AN ABSTRACT OF THE DISSERTATION OF

Shankar Kaji Shakya for the degree of Doctor of Philosophy in Botany and Plant Pathology presented on October 11, 2019.

Title: Phylogeography and Evolution of Effector Genes in *Phytophthora* Species.

Abstract approved:

_____

Niklaus J. Grünwald

Plant pathogens in the genus *Phytophthora* are known to cause disease on field crops, nursery plants, and forest trees. The best known example probably is *Phytophthora infestans*, which triggered the infamous Irish potato famine. Other important *Phytophthora* species include: *P. ramorum* (sudden oak death pathogen), *P. sojae* (soybean root rot pathogen) and *P. cinnamomi* (broad host range root rot pathogen). Knowledge and understanding of a pathogen's host range, geographic distribution, genetic diversity, and evolutionary potential is a key to successful disease management. This thesis focused on using phylogeographic principles combined with genomics tools to answer questions regarding the expansion from the center of origin of *P. infestans*, identifying the putative center of origin for *P. cinnamomi* and evolution of RxLR (Arginine-any amino acid-Leucine-Arginine) effector genes in three *Phytophthora* species.

A pathogen's biology at its center of origin may be different compared to other geographical regions. This has been well demonstrated for the late blight pathogen *P. infestans*.

Toluca Valley of Mexico is considered the center of origin of *P. infestans*, where it reproduces sexually and both mating types exist in a 1:1 ratio. Very few studies have looked into the population structure outside of Toluca Valley, but still in Mexico. We performed a detailed analyses using microsatellite markers and identified a gradient of genetic diversity in *P. infestans* populations outside of the Toluca Valley.

We used evidence of equal mating type frequencies, high genetic diversity, sexual reproduction combined with phylogenetics to identify the putative center of origin of broad host range pathogen *P. cinnamomi* from a global population sample. Our result suggests Asia as the center of origin for the broad host range pathogen *P. cinnamomi* and subsequent migration out of Asia to Australia and Africa.

Phytophthora species secrete numerous effectors during the process of infection which are thought to modulate the host defense system. One such class of effector is RxLR which help promote the infection and colonization of host cells. Some of these RxLR effectors can be targeted by plant resistance genes to initiate a defense response. The question of which of these genes have been conserved or lost during the evolution of *Phytophthora* species is largely unknown. Therefore, we used whole genome sequence data and a comparative genomics approach to test the hypothesis of variation in RxLR gene content in three *Phytophthora* clade 1C species. Our analysis demonstrated both conservation and proliferation of RxLR genes in these species along with high sequence diversity. We also observed emergence of a virulent allele for effector *Avr3a* in the modern isolates of *P. infestans*. Balancing selection has acted on the RxLR genes maintaining the high sequence diversity. The ascertainment bias due to mapping of short sequencing reads to the single reference genome was further quantified by assembling the unmapped reads and predicting genes shared by all isolates of *P. infestans.*

Overall, this thesis broadens our phylogeographic and evolutionary understanding of pathogen *Phytophthora* species. A key to breeding for disease resistance is highly dependent on the effector repertoire of a pathogen and thus our work can help breeders to screen for resistance gene against the late blight pathogen.

©Copyright by Shankar Kaji Shakya
October 11, 2019
All Rights Reserved

Phylogeography and Evolution of Effector Genes in *Phytophthora* Species

by

Shankar Kaji Shakya

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented October 11, 2019
Commencement June 2020

Doctor of Philosophy dissertation of Shankar Kaji Shakya presented on October 11, 2019

APPROVED:

Major Professor, representing Botany and Plant Pathology

Head of the Department of Botany and Plant Pathology

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Shankar Kaji Shakya, Author

ACKNOWLEDGEMENTS

I like to thank my major advisor Niklaus Grünwald for his support and patience during my graduate study. Current and former lab members of Grünwald lab have been very supportive from the very beginning: Brian Knaus, Alex Weisberg, Zhian Kamvar, Javier Tabima, Zach Foster, Nick Carleson, Caroline Press, Meg Larsen, Val Fieland, Karan Fairchild and Kenny Rolfe. Thank you all.

I must thank my wonderful committee members Tom Wolpert, Chris Mundt, Jim Myers and Darell Ross who constantly provided feedback to my work. Faculty, staffs and students of Botany and Plant Pathology department definitely have played a huge role in my success. I am thankful to Lynda Ciuffetti, Joey Spatafora and Ken Johnson for providing me an opportunity to TA plant pathology for two years. I want thank my TA partners Hannah Rivedal and Alex Wong. I also want to thank Andy Jones who provided space for me to work during the government shutdown.

During my Ph.D. I regularly attended the American Phytopathology Society (APS) meetings. I received several supports for the meeting from the Department of Botany and Plant Pathology through Anita Summers travel grant and a travel award from APS. I also want to thank APS for providing me with an APS Experiential award in 2019 which supported my visit to J.R. Simplot Company in Boise, Idaho. Thanks to Nico Champouret for hosting me in Boise.

My computational skills were honed by the trainings provided by CGRB. Thanks to Shawn Oneill and CGRB community for that. I also appreciate the support I received from the USDA family. Thanks to Bob Martin, Inga Zasada, Jerry Weiland, Joyce Loper, Virginia Stockwell.

Being away from home was tough but presence of Nepali family and friends here in Corvallis have made my stay very wonderful. Thanks to Deepak Rayamaghi, Rajendra Soti, Krishna Poudel, Manoj Gurung and Bhim Chaulagain.

Whatever I have achieved wouldn't have been possible without the support, patience and trust from my family members. I want to thank my mother (Tara), my father (Ram), my sisters (Sarita and Sabita) and my wife (Sujata). Our first child Sushank has contributed to thesis by being such a good child so I could focus on writing. Thank you all.

CONTRIBUTION OF AUTHORS

Chapter 2

Shankar K. Shakya, Héctor Lozoya-Saldaña and Niklaus Grünwald conceived the study. Héctor Lozoya-Saldaña and Mercedes María Cuenca-Condoy collected *P. infestans* samples from Mexico. Meg Larsen performed the DNA extraction and microsatellite genotyping. Shankar K. Shakya and Niklaus Grünwald analyzed the data and wrote the manuscript.

Chapter 3

Shankar K. Shakya, Thomas Jung and Niklaus Grünwald conceived the study. Jerry Weiland, Marillia Horta Jung, Cristiana Maia, Andre Drenth, David Guest, Edward Liew, Colin Crane, Bruno Scanu and Thomas Jung contributed to the isolates. Shankar K. Shakya, Brian Knaus and Niklaus Grünwald analyzed the data. Shankar Shakya, Thomas Jung and Niklaus Grünwald wrote the manuscript.

Chapter 4

Shankar K. Shakya, Amirhossein Bahramisharif, Laura Rose, Brian Knaus, Howard Judelson and Niklaus Grünwald conceived the design. Shankar K. Shakya, Amirhossein Bahramisharif and Brian Knaus generated and analyzed the data. Shankar K. Shakya, Amirhossein Bahramisharif, Laura Rose, Brian Knaus and Niklaus Grünwald wrote the manuscript.

Chapter 5

Shankar K. Shakya, Brian Knaus and Niklaus Grünwald conceived, designed, collected, analyzed data and wrote the manuscript.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

Page

LIST OF FIGURES

LIST OF FIGURES (Continued)

Figure Page

LIST OF FIGURES (Continued)

LIST OF TABLES

This dissertation is dedicated to all the people who believed in me.

