolates of *Pythium* species with similar genetic backgrounds can show variation in their sensitivity to fungicides and that genetic variation can help to explain fungicide resistance (Al-Sa'di et al. 2008a; Al-Sa'di et al. 2008b; Garrido 2014). Fungicide resistance has been reported to affect some species' genetic diversity (Lee et al. 2010).

The phylogenetic reconstruction and the population structure analyses indicated a lack of structure. The analyses of population differentiation supported those results with the *Fst* and the AMOVA, in which no significant genetic differentiation was observed among populations. The comparison of the isolates' phylogenetic relationships based on SSR and

SNPs markers was similar. These results suggest that *G. cryptoirregulare* forms a large and diverse population in Oregon and Washington. Further exploration of the population dynamics of *G. irregulare* and *G. cryptoirregulare* using GBS as a next-generation sequencing technique will be done once the rest of the samples finish being genotyped at Oregon State University. Similar patterns in diversity and population structure using SSRs and SNPs markers are expected. In this case, de novo SNP calling can represent a limit for posterior studies; however, for population genetic inferences, it should not represent an issue. In addition, draft genomes of *G. cryptoirregulare* and *G. cryptoirregulare* (Chapter III) were generated and may be useful in the future.

This work provides a better understanding of the genetic diversity in populations of G. *cryptoirregulare* in rhododendrons and suggests movement via contaminating planting material within and across states. Little is known about the movement and distribution of *G. irregulare* and *G. cryptoirregulare* inoculum in nurseries and greenhouses. However, it can be harbored in many potential sources, including infected plant material, plant debris, soil, tools, equipment, potting mixtures, irrigation water, and splashes around potting mixes and other substrates (Al-Sa'di et al. 2008c).

Finally, we want to highlight the importance of the presence of resistant isolates in nurseries. This could mean that fungicides may fail to provide adequate disease control in the future (Weiland et al. 2014). This study may impact the management strategies developed to control diseases in rhododendron nurseries in Oregon and Washington. Better management strategies should be performed for *G. cryptoirregulare* as a pathogen of rhododendron.

5. Acknowledgments

The assistance of Dr. Niklaus Grünwald and Dr. Jerry Weiland from USDA ARS Horticultural Crops Research laboratory at Oregon State University and their teams is greatly appreciated, for providing funding and samples. Thanks to the Center for Genome Research and Biocomputing (CGRB) at Oregon State University for the GBS processing. Viviana Freire, Patricia Calderon and Cassidy Ward from Oklahoma State University helped with technical support. Thanks to the Recombinant DNA and Protein Core Facility at Oklahoma State University for fragment analysis. Some of the computational analysis done in this project was performed at the Oklahoma State University High-Performance Computing Center, supported in part through the National Science Foundation grant OAC-1126330.

Tables

Table 4-1. Information of the isolates included in this study, nursery, production type, year collected, rhododendron cultivar, source material, and state.

Nursery	Sample	Production Type	Year collected	Cultivar	Source Material	Mefenoxam test[d]	State
A	M03R-8[a]	Container	2013	Yaku Princess	Diseased	R	OR
A	M06R-1[a]	Container	2013	English Roseum	Diseased	R	OR
A	M09R[a,b]	Container	2013	Nova Zembla	Diseased	R	OR
A	M13S-1[a]	Container	2013	R. daphnoides	Diseased	R	OR
A	M16R[a]	Container	2013	R. daphnoides	Diseased	R	OR
A	M20S[a]	Container	2013	Elvira	Diseased	R	OR
A	M29R-1[a]	Container	2013	Lee's Dark Purple	Diseased	R	OR
A	M31R-1[a]	Container	2013	PJM	Diseased	R	OR
A	M34S[a]	Container	2013	Roseum Elegans	Diseased	R	OR
A	M35R[a]	Container	2013	Lee's Dark Purple	Diseased	R	OR
A	M38R-4[a]	Container	2013	Lee's Dark Purple	Diseased	I	OR
A	M46R-4[a]	Container	2013	Husky Mania	Diseased	R	OR
A	M62[a]	Container	2013	unknown	Diseased	R	OR
B	P31	Container	2015	unknown	Diseased	S	OR
B	P43	Container	2015	unknown	Diseased	S	OR
B	P53	Container	2015	unknown	Diseased	S	OR
B	P61	Container	2015	unknown	Diseased	S	OR
B	P73Ra	Container	2017	Purpureum Grandiflorum	Diseased	S	OR
B	P81R	Container	2017	Lem's Monarch	Diseased	S	OR
B	P82R	Container	2017	Fastuosum Flore Pleno	Diseased	S	OR
B	P95R	Container	2017	PJM Pink	Diseased	S	OR
B	P98R	Container	2017	Olin O. Dobbs	Diseased	S	OR
B	P103Sa	Container	2017	Ed Darts	Diseased	S	OR
B	O05A[a]	Field	2013	unknown	Diseased	S	OR
B	O14D[a]	Field	2013	unknown	Diseased	S	OR
B	O18A[a]	Field	2013	unknown	Diseased	S	OR
B	O21A	Field	2013	unknown	Diseased	S	OR
B	O22A[a]	Field	2013	unknown	Diseased	S	OR
B	O25A[a]	Field	2013	unknown	Diseased	S	OR

Nursery	Sample	Production Type	Year collected	Cultivar	Source Material	Mefenoxam test[d]	State
B	O26C [a]	Field	2013	unknown	Diseased	S	OR
B	O27A [a]	Field	2013	unknown	Diseased	S	OR
B	O29B [a]	Field	2013	unknown	Diseased	S	OR
B	O46A [a]	Field	2013	unknown	Diseased	S	OR
B	O76 [a]	Field	2015	unknown	Diseased	S	OR
B	O78a[a]	Field	2015	unknown	Diseased	S	OR
B	O108a	Field	2015	unknown	Diseased	S	OR
B	O129a	Field	2015	unknown	Diseased	S	OR
B	O130	Field	2015	unknown	Diseased	S	OR
B	O70b	Field	2015	unknown	Diseased	S	OR
C	S01R-1 [a]	Container	2014	unknown	Diseased	S	OR
C	S03R-1 [a]	Container	2014	unknown	Diseased	NT	OR
C	S05S-1 [a]	Container	2014	unknown	Diseased	S	OR
C	S16R-1 [a]	Container	2014	unknown	Diseased	NT	OR
C	S18R-1 [a]	Container	2014	unknown	Diseased	R	OR
C	S22R-1 [a]	Container	2014	unknown	Diseased	NT	OR
C	S24R-1 [a]	Container	2014	unknown	Diseased	R	OR
C	S25R-1 [a]	Container	2014	unknown	Diseased	NT	OR
C	S27R-1 [a]	Container	2014	unknown	Diseased	NT	OR
C	S28R-1 [a]	Container	2014	unknown	Diseased	R	OR
C	S30R-1 [a]	Container	2014	unknown	Diseased	NT	OR
C	S31S-1	Container	2014	unknown	Diseased	S	OR
C	S33R-1 [a]	Container	2014	unknown	Diseased	NT	OR
C	S36R-1 [a]	Container	2014	unknown	Diseased	NT	OR
C	S39R-1 [a]	Container	2014	unknown	Diseased	R	OR
C	S48R-1 [a]	Container	2014	unknown	Diseased	NT	OR
C	S53S-1	Container	2014	unknown	Diseased	I	OR
C	S63S-1 [a]	Container	2014	unknown	Diseased	NT	OR
C	S64S-1 [a,b]	Container	2014	unknown	Diseased	R	OR
C	S65S-1	Container	2014	unknown	Diseased	NT	OR
C	S66R-1	Container	2014	unknown	Diseased	S	OR
C	F14	Propagation	2016	Cunningham's White	Healthy	NT	OR
C	F15	Propagation	2016	Cunningham's White	Healthy	S	OR
C	F18	Propagation	2016	Cunningham's White	Healthy	R	OR
C	F22	Propagation	2016	Cunningham's White	Healthy	NT	OR
C	F24	Propagation	2016	Cunningham's White	Healthy	NT	OR

Nursery	Sample	Production Type	Year collected	Cultivar	Source Material	Mefenoxam test[d]	State
C	F25	Propagation	2016	Cunningham's White	Healthy	NT	OR
C	F27	Propagation	2016	Cunningham's White	Healthy	S	OR
C	F28	Propagation	2016	Cunningham's White	Healthy	NT	OR
C	F29	Propagation	2016	Cunningham's White	Healthy	S	OR
C	F33	Propagation	2016	Cunningham's White	Healthy	NT	OR
C	F34	Propagation	2016	Cunningham's White	Healthy	R	OR
C	F36	Propagation	2016	Cunningham's White	Healthy	NT	OR
C	F41	Propagation	2016	Yaku Princess	Healthy	R	OR
C	F42	Propagation	2016	Yaku Princess	Healthy	NT	OR
C	F43	Propagation	2016	Yaku Princess	Healthy	NT	OR
C	F44	Propagation	2016	Yaku Princess	Healthy	NT	OR
C	F45	Propagation	2016	Yaku Princess	Healthy	NT	OR
C	F46	Propagation	2016	Yaku Princess	Healthy	NT	OR
C	F47	Propagation	2016	Yaku Princess	Healthy	R	OR
C	F48	Propagation	2016	Yaku Princess	Healthy	NT	OR
C	F410	Propagation	2016	Yaku Princess	Healthy	NT	OR
C	F411	Propagation	2016	Yaku Princess	Healthy	R	OR
C	F61	Propagation	2016	Yaku Princess	Healthy	NT	OR
C	F62	Propagation	2016	Yaku Princess	Healthy	NT	OR
C	F63	Propagation	2016	Yaku Princess	Healthy	S	OR
C	F65	Propagation	2016	Yaku Princess	Healthy	S	OR
C	F66	Propagation	2016	Yaku Princess	Healthy	NT	OR
C	F35	Propagation	2016	Cunningham's White	Healthy	NT	OR
D	V7	Propagation	2016	Grace Seabrook	Healthy	S	OR
D	V12	Propagation	2016	Vulcan's Flame	Healthy	NT	OR
D	V19	Propagation	2016	Vulcan's Flame	Healthy	NT	OR
D	V20	Propagation	2016	Vulcan's Flame	Healthy	S	OR
D	V22	Propagation	2016	Nova Zembla	Healthy	NT	OR
D	V23	Propagation	2016	Nova Zembla	Healthy	NT	OR
D	V24	Propagation	2016	Nova Zembla	Healthy	NT	OR
D	V25	Propagation	2016	Nova Zembla	Healthy	NT	OR
D	V27	Propagation	2016	Nova Zembla	Healthy	NT	OR
D	V29	Propagation	2016	Nova Zembla	Healthy	NT	OR
D	V30	Propagation	2016	Nova Zembla	Healthy	NT	OR
D	V37	Propagation	2016	Solidarity	Healthy	S	OR
D	V48	Propagation	2016	Winsome	Healthy	S	OR

Nursery	Sample	Production Type	Year collected	Cultivar	Source Material	Mefenoxam test[d]	State
D	V51	Propagation	2016	America	Healthy	S	OR
D	V52	Propagation	2016	America	Healthy	NT	OR
D	V55	Propagation	2016	America	Healthy	NT	OR
D	V63	Propagation	2016	cat. Grandiflorum	Healthy	NT	OR
D	V65	Propagation	2016	cat. Grandiflorum	Healthy	NT	OR
D	V66	Propagation	2016	cat. Grandiflorum	Healthy	NT	OR
D	V75	Propagation	2016	Sappho	Healthy	NT	OR
D	V78	Propagation	2016	Sappho	Healthy	R	OR
D	V83	Propagation	2016	Gomer Waterer	Healthy	S	OR
D	V85	Propagation	2016	Gomer Waterer	Healthy	NT	OR
D	V91	Propagation	2016	Anita Gehnrich	Healthy	NT	OR
D	V96	Propagation	2016	Anita Gehnrich	Healthy	S	OR
D	V100	Propagation	2016	Anita Gehnrich	Healthy	NT	OR
C	S68	Container	2016	unknown	Diseased	R	OR
C	S73	Container	2016	unknown	Diseased	NT	OR
C	S74	Container	2016	unknown	Diseased	R	OR
C	S76	Container	2016	unknown	Diseased	NT	OR
C	S86	Container	2016	unknown	Diseased	S	OR
C	S88	Container	2016	unknown	Diseased	R	OR
C	S91	Container	2016	unknown	Diseased	S	OR
C	S101	Container	2016	unknown	Diseased	R	OR
E	B01[a]	Propagation	2014	Yaku Princess	Healthy	NT	WA
E	B07[a]	Propagation	2014	Yaku Princess	Healthy	NT	WA
E	B08[a]	Propagation	2014	Yaku Princess	Healthy	R	WA
E	B11[a]	Propagation	2014	Yaku Princess	Healthy	R	WA
E	B13a[a]	Propagation	2014	Yaku Princess	Healthy	R	WA
E	B19[a]	Propagation	2014	Yaku Princess	Healthy	R	WA
E	B34[a]	Propagation	2014	Cunningham's White	Healthy	R	WA
E	B42[a]	Propagation	2014	Cunningham's White	Healthy	R	WA
E	B43[a]	Propagation	2014	Cunningham's White	Healthy	R	WA
E	B70[a]	Propagation	2016	Cunningham's White	Healthy	R	WA
E	B82[a]	Propagation	2016	Cunningham's White	Healthy	I	WA
E	B85	Propagation	2016	Yaku Princess	Healthy	NT	WA
E	B87[a]	Propagation	2016	Nova Zembla	Healthy	NT	WA
E	B89[a]	Propagation	2016	Nova Zembla	Healthy	NT	WA
E	B94[a]	Propagation	2016	Nova Zembla	Healthy	I	WA

Nursery	Sample	Production Type	Year collected	Cultivar	Source Material	Mefenoxam test[d]	State
E	B84 [a]	Propagation	2016	Yaku Princess	Healthy	NT	WA
G	C003	Field	2017	Purpureum Elegans	Diseased	R	OR
G	C009	Field	2017	Purpureum Elegans	Diseased	R	OR
G	C011	Field	2017	Purpureum Elegans	Diseased	R	OR
G	C013	Field	2017	cat. Boursalt	Diseased	R	OR
G	C014	Field	2017	cat. Boursalt	Diseased	R	OR
G	C017	Field	2017	English Roseum	Diseased	R	OR
G	C064	Field	2017	Anah	Diseased	R	OR
G	C065	Field	2017	cat. Boursalt	Diseased		OR
G	C066	Field	2017	English Roseum	Diseased		OR
G	C068	Field	2017	Calsap	Diseased		OR
None	irr-A2-8 [a]	Control sample	2008	Soil at forest nursery	-		OR
None	irr-C1-1 [a]	Control sample	2008	Soil at forest nursery	-		OR
None	irr-C1-5 [a]	Control sample	2008	Soil at forest nursery	-		OR
None	irr-01-1 [a]	Control sample	2008	Soil at forest nursery	-		OR
None	irr-01-10 [a]	Control sample	2008	Soil at forest nursery	-		OR

[a] =Isolates genotyped with GBS (Genotyping by Sequencing)

[b] = Sample was identified as *G. irregulare* based on ITS sequencing

[c] = Sample appear to be hybrid based on ITS cloning

[d] = Fungicide sensitivity was calculated as EC_{50} values (concentration of mefenoxam required to reduce the culture diameters of each isolate by 50%). E50 values ≥ 100 µg/ml were considered resistant. R: Resistant; S: Sensitive; I: Intermediate; NT: No tested.

Table 4-2. Simple Sequence Repeat (SSR) loci utilized to evaluate *Globisporangium cryptoirregulare* populations (from Lee and Moorman 2008).

SSR loci	Repeat motif	Dye	Primer sequences (5' to '3)	Ta	Size (bp)	Reaction
P50CA1-68	(CA)17	HEX-M13	F: GCTGATCTGCAGTGCACCTA R: GGTAAGGCGATGATGATGCT	56 ºC	138	Multiplex 1
P50TG2-93	(TG)18	PET	F: GCGTGGCTCGCGTCCTTAAA R: TGGGAACTCACACGGAAATGGCTA	60 ºC	113	Multiplex 1
P50GA3-20	(GA)13(GT)11	FAM	F: AGATCCGAAAGGCGATAAGC R: ATCACGCTCGAATAGTTCCTGT	55 ºC	179	Multiplex 1
P50TC2-23	(TC)20	PET	F: CCTGGCTGGTTCATTAGTCTCT R: TGGCTATCTGGATTGGTTTGTA	55 ºC	121	Multiplex 2
P50AG3-30	(AG)18	FAM	F: CGAGCGACGATTTGTAATGCCAGT R: TCAAGGACGGAAACCCTTGTGGAA	60 ºC	117	Multiplex 2
P50CT1-58	(CT)15	NED	F: TCTACCAACACTGAGCGCTAGCAA R: TCGAATGCGCCAGTCAAGCTC	60 ºC	177	Multiplex 2
P50GAA3-42	(GAA)10	NED	F: ACGAACCAACAAACAACAACAC R: ATGATGAACTCCGTAGGCAAGT	56 ºC	322	Multiplex 3
P50AG1-61	(AG)16	HEX-M13	F: AATGTTCAGAAGCGTGGGAAGCAG R: CTCACATTGCTCCACAACCAGTCA	60ºC	246	Multiplex 3

Table 4-3. Population statistics by year, nursery and production type of *G. cryptoirregulare* sampled in Oregon and Washington nurseries. N is the number of individuals sampled. MLG is the number of multilocus genotypes observed. eMLG is the number of expected MLGs at a sample size of 4 based on rarefaction. SE is the standard error of this statistic. Shannon-Wiener Index (H), Stoddart and Taylor's Index (G), Simpson's index (lambda), Simpson's Index after correcting for differing sample sizes (lambda*), evenness (E5), and Nei's unbiased gene diversity (H_{exp}). The standardized index of association r_d was calculated on clone-corrected data and P values were calculated with a null distribution with 999 permutations of the data.

Year	***N***	***MLG***	***eMLG***	***SE***	***H***	***G***	***lambda***	***lambda****	E_5	H_{exp}	r_d	$p.r_d$
2013	23	12	6.32	1.18	2.08	5.57	0.82	0.91	0.66	0.47	0.58	0.001
2014	30	15	6.67	1.21	2.26	5.84	0.83	0.93	0.57	0.62	0.62	0.001
2015	10	4	4.00	0.00	0.94	1.92	0.48	0.75	0.59	0.48	0.86	0.001
2016	69	15	4.52	1.19	1.73	3.55	0.72	0.93	0.55	0.51	0.52	0.001
2017	16	4	2.88	0.78	0.69	1.49	0.33	0.75	0.49	0.33	NaN	NaN
Total	148	40	6.18	1.37	2.55	5.77	0.82	0.96	0.39	0.57	0.52	0.001
Nursery	***N***	***MLG***	***eMLG***	***SE***	***H***	***G***	***lambda***	***lambda****	E_5	H_{exp}	r_d	$p.r_d$
A	13	5	4.31	0.67	1.33	3.19	0.69	0.80	0.79	0.52	0.72	0.001
B	26	13	6.19	1.22	2.08	5.28	0.81	0.92	0.61	0.55	0.55	0.001
C	57	20	6.11	1.30	2.30	5.92	0.83	0.95	0.55	0.62	0.55	0.001
D	26	11	5.43	1.19	1.83	4.07	0.75	0.91	0.59	0.46	0.52	0.001
E	16	4	3.13	0.71	0.82	1.71	0.41	0.75	0.55	0.40	-0.33	1.000
G	10	4	4.00	0.00	0.94	1.92	0.48	0.75	0.59	0.35	NaN	NaN
Total	148	40	6.17	1.35	2.56	5.77	0.82	0.96	0.39	0.57	0.54	0.001
Production type	***N***	***MLG***	***eMLG***	***SE***	***H***	***G***	***lambda***	***lambda****	E_5	H_{exp}	r_d	$p.r_d$
Container	52	20	6.48	1.27	2.38	6.26	0.84	0.95	0.53	0.64	0.56	0.001
Field	26	16	7.30	1.20	2.42	7.68	0.87	0.94	0.65	0.58	0.50	0.001
Propagation	70	16	4.61	1.21	1.77	3.52	0.72	0.94	0.52	0.50	0.53	0.001

Total	148	40	11.94	2.01	2.56	5.77	0.82	0.96	0.39	0.57	0.52	0.001
Mefenoxam test	***N***	***MLG***	***eMLG***	***SE***	***H***	***G***	***lambda***	***lambda****	E_5	H_{exp}	r_d	$p.r_d$
Resistant	43	15	5.88	1.21	2.12	5.62	0.82	0.93	0.62	0.59	0.049	0.001
Sensitive	43	22	7.28	1.25	2.63	8.44	0.88	0.95	0.57	0.65	0.052	0.001
Intermediate	4	3	3.00	0.00	1.04	2.67	0.62	0.66	0.91	0.48	NaN	0.339
No tested	58	14	4.75	1.23	1.77	3.53	0.71	0.92	0.52	0.49	0.069	0.001
Total	148	40	6.18	1.36	2.56	5.77	0.82	0.96	0.39	0.57	0.051	0.001

Table 4-4. Analysis of molecular variance (AMOVA) for clone corrected populations of *G. cryptoirregulare*. Significance of variance was tested from 999 permutations of the data.

Source	**df**[a]	**SS**[b]	**MS**[c]	**Variance**	**P value**	**Nm**[d]
Between year	4	58.11	14.52	9.99%	0.001	1.81
Between isolates within year	45	183.20	4.07	0%		
Within year	50	298.66	5.97	90.01%		
Total	99	539.98		100%		
Source	**df**	**SS**	**MS**	**Variance**	**P value**	**Nm**
Between nursery	5	64.900	12.980	9.57%	0.001	1.94
Between isolates within nursery	51	197.170	3.860	0%		
Within nursery	57	338.820	5.940	90.43%		
Total	113	600.900		100%		
Source	**df**	**SS**	**MS**	**Variance**	**P value**	**Nm**
Between production type	2	19.150	9.570	2.64%	0.003	5.35
Between isolates within production type	49	226.880	4.630	0%		
Within production type	52	307.620	5.910	97.36%		
Total	103	553.660		100%		
Source	**df**	**SS**	**MS**	**Variance**	**P value**	**Nm**
Between mefenoxam	3	13.64	4.54	0.20%	0.195	11.02
Between isolates with mefenoxam	50	243.82	4.87	99.80%		
Within mefenoxam	54	323.69	5.99			
Total	107	581.17				

[a] = Degrees of freedom

[b] = Sum of squared differences

[c] = Mean of the squared differences

[d] = Nm: absolute number of migrants per generation, estimates gene flow from *Fst*. *Nm* > 1 = great gene flow; 0.5 < *Nm* > 1 = weak gene flow, but exchange of alleles may still occur; *Nm* < 0.5 = groups almost genetically isolated (Milgroom 2015).

Table 4-5. Pairwise *Fst* values between clone-corrected populations over five years.

	2013	**2014**	**2015**	**2016**	**2017**
2013	0				
2014	0.020	0			
2015	0.005	0.028	0		
2016	0.013	0.024	0.001	0	
2017	0.488*	0.399*	0.472*	0.434*	0

F_{ST} = fixation index. F_{ST} < 0.05 = low genetic differentiation; 0.05 to 0.15 = moderate; 0.15 to 0.25 = great; and, >0.25 = very great).

* = indicates significant differences.

Table 4-6. Pairwise *Fst* values between clone-corrected populations from six nurseries.

	A	**B**	**C**	**D**	**E**	**G**
A	0					
B	0.044	0				
C	0.012	0.041	0			
D	0.015	0.045	0.022	0		
E	0.015	0.051	0.037	0.003	0	
G	0.469*	0.288*	0.355*	0.474*	0.527*	0

F_{ST} = fixation index. F_{ST} < 0.05 = low genetic differentiation; 0.05 to 0.15 = moderate; 0.15 to 0.25 = great; and, >0.25 = very great).

* = indicates significant differences.

Table 4-7. Pairwise *Fst* values between clone-corrected populations from production types.

	Container	Propagation	Field
Container	0		
Propagation	0.020	0	
Field	0.041	0.097	0

F_{ST} = fixation index. F_{ST} < 0.05 = low genetic differentiation; 0.05 to 0.15 = moderate; 0.15 to 0.25 = great; and, >0.25 = very great).

Table 4-8. Pairwise *Fst* values between clone-corrected populations from results of the mefenoxam test.