Chapter 1

General introduction

1.1 Oomycetes and *Phytophthora* pathogens

Oomycetes are a group of organisms that are closely related to brown algae and diatoms even though their morphology resembles those of fungi (Kamoun 2003; Adhikari et al. 2013). Members of this group have diverse life styles. They range from free living water molds to pathogens of plants, mammals, fish and insects (Kamoun 2006; Martin and Loper 1999; Presser and Goss 2015). The plant pathogenic genus *Phytophthora* in this group causes severe yield loss in field crops, forest trees and threatens plant biodiversity (Haas et al. 2009; Tyler 2007; Grünwald et al. 2012; Hardham and Blackman 2018). *Phytophthora infestans* is a notable pathogen which is responsible for Irish potato famine in the 1840s (Fry et al. 2015). *P. infestans* causes late blight disease in Solanaceous crops like potatoes and tomatoes and costs involved in managing the disease are very high globally (Fry et al. 2015; Haverkort et al. 2008). Other important plant pathogens in the group include *Phytophthora sojae, Phytophthora ramorum* and *Phytophthora cinnamomi*. *P. sojae* is a soybean (*Glycine max*) pathogen and causes stem and root rot (Tyler 2007). It also causes pre and post-emergence damping off under high water saturation. *P. ramorum* causes the disease sudden oak death and is significant on oaks in California, Oregon and Europe (Grünwald et al. 2012). *P. ramorum* also infects Rhododendron in nurseries. *P. cinnamomi* is a broad host range pathogen and is primarily associated with hard woods and nursery crops (Hardham and Blackman 2018). *P. cinnamomi* causes dieback and is well known as the jarrah dieback pathogen (*Eucalyptus marginata*) in Australia (Dell and Malajczuk 1989; Hardham and Blackman 2018). In general, *Phytophthora* reproduces both

sexually and clonally (Erwin and Ribeiro 1996; Fry et al. 2015). Sexual reproduction takes place in presence of two opposite mating types named A1 and A2. This feature is termed heterothallism. Oospores are the result of sexual crossing however some species are self-fertile. The oospore can act as primary inoculum to initiate infections or act as survival structures that remain viable in soil for a number of years (Fry et al. 2015). Sexual reproduction is a primary source of genetic variation in *Phytophthora* species and usually observed at its center of origin or diversity. For example high genetic variation for *P. infestans* is observed in central Mexico which is regarded as its center of origin but can also be observed in parts of northern Europe (Yuen and Andersson 2013; Goss et al. 2014). However, for other species like *P. ramorum* and *P. cinnamomi* the center of origin or diversity is not known. Once an epidemic starts, sporangia (asexual spores) develops on the leaf. These spores serve as secondary inoculum in the disease cycle. Under cold and wet conditions, zoospores which are the motile spores develop inside the sporangia and swarm out into leaf-surface water to start new infections whereas under warm temperature, sporangia can germinate directly to initiate infection. The rate and quantity of sporangia production play a critical role in *Phytophthora* disease epidemics (Shakya et al. 2015).

1.2 *Phytophthora* phylogeny and variation between and within species

Initial inferences on the *Phytophthora* evolutionary history were based on their morphology (Erwin and Ribeiro 1996). This included shape and size of sporangia, mating types and the ability to form oospores (Erwin and Ribeiro 1996). However, with the development of Polymerase Chain Reaction (PCR) and sequencing tools our understanding of placement of certain *Phytophthora* species into a phylogenetic framework significantly improved our understanding of this group. The internal transcribed spacer (ITS) region was first used to create a genus wide phylogeny of *Phytophthora* species (Cooke et al. 2000). Kroon et al. (2004) used

four nuclear and mitochondrial loci and (Blair et al. 2008)) used seven nuclear genetic markers to construct *Phytophthora* phylogenies. Multi locus phylogenies and phylogenies based of whole genome sequences have become more common with an aim to provide the fine scale resolution of *Phytophthora* evolutionary history (McCarthy and Fitzpatrick 2017; Blair et al. 2008). A recent phylogeny from Yang et al. (2017) used seven nuclear genetic markers from 180 *Phytophthora* species and showed a correlation with the evolution of various sporangial forms.

Phylogenetic placement of *Phytophthora* taxa has heavily relied on nuclear and mitochondrial genes; however, variation within *Phytophthora* species exists and it has added a new layer of complexity in understanding the diversity within species and even identifying new species. In case of *P. infestans* , isozyme polymorphisms and later restriction fragment length polymorphism (RFLP) using the moderately repetitive probe RG57 were used to classify *P. infestans* into clonal lineages (Goodwin 1995; Goodwin et al. 1992). These techniques were considered a gold standard to classify *P. infestans* into clonal lineages. Simple Sequence Repeats (SSR) or microsatellites markers are now being widely used and have replaced RFLP markers (Li et al. 2013). SSR markers are used not only to genotype individuals but also to ask questions about their genetic relatedness and mode of reproduction (Wang et al. 2017; Shakya et al. 2018). Similarly, in *P. cinnamomi* SSR markers have been used to identify clusters of clonal lineages in Australia and South Africa. However, the global population structure of the pathogen has not been studied (Linde et al. 1999; Engelbrecht et al. 2017). Kamvar et al. (2015) used microsatellites markers to infer the patterns of multiple introduction of *P. ramorum* into Curry County in Oregon.

Whole genome sequence data and genome wide SNPs are now being used to study the variation between and within species and ask biological questions relevant to genome evolution.

This is critical in understanding how these changes have contributed towards the adaptation to their hosts (Raffaele et al. 2010a; Knaus et al. 2019; Martin et al. 2014; Yang et al. 2017). Fine scale genomic analyses using whole genome sequences were used to identify a shared common ancestor between US-11 and US-18 clonal lineages of *P. infestans* (Knaus et al. 2016). Similarly, identification of rapidly evolving pathogenicity genes, genes necessary for host colonization and use of avirulence genes to breed for disease resistance have driven the majority of the modern genomic research. However, the conservation and diversity of pathogenicity genes in *Phytophthora* at the global population scale is largely unknown (Cooke et al. 2012).

1.3 Effector genes in *Phytophthora* species

Plants have evolved two layers of immune systems to defend against invading pathogens (Jones and Dangl 2006). The first layer of defense consists of receptor-like kinases (RLKs) which recognize conserved pathogen signals called pathogen-associated-molecular-patterns (PAMP). This interaction initiates an immune response called PAMP triggered immunity (PTI) (Jones and Dangl 2006). A second layer of defense consists of nucleotide binding leucine rich repeat (NB-LRR) proteins. These proteins are primarily resistance genes (R genes) which recognize effector molecules from the pathogen that upon recognition trigger immunity called effector triggered immunity (ETI) (Jones and Dangl 2006).

Secretion of effectors by pathogens suppress the basal immunity and promote infection and colonization (Jones and Dangl 2006). One of the characteristic features of the genus *Phytophthora* is that it encodes numerous effector proteins (Haas et al. 2009; Stam et al. 2013). Two major classes of effector proteins have been characterized in *Phytophthora* species. First, RxLR (Arginine-any amino acid- Leucine-Arginine) effectors which contain a conserved RxLR

(Arginine-any amino acid-Leucine-Arginine) domain at the N terminal region of the gene sequence. Some RxLR effectors can be recognized by host resistance genes to initiate a programmed cell death response (Bos et al. 2006). RxLR effector content in a few economically important *Phytophthora* species are known. However, RxLR content in sister species are not known. RxLR effectors are thought to evolve via positive selection and be excessively diverse compared to housekeeping genes (Bos et al. 2010, 2006). A second class of effector proteins in the *Phytophthora* genus are "Crinkler effectors" named after the **CR**inkling and **N**ecrosis (CRN) disease phenotype (Stam et al. 2013; Schornack et al. 2010; Haas et al. 2009). Similar to RxLR effectors, CRN effectors have a signal peptide, a highly conserved functional LFLAK (Leucine-Phenylalanine-Leucine-Alanine-Lysine) motif and a diverse C-terminal functional domain (Stam et al. 2013; Schornack et al. 2010; Haas et al. 2009). The LFLAK motif has been shown to be important for the translocation of crinklers genes.