	Resistant	Sensitive	Intermediate	No tested
Resistant	0			
Sensitive	0.002	0		
Intermediate	0.004	0.000	0	
No tested	0.038	0.030	0.000	0

F_{ST} = fixation index. F_{ST} < 0.05 = low genetic differentiation; 0.05 to 0.15 = moderate; 0.15 to 0.25 = great; and, >0.25 = very great).

Figures

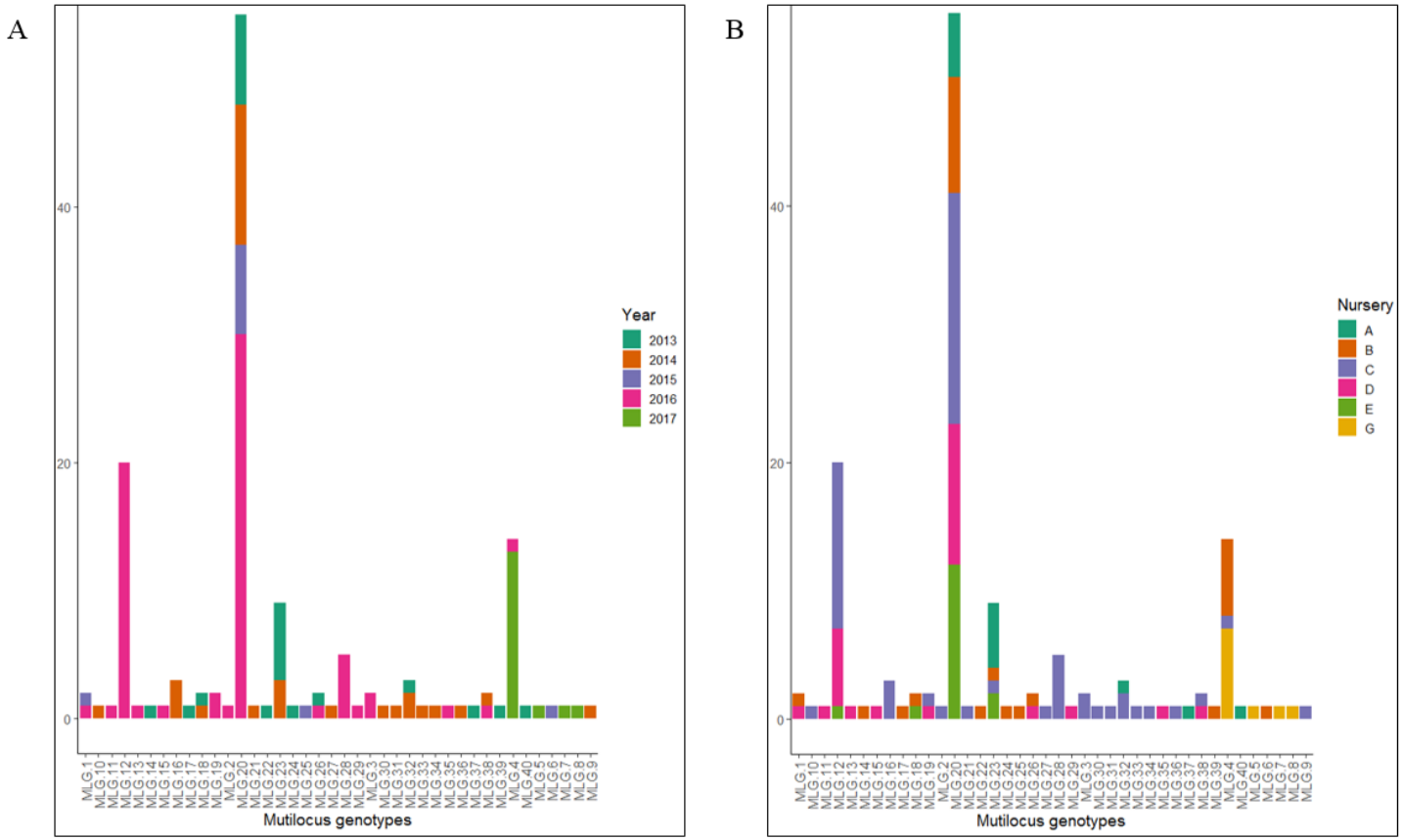


Figure 4-1. Frequency of multilocus genotypes (MLG) detected: **A**. by year; **B**. by nursery.

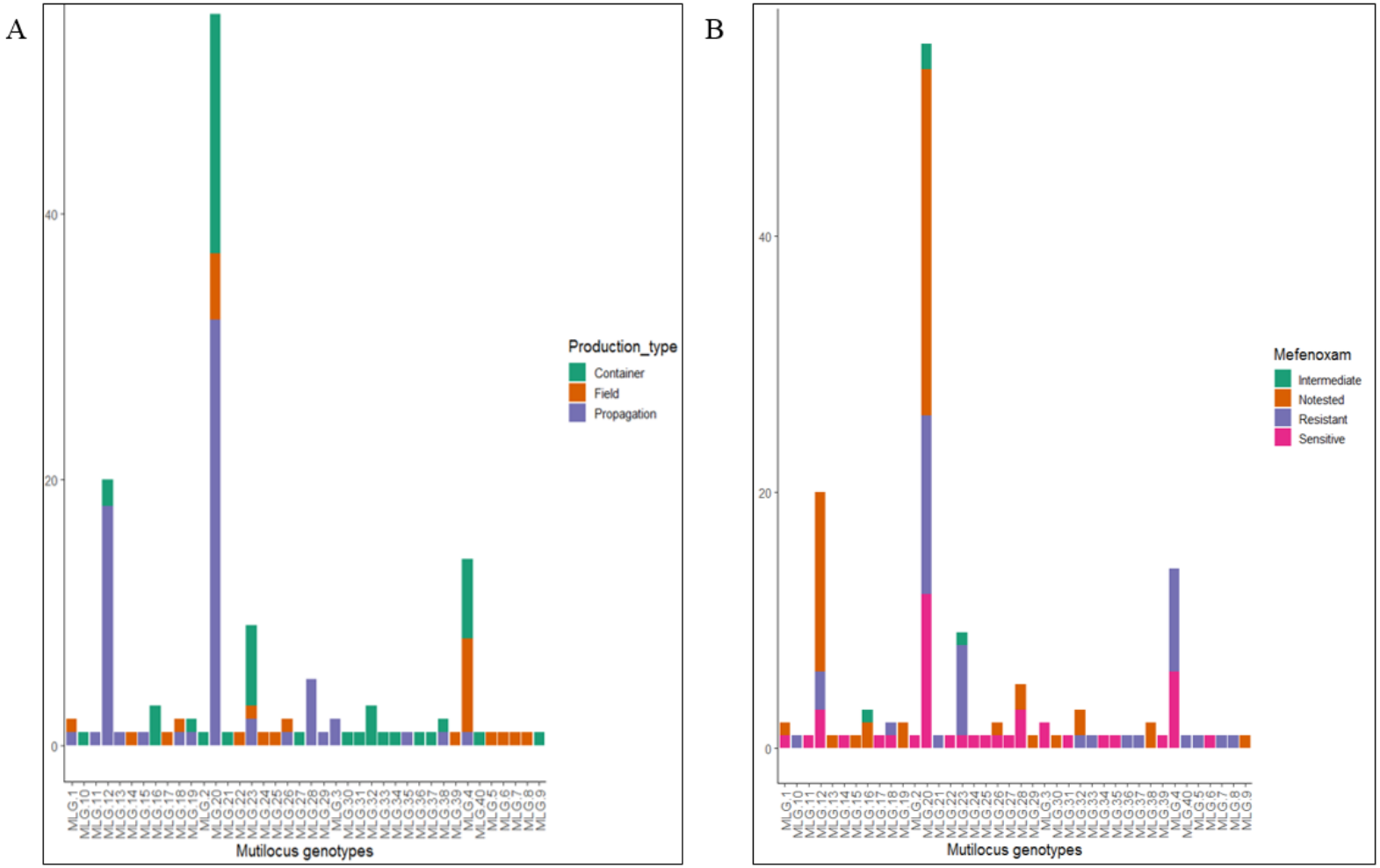


Figure 4-2. Frequency of multilocus genotypes (MLG) detected: **A**. by production type; **B**. by mefenoxam sensitivity.

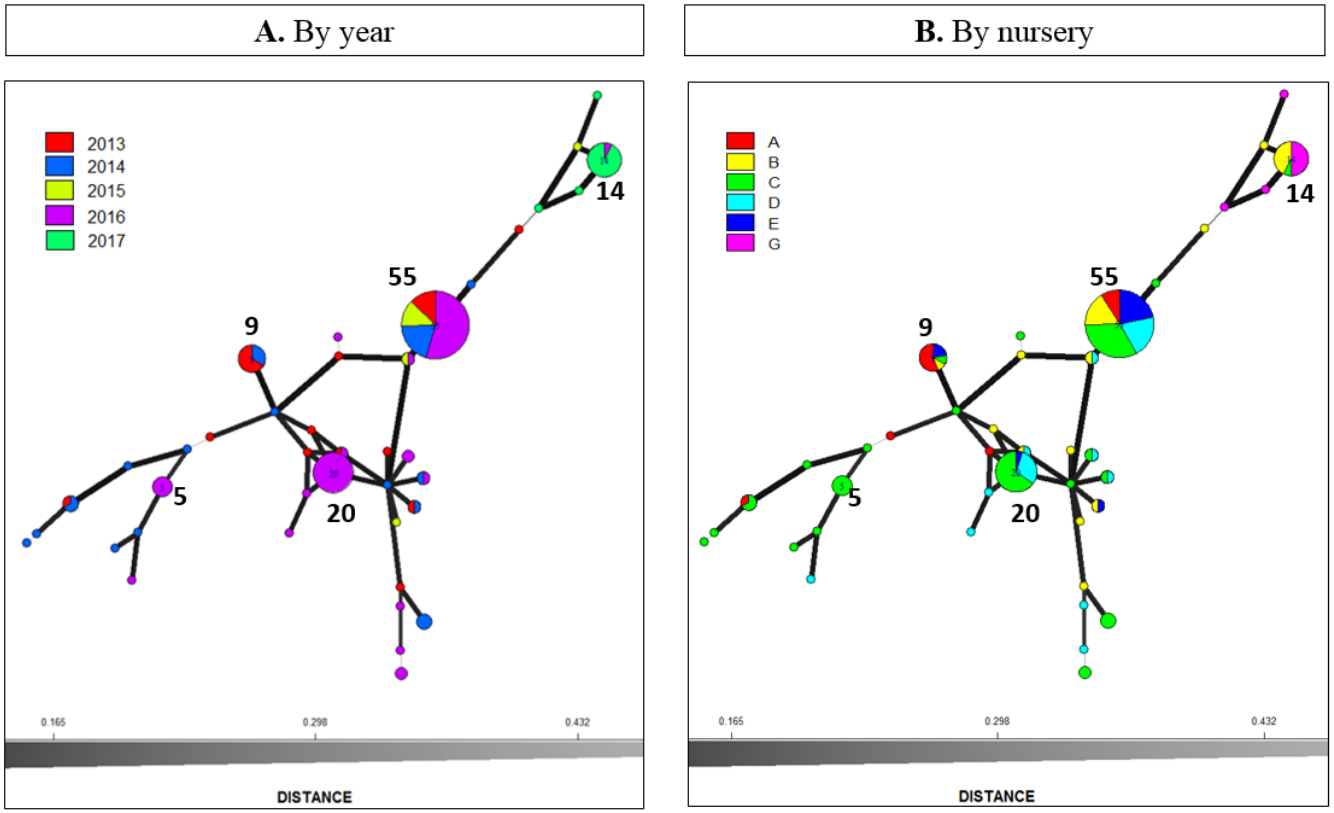


Figure 4-3. Minimum spanning network based on Bruvo's genetic distance for microsatellites markers for *G. cryptoirregulare* populations (A: year; B: nursery). Nodes (circles) represent individual multilocus genotypes. The size of the circle is relative to the number of individuals represented in the data. Nodes more closely related have darker and thicker lines whereas nodes more distantly related have lighter and thinner lines. The 5 most abundant multilocus genotypes are labeled with the number of individuals.

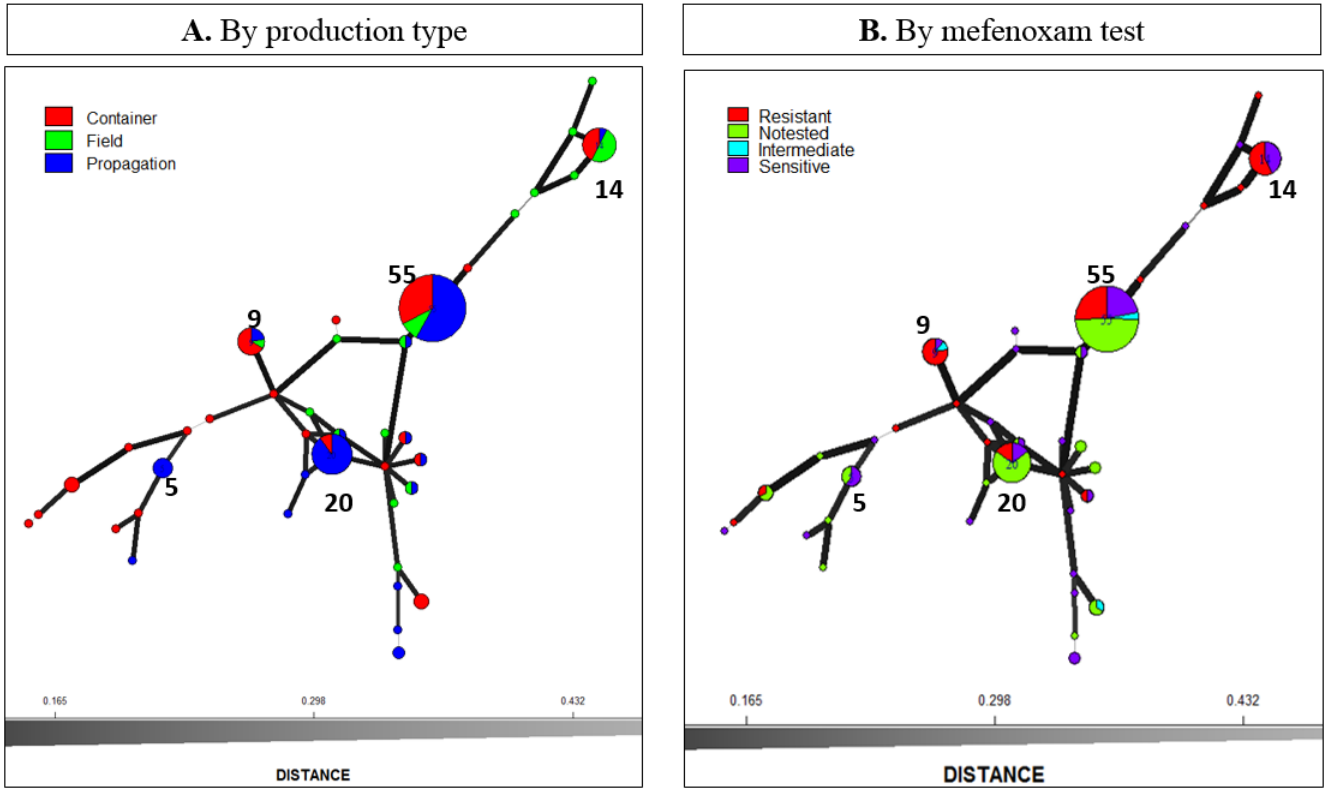


Figure 4-4. Minimum spanning network based on Bruvo's genetic distance for microsatellites markers for *G. cryptoirregulare* populations (A: production type; B: mefenoxam test; C: production type). Nodes (circles) represent individual multilocus genotypes. The size of the circle is relative to the number of individuals represented in the data. Nodes more closely related have darker and thicker lines whereas nodes more distantly related have lighter and thinner lines. The 5 most abundant multilocus genotypes are labeled with the number of individuals.

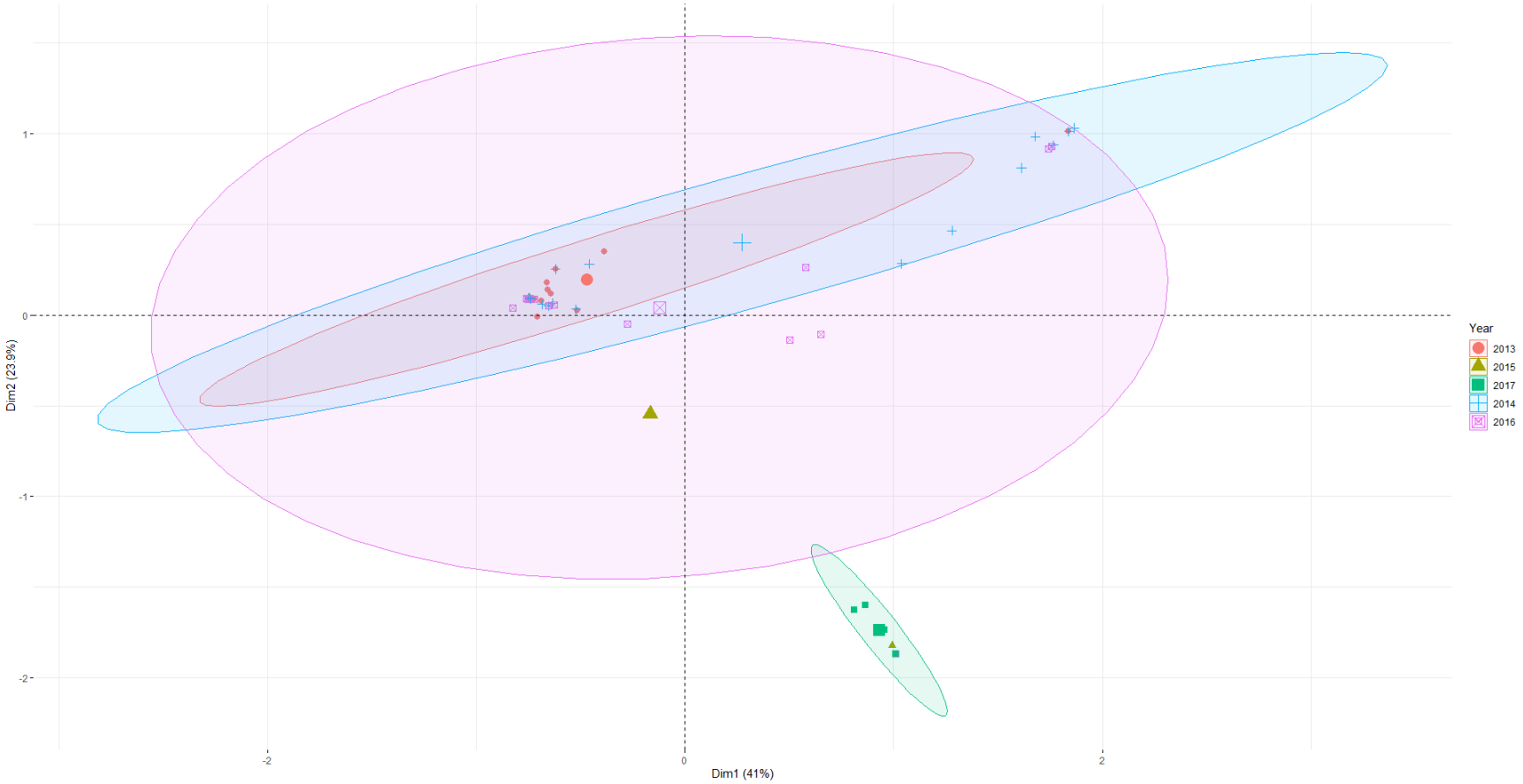


Figure 4-5. Principle component analysis of the *G. cryptoirregulare* isolates over five years. Each axis is labeled with the percentage of total variance explained by that dimension.

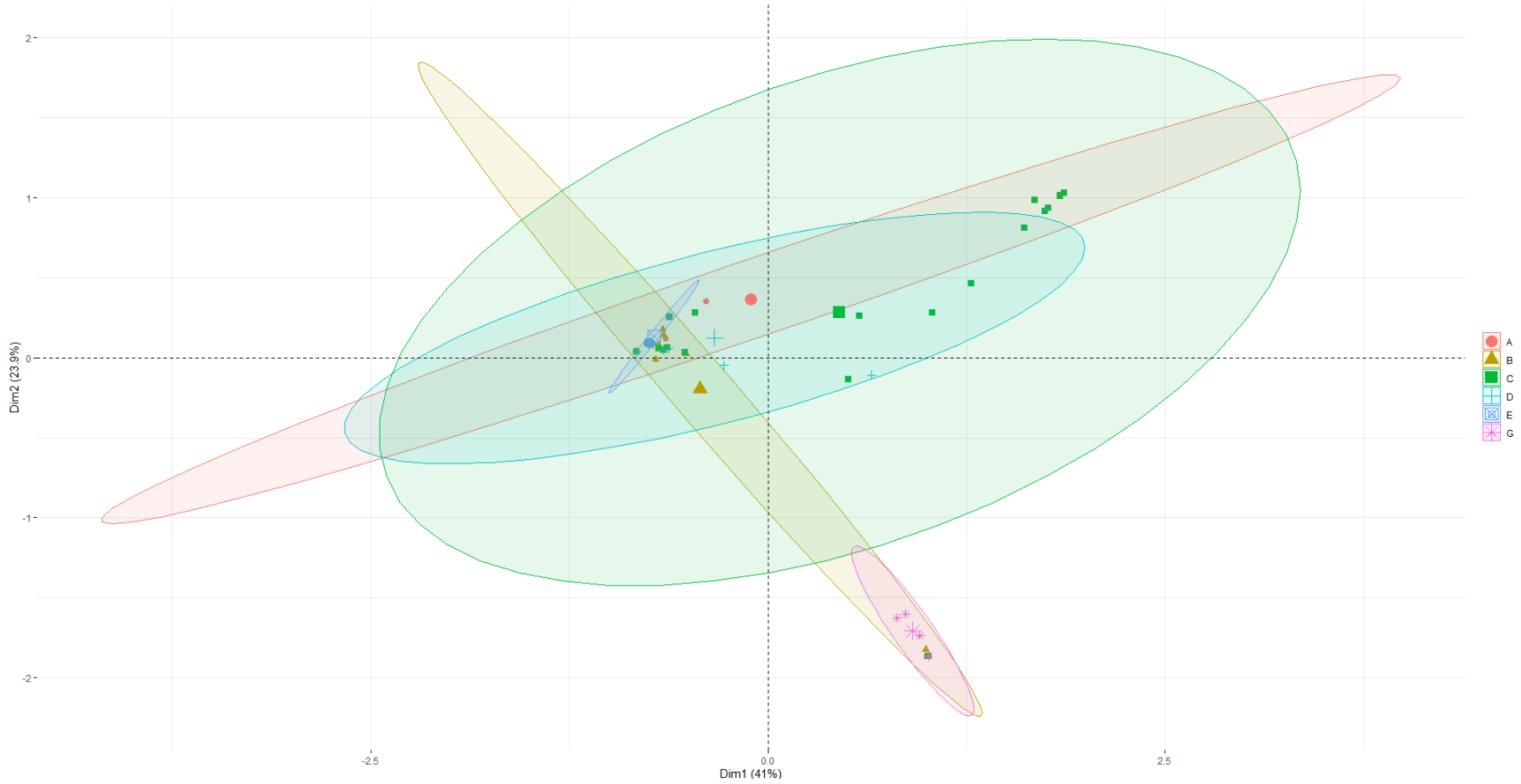


Figure 4-6. Principle component analysis of the *G. cryptoirregulare* isolates in six nurseries. Each axis is labeled with the percentage of total variance explained by that dimension.