Phytophthora species have developed several mechanisms to evade being recognized by plant resistance genes. Some of these include loss of avirulence genes, single nucleotide polymorphisms, and gene expression polymorphisms. *Avr1* is a virulence factor which promotes infection and colonization (Du et al. 2015). When co infiltrated with *R1* resistance gene, *Avr1* elicits a hypersensitive response in *Nicotiana benthamiana*. *P. infestans* isolates virulent on *R1* plants have lost the *Avr1* gene (Du et al. 2015). Similarly, another cytoplasmic effector *Avr3a* confers avirulence in plants carrying the *R3a* gene (Bos et al. 2006). Allelic variation in the *Avr3a* gene determines avirulence. The *Avr3aKI* allele (K80, I103) confers avirulence whereas the *Avr3aEM* (E80, I103) allele is virulent on *R3a* plants. Therefore, understanding the presence/absence and sequence polymorphism in effector genes becomes crucial for breeding

plants for disease resistance against *P. infestans*. Comparison of these effector genes in closely related *Phytophthora* species also sheds light on the evolution of these genes.

1.4 Aim of the dissertation

This dissertation aims at expanding our knowledge and understanding of population diversity, phylogeography of *Phytophthora* species within and outside the center of origin and evolution of RxLR genes in three closely related *Phytophthora* species. To address these questions, various molecular markers including microsatellites, single nucleotide polymorphism (SNP) and whole genome sequence data are being applied.

1.4.1 Summary of chapter 2

Chapter 2 focuses on understanding the population dynamics of *Phytophthora infestans* outside of its center of origin. The Toluca valley of Mexico is considered as the center of origin of *P. infestans* (Goss et al. 2014). This is based on the facts that both mating types (A1 and A2) occur in equal frequencies central Mexico, populations reproduce sexually, populations are genetically diverse, and sister species of *P. infestans* coexist in the Toluca valley. The majority of the previous work has focused on the Toluca Valley. Only one study looked into the population structure of *P. infestans* outside of the Toluca valley (Wang et al. 2017). Here, we aim at testing the hypothesis of a gradient in genetic diversity in *P. infestans* outside of the Toluca valley. We used publicly available data combined with our recent sampling of *P. infestans* in Mexico which was further genotyped using microsatellite markers.

1.4.2 Summary of chapter 3

In chapter 3 we aim to identify the center of origin of the broad host range pathogen *Phytophthora cinnamomi.* Various hypotheses have been proposed regarding the center of origin of *P. cinnamomi*. One hypothesis is that *P. cinnamomi* originated in Asia; however, this has not been tested extensively. Here, we sampled a global population of *P. cinnamomi* and performed genotyping by sequencing (GBS) to test the hypothesis of Asia as a candidate center of origin. This work supports Asia as a candidate center of origin for this pathogen and shows evidence for existence of two dominant pan global, clonal lineages of A2 mating type.

1.4.3 Summary of chapter 4

Phytophthora species secrete RxLR effector to modulate host defenses. These effectors are thought to be evolving continuously to evade recognition by plant resistance genes. Chapter 4 is a comparative genomic study of the evolution of RxLR genes in three closely related *Phytophthora* species. *P. infestans*, *P. mirabilis* and *P. ipomoeae* are species in clade 1c of *Phytophthora*. *P. mirabilis* and *P. ipomoeae* do not cause a significant economic loss and therefore are probably understudied. However, they provide an excellent resource for comparative evolutionary studies. RxLR gene content and variation in *P. ipomoeae* and *P. mirabilis* are not known. Therefore, we identified the core RxLR effectors among the three clade 1c *Phytophthora* species and characterized the patterns of diversity observed in these genes.

1.4.4 Summary of chapter 5

The majority of resequencing projects involve mapping short sequencing reads against a linear reference genome. However, during this process some percentage of reads do not map and

are usually ignored from any kind of analysis (Whitacre et al. 2015; Weide et al. 2016; Gouin et al. 2015; Faber-Hammond and Brown 2016b; Cooke et al. 2012). Chapter 5 focuses on studying the metagenomic composition of *Phytophthora* reads that failed to map to the reference genome. This is achieved by assembling the unmapped reads and gene calling to identify genes that are absent from the reference *P. infestans* T30-4 isolate. This work provides impetus for characterizing unmapped reads in comparative genomic analyses to avoid bias in genes missing by chance due to assembly artifacts.

Chapter 2

Variation in genetic diversity of *Phytophthora infestans* populations in Mexico from the center of origin outwards

Shankar K. Shakya, Meredith M. Larsen, Mercedes María Cuenca Condoy, Héctor Lozoya-Saldaña and Niklaus J. Grünwald

Published in Plant Disease 2018. 102: 1534-1540

DOI: 10.1094/PDIS-11-17-1801-RE

2.1 Abstract

The Toluca valley located in central Mexico is thought to be the center of origin of the potato late blight pathogen *Phytophthora infestans.* We characterized over 500 individuals of *P. infestans* sampled from populations with a geographical distance of more than 400 km in six regions adjacent to the Toluca Valley in three states including Michoacán, Mexico, and Tlaxcala. Our sampling occurred on a predominant East to West gradient and showed significant genetic differentiation. The most Western sampling location found in Michoacán was most differentiated from the other populations. Populations from San Gerónimo, Juchitepec, and Tlaxcala clustered together and appeared to be in linkage equilibrium. This work provides a finer understanding of gradients of genetic diversity in populations of *P. infestans* at the center of origin.

2.2 Introduction

The Toluca valley located in central Mexico is considered to be the center of origin of the potato late blight pathogen *Phytophthora infestans* (Goodwin et al. 1992; Grünwald et al. 2001). This hypothesis is based on several lines of evidence. Populations of *P. infestans* in the Toluca are genetically diverse (Goss et al. 2014; Grünwald et al. 2001; Flier et al. 2003), the A1 and A2 mating types exist in a 1:1 ratio (Grünwald et al. 2001), oospores can readily be observed in various host plant tissues and are ubiquitous (Fernández-Pavía et al. 2004; Flier et al. 2001; Gallegly and Galindo 1958) and populations reproduce sexually based on the observed linkage equilibrium among genetic markers (Goss et al. 2014; Grünwald et al. 2001). This evidence is further supported by the fact that the closest relatives of *P. infestans*, namely *P. ipomoeae* and *P. mirabilis* are only observed in Mexico (Flier et al. 2002; Galindo and Hohl 1985). The only other known close relative, *P. andina*, found in South America is a hybrid species with two ancestral

parents (Goss et al. 2011; Oliva et al. 2010). Finally, Mexico has been described as a center of diversity for tuber bearing *Solanum* species and *P. infestans* has been shown to infect these native species and is thought to have coevolved with the native resistance genes (R) genes described in various hosts such as *S. demissum* and *S. edinense* (Grünwald et al. 2001; Flier et al. 2003; Hijmans and Spooner 2001; Rivera-Peña and Molina-Galan 1989; Rivera-Peña 1990). Knowledge and understanding of a pathogen's center of origin provides insights into its evolutionary potential, genetic diversity, mode of reproduction and gene flow. This information in turn is critical to implementing adaptive disease management strategies and discovery of novel R genes.

Until the 1970s the clonal lineages US-1 and HERB-1 (both of mating type A1) of *P. infestans* were dominant globally (Fry 2008; Yoshida et al. 2013), whereas *P. infestans* populations in central Mexico were shown to have both mating types (Goodwin et al. 1992; Grünwald and Flier 2005). Even though both the A1 and A2 mating types are present in the United States, evidence of sexual recombination is still lacking (Fry et al. 2012; Danies et al. 2013) and populations are described as clonal lineages. Recombinant isolates are occasionally detected in the U.S. but exist for only a short period of time before going extinct (Danies et al. 2014). In the US, we currently distinguish up to 24 clonal lineages (Fry et al. 2012). A recent phylogenomic analysis showed that none of the US clonal lineages are descendants of preexisting lineages (Knaus et al. 2016; Wang et al. 2017). Hence, populations in Mexico are thought to be the main source of novel migrant lineages that establish themselves in the US.

Several studies have investigated diversity of populations in different regions of Mexico. Goodwin et al. (1992) compared the diversity of *P. infestans* between northern and central Mexico based on allozyme loci and DNA fingerprinting and suggested high diversity in

Chapingo. A recent study by Wang et al. (2017) provides novel insights into population structure of *P. infestans* in Mexico. Their results show high genetic diversity and sexual reproduction in the Michoacán, Toluca and Tlaxcala regions based on twelve SSR loci. Here, we expand the range of populations sampled around the Toluca valley to determine the range of sexual populations and whether there is isolation by distance as we sample away from the Toluca valley.

In this study, we analyzed a sample of 517 *P. infestans* isolates sampled from six locations sampled in three states on a predominantly East to West gradient (Figure 2.1). We newly sampled 355 samples and combined these with populations described previously (Wang et al. 2017). The goal of this study was to provide a better understanding of the population structure and diversity of *P. infestans* in central Mexico and adjacent states. We specifically tested several hypotheses including: 1. Is there genetic differentiation among populations of *P. infestans* in central Mexico? 2. Are populations isolated by geographic distance? 3. Are populations that are geographically distant from Central Mexico, such as Tlaxcala and Michoacán, increasingly clonal? Our work provides novel evidence for spatial correlation across this gradient and further refines our understanding of the biology of this pathogen at the center of origin.

2.3 Materials and Methods

2.3.1 Sampling and isolation of *P. infestans*.

In this study, we characterized populations of *P. infestans* sampled at six locations between 1997 and 2016 (Table 2.1; Fig 2.1). We newly sampled 355 isolates in 2015 and 2016. Briefly, isolates were obtained utilizing the protocol described previously (Grünwald et al. 2001) with the slight modification of adding oxytetracycline (20 mg/L) to the selective media (see

Wang et al. 2017). We also included populations sampled previously by Wang et al. (2017) and deposited in Dryad (http://datadryad.org/resource/doi:10.5061/dryad.262qq) (Table 2.1). The Wang et al. (2017) strains were genotyped using the same protocol as described below. Isolates from Toluca in 2015 and Chapingo in 2015 and 2016 were sampled from experimental plots subjected to varying fungicide treatments and are not considered natural populations and should thus be interpreted differently. Most of the isolates from Tlaxcala state were sampled near Villareal and are referred to as the Tlaxcala population hereafter. Isolates originating from the same geographic area, regardless of sampling years, were combined as they were not significantly different based on F_{ST} analysis.

2.3.2 DNA extraction and SSR genotyping.

Isolates were grown in pea broth agar for a week at 18°C and actively growing agar plugs were transferred to 10 ml pea broth for 7-14 days for DNA extraction. DNA was extracted following the manufacturer's protocol with the FastDNA SPIN KIT (MP Biomedicals, United States) and stored at -20°C until used. DNA concentration was measured using a NanoDrop 2000C (Thermo Scientific, United States) and adjusted to 10 ng/µl. Multiplex PCR was performed on twelve polymorphic microsatellite loci as described by

Li et al. (2013). A panel of reference strains (5303, 5304, 5306, 5307 and 5308) with known SSR alleles was included in all PCR and fragment analysis runs as described at http://phytophthora-id.org/ (Grünwald et al. 2011). Allele sizing was done using capillary electrophoresis (ABI 3730, Applied Biosystems, Foster City, CA) at the Center for Genome Research and Biocomputing at Oregon State University and fragments were sized using the GENEMAPPER software (Applied Biosystems) (Wang et al. 2017).

Data preparation

Isolates originating from the same geographical region were pooled into a population regardless of sampling year and study. We also included samples from Toluca and Tlaxcala from Wang et al. (2017) to our recent collection sampled in 2015/16. Populations were checked for varying ploidy levels and missing values. Only strictly diploid genotypes were considered for downstream analysis. The locus D13 had missing values ranging from 5-70% and therefore was removed from the analyses resulting in a final data set with 11 SSR loci.

2.3.4 Genetic diversity and mode of reproduction.

Analysis was conducted on 431 diploid individuals based on eleven SSR loci (e.g., some of the 517 strains sampled were not included due to varying ploidy levels; Table 2.1). Multilocus genotypes (MLG), e.g. the unique combination of alleles at all loci per individual, and expected multilocus genotype (eMLG) based on rarefaction were calculated using the R package *poppr* V.2.3.0 (R Core Team 2016; Kamvar et al. 2014; Grünwald et al. 2017). The Stoddart and Taylor's diversity index (G) was calculated (1988). Nei's unbiased gene diversity (1978) also referred to as expected heterozygosity (Hexp) was calculated for all the diploid individuals per population.

P. infestans is known to reproduce clonally and sexually. The literature suggests that sexual recombination is predominant in the Toluca valley (Goodwin et al. 1992; Grünwald et al. 2003; Grünwald and Flier 2005; Goss et al. 2014). We assessed the mode of reproduction in *P. infestans* across all populations. The standardized index of association (rbarD), a multilocus estimate of the index of association/linkage disequilibrium, was calculated to investigate the mode of reproduction (Agapow and Burt 2001; Kamvar et al. 2014). rbarD was measured on

clone corrected data to remove the bias of resampled MLGs. The expectation of rbarD for a randomly mating population is zero. Any significant deviation from the expectation of zero would suggest clonal reproduction. The significance was tested based on 999 permutations and conducted in the R package *poppr* (Kamvar et al. 2014).

One of the assumptions of Hardy-Weinberg equilibrium (HWE) is random mating. Deviation from HWE would suggest nonrandom mating or population subdivision. We tested the hypotheses of nonrandom mating and population subdivision by calculating HWE per locus per population. HWE was performed using the R package *pegas* V.0.9 and significance was tested using 1,000 permutation events (Paradis 2010).

2.3.5 Population differentiation and structure.

Wright's F_{ST} is a measure of population subdivision for diploid individuals. It is based on differences in allele frequencies between sampled populations (Wright 1949). We calculated pairwise F_{ST} values with the clone corrected *P. infestans* data to test the hypothesis of population subdivision using the R package *strataG* V.1.0.5 (Archer et al. 2017). Statistical significance was calculated with 10,000 permutations events. Analysis of Molecular Variance (AMOVA) was performed to test for genetic structure implemented in the R package *ade4* V.1.7-5 (Dray and Dufour 2007; Excoffier et al. 1992). AMOVA was calculated on clone corrected data using Bruvo's distance (2004) to estimate the variance explained by populations or individuals within populations. Statistical significance was tested using 999 permutations events. The neighbor-joining algorithm based on Bruvo's distance was run to visualize the clustering of *P. infestans* isolates using *poppr* with 1,000 bootstrap replicates and the resulting tree was edited using the R package *ggtree* V.1.4.20 (Yu et al. 2017).

The Bayesian model based clustering algorithm STRUCTURE V.2.3 (Pritchard et al. 2000) was used to infer population structure. The software clusters individuals into K populations using a Markov Chain Monte Carlo (MCMC) approach. Fifteen independent runs were performed with 100,000 iterations/MCMC run after a burn-in period of 20,000 for each value of K ranging from 1 to 6. An admixture model was selected as it assumes mixed ancestry of an individual. The optimum value of K was determined using the Evanno et al. (2005) method. CLUMPP V.1.2 was run to aggregate multiple STRUCTURE runs (Jakobsson and Rosenberg 2007) based on a greedy algorithm and a G.prime pairwise similarity statistic parameter with 10,000 repeats. The R package *strataG* (Archer et al. 2017), providing a wrapper script to run and visualize STRUCTURE data and implementing the Evanno method and CLUMPP analyses, were used.