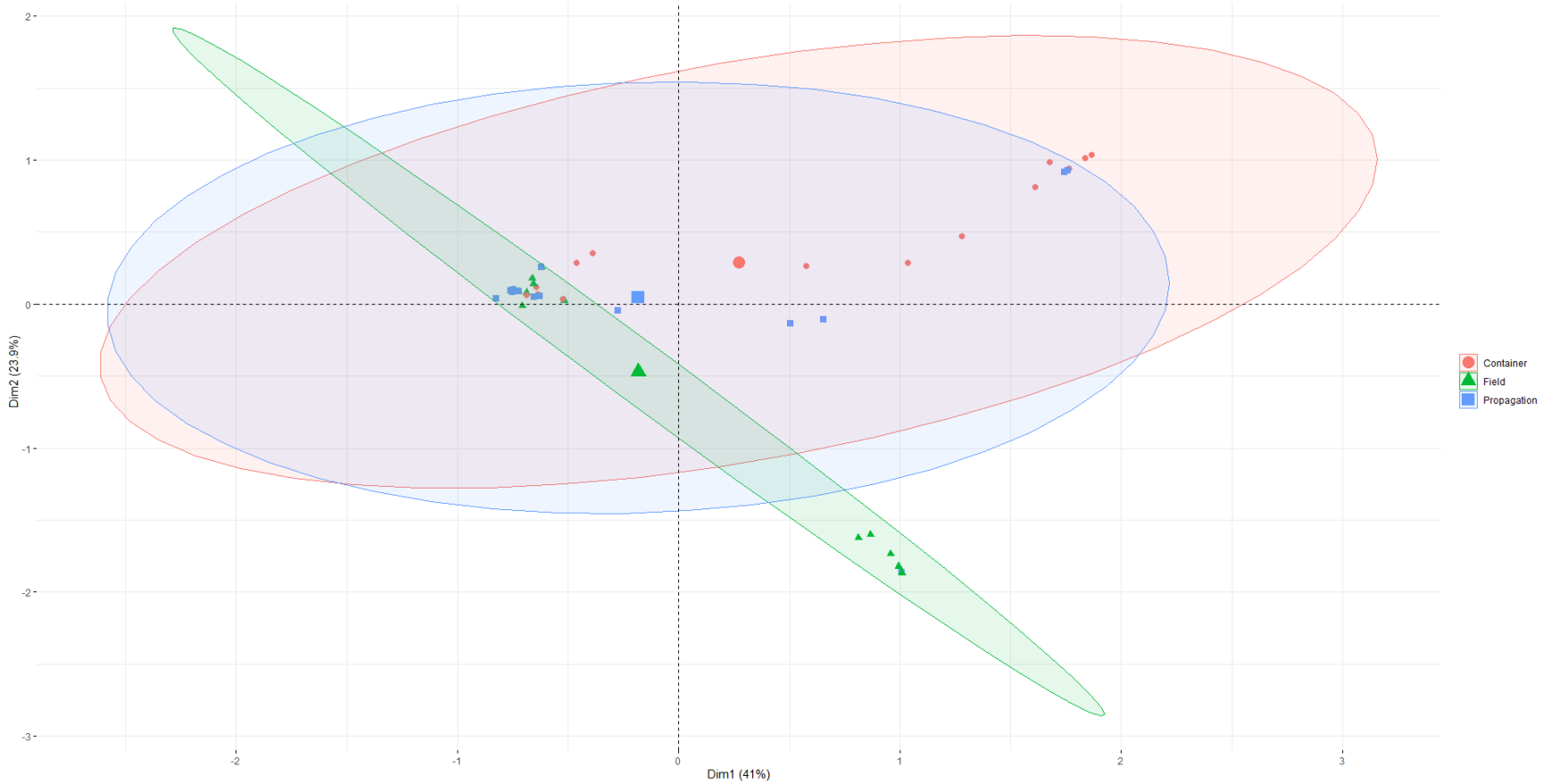


Figure 4-7. Principle component analysis of the *G. cryptoirregulare* isolates in the production systems. Each axis is labeled with the percentage of total variance explained by that dimension.

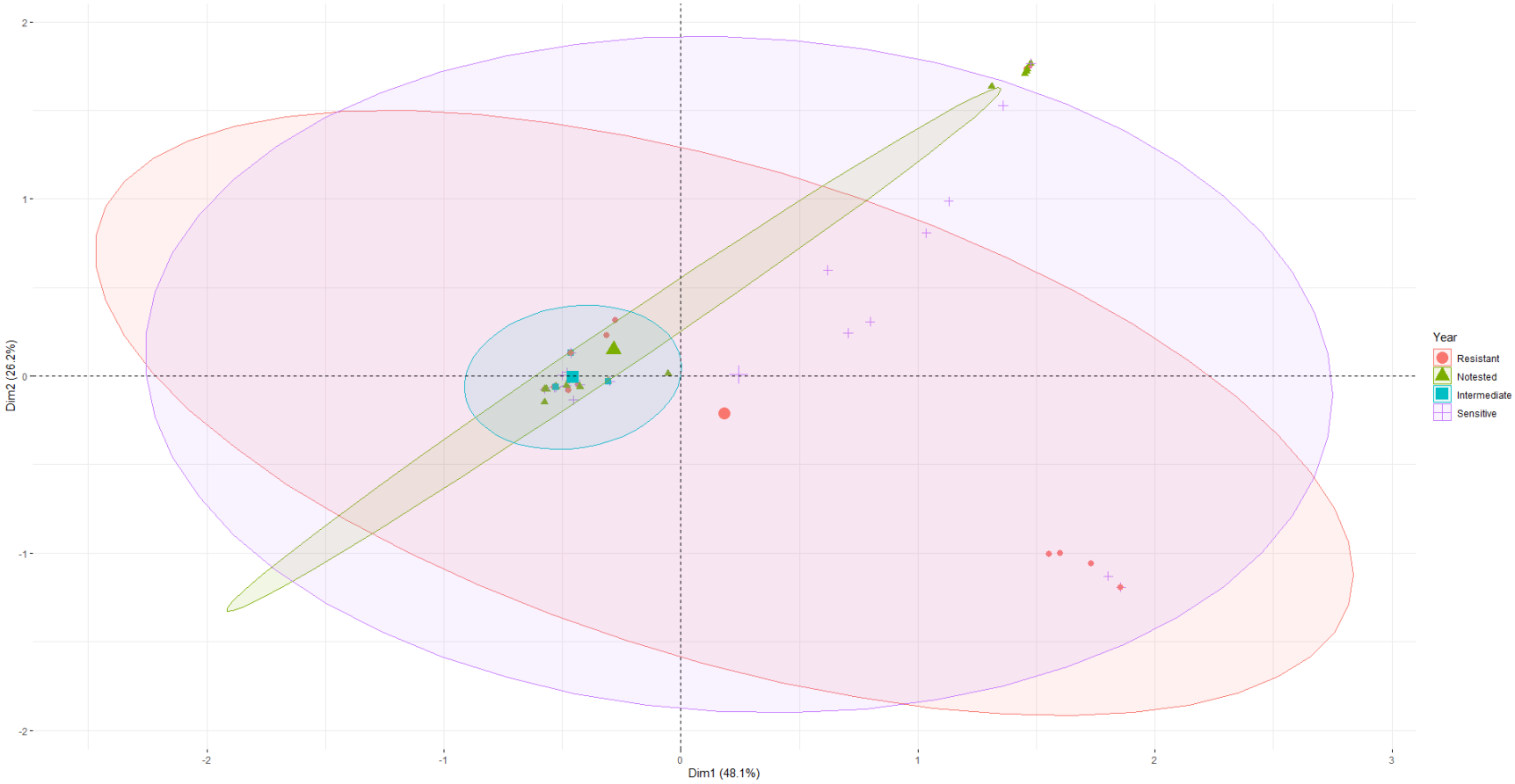


Figure 4-8. Principle component analysis of the *G. cryptoirregulare* isolates by mefenoxam test. Each axis is labeled with the percentage of total variance explained by that dimension.

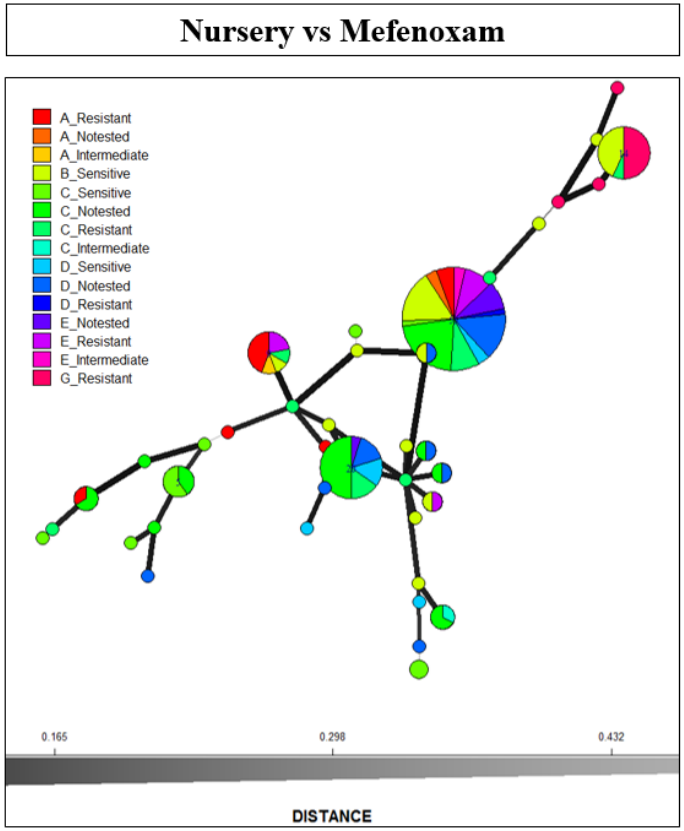


Figure 4-9. Minimum spanning network based on Bruvo's genetic distance for microsatellites markers for *G. cryptoirregulare* populations (by nursery and mefenoxam test). Nodes (circles) represent individual multilocus genotypes. The size of the circle is relative to the number of individuals represented in the data. Nodes more closely related have darker and thicker lines whereas nodes more distantly related have lighter and thinner lines.

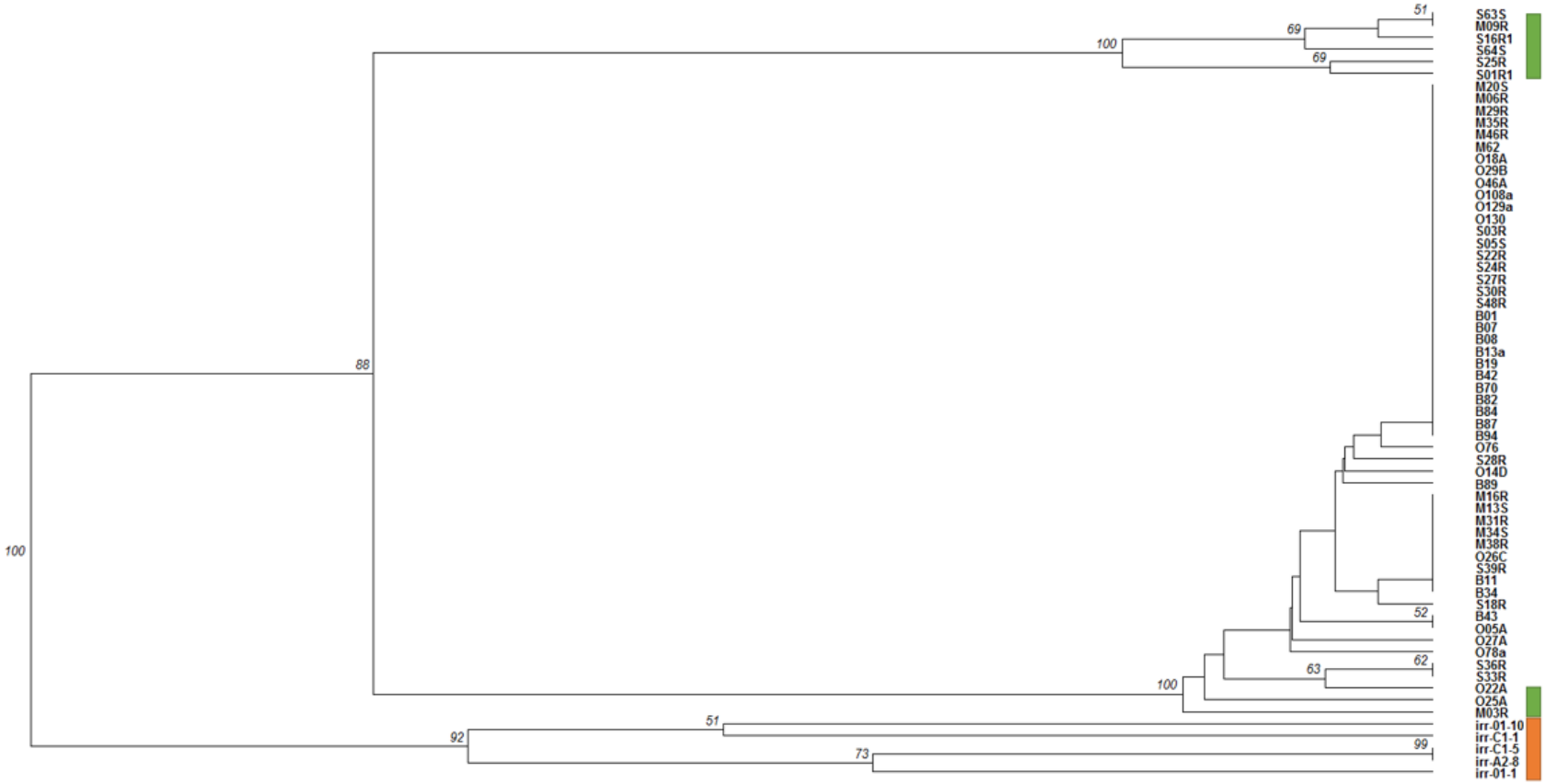


Figure 4-10. Relationships among populations of *G. cryptoirregulare*. The neighbor-joining tree was constructed in R from 8 SSR markers in 64 isolates. Reference isolates from *G. irregulare* are colored in orange. Samples that gave ambiguous identifications based on ITS sequencing are colored green.

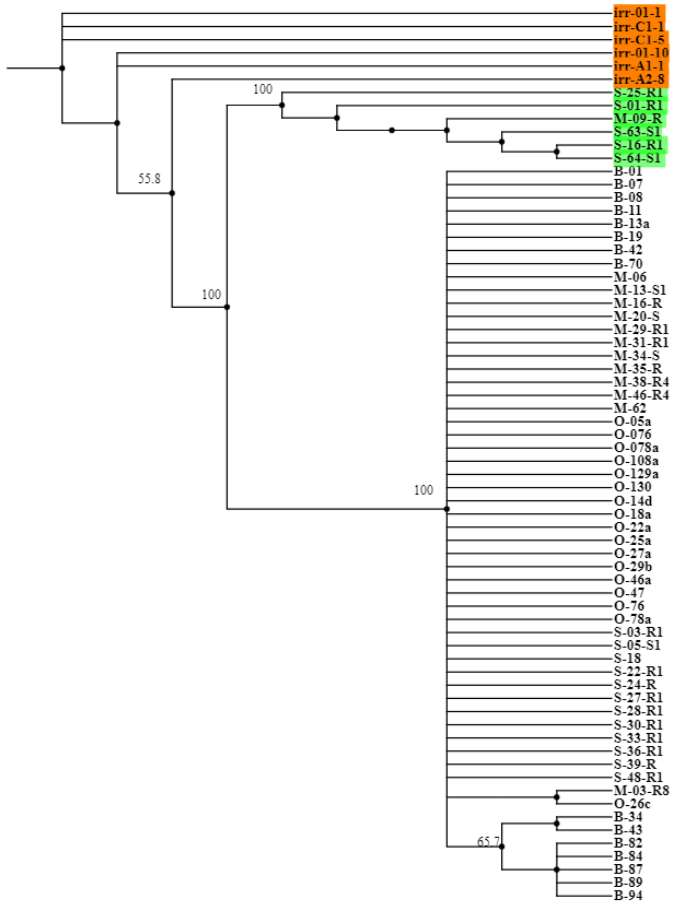


Figure 4-11. Relationships among populations of *G. cryptoirregulare*. The neighbor-joining tree was constructed in R from 663 SNPs in 64 isolates. Reference isolates from *G. irregulare* are colored in orange. Samples that gave ambiguous identifications based on ITS sequencing are colored green.

CHAPTER V

DETECTION OF MULTIPLE OOMYCETES IN METAGENOMIC DATA BY USING E-PROBE DETECTION OF NUCLEIC ACID ANALYSIS (EDNA)

Abstract

Species identification of plant pathogenic oomycetes based on morphology is challenging because of the limited number of variable characters available. Hence, time-demanding pure culture isolation and ITS sequencing are routinely conducted for diagnostic purposes. E-probe Diagnostic Nucleic acid Analysis (EDNA), a bioinformatic pipeline that couples high-throughput sequencing (HTS) and metagenomic sequence analysis, was used to detect a variety of plant pathogenic oomycetes in metagenomic data. The EDNA pipeline, implemented through the new Microbial Finder (MiFi^{TM}) graphical user interface, was used to design short pathogen-specific sequences (e-probes) with a range of sizes from 40 to 80 nucleotides, and tested on metadata containing a known host and pathogenic oomycete genomes. To avoid false positives and increase target specificity, e-probes with similarity to sequences of non-target species in NCBI's nucleotide database were removed. The designed e-probes were tested *in silico* using HST and MiFi^{TM} on

spiked metagenomes with different target pathogen and host sequence levels. *In vitro* testing was performed using nanopore sequencing (MinION™; Oxford Nanopore Technologies) coupled with MiFi™ to analyze sequences of metasamples prepared using genomic DNA of tomato (*Solanum lycopersicum*) spiked with different percentages of *Phytophthora nicotianae* genomic DNA, including a negative control (healthy tomato DNA). A metasample was considered positive for a pathogen's presence when a significant number of e-probe hits were found in a metagenome. EDNA detected the target pathogens in metadata with high accuracy. Thus, it may be concluded that MiFi™ it is a powerful tool for effective detection of oomycetes from metasamples.

1. Introduction

Oomycetes are diploid eukaryotic organisms that include hundreds of plant pathogenic species that cause diverse diseases and significant losses across a wide range of crops worldwide (Agrios 2005). Conventional techniques to identify and detect plant pathogenic oomycetes have relied upon morphological, serological, and molecular approaches, which are often time-consuming and can simultaneously identify a limited number of pathogens (Schroeder et al. 2013; Wakeham and Pettitt 2017). Advance in DNA sequencing has facilitated the identification of oomycetes, particularly the PCR-based technologies that use DNA barcodes such as the ITS (internal transcribed spacer) of the nuclear ribosomal DNA and/or the mitochondria-encoded cytochrome c oxidase subunits I and II (*Cox*1, *Cox*2) (Robideau et al. 2011; Schroeder et al. 2013; White et al. 1990). However, individual DNA barcodes cannot consistently differentiate between closely related species and locus specific primers often provide amplification across a

broad range of groups (Malapi-Wight et al. 2016; Schroeder et al. 2013). Hence, fast and accurate methods for detecting and discriminate among multiple pathogenic oomycetes are required to implement effective disease management (Agrios 2005; Arafa and Shirasawa 2018; Schroeder et al. 2013).

High Throughput Sequencing (HTS) allows the massively parallel sequencing of DNA. Its development has made more feasible and affordable the sequencing of many organisms, including oomycetes (Kircher et al. 2011; Van Dijik et al. 2014). HTS and metagenomic diagnostic methods provide tools for identifying multiple pathogens from a single sample without isolating pure cultures or specific gene-targeted sequencing (Adams et al. 2009; Mardis 2008). E-probe Diagnostic Nucleic acid Analysis (EDNA) is a computational pipeline that couples metagenomics and HTS to detect plant pathogens (Espíndola et al. 2015; Stobbe et al. 2013). EDNA has proven ability to detect viral, bacterial, fungal, and oomycete pathogens in simulated data sets from host metagenomes (Espíndola et al. 2015; Stobbe et al. 2013, 2014). The Microbial Finder (MiFi^{TM}) graphical user interface (Espíndola, not published), implements the EDNA pipeline and includes the *MiProbe* tool, for the identification and selection of pathogen-specific short sequences named e-probes, as well as e-probe database curation; and the MiDetect tool for the detection and identification of target organisms from metagenomic samples.

This study's objective was to use the EDNA approach as implemented through the MiFi^{TM} platform to develop a validated database for oomycete diagnostics. E-probes were designed using genomes generated from public databases and unreleased drafts genomes (*Globisporangium irregulare* and *G. cryptoirregulare)* for eighteen plant pathogenic

oomycetes selected based on their scientific and economic importance (Kamoun et al. 2015; Kerio et al. 2019). The curated e-probes were used for *in silico* detection of the oomycetes of interest using HTS and MiFi^{TM}. Later, *in vitro* validation was conducted using Nanopore sequencing (Oxford Nanopore Technologies) coupled with MiFi^{TM} for the detection of *Phytophthora nicotianae* in spiked tomato (*Solanum lycopersicum*) DNA metasamples.

2. Materials and Methods

2.1. Oomycete pathogens and genome sequences

The organisms studied here represent various oomycete pathogens with scientific and economic importance (Kamoun et al. 2015; Kerio et al. 2019). Each target pathogen was selected along with phylogenetically close-relative organisms and a plant host genome used as background (Table 5-1 and Table 5-2). The genome sequence of each oomycete pathogen and their host were obtained from the National Center for Biotechnology Information (NCBI) genome database, GenBank, or generated de novo if needed. *G. irregulare* and *G. cryptoirregulare* genomes were contributions by Dr. Hai Nguyen from the Ottawa Research and Development Center, Agriculture and Agri-Food in Canada.

2.2. *In silico* assessment of EDNA for the detection of oomycete plant pathogens

2.2.1. E-probe design

E-probes were designed with lengths of 40, 60, and 80 nucleotides (nt) using the MiProbe tool of MiFi^{TM} (Figure 5-1). E-probes were designed using a modified version of the Tools for Oligonucleotide Fingerprint Identification (TOFI) software. TOFI

analysis consisted of three stages: pairwise comparison of the target genome with non-target genomes, identification of candidate e-probes, and a Basic Local Alignment Search Tool (BLAST) search using the NCBI's Nucleotide database to keep only target-specific e-probes (Espíndola et al. 2015; Stobbe et al. 2013; Vijaya et al. 2008). We used one oomycete genome as the target and near neighbor genomes that consisted of a multiFASTA file with all the concatenated neighbor genome sequences (Table 5-1). Neighbors were chosen using the phylogenomic analyses of oomycete species by McCarthy and Fitzpatrick (2017) and McGowan and Fitzpatrick (2020).

The resulting e-probes went through a manual curation step comparing them against the NCBI nucleotide database at an e-value of $1x10^{-9}$ to eliminate e-probes hitting non-target pathogens. The curation was conducted using resources available through the High-Performance Computing Center at Oklahoma State University. The species-specific curated e-probe datasets were uploaded to MiFi^{TM} platform.

2.2.2. Simulation of Mock Sample Sequencing Databases (MSSDs)

MSSDs were generated at known percentages of the targeted oomycete reads to the background host reads (Table 5-2 and Table 5-3) with the software MetaSim® (Richter et al. 2008). Metasim simulated ten million read MSSDs using an algorithm to mimic the Illumina sequencing platform's (Illumina, San Diego, CA) nucleotide substitution error rate. Mock databases containing only the pathogen genome reads were the positive controls. Mock databases containing only the host genome reads were the negative

controls. Three replicates for each level of pathogen abundance (Table 5-3) were simulated and uploaded to the MiFi^{TM} platform.

2.2.3. Detection with MiFi^{TM}

The targeted pathogen's presence was detected when the pathogen-specific e-probes aligned with the MSSDs (hits). Detection was conducted with the *MiDetect* tool of MiFi^{TM}. *MiDetect* conducted pairwise sequence alignments between e-probes and MSSDs using BLASTn with an e-value of $1x10^{-9}$. E-probe hit scores on parsed alignments of metagenomes were compared with decoy e-probe hit scores by pairwise T-student tests. Decoy e-probe sets are the reverse sequences of the final e-probe set, used as internal control for random matching, to avoid false positives within the metagenomes (Stobbe et al. 2013). Scores of e-probes sets and decoy sets were compared, and the statistical support of a decision was made based on the p-value: positive (p-value ≤ 0.05) and suspect or negative (p-value ≥ 0.05). Although the user can define the e-value and number of hits per quantification, the recommended $1x10^{-9}$ p-value and 250 hits were maintained (Espíndola et al. 2015; Stobbe et al. 2013).

2.2.4. Sensitivity analysis

Sensitivity tests were conducted based on e-probe lengths, and on pathogen read abundances to identify optimal e-probe length and limit of detection for each pathogen. Sensitivity was calculated by registering the number of observations (up to 3) at each level of read abundance, with the three replicates, and using Equation 1:

$$\text{Sensitivity} = \frac{\text{number of observed positives}}{\text{total observed positives + false negatives}} * 100 \quad (1)$$

Observed positives were positive calls when the MSSDs contained target pathogen reads. False negatives were negative calls in MSSDs that were known to contain pathogen reads.

2.3. EDNA/MiFi validation using spiked tomato metagenome sequencing databases generated with the Oxford Nanopore Technology (ONT) MinION™ sequencer

2.3.1. Samples

The *Phytophthora nicotianae* - *Solanum lycopersicum* L. (tomato) pathosystem was selected for *in vitro* validation of the EDNA pipeline as implemented in MiFi. Three *Ph. nicotianae* isolates provided by The Plant Disease and Insect Diagnostic Laboratory (PDIDL) at Oklahoma State University (Table 5-4) were plated on modified PARP medium [pea agar medium (120g/L peas; 15 g/L agar) with, pimaricin (0.4ml/L), ampicillin (250mg/L), rifampicin (1ml/L), and PCNB (5ml/L)]. Mycelia from solid media were transferred to pea broth for two weeks at room temperature for mycelia growth. Mycelia mats were harvested, placed in 1.5 ml microtubes, frozen at -80 ºC overnight, and lyophilized.

2.3.2. DNA extraction and species identification

DNA was extracted from tomato cv 'Celebrity' leaf tissue and *Ph. nicotianae* mycelia using the Qiagen DNeasy Plant Mini kit (Germantown, MD) with modifications (Abad and Bienapfl 2019). DNA concentration was determined using a Nanodrop

spectrophotometer ND 1000 (Thermo Fisher Scientific, Waltham, MA, USA). Total double-stranded DNA concentration was assessed with the Qubit 3.0 fluorimeter (Thermo Fisher Scientific) using the Qubit ® dsDNA BR (Broad Range) assay kit. The integrity of the genomic DNA was verified by electrophoresis in a 0.8% agarose gel.

Sample identity was confirmed by PCR and Sanger sequencing of the ITS region and the COI (cytochrome c oxidase subunit I) gene. DNA amplification of the ITS region was performed with the tomato samples and *Ph. nicotianae* isolates using primers ITS 5 (5'-GGAAGTAAAAGTCGTAACAAGG-3') and ITS 4 (5'-TCCTCCGCTTAYYGATATGC-3'), and following previously reported PCR conditions (White et al. 1990). Amplification of the COI gene was performed only on *Ph. nicotianae* isolates using primers COIF-1 (5'-TCAWCWMGATGGCTTTTTTCAAC-3') and COIR-1 (5'-RRHWACKTGACTDATRATACCAAA-3') and following previously reported PCR conditions (Robideau et al. 2011). Sequencing was performed using the same primers in the initial PCR steps. Sequencing reactions products were run on an ABI 3100 DNA sequencer (Applied Biosystems) at the Recombinant DNA and Protein Core Facility at Oklahoma State University.

2.3.3. DNA spiking

The DNA spiking experiment was conducted to simulate conditions where the oomycete DNA was present in small concentrations in an infected plant host's metagenomic DNA. DNA stock solutions of *Ph. nicotianae* and tomato were prepared at concentrations of 100 ng/µL. Serial dilutions of *Ph. nicotianae* DNA diluted in tomato

DNA dilutions were prepared 10 µL volumes, at a total concentration of 100 ng/µL. Dilutions contained in total 100 ng, 10 ng, or 1 ng *Ph. nicotianae* DNA, and 900 ng, 990 ng, and 999 ng of tomato DNA respectively. This generated pathogen to host DNA ratios of 1:10, 1:100, and 1:1,000 (Figure 5-2) to simulate the detection limits in-silico. The two lower dilutions were used in subsequent assays due to budget restrictions.

Two controls were included in the experimental design. First, a positive control containing 100% *Ph. nicotianae* DNA (100 ng μL^{-1}) was equivalent to 100% pathogen read abundance *in silico.* Second, the negative control, containing 100% tomato DNA (100 ng µL-1), was equivalent to 0% pathogen abundance *in silico*. The first experimental "treatment" was the 1:100 pathogen/host DNA dilution (1% pathogen abundance). The second experimental treatment was the 1:1,000 pathogen/host DNA dilution (0.1 % abundance) (Table 5-5). Three biological repetitions were sequenced for each control and treatment.

2.3.4. Nanopore library preparation and sequencing

Nanopore sequencing was performed on the 12 samples described above (2 controls, 2 pathogen: host dilutions, 3 biological replicates each). Metagenomic libraries were prepared using Oxford SQK-RBK004 rapid barcoding sequencing kits following the manufacturer's instructions (Oxford Nanopore Technologies, Oxford, UK). DNA was purified using Agencourt AMPure XP beads (Beckman Coulter, Indianapolis, IN), using 600 ng per sample as input to generate MinION™ Oxford Nanopore libraries. Six barcoded libraries were pooled before adding the adapters. Libraries were stored at -20 °C until they were loaded into MinION flow cells (FLO-MIN106D R9; Oxford Nanopore

Technologies, Oxford, UK). Two flow cells, each with six barcoded libraries (Table 5-5), were loaded with 75 µL of sequencing mix, and assembled in the MinION™ sequencer (MK 1B version; Oxford Nanopore Technologies, Oxford, UK). Sequencing runs were executed with MinKNOW UI (version 20.06.4) software for a total of 16 hours.

2.3.5. Detection using MiFi™ oomycete database

The raw fast5 MinION reads were base called and demultiplexed using Guppy version 3.2.4 (Oxford Nanopore Technologies, Oxford, UK) to produce fastq files. The fastq files were merged into one for each metagenome/barcode (n=12) and uploaded to MiFi™ platform. The metagenomes (Table 5-5) were queried using the curated e-probes. The recommended e-value and number of hits for quantification $1x10^{-9}$ p-value and 250 hits, respectively, were used, as done previously for the *in silico* validation. Detection was described as mentioned above to determine the sensitivity and detection limits.

2.4.EDNA validation with published SRA data

The SRX3160109 biosample (Espíndola, not published) deposited in the Sequence Read Archive (SRA) was retrieved from NCBI and was queried using the e-probe sets designed for *Phytophthora ramorum*. The original sample, sequenced and submitted by Espíndola (2015), consisted of Rhododendron leaves infected with *Phytophthora ramorum*. The metagenomic raw fastq file from SRA was directly uploaded to MiFi platform, and the detection was conducted with the MiDetect tool. The metagenome was also queried against the e-probe sets for all the other available *Phytophthora* species to assess their specificity.

3. Results

3.1. E-probe design

The number of e-probes designed at each length (40 nt, 60 nt, and 80 nt) varied between the pathogens (Figure 5-3 and Figure 5-4). The number of e-probes increased as the e-probe length decreased. E-probes of 40 nt in length produced the highest number of e-probes, while the e-probes of 80 nucleotides produced the lowest number of e-probes.

For 40 nt e-probes, *Ph. sojae* and *G. irregulare* generated the highest and smallest numbers of e-probes, respectively (Figure 5-3 and Figure 5-4). For 60 nt e-probes, *Ph. pinifolia* and *Ph. fragariae* generated the highest and smallest numbers of e-probes, respectively (Figure 5-3 and Figure 5-4). For 80 nt e-probes, *Ph. infestans* and *G cryptoirregulare* generated the highest and smallest numbers of e-probes, respectively (Figure 5-3 and Figure 5-4).

3.2. *In silico* detection with MiFi™

MiFi detected all the pathogens in the corresponding 10 million read MSSD using the species-specific e-probe sets designed. However, the detection limit varied depending on the e-probe length and the pathogen. For 40 nt e-probes, the limits of positive detection were in the range of 0.001% - 0.01% of pathogen read abundance. Nine of the pathogens (*H. arabidopsis, Ph. agathidicida, Ph. infestans, Ph, nicotianae, Ph. kernoviae, Ph. pinifolia, Ph. ramorum, Ph. melonis,* and *G. irregulare*) were detected at 0.001% of pathogen read abundance, which means that at least 100 pathogen reads were present in the MSSDs queried with the e-probe sets (Figure 5-5 and Figure 5-6). The other nine

species (*Ph. fragariae, Ph. lateralis, Ph. sojae, Ph. cambivora, Ph. citricola, Pl. viticola, Py. aphanidermatum, Py. oligandrum,* and *G. cryptoirregulare*) were detected at 0.01% pathogen read abundance, which means that at least 1,000 pathogen reads were present in the MSSDs queried with the e-probe sets (Figure 5-5 and Figure 5-6).

For 60 nt e-probes, the limits of positive detection were in the range of 0.001% - 0.1% of pathogen read abundance. *Ph. infestans* was the only pathogen detected at 0.001% pathogen read abundance (100 reads present). Most of the pathogens (*Ph. agathidicida, Ph. lateralis, Ph. sojae, Ph. nicotianae, Ph. pinifolia, Ph. melonis, Ph. ramorum, Ph. kernoviae, Ph. citricola, Pl. viticola, Py. oligandrum,* and *G. irregulare*) were detected at 0.01 % of pathogen read abundance (Figure 5-5 and Figure 5-6). The other five pathogens (*H. arabidopsis, Ph. fragariae, Ph. cambivora, Py. aphanidermatum* and *G. cryptoirregulare*) were detected at 0.1 % of pathogen read abundance, which means that at least 10,000 pathogen reads were present in the MSSDs queried with the e-probe sets (Figure 5-5 and Figure 5-6).

For 80 nt e-probes, the limits of positive detection were in the range of 0.01% - 1% of pathogen read abundance. *Ph. infestans* and *Ph. nicotianae* were detected at 0.01%. Most of the pathogens (*Ph. agathidicida, Ph. lateralis, Ph. sojae, Ph. pinifolia, Ph, melonis, Ph. kernoviae, Ph. citricola, Py. aphanidermatum, Py. oligandrum, G. irregulare,* and *G. cryptoirregulare*) were detected at 0.1 % of pathogen abundance (Figure 5-5 and Figure 5-6). The other five pathogens (*H. arabidopsis, Ph. fragariae, Ph. cambivora, Ph. ramorum,* and *Pl. viticola*) were detected at 1 % of pathogen abundance, which means

that at least 100,000 pathogen reads were present in the MSSDs queried with the e-probe sets (Figure 5-5 and Figure 5-6).

Detection at very low pathogen read abundance (0.0001 % = at least ten pathogen reads in the MSSD) was not achieved. The positive and negative controls gave the expected diagnostic results at all the pathogen or host abundances tested.

3.3. *In silico* sensitivity

In silico sensitivity decreased as the e-probe length increased. A sensitivity of 100 % was obtained in the range of 0.01 % – 0.1 % of pathogen read abundance (1,000-10,000 pathogen reads in the metagenomes) when the 40 nt e-probes were used. A sensitivity of 100 % was achieved in the range of 0.01 % – 1 % of pathogen reads abundance (10,000-100,000 pathogen reads in the metagenomes) when using the 60 nt e-probes. In the case of 80 nt e-probes, a sensitivity of 100 % was obtained in all cases at 1 % of pathogen reads abundance (at least 100,000 pathogen reads in the metagenomes) only (Figure 5-7).

Sensitivity at very low pathogen read abundances (0.001 %: -0.0001%; at least 1,000 pathogen reads in the metagenomes) was not obtained (Figure 5-7).

3.4. *In vitro* detection and sensitivity with Oxford Nanopore Sequencing and MiFi™

Metagenomes generated by MinION sequencing had different read numbers, but all of them produced the long reads expected with Nanopore sequencing technology (Table 5-6). MiFi™ detected the presence of *Ph. nicotianae* in the metagenomes using the e-probes sets designed. Positive diagnostic calls were achieved with all the e-probe sets (40

nt, 60 nt and 80 nt) in the positive control and the two dilution treatments (1 % abundance and 0.1 % abundance). However, the number of matches and the p-values decreased with the e-probe length (Table 5-7).

In the negative controls, expected negative diagnostic calls were obtained except for one replicate (metagenome 12) that gave a positive call with 40 nt e-probes (Figure 5-8). Negative controls gave suspect results since matches were found in the metagenomes when using e-probes of 40 nt and 60 nt (Table 5-7). No matches were found when the metagenomes (4, 8, and 12) were queried with e-probes of 80 nt (Table 5-7).

A sensitivity of 100 % was obtained in all the abundance tested levels except when using the 40 nucleotides e-probes with the negative controls. A 66.66 % sensitivity was obtained when the tomato metagenomes were queried with the 40 nt e-probe set (Figure 5-9).

3.5. EDNA validation with published SRA data

Phytophthora ramorum was detected in the SRA metagenome when queried with the 40 nt, 60 nt, and 80 nt e-probe sets. P-values and number of matches decreased as the e-probe length increased (Table 5-8). Other *Phytophthora* species were found when querying the metagenome with e-probe sets of 40 nt and 60 nt e-probes, which were considered false positives. Nonetheless, negative diagnostic calls were obtained for all the other *Phytophthora* species, except *Ph. ramorum*, when the metagenome was queried with the 80 nt e-probes (Table 5-8). No false negatives were observed with the 80 nt e-probes.

4. Discussion

Diseases caused by oomycetes are responsible for significant losses across a wide range of crops worldwide, highlighting the need for rapid and accurate identification of potential pathogens (Agrios et al. 2005; Arafa and Shirasawa 2018; Schroeder et al. 2013; Wakeham and Pettitt 2017). HTS has shown its applicability for identifying pathogens using whole genomes, but such applications are limited due to bioinformatic analyses required (Bronzato et al. 2018; Chalupowicz et al. 2019). One promising approach that uses HTS and metagenomic sequencing data to detect pathogens is termed E-probe Diagnostic of Nucleic acid Analysis (EDNA). EDNA detects the presence or absence of target pathogens in shotgun metagenomic data by using signature pathogen-associated sequences, named e-probes (Stobbe et al. 2013). E-probes are short species-specific markers that match only the organism of interest in the metagenomic sequence data, reducing the computational time needed to detect pathogenic microbes (Blagden et al. 2016; Espíndola et al. 2015; Stobbe et al. 2013, 2014).

E-probes were designed and used to detect the presence or absence of the oomycete species in simulated and real raw sequence data. E-probes were designed in non-conserved regions of the targeted genomes and curated against phylogenetically related near neighbors (Stobbe et al. 2013). Stobbe et al. (2013) found that the number of e-probes at each length varied among the pathogens studied. This variation was due to the number of near neighbors' closeness, the e-probe length, and the target genome size (Espíndola et al. 2015). A previous study found that e-probe optimal sizes for fungal and

oomycete plant pathogens ranged from 40 nt to 80 nt (Espíndola et al. 2015), therefore, the e-probes designed in this study were in that range.

The EDNA pipeline implemented on MiFi^{TM} (EDNA-MiFi) platform successfully detected the oomycete species of interest in the simulated sequence datasets. However, the detection limit and sensitivity varied depending on the e-probe length and the pathogen. Overall, *in silico* detection with 100% sensitivity was obtained at 0.01 % - 1% pathogen read abundances with the three e-probe lengths tested. When there was a small number of e-probes, the chances to detect the pathogen were reduced. On the other hand, a large number of e-probes could increase data processing time (Espíndola et al. 2015). The number of matches (total number of e-probes that had one or more hits) varied depending on pathogen abundance and the number of e-probes, and was inversely correlated with the e-probe length. Stobbe et al. (2013) found similar results and suggested improving the limit of detection by adjusting the number of e-probes and their length. Therefore, e-probes using more than one neighbor genome were designed to generate databases with thousands of e-probes 40 and 60 nucleotides long, and with hundreds of 80 nucleotide long e-probes, to improve sensitivity, specificity, and data processing time.

Validation *in vitro* of EDNA - MiFi by coupling MiFi^{TM} with Nanopore MinION^{TM} sequencing successfully identified *Phytophthora nicotianae* in the sequenced metagenomes. Using the DNA barcoding sequencing kit of Oxford Nanopore Technologies (ONT) allowed to sequence multiple samples in one run. Even though sequencing produced an uneven number of reads per metagenome, this did not affect the

positive detection of *Ph. nicotianae*. The variability in the number of reads could result from failed barcode labeling and technical conditions that could be improved in future studies (Chalupowicz et al. 2019). After demultiplexing, the reads generated with the MinION™ sequencer were directly uploaded to MiFi™ and queried with the curated e-probe sets, followed by successful detection of the pathogen by MiDetect. These steps and outcome validated the hypothesis that EDNA could allow rapid detection that did not require post-sequencing processing or analysis of the metagenome (Stobbe et al. 2013; Espíndola et al. 2015).

MiDetect provided the expected positive calls with the studied pathogen abundances, including the positive controls. Nevertheless, further analysis should explore more pathogen abundance levels to find the average limit of detection and optimal e-probe length when using the MinIONTM sequencing specifically. The spiking experiment demonstrated that serial dilution of the pathogen DNA in host DNA could be used as a method to test pathogen detection using samples with progressively lower pathogen DNA. However, further tests must be conducted using lower pathogen DNA concentrations to determine the actual limit of detection of MiDetect.

As expected, matches were found in all positive calls except for ambiguous results when the metagenomes that corresponded to the tomato samples (negative controls) were used. E-probes of 40 nt and 60 nt hit in metagenomes 4 and 12. Although the final diagnostic call was negative for the presence of *Ph. nicotianae* based on the statistical support, those matches were taken as suspect results. The presence of the pathogen was detected when hits were reported. Nonetheless, it was important to consider that each e-

probe could have multiple hits (matches). In this study, the limits of high-quality matches (HQM) to confirm a positive or negative call (Espíndola et al. 2015) were not addressed due to the limited sample size. Not surprisingly, the e-probe sets of 80 nt produced some negative calls on the negative controls, which can be explained by the high specificity of the e-probe sets at this length. Nonetheless, no statistically supported false negatives were observed.

EDNA-MiFi was challenged to discriminate pathogen detection at the genus and species level by querying a publicly available SRA dataset. This part of the study showed that 80 nt e-probe sets were highly specificity. While 40 nt and 60 nt correctly diagnosed infection of rhododendron with a *Phytophthora*, these e-probe sets incorrectly attributed the species to diverse *Phytophthora* species, while the 80 nt e-probe set correctly identified the pathogen as *Ph. ramorum*, which was the expected result, given that the SRA metagenome was obtained from rhododendron tissue infected with *Ph. ramorum* under experimental conditions (Espíndola et al. 2015). The positive calls for other *Phytophthora* species could be due to similarity of short genome segments between species, an issue that was corrected using 80 nt long e-probes, which by being longer were also more specific. The SRA metagenome was also queried with e-probes designed for the other oomycete species (*Pythium oligandrum, Pythium aphanidermatum, Globisporangium irregulare, Globisporangium cryptoirregulare, Plasmopara viticola,* and *Hyaloperonospera arabidopsis*); in all these cases results were negative (data are not shown). EDNA-MiFi maintained a highly specific detection thanks to the highly species-specific e-probes designed (Espíndola et al. 2015).

The MinION™ technology has many advantages, specially its portability and real-time applications. The MinION™ sequencer can be easily carried for use on site, and it produces long reads that can be called in real-time (Jain et al. 2016; Laver et al. 2015). This real-time sequencing capability implies that enough data can be obtained within minutes to make an identification. One down-side of the MinION™ is the high error rate of the sequences; however, for diagnostic purposes, the error rate may not affect the detection significantly, depending on the tested pathogen (Besser et al. 2018; Laver et al. 2015). Viral and bacterial pathogens represent a challenge because their small genomes produce relatively small e-probe data sets. However, that issue was not expected for eukaryotes since their large genomes allow the compilation of large data sets of highly specific e-probes, a hypothesis that was confirmed in this study.

In conclusion, this study demonstrated that EDNA- MiFi can successfully diagnose plant infection by plant pathogenic oomycetes. Our study shows outstanding potential of the EDNA-MiFi coupled with Nanopore sequencing for a rapid plant and, potentially, on-site pathogen diagnosis. Future studies must evaluate the potential field use of these coupled technologies.

5. Acknowledgments

The assistance of Dr. Z. Gloria Abad from USDA-APHIS-PPQ, CPHST Beltsville Laboratory and her team is greatly appreciated, for providing funding and some of the *Phytophthora* genomes used in this study. Thanks to Jennifer Olson, Director of the Plant Disease and Insect Diagnostic Laboratory, and her team, especially Sara Wallace, for

providing the *Phytophthora nicotianane* isolates studied. Some of the computational analysis done in this project was performed at the Oklahoma State University High-Performance Computing Center, supported in part through the National Science Foundation grant OAC-1126330.

Tables

Table 5-1. Genomic information used for the e-probes design. Pathogens, strain/isolate, GenBank accession numbers of the targeted pathogens and list of the near neighbors.

Pathogen	strain/isolate	GenBank	Neighbors[a]
Phytophthora sojae	P6497	AAQY00000000.2	*Ph. pisi, Ph. cinnamomi, Ph. fragariae, Ph. ramorum, Ph. lateralis, Ph. rubi*
Phytophthora infestans	T30-4	AATU00000000.1	*Ph. nicotianae, Ph. cactorum, Ph. capsici, Ph. citricola, Ph. multivora, Ph. parasitica, Ph. plurivora, Ph. pluvialis, Pl. halstedii, Pl. viticola*
Plasmopara viticola	INRA-PV221	MBPM00000000.2	*Pl. halstedii, Bremia lacttuca, Pl. muralis, Ph. nicotianae, Ph. parasitica, Ph. Infestans*
Pythium aphanidermatum	DAOM BR444	AKXX00000000.2	*Py. arrhenomanes, G. irregulare, G. ultimum, Phytopythium vexans, Pilasporangium apinafurcum, Py. insidiosum, Py. iwayamai, Py. oligandrum*
Pythium oligandrum	ATCC 38472_TT	SPLM00000000.1	*Py. arrhenomanes, G. irregulare, G. ultimum, Pp. vexans, Ps. apinafurcum, Py. insidiosum, Py. iwayamai, Py. aphanidermatum*
Hyaloperonospora arabidopsis	Emoy2	ABWE00000000.2	*Ph. agathidicida, Peronospora belbahrii, Pseudoperonospora cubensis, Peronospora effusa, Pseudoperonospora humuli, Peronospora tabacina, Ph. torara, Sclerospora graminícola*
Phytophthora agathidicida	NZFS 3770	LGTS00000000.1	*Ph. pluvialis, Hyaloperonospora arabidopsis, Pl halstedii, Ph. multivora, Ph. pluvialis, Pl. viticola*
Phytophthora lateralis	CBS 168.42	AWVV00000000.2	*Ph. ramorum, Ph. cryptogea, Ph. pinifolia, Ph. pisi, Ph. rubi*
Phytophthora fragariae	CBS 209.46	MWJK00000000.1	*Ph. rubi, Ph. pisi, Ph. ramorum, Ph. sojae*
Phytophthora ramorum[b]	Ex-type CPHST BL 55G	RYEP00000000.1	*Ph. lateralis, Ph. cryptogea, Ph. fragariae, Ph. pinifolia, Ph. ramorum, Ph. rubi. Ph. syringae*
Phytophthora kernoviae[b]	P19827 CPHST BL 91	VKKV00000000.1	*Ph. boehmeriae, Ph. cinnamomi, Ph. cryptogea, Ph. lateralis, Ph. rubi, Ph. sojae*
Phytophthora citricola[b]	P0716 CPHST BL 34	VMRO00000000.1	*Ph. plurivora, Ph. capsici, Ph. multivora*
Phytophthora melonis[b]	CPHST BL 23	VXDT00000000.1	*Ph. sojae, Ph. agathidicida, Ph. cambivora, Ph. cinnamomi, Ph. fragariae, Ph. pinifolia, Ph. rubi*
Phytophthora cambivora[b]	CBS 114087	AUVH00000000.1	*Ph. sojae, Ph. agathidicida, Ph. melonis, Ph. cinnamomi, Ph. fragariae, Ph. pinifolia, Ph. rubi, Ph. pisi*
Phytophthora pinifolia	CBS 122922	AWVW00000000.2	*Ph. fragariae, Ph. sojae, Ph. agathidicida, Ph. melonis, Ph. cinnamomi, Ph. fragariae, Ph. cambivora, Ph. rubi*

Pathogen	strain/isolate	GenBank	Neighbors[a]
Phytophthora nicotianae[b]	BL162	JAAKBE000000000.1	*Ph. infestans, Ph. cactorum, Ph. parasitica*
Globisporangium irregulare	LEV1481	This study	*Py. arrhenomanes, G. cryptoirregulare, G. ultimum, Pp. vexans, Py. insidiosum, Py. iwayamai, Py. aphanidermatum, Py. oligandrum*
Globisporangium cryptoirregulare	LEV4534	This study	*Py. arrhenomanes, G. irregulare, G. ultimum, Pp. vexans, Py. insidiosum, Py. iwayamai, Py. aphanidermatum, Py. oligandrum*

[a] = Neighbors selected for each pathogen are the representative genomes in GenBank database.