Discriminant Analysis of Principal Component (DAPC) is a model-free approach to infer clusters of populations, relative population membership, and assign individuals to populations (Jombart et al. 2010). Proportional assignment of an individual to populations is provided by the retained discriminant functions. DAPC was conducted using the R package *adegenet* V. 2.0.1 (Jombart 2008).

Long-distance dispersal of *P. infestans* sporangia is limited due to desiccation and susceptibility to radiation (Mizubuti et al. 2000). Thus, isolation by distance might be observed as individuals cannot migrate freely over long distances (e.g., >1-10km). We tested for correlation in *P. infestans* populations in Mexico using pairwise F_{ST} and pairwise geographic distance. We also performed Mantel's test implemented in the R package *ade4* with 1,000 permutations (Mantel 1967; Dray and Dufour 2007). All code and data used in this study are

deposited on github (https://github.com/grunwaldlab/MexicoPopulations) and OSF (Shakya et al. 2017; https://osf.io/hbxvj/).

2.4 Results

2.4.1 Variation in genetic diversity and reproduction in *P. infestans* populations

A total of 431 diploid *P. infestans* isolates (after removing polyploid strains from a total of 517) sampled in six regions of Mexico were analyzed using eleven highly polymorphic microsatellite loci (Table 2.1). The majority of isolates from Chapingo that were sampled in experimental plots, rather than grower fields, were identified as clones. Expected heterozygosity ranged from 0.44 (Chapingo) to 0.57 (Tlaxcala) (Table 2.2). Stoddart and Taylor's diversity, which is a measure of evenness and richness of MLGs, was lowest for Chapingo (5.45) and highest for Tlaxcala (65.20).

To infer the mode of reproduction, the standardized index of association (rbarD) was calculated for clone corrected data. Populations from Chapingo and Tlaxcala showed significant deviations in rbarD values from the null expectation supporting clonal reproduction (Table 2.3). rbarD values for all other regions were not significantly different from zero indicating linkage equilibrium or presence of sexual reproduction. In each population, between 5 to 11 loci were considered to be in Hardy-Weinberg equilibrium (Table 2.3).

2.4.2 Genetic differentiation among populations of *P. infestans* in central Mexico

Pairwise F_{ST} values were calculated to estimate differentiation among populations. These values ranged from 0.01 to 0.11 indicating that populations show low to modest levels of differentiation (Table 2.4). The Michoacán population was most differentiated from the other

populations in Mexico. The highest differentiation was between populations from Michoacán and San Gerónimo (central Mexico). Populations that were geographically close had small but significant F_{ST} values. San Gerónimo and Tlaxcala had the smallest pairwise F_{ST} value of 0.01. The results of population differentiation using F_{ST} are well supported by our AMOVA analysis. AMOVA on clone corrected *P. infestans* populations based on Bruvo's distance revealed 11% of variation between populations (Table 5, P = 0.001). This is similar to our highest pairwise F_{ST} value of ~12%. Approximately 90% of the genetic variation was explained due to variation within populations rather than between populations. Our data support the hypothesis of a gradient of increasing differentiation as we move west towards Michoacán from Toluca. However, at the Eastern front of our sample areas we do not yet see this effect and further sampling East of Tlaxcala is likely needed.

2.4.3 Population structure and isolation by distance

We performed Bayesian clustering using STRUCTURE to determine population relationships. The STRUCTURE analysis identified two likely solutions for K based on Evanno's method. The value of ΔK had two peaks at K = 2 and K = 4 (Fig 2.2). For K = 2, Michoacán was identified as a first group and the rest of the populations as a second group. STRUCTURE results for K = 4 identified Michoacán, Chapingo and Juchitepec as distinct clusters and Toluca, Tlaxcala and San Gerónimo as one cluster (Fig 2.3). We also conducted neighbor-joining analysis based on Bruvo's distance. In this tree, isolates from Michoacán formed a cluster whereas isolates from other regions were more or less randomly distributed confirming the STRUCTURE analysis.

To resolve the discrepancy in the number of clusters based on STRUCTURE, we performed a Discriminant Analysis of Principal Components (DAPC) which is a model free approach to identify clusters. Clustering of isolates using DAPC was consistent with Evanno's result of K = 4 (Fig 2.4). Michoacán, Juchitepec and Chapingo each formed an independent cluster in the DAPC analysis whereas the Toluca, Tlaxcala and San Gerónimo populations formed a fourth cluster.

Geographic and genetic distance were significantly and positively correlated (Fig 2.5A) ($P < 0.05$). However, this correlation between geographic and genetic distances may be confounded with population structure. Mantel's test indicated no significant isolation by distance (Fig 2.5B). Two clusters were observed for Mantel's test (Fig 2.5B) that resulted in K=2 clusters identified above: Michoacán vs. all other populations. When removing the Michoacán populations, no spatial correlation was observed between populations of central Mexico and Tlaxcala. Our analysis does not demonstrate significant isolation by distance, but suggests that the Michoacán population is distinct compared to the other populations in line with a gradient of gene flow at the Western border of our sampling range.

2.5 Discussion

In this study, we investigated the population structure of *P. infestans* beyond the well-characterized Toluca valley region, considered to be the center of origin. Our work built on the previous work by Wang et al. (2017) by adding samples from regions in central Mexico other than the Toluca valley. Our work suggests that populations in central Mexico show high genotypic diversity beyond the center of origin previously described to be located in the Toluca valley. Similar results were recently reported by Wang et al. (2017). Our analyses suggest

presence of both sexual and asexual reproduction in central Mexico. We found that, in the range sampled, populations on the Western edge in Michoacán are the most differentiated, yet still genetically diverse and sexual in nature. Removing the Michoacán population resulted in three main clusters: Chapingo, Juchitepec, and a third cluster containing San Gerónimo, Tlaxcala, and Toluca.

The populations we sampled in Chapingo (2015 & 2016) showed low genotypic diversity and clonality. STRUCTURE and DAPC analyses suggest that the Chapingo population is a group by itself. We do not currently know why Chapingo populations show a different pattern; these plots were not inoculated and did receive various fungicide treatments.

Tlaxcala is a state east of central Mexico. The *P. infestans* population in this state has been reported to reproduce sexually based on twelve microsatellite loci (Wang et al. 2017). However, our analyses of the Tlaxcala populations based on eleven microsatellite loci suggest clonal reproduction. We did not include locus D13 in our analyses because of the high percentage of missing data. We checked for a putative effect of missing data in Wang et al. (2017) and found that inclusion of locus D13 (high percentage of missing data) actually suggests sexual reproduction.

We detected population differentiation based on pairwise F_{ST} values. Small F_{ST} values can be observed due to recent common ancestry or because of ongoing migration. Different genetic markers will result in different estimates of F_{ST}. A high F_{ST} value between Michoacán and the other populations suggests limited or no gene flow between these regions. One way migration has been reported from Michoacán to Toluca valley in Wang et al. (2017).

We expected to see isolation by distance for the populations sampled on a predominant East to West gradient. However, with the exception of the Michoacán population, no isolation by distance was apparent. F_{ST} values did not show a correlation with distance except when Michoacán was included. This indicates that more sampling is necessary between Michoacán and other regions to test for spatial correlation.

Our understanding of variation in *P. infestans* ploidy is limited. However, a recent study by Li et al. (2016) suggested that sexually reproducing *P. infestans* might be diploid while clonal lineages might be triploid. In this study, we only scored *P. infestans* isolates as being diploid for SSR markers. However, in genotyping of these isolates we did notice some apparently trisomic or even tetrasomic loci for strains that were excluded from our analysis. Eighty per cent of isolates genotyped in 2015/16 from Mexico were scored as strictly diploid. A similar pattern of diploidy and trisomy was also reported by Wang et al. (2017). Because the pattern of inheritance of SSR markers is complex, this poses a distinct challenge in performing genetic analyses for populations that are of varying ploidy levels. We chose to eliminate non-diploid strains from our analysis, but other paths can be considered (Grünwald et al. 2017).