[b] = Genomes were provided by Dr. Gloria Abad, USDA-APHIS-PPQ, CPHST Beltsville Laboratory.

Table 5-2. Genomic information used for the generation of metagenomes. Pathogens and its selected host, strain/isolate and their GenBank accession numbers.

Pathogen	strain/isolate	GenBank	Host	strain/cultivar	GenBank
Phytophthora sojae	P6497	AAQY00000000.2	*Glycine max* (Soybean)	Williams 82	ACUP00000000.4.
Phytophthora infestans	T30-4	AATU00000000.1	*Solanum tuberosum* (Potato)	DM1-3 516 R44	AEWC00000000.1
Plasmopara viticola	INRA-PV221	MBPM00000000.2	*Vitis vinifera* (Grape)	PN40024	CAAP00000000.3
Pythium aphanidermatum	DAOM BR444	AKXX00000000.2	*Cucumis sativus* (Cucumber)	9930	ACHR00000000.2
Pythium oligandrum	ATCC 38472_TT	SPLM00000000.1	*Solanum lycopersicum* (Tomato)	Heinz 1706	AEKE00000000.3
Hyaloperonospora arabidopsis	Emoy2	ABWE00000000.2	*Arabidopsis thaliana* (Thale cress)	TAIR10.1	GCA_000001735.2
Phytophthora agathidicida	NZFS 3770	LGTS00000000.1	*Picea abies var. abies* (Norway spruce)	ERS2564098	UETF00000000.1
Phytophthora lateralis	CBS 168.42	AWVV00000000.2	*Picea abies var. abies* (Norway spruce)	ERS2564099	UETF00000000.1
Phytophthora fragariae	CBS 209.46	MWJK00000000.1	*Fragaria* × *ananassa* (Strawberry)	FAN_r1.1	BATT00000000.1
Phytophthora ramorum	Ex-type CPHST BL 55G	RYEP00000000.1	*Rhododendron williamsianum*	RSF 1966-606	QEFC00000000.1
Phytophthora kernoviae	P19827 CPHST BL 91	VKKV00000000.1	*Rhododendron williamsianum*	RSF 1966-606	QEFC00000000.1
Phytophthora citricola	P0716 CPHST BL 34	VMRO00000000.1	*Rhododendron williamsianum*	RSF 1966-607	QEFC00000000.2
Phytophthora melonis	CPHST BL 23	VXDT00000000.1	Cucumis sativus (Cucumber)	9930	ACHR00000000.2
Phytophthora cambivora	CBS 114087	AUVH00000000.1	*Fagus sylvatica (*European beech*)*	Fs_037	QCXR00000000.1
Phytophthora pinifolia	CBS 122922	AWVW00000000.2	*Pinus taeda(*Loblolly pine)	-	APFE000000000.3
Phytophthora nicotianae	BL162	JAAKBE000000000.1	*Solanum lycopersicum* (Tomato)	Heinz 1706	AEKE00000000.3
Globisporangium irregulare	LEV1481	This study (Chapter III)	*Solanum lycopersicum* (Tomato)	Heinz 1706	AEKE00000000.3
Globisporangium cryptoirregulare	LEV4534	This study (Chapter III)	*Solanum lycopersicum* (Tomato)	Heinz 1706	AEKE00000000.3

Table 5-3. Pathogen and host abundances in Mock Sample Sequencing Databases (MSSDs).

Abundance level	Pathogen read abundance	Host read abundance	Total expected reads	
1	100%	0%	10’000.000[a]	Positive control
2	10%	90%	1’000.000[a]	-
3	1%	99%	100.000[a]	-
4	0.1%	99.9%	10.000[a]	-
5	0.01%	99.99%	1.000[a]	-
6	0.001%	99.999%	100[a]	-
7	0.0001%	99.9999%	10[a]	-
8	0%	100%	10.000.000 [b]	Negative control

[a] = Expected reads for the pathogen

[b] = Expected reads for the host

Table 5-4. *Phytophthora nicotianae* isolates used for *in vitro* validation.

Isolate id	State	Plant Material
Pnic_3635	TX	Boxwood, *Buxus*
Pnic_3645	TX	Bougainvillea, *Bougainvillea glabra*
Pnic_4523	TX	Leadwort., *Plumbago*

Table 5-5. *Phytophthora nicotianae* and tomato samples used for the DNA spiking experiment and Nanopore Sequencing with the MinION™ device.

Metagenome	DNA combination	Abundance/concentration (ng/µL)	Dilution	Code[b]	Flow cell
Metagenome 1	Pnic_3635	100	-	RB01	Flow cell 1
Metagenome 5	Pnic_3645	100	-	RB05	Flow cell 1
Metagenome 9	Pnic_4523	100	-	RB09	Flow cell 2
Metagenome 4	Tom1[a]	100	-	RB04	Flow cell 1
Metagenome 8	Tom2[a]	100	-	RB08	Flow cell 2
Metagenome 12	Tom3[a]	100	-	RB012	Flow cell 2
Metagenome 2	Pnic_3635 + Tom1	1	1:100	RB02	Flow cell 1
Metagenome 6	Pnic_3645 + Tom2	1	1:100	RB06	Flow cell 1
Metagenome 10	Pnic_4523 + Tom3	1	1:100	RB10	Flow cell 2
Metagenome 3	Pnic_3635 + Tom1	0.1	1:1,000	RB03	Flow cell 1
Metagenome 7	Pnic_3645 + Tom2	0.1	1:1,000	RB07	Flow cell 2
Metagenome 11	Pnic_4523 + Tom3	0.1	1:1,000	RB011	Flow cell 2

[a] = Tom1, Tom2 and Tom3 are DNA samples extracted from tomato leaves.

[b] = IDs for the fragmentation mix as in the rapid barcoding kit (Oxford Nanopore)

Table 5-6. Read numbers and sequence length generated after Nanopore sequencing for each metagenome.

Metagenome	Sample	Read number	Sequence length
Metagenome 1	Pnic_3635	13914	69-15705
Metagenome 5	Pnic_3645	77030	121-25467
Metagenome 9	Pnic_4523	2618	66-18057
Metagenome 2	Pnic_3635 + Tom1 (1:100)	484345	102-40053
Metagenome 6	Pnic_3645 + Tom2 (1:100)	588297	74-34433
Metagenome 10	Pnic_4523 + Tom3 (1:100)	106810	68-60468
Metagenome 3	Pnic_3635 + Tom1 (1:1000)	524171	74-78826
Metagenome 7	Pnic_3645 + Tom2 (1:1000)	83916	33-15248
Metagenome 11	Pnic_4523 + Tom3 (1:1000)	278219	67-45692
Metagenome 4	Tom1	844157	101-36963
Metagenome 8	Tom2	127322	68-34329
Metagenome 12	Tom3	149314	69-106679

Table 5-7. Results of the pairwise T-test with the metagenomes generated with Oxford Nanopore Sequencing. P-values and number of matches are depicted for each e-probe length. Values in red are negative/suspect diagnostic calls.

Metagenome	Sample	40 nt e -probe		60 nt e -probe		80 nt e -probe	
		p-value	# match	p-value	# match	p-value	# match
Metagenome1	Pnic_3635	8.99E-176	122083	6.00E-105	6956	1.20E-36	895
Metagenome5	Pnic_3645	0	346860	6.71E-117	53868	4.23E-44	22356
Metagenome9	Pnic_4523	9.11E-299	10815	1.09E-74	744	3.70E-36	383
Metagenome2	Pnic_3635 + Tom1 (1:100)	2.50E-134	87423	6.21E-92	8346	6.32E-20	1566
Metagenome6	Pnic_3645 + Tom2 (1:100)	1.07E-226	60059	2.33E-91	12986	1.25E-43	4608
Metagenome10	Pnic_4523 + Tom3 (1:100)	6.10E-263	10288	1.87E-63	1740	9.71E-37	697
Metagenome3	Pnic_3635 + Tom1 (1:1000)	1.00E-209	5708	3.68E-48	584	2.21E-06	42
Metagenome7	Pnic_3645 + Tom2 (1:1000)	2.83E-160	2995	8.95E-47	260	5.07E-05	18
Metagenome11	Pnic_4523 + Tom3 (1:1000)	4.81E-139	930	7.90E-18	203	3.06E-09	137
Metagenome4	Tom1	-	4[s]	-	3[s]	-	0
Metagenome8	Tom2	-	0	-	0	-	0
Metagenome12	Tom3	2.23E-73	331	-	1[s]	-	0

[s] = Suspect

- = Not computed

Table 5-8. Detection with MiFi in raw SRA data (SRX3160109) using *Ph. ramorum* e-probe sets and *Phytophthora* e-probe sets to measure specificity. P-values and number of matches are depicted for each e-probe length. Values in red are negative/suspect diagnostic calls.

Pathogen	40 e-probe			60 e-probe			80 e-probe		
	Diagnostic	**p-value**	**# hits**	**Diagnostic**	**p-value**	**# hits**	**Diagnostic**	**p-value**	**# hits**
Ph.ramorum	Positive	3.54E-154	22136	Positive	5.25E-51	1313	Positive	2.55E-09	84
Ph.sojae	Positive	6.17E-87		Negative	-	0	Negative	-	0
Ph.infestans	Positive	4.62E-93		Negative	-	0	Negative	-	0
Ph.agathidicida	Positive	7.69E-89		Negative	-	0	Negative	-	0
Ph.lateralis	Positive	6.51E-126		Positive	4.25E-47		Negative	-	0
Ph.fragariae	Negative	-	0	Negative	-	0	Negative	-	0
Ph.kernoviae	Negative	-	0	Negative	-	0	Negative	-	0
Ph.citricola	Positive	2.89E-90		Negative	-	0	Negative	-	0
Ph.melonis	Negative	-	0	Negative	-	0	Negative	-	0
Ph.cambivora	Positive	3.27E-178		Positive	8.25E-76		Negative	-	0
Ph.pinifolia	Positive	3.71E-105		Positive	5.25E-68		Negative	-	0
Ph.nicotianae	Negative	-	0	Negative	-	0	Negative	-	0

Figures

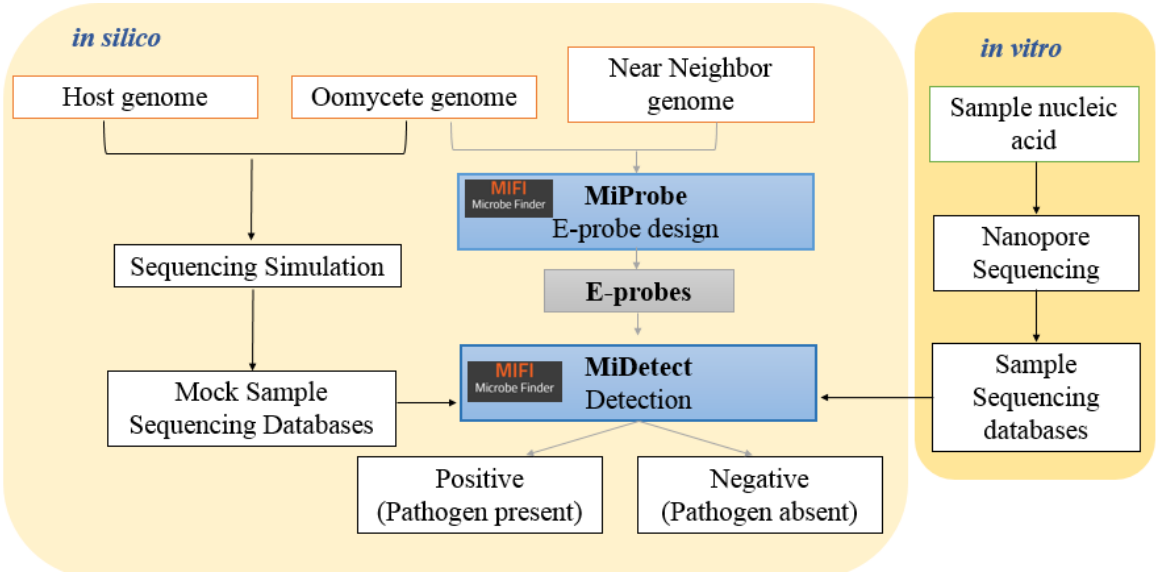


Figure 5-1. EDNA pipeline for *in silico* and *in vitro* diagnostics of oomycete plant pathogens used in this study. Adapted from Espíndola et al. (2015).

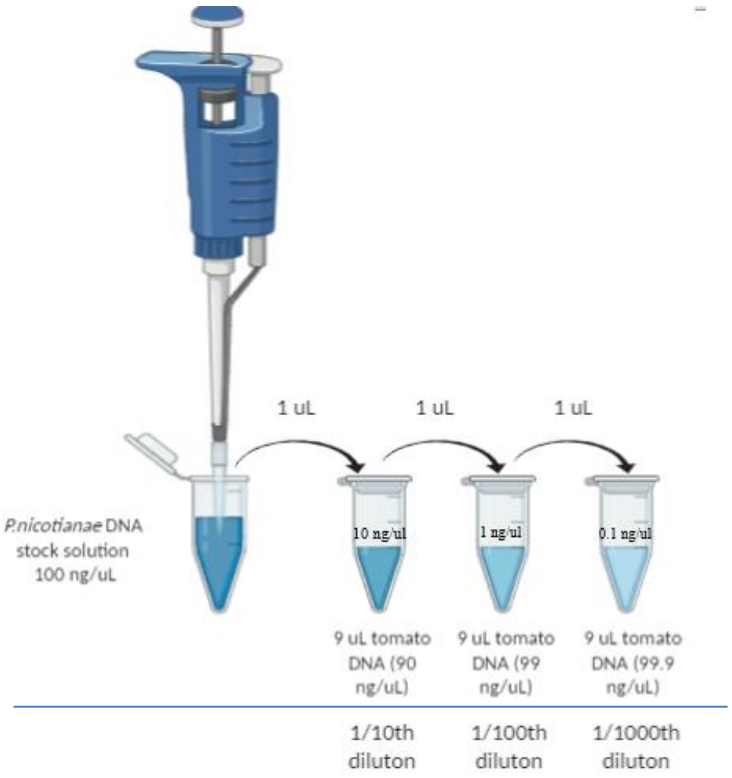


Figure 5-2. 10-fold serial dilutions for the DNA spiking experiment. DNA samples were *Ph. nicotianae* DNA and Tomato DNA diluted from 100 ng/µl to 0.1ng/µl in a final volume of 10 µl.

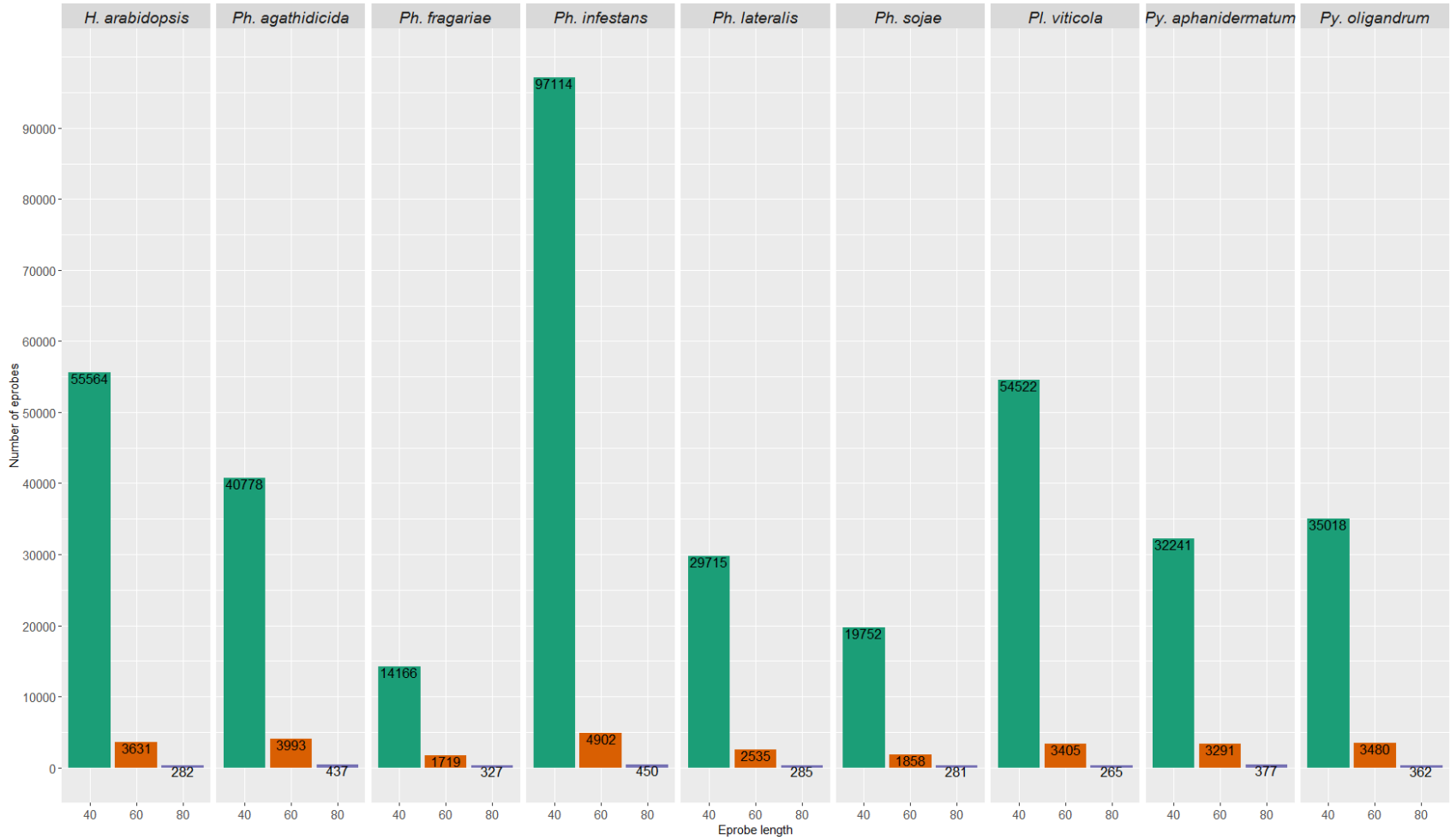


Figure 5-3. Number of e-probes designed among the first group of pathogens (n=9) and e-probe lengths. The number of e-probes at each length is depicted in each bar. Labels key: Py: *Pythium*, H: *Hyaloperonospera*, Ph: *Phytophthora*, Pl: *Plasmopara*

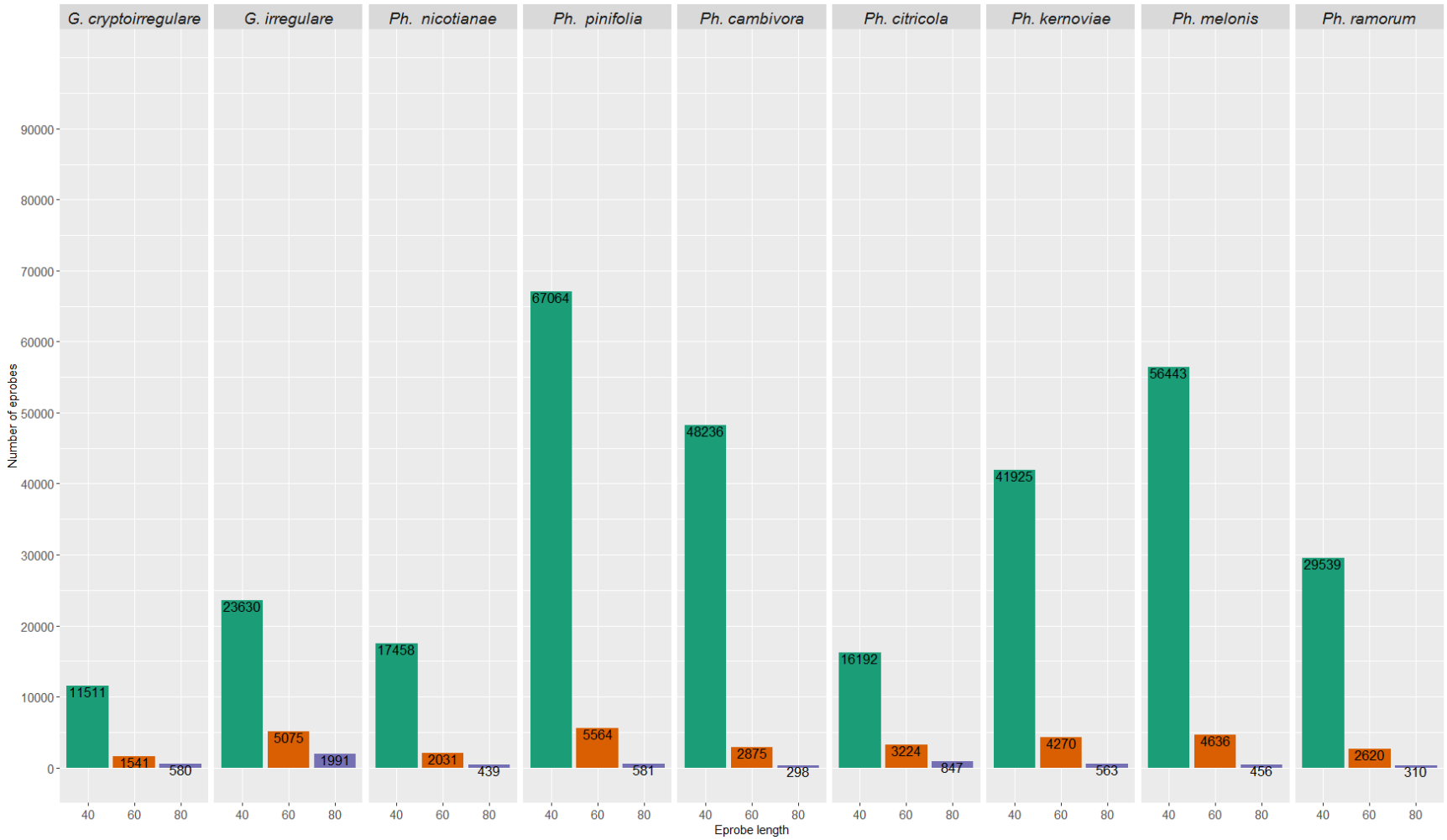


Figure 5-4. Number of e-probes designed among the second group of pathogens (n=9) and e-probe lengths. The number of e-probes at each length is depicted in each bar. Labels key: G: *Globisporangium;* Ph: *Phytophthora*, Pl: *Plasmopara*.

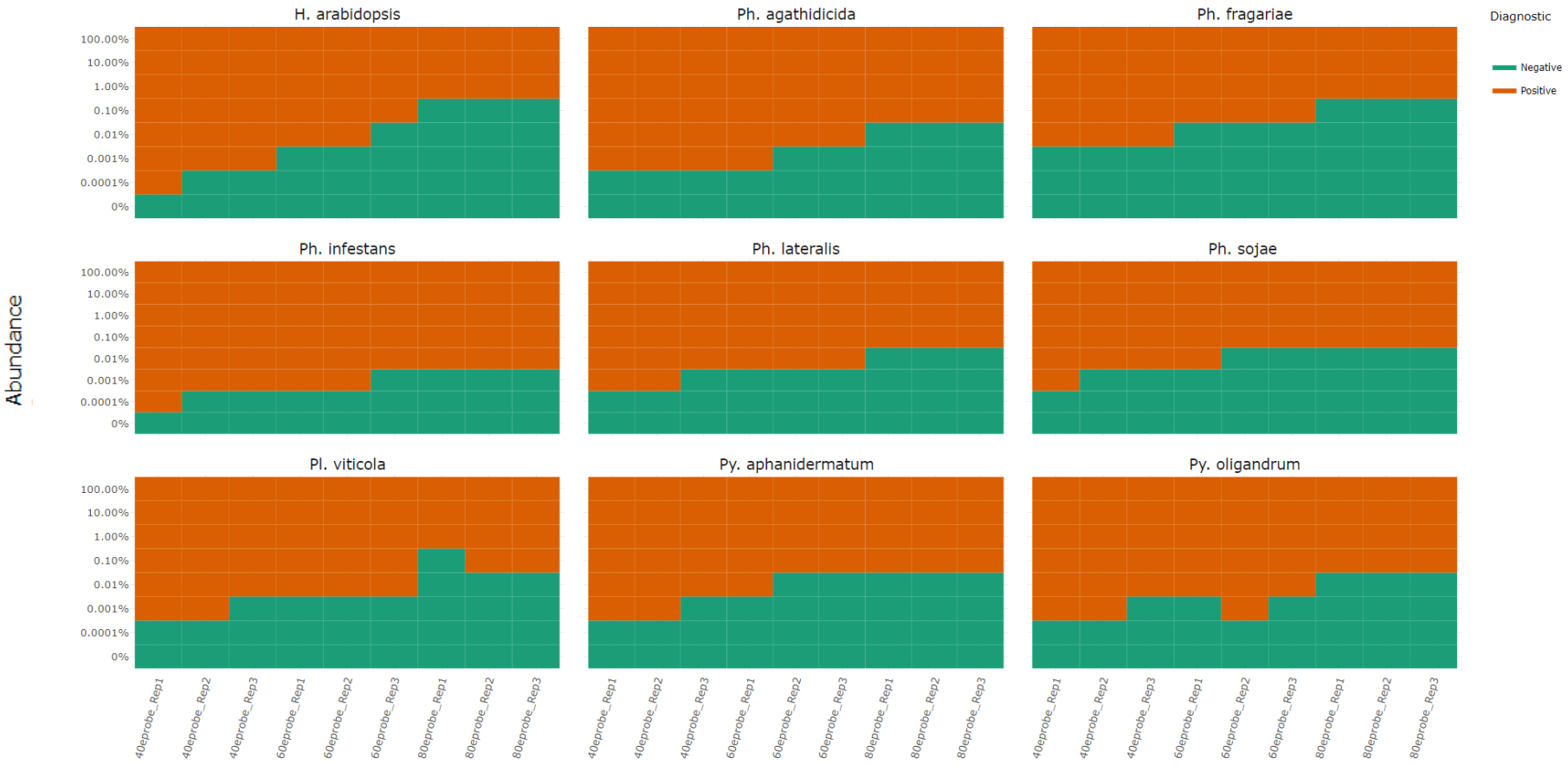


Figure 5-5. Matrix depicting the *in silico* diagnostic results among the first group of pathogens (n=9) for each e-probe length, its replicates and pathogen abundance levels. The metagenomes for each level of pathogen abundance were generated with three replicates (Rep 1, Rep2, and Rep3) and queried against the e-probe lengths (40, 60, and 80). A positive diagnostic result reflects the presence of the pathogen in a given metasample. A negative diagnostic result reflects the absence of the pathogen in a given metasample. Labels key: Py: *Pythium*, H: *Hyaloperonospera*, Ph: *Phytophthora*, Pl: *Plasmopara.* Color key: green = negative, or absence of the pathogen, orange = positive, or presence of the pathogen.

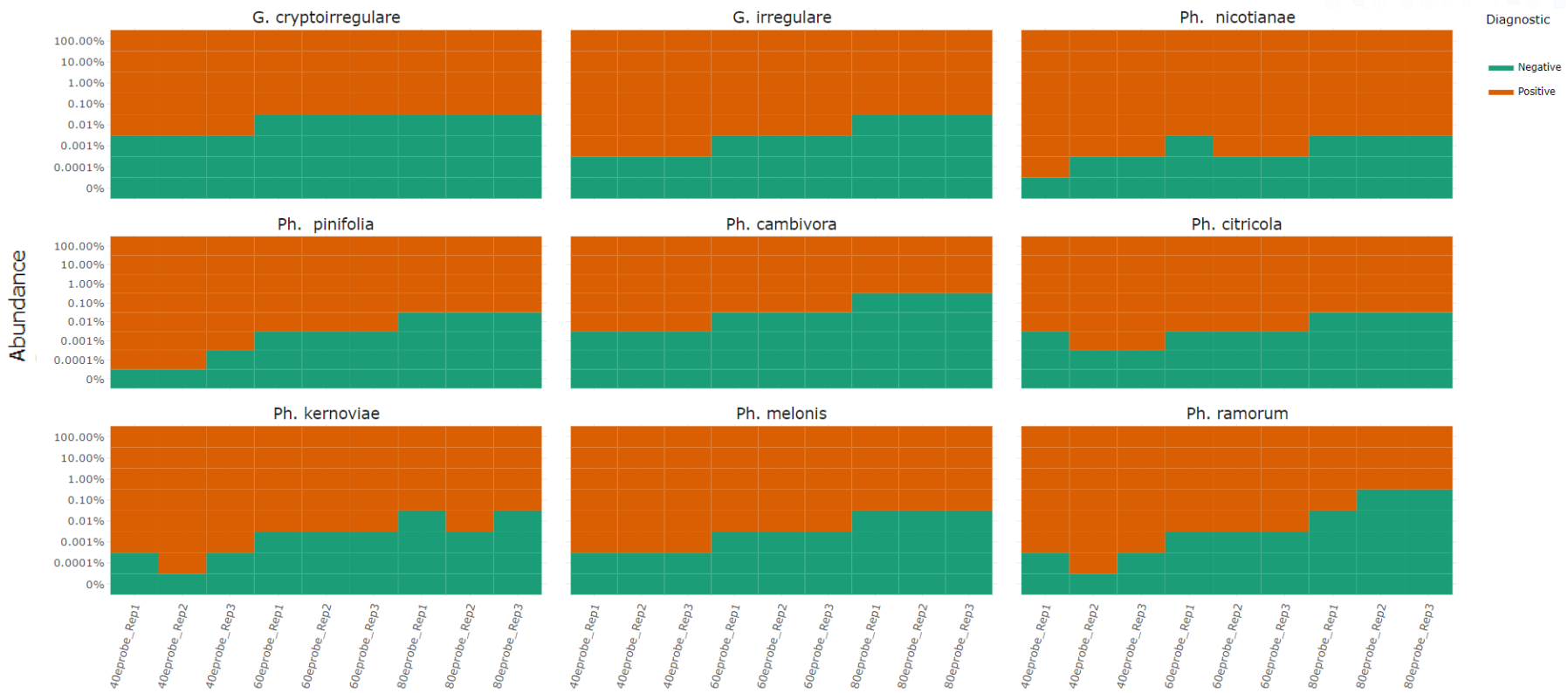


Figure 5-6. Matrix depicting the *in silico* diagnostic results among the second group of pathogens (n=9) for each e-probe length, its replicates and pathogen abundance levels. The metagenomes for each level of pathogen abundance were generated with three replicates (Rep 1, Rep2, and Rep3) and queried against the e-probe lengths (40, 60, and 80). A positive diagnostic result reflects the presence of the pathogen in a given metasample. A negative diagnostic results reflects the absence of the pathogen in a given metasample. Labels key: Py: *Pythium*, H: *Hyaloperonospera*, Ph: *Phytophthora*, Pl: *Plasmopara.* Color key: green = negative, or absence of the pathogen, orange = positive, or presence of the pathogen.

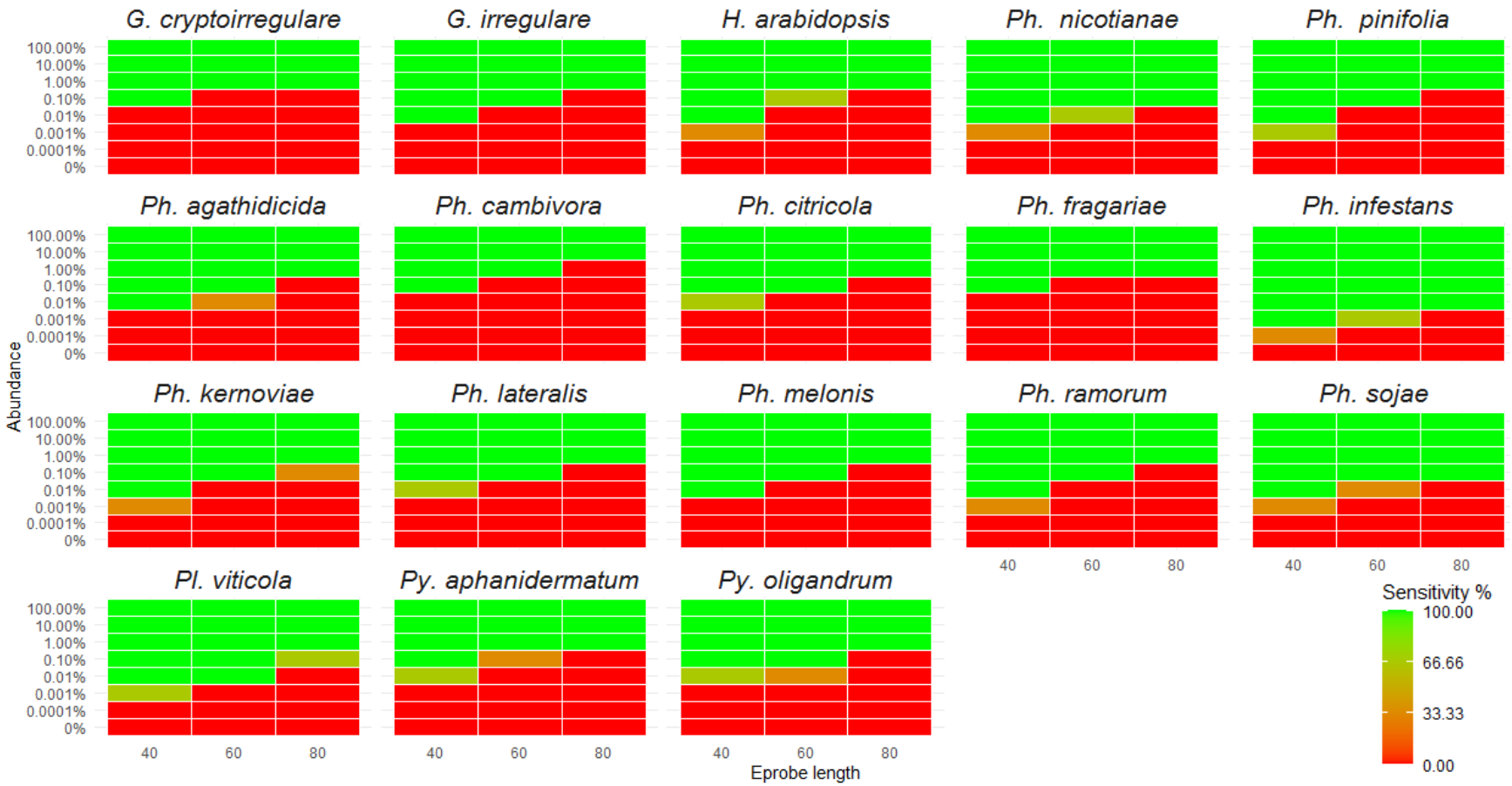


Figure 5-7. Heat-maps representing the *in silico* sensitivity in relation with the e-probe length and the pathogen read abundance. Color key: red to green correspond to low (0 %) to high (100 %) sensitivity.

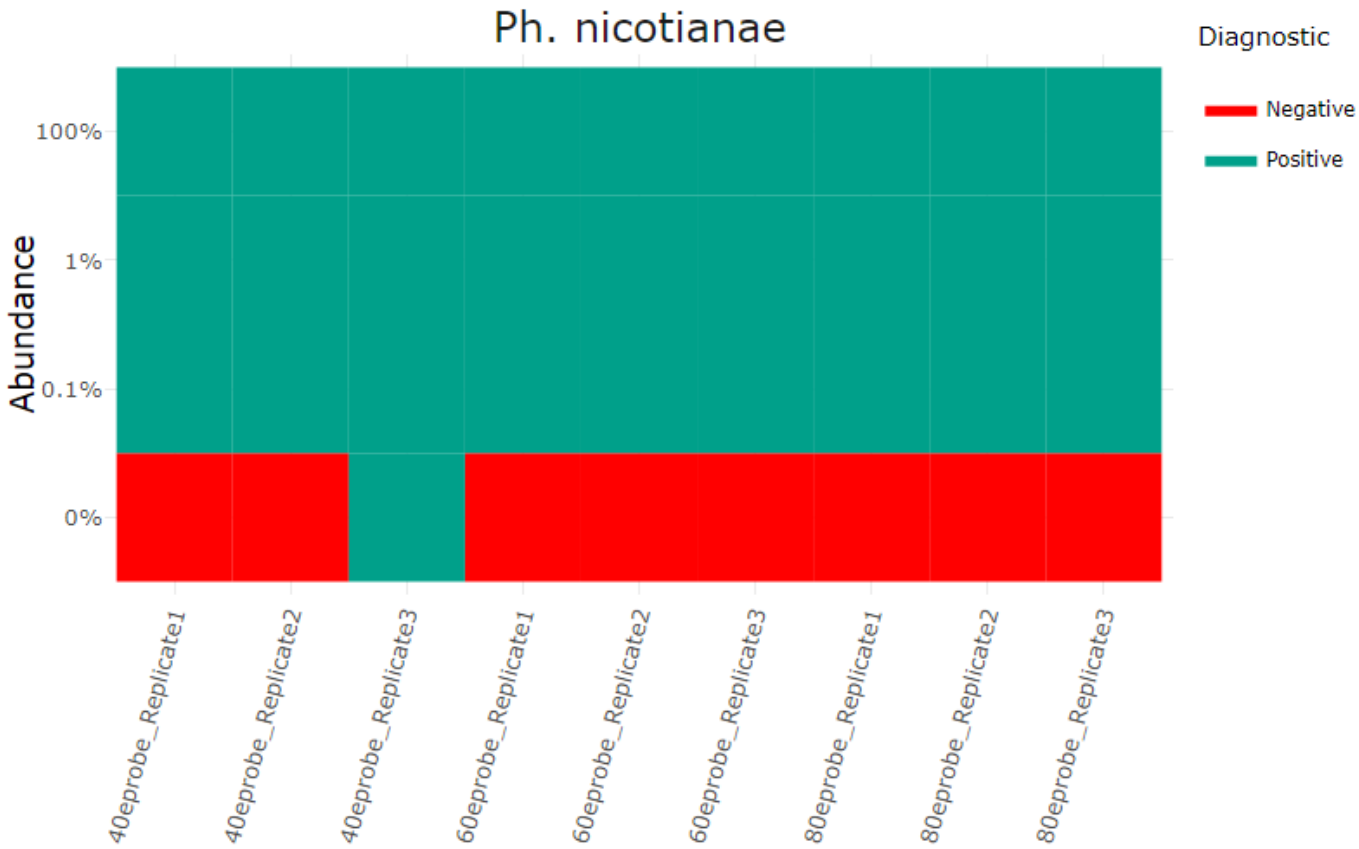


Figure 5-8. Matrix depicting the *in vitro* diagnostic results for each e-probe length, its replicates and pathogen abundance levels. The metagenomes for each level of pathogen abundance were generated with three replicates (Replicate 1, Replicate 2, and Replicate 3) and queried against the e-probe lengths (40, 60, and 80). A positive diagnostic result reflects the presence of the pathogen in the metasample generated with Nanopore sequencing. A negative diagnostic result reflects the absence of the pathogen in a given metasample generated with Nanopore sequencing. Color key: red = negative, or absence of the pathogen; green = positive, or presence of the pathogen.

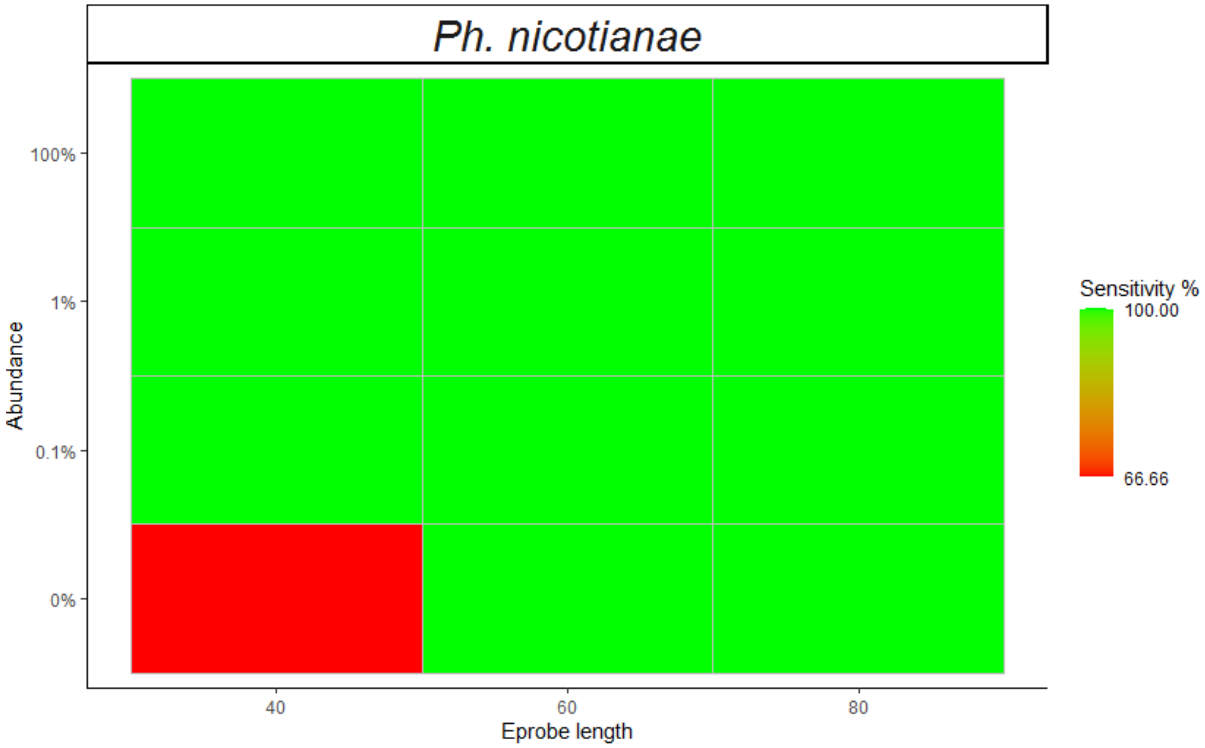


Figure 5-9. Heat-map representing the *in vitro* sensitivity in relation with the e-probe length and the pathogen abundance. Metagenomes were generated with Oxford Nanopore Sequencing. Color key: red to green correspond to low (66.66%) to high (100%) sensitivity.

REFERENCES

Abad, Z.G., and Bienapfl, J.C. 2019. Qiagen DNeasy Plant Mini Kit extraction procedure for oomycetes and fungi. Document control number: SOP-PID-02.01 from Abad, Z.G., Burgess T., Bienapfl J.C., Redford A.J., Coffey M., and Knight L. 2019. IDphy: Molecular and morphological identification of Phytophthora based on the types. USDA APHIS PPQ S&T Beltsville Lab, USDA APHIS PPQ S&T ITP, Centre for Phytophthora Science and Management, and World Phytophthora Collection. http://idtools.org/id/phytophthora/index.php.

Adams, I. P., Glover, R. H., and Monger, W. A. 2009. Next-generation sequencing and metagenomic analysis: a universal diagnostic tool in plant virology. Molecular Plant Pathology. 10:537-545.

Adhikari, B.N., Hamilton, J.P., Zerillo, M.M., Tisserat, N., Levesque, C.A., and Buell, C.R. 2013. Comparative Genomics Reveals Insight into Virulence Strategies of Plant Pathogenic Oomycetes. PLOS One. 8: e75072.

Adl, S. M., Simpson, G., and Lane, C. E. 2012. The revised classification of eukaryotes. Journal of Eukaryotic Microbiology. 59:429-493.

Agrios, G. N. 2005. Plant pathology. Elsevier Academic Press Amsterdam, The Netherlands.

Ali-Shtayeh, M. S., McDonald, J. D., and Kabashima, J. 1991. A method for using commercial ELISA tests to detect zoospores of *Phytophthora* and *Pythium* species in irrigation water. Plant Disease. 75:302-311.

Alkan, C., Sajjadian, S., and Eichler, E. 2011. Limitations of next-generation genome sequence assembly. Nature Methods. 8:61-5.

Allen, T. W., Martinez, A., and Burpee, L. L. 2004. Pythium blight of turfgrass. The Plant Health Instructor. DOI:10.1094/PHI-I-2004-0929-01.

Al-Sa'di, A., Drenth, A., Deadman, M., and Aitken, E. 2008a. Genetic diversity, aggressiveness and metalaxyl sensitivity of *Pythium aphanidermatum* populations infecting cucumber in Oman. Plant pathology 57:45-56.

Al-Sa'di, A., Drenth, A., Deadman, M., DeCock, A., Al-Said, F., and Aitken, E. 2008b. Genetic diversity, aggressiveness and metalaxyl sensitivity of *Pythium spinosum* infecting cucumber in Oman. Journal of phytopathology 156:29-35.

Al-Sa'di, A., Drenth, A., Deadman, M., Al-Said, F., Khan, I., and Aitken, E. 2008c. Potential sources of Pythium inoculum into greenhouse soils with no previous history of cultivation. Journal of Phytopathology 156:502-505.

Anisimova, M., Gil, M., Dufayard, J.F., Dessimoz, C., and Gascuel, O. Survey of Branch Support Methods Demonstrates Accuracy, Power, and Robustness of Fast Likelihood-based Approximation Schemes. Systematic Biology. 60:685–699.

Arafa, R. A., and Shirasawa, K. 2018. Technical review of molecular markers and next generation sequencing technology to manage plant pathogenic oomycetes. African Journal of Biotechnology. 17:369-379.

Ari, S., and Arikan, M. 2016. Next-Generation Sequencing: Advantages, Disadvantages, and Future. In *Plant Omics: Trends and applictions*, eds. Rehman, K., Tombuloglu, H., and Tombuloglu, G. Switzerland: Springer.

Asano, T., Senda, M., Suga, H., and Kageyama, K. 2010. Development of Multiplex PCR to Detect Five *Pythium* Species Related to Turfgrass Diseases. Journal Phytopathology. 158:60-615.

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

Avila, F. J., Yuen, G. Y., and Klopfenstein, N. B. 1995. Characterization of *Pythium ultimum*-specific antigen and factors that affect its detection using a monoclonal antibody. Phytopathology. 85:1378-1387.

Bala, K., Robideau, G. P., Lévesque, C. A., Abad, Z. G., Lodhi, A. M., Shahzad, S., et al. 2010. *Phytopythium* Abad, de Cock, Bala, Robideau, Lodhi and Lévesque, gen. nov. and *Phytopythium sindhum* Lodhi, Shahzad, and Lévesque, sp. nov. 2010. Persoonia. 24:136-137.

Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S. et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology. 19: 455-477.

Bar, D.J.S., Warwick, S. I., and Desaulniers, N. L. 1997. Isozyme variation, morphology, and growth response to temperature in *Pythium irregulare*. Canadian Journal of Botany. 75:2073-2081.

Barkely, N.A., Dean, R.E., Pittman, R.N., Wang, M.L., Holbrook, C.C. et al. 2007. Genetic diversity of cultivated and wild-type peanuts evaluated with M13-tailed SSR markers and sequencing. Genetics Research. 89:93-106.

Baten, M. A., Mingzhu, L., Motohashi, K., Ishiguro, Y., Rahman, M. Z., Suga, H. 2015. Two new species, *Phytopythium iriomotense* sp. nov. and *P. aichiense* sp. nov., isolated from river water and water purification sludge in Japan. Mycological Progress. 14:2-12.

Beakes G. W., Honda D., Thines M. 2014. 3 Systematics of the Straminipila: Labyrinthulomycota, Hyphochytriomycota, and Oomycota. In *The Mycota (A Comprehensive Treatise on Fungi as Experimental Systems for Basic and Applied Research)*, eds. McLaughlinm, D and Spatafora, J. Berlin, Heidelberg: Springer.

Beakes, G. W., and Sekimoto, S. 2009. The evolutionary phylogeny of Oomycetes-insights gained from studies of holocarpic parasites of algae and invertebrates. In *Oomycete Genetics and Genomics*, eds. : Lamour, K and Kamoun, S. New York: John Wiley & Sons In, p. 1-24.

Bebber, D. P., and Gurr, S. J. 2015. Crop-destroying fungal and oomycete pathogens challenge food security. Fungal Genetics and Biology. 74:62- 64.