We studied the population structure of *P. infestans* across a gradient from East to West within central Mexico. The population structure observed for *P. infestans* in central Mexico suggests gene flow and sexual reproduction beyond the Toluca Valley. We observed high differentiation at the Western edge of the populations sampled while on the Eastern edge further sampling is indicated. This work provides a finer understanding of gradients of genetic diversity in populations of *P. infestans* at the center of origin.

Table 2.1 Samples of *P. infestans* collected and genotyped from Mexico using eleven polymorphic microsatellite loci. Our analysis combined newly sampled population with those analyzed recently by Wang et al. (2017).

Region	**Year**	**Total isolates**	**Diploid**	**Trisomic**	**Tetrasomic**	**Reference**
Michoacán	2007-2010	76	71	5	0	Wang et al. (2017)
Chapingo	2015	44	43	1	0	This study
Chapingo	2016	69	67	2	0	This study
Toluca	1997	31	24	7	0	Wang et al. (2017)
Toluca	2015	23	22	1	0	This study
San Gerónimo	2015	55	39	10	6	This study
San Gerónimo	2016	8	6	2	0	This study
Juchitepec	2015	33	20	10	3	This study
Juchitepec	2016	57	48	8	1	This study
Tlaxcala	2007	55	50	5	0	Wang et al. (2017)
Tlaxcala	2015	66	41	24	1	This study
Total	1997-2016	517	431	75	11	-

Table 2.2 Population statistics for diploid genotypes for *P. infestans* populations by region in Mexico. Total number of samples (N), observed multilocus genotype (MLG), expected multilocus genotype (eMLG), Stoddart and Taylor's diversity (G), and expected heterozygosity (Hexp).

Region	**N**	**MLG**	**eMLG**	**G**	**Hexp**
Michoacán	71	43	29.86	17.3	0.54
Chapingo	110	17	11.22	5.5	0.44
Toluca	46	38	37.26	29.4	0.45
San Gerónimo	45	22	22.00	9.3	0.55
Juchitepec	68	28	21.96	12.8	0.56
Tlaxcala	91	76	41.05	65.2	0.57

Table 2.3 Linkage disequilibrium among loci based on the standardized index of association (rbarD) and number of loci in Hardy-Weinberg equilibrium after clone correction for *P. infestans* in different regions of Mexico. Total number of samples (N), observed multilocus genotype (MLG), the standardized index of association (rbarD), the P value for rbarD, and the number of loci in Hardy Weinberg equilibrium.

Region	**N**	**MLG**	**rbarD**	**P-value**	**Loci under HWE**
Michoacán	43	43	0.010	0.133	5
Chapingo	17	17	0.139	0.001	6
Toluca	38	38	-0.006	0.725	7
San Gerónimo	22	22	0.015	0.125	9
Juchitepec	28	28	-0.006	0.724	11
Tlaxcala	76	76	0.016	0.006	7

Table 2.4 Pairwise FST values for clone corrected *P. infestans* populations from regions in Mexico. Values significant (alpha = 0.05) based on 10,000 permutations are marked with '*'.

	Chapingo	**Juchitepec**	**San Gerónimo**	**Toluca**	**Tlaxcala**	**Michoacán**
Chapingo	-					
Juchitepec	0.05*	-				
San Gerónimo	0.06*	0.05*	-			
Toluca	0.04*	0.06*	0.07*	-		
Tlaxcala	0.05*	0.04*	0.01*	0.06*	-	
Michoacán	0.06*	0.06*	0.11*	0.10*	0.10*	-

Table 2.5 Analysis of Molecular Variance (AMOVA) for clone corrected, diploid *P. infestans* populations based on Bruvo's genetic distance.

Source	**df**[a]	**SS**[b]	**MSS**[c]	**% variance**
Between regions	5	4.39	0.88	10.11
Within regions	218	38.50	0.18	89.89
Total	223	42.89	0.19	100.00

[a] Degrees of freedom

[b] Sum of squares

[c] Mean sum of squares

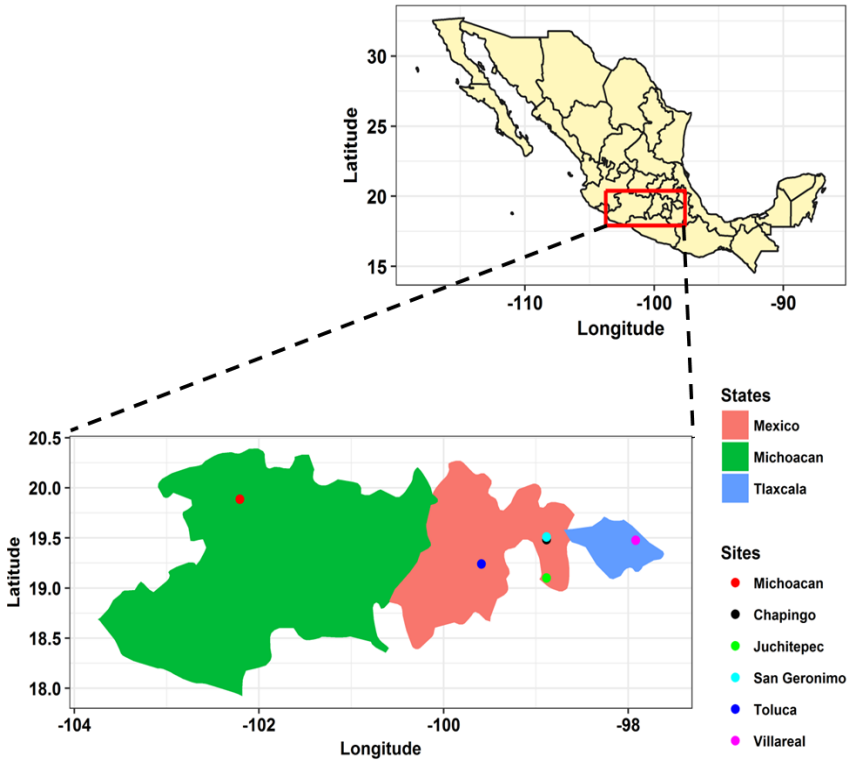

Figure 2.1 Map of Mexico (top) and enlarged map of the Mexican states where population of *P. infestans* were sampled. The colored dots represent the areas where samples were collected.

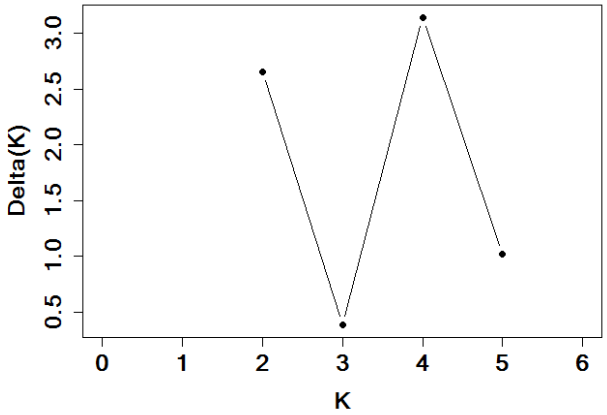

Figure 2.2 Mean likelihood value and delta K plots for each inferred cluster K of *P. infestans* using Evanno's method (Evanno et al. 2005).