Bedard, J. E. J., Schurko, A. M., de Cock, A.W.A.M., and Klanssen, G. R. 2006. Diversity and evolution of 5S rRNA gene family organization in *Pythium*. Mycological Research. 110:86-95.

Belbahri, L., McLeod, A., Paul, B., Calmin, G., Moralejo, E., Spies, C.F.J., et al. 2008. Intraspeci¢c andwithin-isolate sequence variation in the ITS rRNA gene region of *Pythium mercuriale* sp. nov. (*Pythiaceae*). FEMS. Microbiology letters. 284: 17-27.

Bennett, P., and Stone, J. 2016. Assessments of population structure, diversity, and phylogeography of the Swiss needle cast fungus (*Phaeocryptopus gaeumannii*) in the US Pacific Northwest. Forests 7:14.

Besser, J., Carleton, H.A., Gerner-Smidt, P., Lindsey, R.L., and Trees, E. 2018. Next-generation sequencing technologies and their application to the study and control of bacterial infections. Clinical Microbiology and Infection. 24:335-341.

Binagwa, P. H., Bonsi, C. K., Msolla, S. N., and Ritte, I. I. 2016. Morphological and molecular identification of *Pythium* spp. isolated from common beans (*Phaseolus vulgaris*) infected with root rot disease. African Journal of Plant Science. 10:1-9.

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. Journal of Food Protection. 4:574-581.

Blair, J. E., Coffey, M. D., Park, S-Y., Geiser, D. M., and Kang, S. 2008. A multi-locus phylogeny for *Phytophthora* utilizing markers derived from complete genome sequences. Fungal Genetics and Biology. 45:266-277.

Briard, M., Dutertre, M., Rouxel, F., and Brygoo, Y. 1995. Ribosomal RNA sequence divergence within the Pythiaceae. Mycological Research. 99:1119-1127.

Bronzato, A., Sherman, D., Gopakumar, A., Wilson, V., Schneider, W., and King, J. 2018. Nanopore sequencing as a surveillance tool for plant pathogens in plant and insect tissues. Plant Disease. 102:1648-1652.

Bruen, T. C., Philippe, H., and Bryant, D. 2006. A simple and robust statistical test for detecting the presence of recombination. Genetics. 172: 2665–2681.

Bruvo, R., Michiels, N. K., D'Souza, T. G., and Schulenburg, H. 2004. A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. Molecular Ecology. 13:2101-2106.

Burki, F. 2014. The Eukaryotic Tree of Life from a Global Phylogenomic Perspective. Cold Spring Harbor Perspectives in Biology. 6:a016147.

Cai, L., Giraud, T., Zhang, N., Begerow, D., Cai, G, and Shivas, R.G. 2011. The evolution of species concepts and species recognition criteria in plant pathogenic fungi. *Fungal Diversity.* 50 (1): 121-133.

Capella-Gutierrez, S., Silla-Martinez, J. M., Gabaldon, T. 2009. trimAL: A tool for automated alignemtn trimming in large-scale phylogenetic analysis. Bioinformatics. 25: 1972-1973.

Carleson, N. C., Fieland, V. J., Scagel, C. F., Weiland, J. E., and Grünwald, N. J. 2019. Population Structure of *Phytophthora plurivora* on Rhododendron in Oregon Nurseries. Plant Disease. 103:1923-1930.

Carstens, B.C., and Knowles, L.L. 2007. Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from Melanoplus grasshoppers. Systematic Biology. 56: 400–411.

Cavalier-Smith, T. 2018. Kingdom Chromista and its eight phyla: a new synthesis emphasizing periplastid protein targeting, cytoskeletal and periplastid evolution, and ancient divergences. Protoplasma. 255:297-357.

Cavalier-Smith, T. 1998. A revised six-kingdom system of life. Biological Reviews. 73:203-226.

Cavalier-Smith, T. 1981. Eukaryote kingdoms: seven or nine?. Biosystems. 14:461-481.

Chalupowicz., L., Dombrovsky, A., Gaba, V., Luria, N., Reuven, M., Beerman, A., et al. 2019. Diagnosis of plant diseases using the Nanopore sequencing platform. Plant Pathology. 68:229-238.

Chao, Z., Rabiee, M., Sayyari, E., and Mirarab. S. 2018. ASTRAL-III: Polynomial Time Species Tree Reconstruction from Partially Resolved Gene Trees. BMC Bioinformatics. 19: 153.

Chen, W., Schneider, R. W., and Hoy, J. W. 1992. Taxonomic and Phylogenetic Analyses of ten *Pythium* species using isozyme polymorphisms. Phytopathology. 82:1234-1244.

Chen, W., Hoy, J. W., and Schneider, R. W. 1991. Comparisons of soluble proteins and isozymes for seven *Pythium* species and applications of the biochemical data to *Pythium* systematics. Mycological Research. 95:548-555.

Choi, Y. J., Neakes, G., Glockling, S., Kruse, J., Nam, B., Nigrelli, L., et al. 2015. Towards a universal barcode of oomycetes - a comparison of the *cox*1 and *cox*2 loci. Molecular Ecology Resources. 15:1275-1288.

Collard, B.C.Y., Jahufer, M.Z.Z., Brouwer, J. B., and Pang, E.C.K. 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica.142:169-196.

Cooke, D.E.L., Drenth, A., Duncan, J. M., Wagels, G., and Brasier, M. 2000. A molecular phylogeny of *Phytophthora* and related oomycetes. Fungal Genetics and Biology. 30:17-32.

Daughtrey, M. 2011. Disease Management in Floral and Nursery Crops. Cornell University, Riverhead, NY.

Daughtrey, M., and Miller, W. 2009. Reducing root rot, bulb rot and vascular wilt disease losses in floral crops by improvements in pathogen tracking and management. ARS Floriculture and Nursery Research Initiative Research Reports. 1-7.

Daughtrey, M. L., Wick, R. L., and Peterson, J. L. 1995. Compendium of flowering potted plant diseases. St. Paul, MN: APS Press.

De Cock, A.W.A.M., Lodhi, A. M., Rintoul, T. L., Bala, K., Robideau, G. P., Abad, Z. G., et al. 2015. *Phytopythium:* molecular phylogeny and systematics. Persoonia. 34:25-39.

Degnan, J.H., and Rosenberg, N.A. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends in Ecology and Evolution. 24: 332-340.

De Jesus, A. L., Goncalves, D. R., Rocha, S.C.O., Marano, A. V., Jeronimo, G. H., De Souza, J. I., et al. 2016. Morphological and phylogenetic analyses of three *Phytopythium* species (Peronosporales, oomycota) from Brazil. Cryptogamie Mycologie. 37:117-128.

De Maio., N., Shaw, L., Hubbard, L., George,S., Sanderson, N., Swann, J., et al. 2019. Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes. Microbial Genomics. 5(9).

Del Castillo-Munera, J., and Hausbeck, M. K. 2016. Characterization of *Pythium* Species Associated With Greenhouse Floriculture Crops in Michigan. Plant Disease. 100:569-576.

Del Castillo-Munera, J., Quesada-Ocampo, L. M., Rojas, A., Chilvers, M. I., and Hausbeck, M. K. 2019. Population structure of *Pythium ultimum* from greenhouse floral crops in Michigan. Plant Disease. 103:859-867.

Derevnina, L., Petre, B., Kellner, R., Dagdas, Y. F., Sarowar, M. N., Giannakopoulou, A., et al. 2016. Emerging oomycete threats to plants and animals. Philosophical Transactions of the Royal Society B. 371:20150459.

Dick M. W., Vick, M. C., Gibbings, J. G., Hederson, T. A., and Lopez-Lastra, C. C. 1999. 18S rDNA for species of Leptolegnia and other Peronosporomycetes: justification for the subclass taxa Saprolegniomycetidae and Peronosporomycetidae and division of the Saprolegniaceae sensu lato into the Leptolegniaceae and Saprolegniaceae. Mycological Research. 103:1119-1125.

Dick, M. W. 1991. Keys to *Pythium*. Mycologia. 83:386-387.

Edgar, R. C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. 32: 1792-1797.

Edwards, S. V. 2009. Is a new and general theory of molecular systematics emerging? Evolution. 63:1-19.

Eggertson, Q. 2012. Resolving the *Pythium ultimum* species complex. Carleton Universtiy (Master thesis). Ottawa, Ontario.

Eisen, J. A., and Frase, C. M. 2003. Phylogenomics: Intersection of Evolution and Genomics. Science. 300:1706-1707.

Eisen, J. A. 1998. Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. Genome Research. 8:163-167.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto K., Buckler, E. S., et al. 2011. A robust, simple genotyping by sequencing (GBS) approach for high diversity species. PLOS ONE. 6:e19379.

Escobar-Zepeda, A., Vera-Ponce de Leon, A., and Sanchez-Flores, A. 2015. The Road to Metagenomics: From Microbiology to DNA Sequencing Technologies and Bioinformatics. Frontiers in Genetics. 6:348.

Espindola, A. S.; Schneider, W.; Cardwell, K. F.; Carrillo, Y., Hoyt, P. R., Marek, S. M., et al. 2018. Inferring the presence of aflatoxin-producing *Aspergillus flavus* strains using RNA sequencing and electronic probes as a transcriptomic screening tool. PLOS ONE. 13:e0198575.

Espindola, A. S., 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. International Journal of Data Mining and Bioinfomatics.12:115-128.

Felsenstein, J. 2005. PHYLIP (Phylogeny Inference Package) version 3.6. Department of Genome Sciences, University of Washington, Seattle, Washington.

Feng, W., Hieno, A., Kusunoki, M., Suga, H., and Kageyama, K. 2019. LAMP detection of four plant pathogenic oomycetes and its application in Lettuce fields. Plant Disease. 103:298-307.

Feng, W., Ishiguro, Y., Hotta, K., Watanabe, H., Suga, H., and Kageyama, K. 2015. Simple detection of *Pythium irregulare* using loop-mediated isothermal amplification assay. FEMS Microbiology letters. 362:1-8.

FRAC. 2018. List of plant pathogenic organisms resistant to disease control agents. Available from: https://www.frac.info/docs/default-source/publications/list-of-resistant-plant-pathogens/list-of-resistant-plant-pathogenic-organisms_may-2018.pdf?sfvrsn=a2454b9a_2

Francis, D. M., and St. Clair, D. A. 1993. Outcrossing in the homothallic oomycete, *Pythium ultimum* detected with molecular markers. Current Genetics. 24:100-106.

Fry, W. E., and Grünwald, N. J. 2010. Introduction to Oomycetes. The Plant Health Instructor. DOI:10.1094/PHI-I-2010-1207-01.

Fukuta, S., Takahashi, R., Kuroyanagi, S., Ishiguro, Y., Miyake, N., Nagai, H., et al. 2014. Development of loop-mediated isothermal amplification assay for the detection of *Pythium myriotylum*. Letters in Applied Microbiology. 59:49-57.

Fukuta, S., Takahashi, R., Kuroyanagi, S., Miyake, N., Nagai, H., Suzuki, H., et al. 2013. Detection of *Pythium aphanidermatum* in tomato using loop-mediated isothermal amplification (LAMP) with species-specific primers. European Journal of Plant Pathology. 136:689-701.

Gandhi, S. R., and Weete, J. D. 1991. Production of the polyunsaturated fatty acids arachidonic acid and eicosapentaenoic acid by the fungus *Pythium ultimum.* Journal of General Microbiology. 137:1825-1830.

Garrido, P. 2014. Molecular characterization of *Pythium* populations in ornamental greenhouse and nursery crops. Oklahoma State University (Thesis). Stillwater, Oklahoma.

Garzón, C. D., Yánez, J. M., and Moorman, G. W. 2007. *Pythium cryptoirregulare*, a new species within the *P. irregulare* complex. Mycologia. 99:291-301.

Garzón, C. D., Geiser, D. M., and Moorman, G. W. 2005a. Amplified fragment length polymorphism analysis and internal transcribed spacer and *cox* II sequences reveal a species boundary within *Pythium irregulare*. Phytopathology. 95:1489-1498.

Garzon, C. D., Geiser, D. M., and Moorman, G. W. 2005b. Diagnosis and Population Analysis of *Pythium* Species Using AFLP Fingerprinting. Plant Disease. 89:81-89.

Gómez-Alpizar, L., Saalau, E., Picado, I., Tambong, J. T., and Saborio, F. 2010. A PCR-RFLP assay for identification and detection of *Pythium myriotylum*, causal agent of the cocoyam root rot disease. Letters in Applied Microbiology. 52:185-192.

Goss, E. M., Cardenas, M. E., Myers, K., Forbes, G. A., Fry, W. E., Restrepo, S., et al. 2011. The plant pathogen *Phytophthora andina* emerged via hybridization of an unknown *Phytophthora* species and the Irish potato famine pathogen, *P. infestans*. PLOS ONE. 6:e24543.

Goudet, J. 2004. Hierfstat , a package for R to compute and test hierarchical F-statistics. Molecular Ecology Resources. 5:184-186.

Grünwald, N. J., McDonald, B., and Milgroom, M. 2016. Population Genomics of Fungal and Oomycete Pathogens. Phytopathology. 54:323-346.

Grünwald, N. J., and Goss, E. 2011. Evolution and population genetics of exotic and re-emerging pathogens: Novel tools and approaches. Phytopathology. 49:249-267.

Grünwald, N. J., Goodwin, S. B., Milgroom, M. G., and Fry, W. E. 2003. Analysis of genotypic diversity data for populations of microorganisms. Phytopathology. 93:738-746.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, V., and Gascuel, O. 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Systematic Biology. 59:307–321.

Gurevich, A.; Saveliev, V.; Vyahhi, N. and Tesler, G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 29: 1072-1075.

Harvey, P. R., Butterworth, P. J., Hawke, B. G., and Pankhurst, C. E. 2001. Genetic and pathogenic variation among cereal, medic and sub-clover isolates of *Pythium irregulare.* Mycological Research. 105:85-93.

Haas, B. J., Kamoun, S., O'Neill, M. C, Jiang, R.H.Y., and Handsaker, R. E., et al. 2009. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans.* Nature. 461:393-398.

He, J., Zhao, X., Laroche, A., Lu, Z-X., Liu, H., and Li, Z. 2014. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Frontiers in Plant Science. 5:1-8.

Hedrick, P. W. 2011. Genetics of populations. Jones & Bartlett Learning. Fourth edition. T, Arizona.

Hendrix, F. F. and Campbell, W. A. 1973. Pythiums as plant pathogens. Annual Review of Phytopathology. 11:77-98.

Hermansen, A., Herrero, M.L., Gauslaa, E., Razzaghian, J., Nerstad, R., and Klemsdal. S.S. 2006. *Pythium* species associated with cavity spot on carrots in Norway. Annals of Applied Biology. 150: 115-121.

Ho, H. H. 2018. The taxonomy and biology of *Phytophthora* and *Pythium*. Journal of Bacteriology and Mycology: Open Access. 6:40-45.

Hoang, D., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Molecualr Biology and Evolution. 35:518-522.

Huelsenbeck, J., Bull, J.J., and Cunningham, C. 1996. Combining data in phylogenetic analysis. Trends Ecology and Evolution. 11:152–158.

Huson, D. H., Mitra, S., Ruscheweyh, H-J., Weber, N., and Schuster, S. C. 2011. Integrative analysis of environmental sequences using MEGAN4. Cold Spring Harbor Laboratory Press. 11:1552-1560.

Huson, D., and Bryant, D., Application of Phylogenetic Networks in Evolutionary Studies. 2006. Molecular Biology and Evolution. 23: 254–267.

Huson, D.H. 1998. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics. 14: 68-73.

Huzar-Novakowiski, J., and Dorrance, A. E. 2018. Genetic diversity and population structure of *Pythium irregulare* from Soybean and Corn production fields in Ohio. Plant Disease. 102:1989-2000.

Hyde, K. D., Henrik Nilsson, R., Aisyah Alias, S., Ariyawansa, H. A., Blair, J. E., Cai, L., et al. 2014. One stop shop: backbones trees for important phytopathogenic genera: I. Fungal Diversity. 67:21-125.

Ioos, R., Andrieux, A., Marcais, B., and Frey, P. 2006. Genetic characterization of the natural hybrid species *Phytophthora alni* as inferred from nuclear and mitochondrial DNA analyses. Fungal Genetics and Biology. 43:511-529.

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.* Journal of General Plant Pathology. 79:350-358.

Jain, M., Fiddes, I.T., Miga, K.H., Olsen, H. E., Paten, B., and Akeson, M. 2015. Improved data analysis for the MinION nanopore sequencer. Nature Methods. 12; 351-355.

Jiang, W., Chen, S.Y., Wang, H., Li, D.Z., and Wiens, J.J. 2014. Should genes with missing data be excluded from phylogenetic analyses? Molecular Phylogenetics and Evolution. 80:308-18.

Jombart, T., and Ahmed, I. 2011. Adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics. 27:3070-3071.

Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. 24:1403-1405.

Junier, T., and Zdobnov, E.M. 2010. The Newick Utilities: High-throughput Phylogenetic tree Processing in the UNIX Shell. Bioinformatics. 26:1669-1670.

Kageyama, K. 2014. Molecular taxonomy and its application to ecological studies of *Pythium* species. Journal of General Plant Pathology. 80:314-326.

Kageyama, K., Nakashima, A., Kajihara, Y., Suga, H., and Nelson, E.B. 2005. Phylogenetic and morphological analyses of *Pythium graminicola* and related species. Journal of General Plant Pathology. 71:174-182.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., and Jermiin, L. S. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nature Methods. 14: 587-589.

Kammarnjesadakul, P., Palaga, T., Sritunyalucksana, K., Mendoza, L., Krajaejun, T., Vanittanakom, N., et al. 2011. Phylogenetic analysis of *Pythium insidiosum* Thai strains using cytochrome oxidase II (*COX* II) DNA coding sequences and internal transcribed spacer regions (ITS). Medical Mycology. 49:289-295.

Kamoun, S., Furzer, O., Jones, J. D., Judelson, H. S., Ali, G. S., Dalio, R. J., et al. 2015. The Top 10 oomycete pathogens in molecular plant pathology. Molecular plant pathology. 16:413-434.

Kamoun, S. 2003. Molecular genetics of pathogenic oomycetes. Eukaryotic cell. 2:191-1999.

Kamoun, S., Huitema, E., and Vleeshouwers, G. A. A. 1999. Resistance to oomycetes: a general role for the hypersensitive response?. Trends in Plant Science Perspectives. 4: 196-200.

Kamvar, Z. N., Brooks, J. C., and Grünwald, N. J. 2015a. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Frontiers in Genetics. 6: 208.

Kamvar, Z. N., Larsen, M. M., Kanaskie, A. M., Hansen, E. M., and Grünwald, N. J. 2015b. Spatial and temporal analysis of populations of the Sudden Oak Death pathogen in Oregon forests. Phytopathology. 105:982-989.

Kamvar, Z. N., Tabima, J. F., and Grünwald, N. J. 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ. 2:e281.

Karlin, S. 2012. Population genetics and ecology. Elsevier.

Kassambara, A. Fastqcr: Quality Control of Sequencing Data. Available online: https://archive.st/archive/2020/3/cran.r-project.org/ln3a/cran.r-project.org/web/packages/fastqcr/index.html.

Kassambara, A., and Mundt, F. 2017. Factoextra: Extract and Visualize the Results of Multivariate Data Analyses.

Katoh, K., and Standley, D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecualr Biology and Evolution. 30: 772-80.

Keiro, S., Daniels, H.A., Gomez-Gallego, M., Tabima, J.F., Lenz, K.L., and Sondreli, K.L. et al. 2019. From genomes to forest management – tackling invasive *Phytophthora* species in the era of genomics. Canadian Journal of Plant Pathology. 42:1-29.

Keller, O., Kollmar, M., Stanke, M., and Waack, S. 2011. A novel hybrid gene prediction method employing protein multiple sequence alignments. Bioinformatics. 27:757-63.

Kernaghan, G., Reeleder, R. D., and Hoke, S.M.T. 2008. Quantification of *Pythium* populations in ginseng soils by culture dependent and real-time PCR methods. 40:447-455.

Kircher, M., Heyn, P., and Kelso, J. 2011. Addressing challenges in the production and analysis of illumina sequencing data. BMC Genomics. 12:382.

Kirk, P., Cannon, P., David, J., and Stalpers, J. 2001. Ainsworth & Bisby's Dictionary of the Fungi, 2001. CAB International. Wallingford UK.

Klassen, G.R., Balcerzak, M., Cock, A.W.A.M. 1996. 5S ribosomal RNA gene spacers as species specific probes for eight species of *Pythium.* Phytopathology. 86:581-587.

Klemsdal, S. S., Herrero, M. L., Wanner, L. A., Lund, G., and Hermansen, A. 2008. PCR-based identification of *Pythium* spp. causing cavity spot in carrots and sensitive detection in soil samples. Plant Pathology. 57:877-886.

Knaus, B. J., and Grünwald, N. J. 2017. VCFR: a package to manipulate and visualize variant call format data in R. Molecular Ecology Resources. 17:44-53.

Ko, W. H., Lin, M. J., Hu, C. Y., and Ann, P. J. 2010. *Aquaperonospora taiwanensis* gen. et sp. nov. In Peronophythoraceae of Peronosporales. Botanical studies. 51:343-350.

Kriventseva, E., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F., et al. 2018. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic Acids Research. 47:D807–D811.

Kubatko, L.S., and Degnan, J.H. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Systematic Biology. 56: 17-24.

Kyuchukova, M. A., Buttner, C., Gabler, J., Bar-Yosef, B., Grosch, R., and Klaring, H. P. 2006. Evaluation of a method for quantification of *Pythium aphanidermatum* in cucumber roots at different temperatures and inoculum densities. Journal of Plant Diseases and Protection. 113:113-119.

Lamour, K. W., Win, J., and Kamoun, S. 2007. Oomycete genomics: new insights and future directions. FEMS Microbiology letters. 274: 1-8.

Lanier, H., and Knowles, L. 2012. Is Recombination a Problem for Species-Tree Analyses?, Systematic Biology. 61: 691–701.

Latijnhouwers, M., de Wit P. J., and Govers, F. 2003. Oomycetes and fungi: similar weaponry to attack plants. Trends in Microbiology. 11:462-469.

Laver, T., Harrison, J., O'Neill, P.A., Moore, K., Farbos, A. Paszkiewicz, K., et al. 2015. Assesing the performance of the Oxford Nanopore Technologies MinION. Biomolecular Detection and Quantification. 3:1-8.

Lazreg, F., Belabid, L., Sanchez, J., Gallego, E., and Garrido-Cardenas, J. A. 2013. First Report of *Globisporangium ultimum* causing Pythium Damping-Off on Aleppo Pine in Algeria, Africa, and the Mediterranean Region. Plant Disease. 97:1111-1117.

Leclerc, M. C., Guillot, J., and Deville, M. 2000. Taxonomic and phylogenetic analysis of Saprolegniaceae (Oomycetes) inferred from LSU rDNA and ITS sequence comparisons. Antonie Van Leeuwenhoek. 77:369–377.

Lee, S., Garzón, 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. Current Genetics. 53:81-93.

Lee, Y. S. 1993. Comparisons of isozyme patterns in *Pythium* species and application to *Pythium* systematics. The Korean Journal of Mycology. 21:293-300.

Le Floch, G., Tambong, J., Vallance, J., Tirilly, Y., Lévesque, A., and Rey, P. 2007. Rhizosphere persistence of three *Pythium oligandrum* strains in tomato soilless culture assessed by DNA macroarray and real-time PCR. FEMS Microbiology Ecology. 61:317-326.

Lendenmann, M. H., Croll, D., and McDonald, B.A. 2015. QTL mapping of fungicide sensitivity reveals novel genes and pleiotropy with melanization in the pathogen *Zymoseptoria tritici.* Fungal Genetics and Biology. 80:53-67.

Letunic, I., and Bork, P. 2007. Interactive Tree of Life (iTOL): An online tool for phylogenetic tree display and annotation. Bioinformatics. 23: 127-128.

Lévesque, C. A. 2011. Fifty years of oomycetes: from consolidation to evolutionary and genomic exploration. Fungal Diversity. 50:35-46.

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

Lévesque, C. A. Harlton, C. E., Cock, A.W.A.M. 1997. Identification of some oomycetes by reverse dot blot hybridization. Phytopathology. 88:213-222.