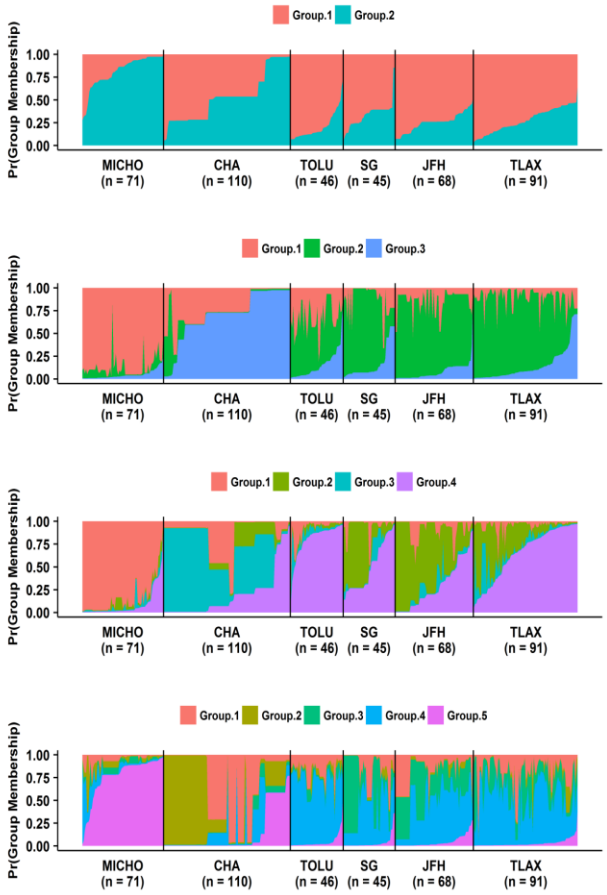

Figure 2.3 Bayesian clustering of diploid *P. infestans* genotypes based on the admixture ancestry model using STRUCTURE. The optimum value of K inferred with the Evanno's method was 2 (Evanno et al. 2005). Shown are STRUCUTRE plots for K=2 to 5 from top to bottom. MICHO = Michoacán; CHA = Chapingo; TOLU = Toluca; SG = San Gerónimo; JFH = Juchitepec and TLAX = Tlaxcala.

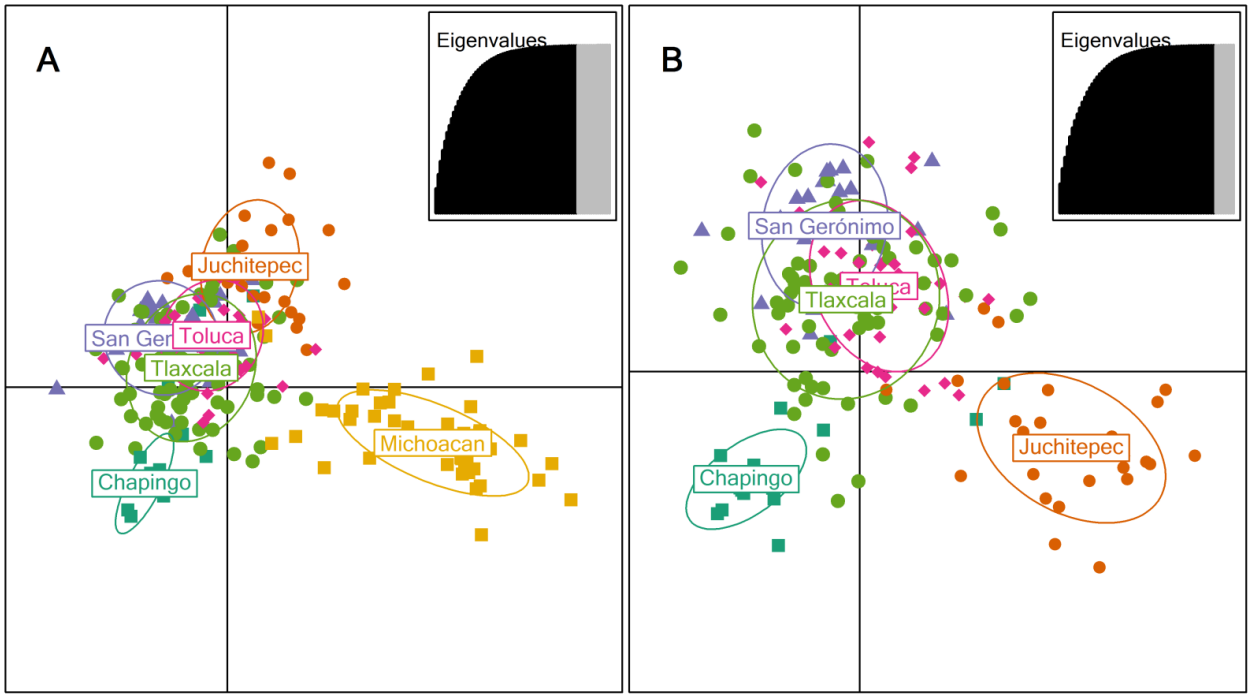

Figure 2.4 Discriminant Analysis of Principal Components (DAPC) of *P. infestans* populations from six regions in Mexico. A. All populations and B. excluding the Michoacán population.

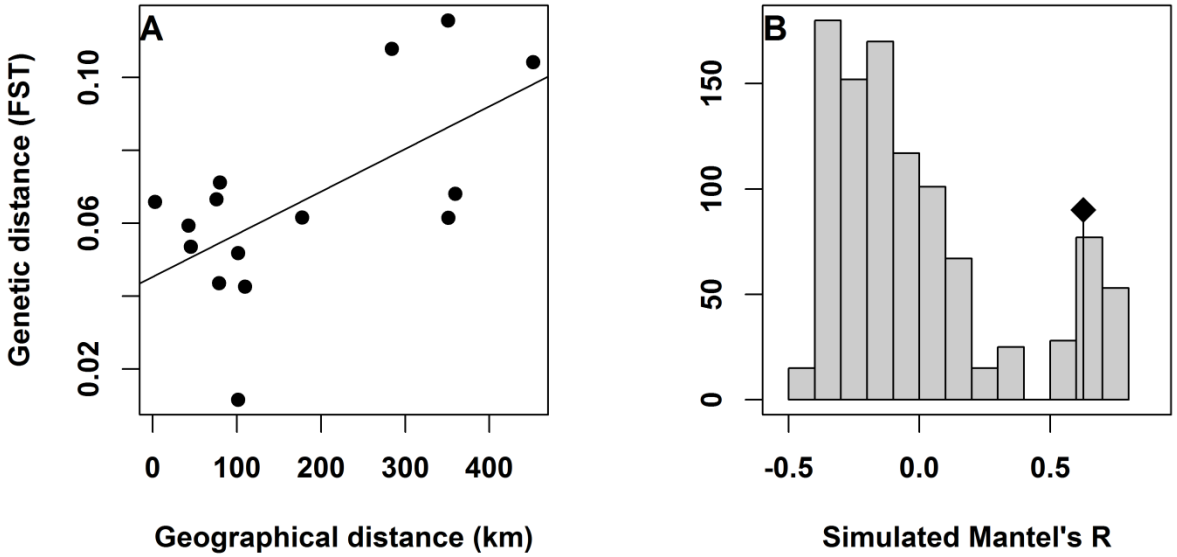

Figure 2.5 A. Plot of pairwise genetic distance (FST values) versus geographic distance between *P. infestans* populations. Based on the adjusted R^2, 34% of the variance in FST is explained by geographical distance (P = 0.0123). B. Test for isolation by distance (Mantel's test) with 1,000 permutations (P = 0.11).