Lévesque, C. A., Beckenbach, K., Baillie, D. L., and Rahe, J. E. 1993. Pathogenicity and DNA restriction fragment length polymorphisms of isolates of *Pythium* spp. from glyphosate treated seedlings. Mycological Research. 97:307-312.

Li, M., Ishiguro, Y., Otsubo, K., Suzuki, H., Tsuji, T., Miyake, N., et al. 2014. Monitoring by real-time PCR of three water-borne zoosporic *Pythium* species in potted flower and tomato greenhouses under hydroponic culture systems. European Journal of Plant Pathology. 140:229-242.

Li, M., Senda, M., Komatsu, T., Suga, H., and Kageyama, K. 2010. Development of real-time PCR technique for the estimation of population density of *Pythium intermedium* in forest soils. Microbiological Research. 165:695-705.

Lievens, B., Brouwer, M., Vanachter, A.C.R.C., Cammue, B.P.A., and Thomma, B.P.H.J. 2006. Real-time PCR for detection and quantification of fungal and oomycete tomato pathogens in plant and soil samples. Plant Science. 171:155-165.

Lievens, B., Brouwer, M., Vanachter, A.C.R.C., Lévesque, C. A., Cammue, B.P.A., and Thomma, B.P.H.J. 2005. Quantitative assessment of phytopathogenic fungi in various substrates using a DNA macroarray. Environmental Microbiology. 7:1698-1710.

Liu, F., Wang, M., Damm, U., Crous, P.W., and Cai, L. 2016. Species boundaries in plant pathogenic fungi: a *Colletotrichum* case study. BMC Evolutionary Biology. 16:81.

Lu, F., Glaubitz, J., Harriman, J., Casstevens, T., and Elshire, R. 2012. TASSEL 3.0 Universal Network Enabled Analysis Kit (UNEAK) pipeline documentation.

Ludwig, J. A., and Reynolds, J. F. 1988. Statistical ecology: a primer in methods and computing. John Wiley & Sons.

Lyons, N. F., and White, J. G. 1992. Detection of *Pythium violae* and *Pythium sulcatum* in carrots with cavity spot using competition ELISA. Annals of Applied Biology. 120:235-244.

MacDonald, J. D., Stites, J., and Kabashima, J. 1990. Comparison of serological and culture methods for detecting species of *Phytophthora, Pythium* and *Rhizoctonia* in ornamental plants. Plant Disease. 74:655-659.

Malapi-Wight, M., Salgado-Salazar, C., Demers, J., Clement, D., Rane, K., and Crouch, J. 2016. Sarcococca Blight: Use of Whole-Genome sequening for fungal plant disease diagnosis. Plant Disease. 100:1093-1100.

Mallet, J., Besansky, N., and Hahn, M.W. 2016. How reticulated are species? BioEssays. 38: 140–149.

Mardis, E.R. 2008. The impact of next-generation sequencing technology on genetics. Trends in Genetics. 24:133-141.

Margulis, L., and Chapman, M. J. 2009. Kingdoms and domains: An illustrated guide to the phyla of life on earth. W. H. Freeman and Company. New York.

Martin, F. 2009. Pythium Genetics. In *Oomycete Genetics and Genomics*, eds, Lamour, K and Kamoun, S. New York: John Wiley & Sons Inc.

Martin, F.N. 2000. 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 Porspects for Biological control. Critical Reviews in Plant Science. 18: 111-181.

Martin, F. N., and Kristler, H. C. 1990. Species-Specific Banding Patterns of Restriction Endonuclease-Digested Mitochondrial DNA from the Genus *Pythium.* Experimental Mycology. 14:32-46.

Matari, N. H., and Blair, J. E. 2014. A multilocus timescale for oomycete evolution estimated under three distinct molecular clock models. BMC Evolutionary Biology. 14:101.

Matsumoto, C., Kageyama, K., Suga, H., and Hyakumachi, M. 2000. Intraspecific DNA polymorphisms of *Pythium irregulare*. Mycological Research. 104:1333-1341.

Matsumoto, C., Kageyama, K., Suga, H., and Hyakumachi, M. 1999. Phylogenetic relationships of *Pythium* species based on ITS and 5.8S sequences of the ribosomal DNA. Mycoscience. 40:321-331.

McCarthy, C.G.P., and Fitzpatrick, D. A. 2017. Phylogenomic Reconstruction of the Oomycete Phylogeny Derived from 37 Genomes. mSphere. 2:e00095-17.

McGowan, J., Fitzpatrick D.A. 2020. Recent advances in oomycete genomics. Advances in Genetics. 105:175-228.

Mendoza, L. 1998. *Pythium insidiosum*. In T*opley & Wilson's microbiology and microbial infections*, eds. L. Collier, A, Balows, and M. Sussman. London, England. p. 617-630.

Middleton, J. T. 1943. The taxonomy, host range and geographic distribution of the genus *Pythium.* Memoirs of the Botanical Society. 20:1-171.

Milgroom, M. G. 2015. Population Biology of Plant Pathogens: Genetics, Ecology, and Evolution. St. Paul, MN. APS Press.

Milgroom, M. G., Jimenez-Gasco, M., Olivares-Garcia, C., Drott, M. T., and Jimenez-Diaz, R. 2014. Recombination between Clonal Lineages of the Asexual Fungus *Verticillium dahliae* Detected by Genotyping by Sequencing. PLOS ONE. 9:e106740.

Milgroom, M. G., and Peever, T. L. 2003. Population biology of plant pathogens: the synthesis of plant disease epidemiology and population genetics. Plant Disease. 87:608-617.

Milgroom, M. G. 1996. Recombination and the multilocus structure of fungal populations. Annual Reviews Phytopathology. 34: 457-477.

Mirarab, S., and Warnow, T. 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics. 31:i44–i52.

Mirarab, S., Reaz, R., Bayzid, M., Zimmermann, T., Swenson, M., and Warnow, T. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics. 30: i541–i548.

Momand, J., and McCurdy, A. 2017. Concepts in Bioinformatics and Genomics. Oxford University Press. New York, US.

Moorman, G., and Kim, S. 2004. Species of *Pythium* from greenhouses in Pennsylvania exhibit resistance to propamocarb and mefenoxam. Plant Disease. 88:630-632.

Moorman, G. W., Kang, S., Geiser, D. M., Kim, H. 2002. Identification and characterization of *Pythium* species associated with greenhouse floral crops in Pennsylvania. Plant Disease. 86:1227-1231.

Moralejo, E., Clemente, A., Descals, E., Belbahri, L., Calmin, G., Lefort, F., et al. 2008. *Pythium recalcitrans* sp. nov. revealed by multigene phylogenetic analysis. Mycologia. 100:310-319.

Murillo-Williams, A., and Pedersone, P. 2008. Early Incidence of Soybean Seedling Pathogens in Iowa. Agronomy Journal. 100:1481-1487.

Nguyen, H.D.T., Humpries, Z., Dodge, A., Findlay, W., and Rintoul, T.L. 2018. Evaluating the nuclear gene regions for phylogenetic species delimitation in the *Pythium irregulare* species complex. APS National Meeting Poster.

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., and Minh, B. Q. 2015. IQ-TREE: A fast and stocastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution. 32: 268-274.

Njambere, E. N., Clarke, B. B., and Zhang, N. 2011. Dimeric oligonucleotide probes enhance diagnostic macroarray performance. Journal of Microbiological Methods. 86:52-61.

Nzungize, J. R., Lyumugabe, F., Busogoro, J. P., and Baudoin, J. P. 2012. Pythium root rot of common bean: biology and control methods. A review. Biotechnologie, Agronomie, Société et Environnement. 16:405-412.

Nzungize, J., Gepts, P., Buruchara, R., Buah, S., Ragama, P., Busogoro, J. P., and Baudoin, J. P. 2011. Pathogenic and molecular characterization of *Pythium* species inducing root rot symptoms of common bean in Rwanda. African Journal of Microbiology Research. 5:1169-1181.

Okubara, P. A., Schroeder, K. L., and Paulitz, T. C. 2005. Real-time polymerase chain reaction: applications to studies on soilborne pathogens. Canadian Journal of Plant Pathology. 27:300-313.

Paradis, E., and Schliep, K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 35: 526-528.

Patterson, DJ. 1989. Stramenopiles: chromophytes from a protistan perspective. In *The Chromophyte algae: problems and perspectives*, eds. Green, J. C, Leadbeater, B.S.C and Diver, W. L. Oxford Clarendon Press. p. 357-379.

Patwardhan, A., Ray, S., and Roy, A. 2014. Molecular Markers in Phylogenetic Studies-A Review. Journal of Phylogenetics and Evolutionary Biology. 2:1-9.

Paul, B., and Masih, I. 2000. ITS1 region of the nuclear ribosomal DNA of the mycoparasite *Pythium periplocum*, its taxonomy, and its comparison with related species. FEMS microbiology letters. 189:61-65.

Perneel, M., Tambong, J. T., Adiobo, A., Floren, C., Saborio, F., Lévesque, A., et al. 2006. Intraspecific variability of *Pythium myriotylum* isolated from cocoyam and other host crops. Mycological Research. 110:583-593.

Pettitt, T. R., Wakeham, A. J., Wainwright, M. F., and White, J.G. 2002. Comparison of serological, culture, and bait methods for detection of *Pythium* and *Phytophthora* zoospores in water. Plant Pathology. 51:720-727.

Philippe, H., Delsuc, F., Brinkmann, H., and Lartillot, N. 2005. Phylogenomics. Annual Review of Ecology, Evolution and Systematics. 36:541–562.

Pielou, E. 1975. Ecological diversity. New York: John Wiley & Sons.

Pop, M. 2009. Genome assembly reborn: recent computational challenges. Briefings in Bioinformatics. 10:354-366.

Poppe, S., Dorsheimer, L., Happel, P., and Stukenbrock, E. H. 2015. Rapidly evolving genes are key players in host specialization and virulence of the fungal wheat pathogen *Zymoseptoria tritici* (*Mycosphaerella graminicola*).PLOS Pathogones. 11:e1005055.

Pringsheim, N. 1858. Beiträge zur Morphologie and Systematik der Algen. 2. Die Saprolegnieen. Jahrbücher für Wissenschaftliche Botanik. 1:284-306.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Raffaele, S., Farrer, R. A., Cano, L. M., Studholme, D. J., MacLean, D., Thines, M., et al. 2010. Genome evolution following host jumps in the Irish potato famine pathogen lineage. Science 330:1540-43.

Rafin, C., Brygoo, Y., and Tirilly, Y. 1995. Restriction analysis of amplified ribosomal DNA of *Pythium* spp. isolated from soilless culture systems. Mycological Research. 99:277-281.

Rafin, C., Nodet, P., and Tirilly, Y. 1994. Immuno-enzymatic staining procedure for *Pythium* species with filamentous non-inflated sporangia in soilless cultures. Mycological Research. 98:535-541.

Restrepo, S., Tabima, J. F., Mideros, M. F., Grünwald, N. J., and Matute, D. R. 2014. Speciation in Fungal and Oomycete Plant Pathogens. Phytopathology. 52:289-316.

Richards, T. A., Soanes, D. M., Jones, D. M., Vasieva, O., Leonard, G., Paszkiewicz, K., et al. 2011. Horizontal gene transfer facilitated the evolution of plant parasitic mechanisms in the oomycetes. PNAS. 108:5258-15263.

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

Riethmüller, A., WeiB, M., and Oberwinkler, F. 1999. Phylogenetic studies of Saprolegniomycetidae and related groups based on nuclear large subunit DNA sequences. Canadial Journal of Botany. 77:1790-1800.

Ritcher, D. C., Ott, F., Auch, A. F., Schmid, R., and Huson, D. H. 2008. MetaSim-A Sequencing Simulator for Genomics and Metagenomics. PLOS ONE. 3:e3373.

Robideau, G. P., de Cock, A.W.A.M., Coffey, M. D., Voglmayr, S., Brouwer, H., Bala, A. K., et al. 2011. DNA barcoding of oomycetes with cytochrome c oxidase subunit I and internal transcribed spacer. Molecular Ecology Resources. 11:1002-1011.

Rokas, A., Williams, B.L., King, N., Carroll, S.B., 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature. 425: 798–804.

Ruggiero, M. A., Gordon, D., Orrell, T. M., Bailly, N., Bourgoin, T., Brusca, R. C., et al. 2015. Correction: a higher level classification of all living organisms. PLOS ONE. 10:e0130114.

Rujirawat, T., Patumcharoenpol, P., Lohnoo, T., Yingyong, W., Kumsang, Y., Payattikul, P., et al. 2018. Probing the Phylogenomics and Putative Pathogenicity Genes of *Pythium insidiosum* by Oomycete Genome Analyses. Scientific reports. 8:4135.

Sanger, F., Nicklen, S., and Coulson, A. R. 1997. DNA sequencing with chain-terminating inhibitors. PNAS. 74:5463-5467.

Sayyari, E., and Mirarab, S. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Molecular Biology Evolution. 33: 1654-1668.

Schardl, C.L., and Craven, K.D. 2003. Interspecific hybridization in plant-associated fungi and oomycetes: a review. Molecualr Ecology. 12: 2861-2873.

Schroeder, K. L., Martin, F. N., de Cock, A. W., Lévesque, C. A., Spies, C. F., Okubara, P. A., and Paulitz, T. C. 2013. Molecular detection and quantification of *Pythium* species: evolving taxonomy, new tools, and challenges. Plant Disease. 97:4-20.

Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments. Nature Biotechnology. 18: 233–234.

Schuster, S. 2008. Next-generation sequencing transforms today's biology. Nature Methods. 5:16-18.

Schwarz, R.H. 1978. Estimating the dimension of a model. The Annals of Statistics. 6: 461-464.

Shannon, C. E. 1949. The Mathematical Theory of Communication. Weaver; W. Urbana.IL.

Shen, D., Li, Q., Yu, J., Zhao, Y., Zhu, Y., Xu, H., and Dou, D. 2017. Development of a loop-mediated isothermal amplification method for the rapid detection of *Pythium ultimum.* Australasian Plant Pathology. 46:571-576.

Short, D.P.G., Gurung, S., Hu, X., Inderbitzin, P., and Subbarao, K. V. 2014. Maintenance of sex-related genes and the co-occurrence of both mating types in *Verticillium dahliae*. PLOS ONE. 9:e112145.

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

Sonah, H., Bastien, M., Iquira, E., Tardivel, A., Legare, G., Boyle, B., et al. 2013. An Improved Genotyping by Sequencing (GBS) Approach Offering Increased Versatility and Efficiency of SNP Discovery and Genotyping. PLOS ONE. 8:e54603.

Spies, C. F., Mazzola, M., Botha, W. J., Langenhoven, S. D., Mostert, L., and Mcleod, A. 2011. Molecular analyses of *Pythium irregulare* isolates from grapevines in South Africa suggest a single variable species. Fungal biology. 115:1210-1224.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30: 1312-1313.

Stewart, E. L., Croll, D., Lendenmann, M. H., Sanchez-Vallet, A., and Hartmann, F. E, et al. 2016. QTL mapping reveals complex genetic architecture of quantitative virulence in the wheat pathogen *Zymoseptoria tritici*. bioRxiv.

Stewart, J.E., Timmer, L.W., Lawrence, C.B., Pryor, B.M.,and Peever, T.L. 2014. Discord between morphological and phylogenetic species boundaries: incomplete lineage sorting and recombination results in fuzzy species boundaries in an asexual fungal pathogen. BMC Evolutionary Biology. 14:38.

Stinchcombe, J., and Hoekstra, H. 2008. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. Heredity. 100:158-170.

Stobbe, A. H., Schneider, W. L., Hoyt, P. R., and Melcher, U. 2014. Screening Metagenomic data for viruses using the E-probe Diagnostic Nucleic Acid Assay. Phytopathology. 104:1125-1129.

Stobbe, A. H., Daniels, J., Espindola, A. S., Verma, R., Melcher, U., Ochoa-Corona, F., et al. 2013. E-probe Diagnostic Nucleic acid Analysis (EDNA): A theoretical approach for handling of next generation sequencing data for diagnostics. Journal of Microbiological Methods. 94:356-66.

Stoddart, J. A., and Taylor, J. F. 1988. Genotypic diversity: estimation and prediction in samples. Genetics. 118:705-711.

Tabima, J., Coffey, M.D., Zazada, I., and Grunwald, N.J. 2018. Populations of Phytophtthora rubi show little differentiation and high rates of migration amog states in the western United States. MPMI. 31: 614-622.

Takahashi, R., Fukuta, S., Kuroyanagi, S., Miyake, N., Nagai, H., Kageyama, K., and Ishiguro, Y. 2014. Development and application of a loop-mediated isothermal amplification assay for rapid detection of *Pythium helicoides*. FEMS Microbiology letters. 355:28-35.

Talas, F., Kalih, R., Miedaner, T., and McDonald, B. A. 2016. Genome-wide association study identifies novel candidate genes for aggressiveness, deoxynivalenol production, and azole sensitivity in natural field populations of *Fusarium graminearum*. Molecular Plant-Microbe Interaction. 29:417-430.

Talas, F., and McDonald, B. A. 2015. Genome-wide analysis of *Fusarium graminearum* field populations reveals hotspots of recombination. BMC Genomics. 16:996.

Tambong, J. T., de Cock, A.W.A.M., Tinker, N. A., and Lévesque, C. A. 2006. Oligonucleotide Array for Identification and Detection of *Pythium* Species. Applied and Environmental Microbiology. 72:2691-2706.

Taylor, J.W., Jacobson, D. J., Kroken, S., Kasuga, T., Gieser, D. M. et al. 2000. Phylogenetic Species Recognition and Species Concepts in Fungi. Fungal genetics and Biology. 31: 21-32.

Tyler, B. M., Tripathy, S., Zhang, X., Dehal, P., Jiang, R. H. Y., Aerts, A., et al. 2006. *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science. 313:1261e1266.

Uzuhashi, S., Tojo, M., and Kakishima, M. 2010. Phylogeny of the genus *Pythium* and description of new genera. Mycoscience. 51:337-365.

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

Van Dijik, E.L., Auger, H., Jaszczyszyn, Y., and Thermes, C. 2014. Ten years of next-generation sequencing technology. Trends in Genetics 30:418-426.

Van Orsouw, N. J., Hoger, R.C.J., Janssen, A., Yalcin, F., Snoeijers, S. Vestege, E., et al. 2007. Complexity Reduction of Polymorphic Sequences (CRoPSTM): A Novel Approach for Large-Scale Polymorphism Discovery in Complex Genomes. PLOS ONE. 11:e1172.

Verma, M., Kulshrestha, S., and Puri, A. 2016. Genome Sequencing. In: Keith, J. Bioinformatics Volume I: Data, Sequence Analysis, and Evolution. Second edition. Humana Press. Melbourne, Australia. pp. 3-34.

Vijaya Satya, R., Zavaljevski, N., Kumar, K. and Reifman, J. 2008a. A high-throughput pipeline for designing microarray-based pathogen diagnostic assays. BMC Bioinformatics. 9:185.

Vijaya, R., Zavaljevski, N., Kumar, K., Bode, E., Padilla, S., Wasieloski, L., et al. 2008b. *In silico* microarray probe design for diagnosis of multiple pathogens. BMC Genomics. 9:496.

Vijayan, P., Shockey, J., Lévesque, C. A., Cook, R. J., and Browse, J. 1998. A role for jasmonate in pathogen defense of *Arabidopsis.* PNAS. 95:7209-7214.

Villa, N.O., Kageyama, K., Asano, T., Suga, H. 2006. Phylogenetic relationships of *Pythium* and *Phytophthora* species based on ITS rDNA, cytochrome oxidase II and β-tubulin gene sequences. Mycologia. 98:410-422.

Wakeham, A. J., and Pettitt, T. R. 2017. Diagnostic tests and their application in the management of soil- and water-borne oomycete pathogen species. Annals of Applied Biology. 170:45-67.

Wang, P. H., Wang, Y. T., and White, J. G. 2003. Species-specific PCR primers for *Pythium* developed from ribosomal ITS1 region. Letters in Applied Microbiology. 37:127-132.

Wang, P. H., and White, J. G. 1997. Molecular characterization of *Pythium* species based on RFLP analysis of the internal transcribed spacer region of ribosomal DNA. Physiological and Molecular Plant Pathology. 51:129-143.

Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., et al. 2018. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. Molecular Biology and Evolution. 35:543-548.

Waterhouse, R.M., Tegenfeldt, F, Li J., Zdobnov, E.M., and Kriventseva, E.V. 2013. OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. Nucleic Acids Research.41: D358–D365.

Weiland, J. E., Scagel, C. F., Grünwald, N. J., Davis, E. A., Beck, B. R., Foster, Z. S. L., and Fieland, V. J. 2020. Soilborne *Phytophthora* and *Pythium* diversity from

rhododendron in propagation, container, and field production systems of the Pacific Northwest. Plant Disease. In press.

Weiland, J. E., Scagel, C. F., Grünwald, N. J., Davis, E. A., Beck, B. R., and Fieland, V. J. 2018. Variation in disease severity caused by *Phytophthora cinnamomi, P. plurivora,* and *Pythium cryptoirregulare* on two rhododendron cultivars. Plant Disease. 102:2560-2570.

Weiland, J. E., Garrido, P., Kamvar, Z. N., Espindola, A. S., Marek, S. M., Grünwald, N. J., and Garzon, C. D. 2015. Population Structure of *Pythium irregulare, P. ultimum,* and *P. sylvaticum* in Forest Nursery Soils of Oregon and Washington. Phytopathology. 105:684-694.

Weiland, J. E., Santamaria, L., and Grünwald, N. J. 2014. Sensitivity of *Pythium irregulare, P. sylvaticum,* and *P. ultimum* from Forest nurseries to mefenoxam and fosetyl-Al, and control of Pythium Damping-off. Plant Disease. 98:937-942.

Weiland, J.E. 2011. Influence of isolation method on recovery of *Pythium* species from forest nursery soils in Oregon and Washington. Plant Disease. 95:547-553.

White, T. J., Bruns, T., Lee, S., and Taylor, J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: PCR Protocols, a Guide to Methods and Applications. Academic Press. San Diego, USA. pp 315-322.

Wick, R., Judd, L., Gorrie, C., and Holt, K. 2017. Completing bacterial genome assemblies with multiplex MinION sequencing. Microbial Genomics. 3:1-7.

Xi Z, L., Rest, J.S., Davis, C.C. 2014. Coalescent versus concatenation methods and the placement of Amborella as sister to water lilies. System Biology. 63: 919-32.

Yin, L., An, Y., Qu, J., Li, X., Zhang, Y., Dry, I., Wu, H., and Lu, J. 2017. Genome sequence of *Plasmopara viticola* and insight into the pathogenic mechanism. Scientific Reports. 7:46553.

Yuen, G. Y., Xia, J. Q., and Sutula, C. L. 1998. A Sensitive ELISA for *Pythium ultimum* Using Polyclonal and Species-Specific Monoclonal Antibodies. Plant Disease. 82:1029-1032.

Zhang, C., Scornavacca, C., Molloy, E. K., and Mirarab, S. 2019. ASTRAL-Pro: quartet-based species tree inference despite paralogy. bioRxiv. doi: https://doi.org/10.1101/2019.12.12.874727.

Zhou, W., Motohashi, K., Suga, H., Fukui, H., and Kageyama, K. 2009. Development of microsatellite markers for *Pythium helicoides*. FEMS microbiology letters. 293:85-91.

Zody, M., and Nusbaum, C. 2009. Strategy and tactics for genome sequencing. In *Oomycete Genetics and Genomics*, eds. Lamour, K and Kamoun. New York: John Wiley & Sons Inc.

VITA

Maria Fernanda Proaño Cuenca

Candidate for the Degree of

Doctor of Philosophy

Dissertation: GENOMIC RESOURCES FOR PHYLOGENETICS, POPULATION GENETICS AND DETECTION OF OOMYCETE PLANT PATHOGENS

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, 2020.

Completed the requirements for the Master of Science in Entomology and Plant Pathology at Oklahoma State University, Stillwater, Oklahoma in 2016.

Completed the requirements for the Bachelor of Science in Biotechnology Engineering at Universidad de las Fuerzas Armadas ESPE, Sangolquí, Ecuador in 2014.

Experience:

Graduate Research Assistant and Teaching Assistant, Department of Entomology and Plant Pathology, Oklahoma State University, Stillwater, Oklahoma, from January 2015 to December 2020.

Professional Memberships:

American Phytopathological Society (APS)
Oomycete Molecular Genetics Network (OMGN)