Chapter 3

Narrowing in on the origin of the invasive jarrah dieback pathogen *Phytophthora cinnamomi*

Shankar K. Shakya, Niklaus J. Grünwald, Brian J. Knaus, Jerry E. Weiland, Valerie J. Fieland, Marilia Horta Jung, Cristiana Maia, André Drenth, David I. Guest, Edward C.Y. Liew, Coline Crane, Bruno Scanu, and Thomas Jung

Target journal: Molecular Ecology

3.1 Abstract

Various hypotheses have been proposed regarding the origin of the plant pathogen *Phytophthora cinnamomi*. *P. cinnamomi* is a devastating pathogen associated with agricultural and forest ecosystems worldwide. The pathogen is best known for killing jarrah (*Eucalyptus marginata*) trees in Western Australia including other native host plants. We conducted a phylogeographic analysis of populations of this pathogen sampled in Asia, Europe, South Africa, South America, and North America. Based on genotyping-by-sequencing, we observed highest genotypic diversity in Taiwan and Vietnam, followed by Australia and South Africa. Furthermore, mating type ratios were in equal proportions in Asia as expected for a sexual population. Ancestral state reconstruction supports Taiwan as the ancestral group indicating that this region might be the center of origin for this pathogen. The Australian and South African populations appear to be a secondary center of diversity following migration from Taiwan or Vietnam. Our work also identified two pan global clonal lineages of A2 mating type found on all continents and supported by the prior literature.

3.2 Introduction

Phytophthora cinnamomi is a widely distributed devastating and invasive pathogen affecting hundreds of plant species worldwide (Erwin and Ribeiro 2005; Eggers et al. 2012; Balci et al. 2007; Hardham 2005). This pathogen is considered to be a major invasive threat to native forest tree species in Australia and South Africa and agricultural and forest ecosystems worldwide (Pratt and Heather 1973; Pratt et al. 1973; Cahill et al. 2008; Shearer et al. 2007). One of the most devastating epidemics of *P. cinnamomi* is dieback of jarrah trees (*Eucalyptus marginata*) in Western Australia (Figure 3.1). The Australian government has labeled *P.*

cinnamomi a “key threatening process to Australia’s biodiversity” in schedule 4 of the Biodiversity Conservation Act, 2016 (https://www.environment.nsw.gov.au/resources/threatenedspecies/1ASched20180427.pdf). *P. cinnamomi* is a broad host range pathogen also causing oak tree mortality (Eggers et al. 2012; Balci et al. 2007), root rot in avocado, citrus, orchard and nursery crops (Beaulieu et al. 2017).

P. cinnamomi is an oomycete, a group of water molds that are most closely related to brown algae (Erwin and Ribeiro 2005). This particular *Phytophthora* species is heterothallic requiring the two mating types A1 and A2 for sexual reproduction via formation of oospores (Erwin and Ribeiro 2005). To date both mating types have been found around the world yet populations described to date are clonal (Linde et al. 1997, 1999; Eggers et al. 2012; Engelbrecht et al. 2017; Beaulieu et al. 2017).

The origin of *P. cinnamomi* is still a topic of debate however there is an increasing consensus that *P. cinnamomi* might have originated in Southeast Asia (Arentz and Simpson 1986; Ko et al. 1978; Zentmyer 1988; Shepherd 1975; Arentz 2017; Pratt and Heather 1973). The pathogen was first described by (Rands 1922) from stripe cankers of the cinnamon tree, *Cinnamomum burmanii*, in Sumatra (Rands 1922). *C. burmanii* is considered a plant native to Southeast Asia and Indonesia. This might have led earlier researchers to infer that *P. cinnamomi* is probably native to Southeast Asia. Crandall and Gravatt (1967) first advanced the hypothesis that *P. cinnamomi* is native to Southeast Asia. This hypothesis was based on the fact that native plants in these regions were resistant to the pathogen. Ko et al. (1978) provided further evidence to support Southeast Asia as a center of origin based on observed mating type (A1:A2) ratios and suggested Taiwan could lie within the center of origin of the pathogen. Alternate hypothesis for the origin of *P. cinnamomi* include those put forwarded by Shepherd (1975) suggesting New

Guinea and Indonesia's Sulawesi island (formerly Celebes) as a center based on mating type ratios and absence of disease in the region. Arentz and Simpson (1986) provided some support for Papua New Guinea as a center of origin based on observed genetic diversity using isozymes. (Pratt and Heather 1973) argued that *P. cinnamomi* is native to Australia as it was found to be associated with both disturbed and undisturbed geographic regions. However, *P. cinnamomi* isolates from Australia and South Africa were shown to be genetically identical suggesting a potential introduction from the same source with low genotypic diversity (Old et al. 1988; Linde et al. 1999, 1997).

Identification of a center of origin relies on a range of approaches (Grünwald and Flier 2005; Stukenbrock and McDonald 2008; Grünwald et al. 2016). Mating type ratios are expected to follow a 1:1 ratio and molecular markers are unlinked (Grünwald et al. 2001). Hence, populations are expected to reproduce sexually (Goss et al. 2014). Genotypic diversity is highest at the center of origin and declines as one moves out of the center of origin (Shakya et al. 2018). Pathogens are found on native plants but do not necessarily cause a lot of disease (Flier et al. 2003). This has been well demonstrated for the Irish famine pathogen, another *Phytophthora* pathogen, at its center of origin in Mexico based on support from root state probabilities for phylogeographic analyses (Goss et al. 2014). Similar criteria were used by O'Hanlon et al. (2018) to infer the center of origin of chytrid fungi. East Asia was identified as a geographic hotspot for the fungus suggesting its center of origin.

Genotyping by sequencing (GBS) is a technique for obtaining genome-wide single nucleotide polymorphisms (SNPs) using restriction, barcoding and high throughput sequencing (Elshire et al. 2011). This technique has been used successfully to ask hypothesis driven questions about key demographic processes such as gene flow, migration, bottlenecks, mating

type systems and other evolutionary processes (Grünwald et al. 2016). Tabima et al. (2018) has shown the genetic differentiation and migration between *Phytophthora rubi* populations in United States using GBS. The same technique has been applied to study other plant pathogens like nematodes to study their phylogeny (Rashidifard et al. 2018).

We sampled populations of *P. cinnamomi* from Asia (Taiwan and Vietnam), Europe (France, Italy Hungary, United Kingdom, Portugal and Spain), North America (USA and Dominican Republic), Africa (South Africa, Algeria and Tunisia) and Australia to infer the possible center of origin and global population structure of *P. cinnamomi*. We used genotyping by sequencing to determine genetic diversity. *P. cinnamomi* populations are shown to be genetically less diverse with the existence of few genotypes based of different marker systems (Old et al. 1984, 1988; Chang et al. 1996). We tested the hypothesis that Asia is the center of origin of *P. cinnamomi* and migration from Asia has led to clonal expansion of *P. cinnamomi* in rest of the world.

3.3 Materials and Methods

3.3.1 Isolates of *P. cinnamomi*

P. cinnamomi isolates were collected by extensive sampling of soil and roots from multiple hosts. The details of the isolates used in the study are summarized in Table 3.1. Isolates were identified to species level based on ITS sequencing. *P. cinnamomi* isolates from the US were primarily sampled from *Rhododendron* in ornamental nurseries in Oregon.

3.3.2 Genome sequencing, SNP calling and filtering

DNA extraction, library preparation and sequencing were performed by the Center for Genome Research Biocomputing (CGRB) at Oregon State University. DNA was extracted using Qiagen DNEasy plant kit following the manufacturer's instruction. Digestion of genomic DNA was done using two restriction enzymes *Pst*I and *Msp*I and sequenced using the Illumina Hiseq 3000 (San Diego, CA) with 100 bp single end (SE) read in three different runs. Raw fastq reads were demultiplexed using sabre (https://github.com/najoshi/sabre). The *P. cinnamomi* reference genome v.1.0 was downloaded from the Joint Genome Institute (https://genome.jgi.doe.gov/Phyci1/Phyci1.home.html). Reads were mapped against the reference using bowtie2 v2.3.2 (Langmead and Salzberg 2012). The resulting SAM (Sequence Alignment Map) file was converted to bam, sorted and indexed using samtools v1.3 (Li et al. 2009). Genomic vcf (gvcf) files were produced for each sample and variants were called using GATK HaplotypeCaller v3.7(McKenna et al. 2010). Variants were filtered based on the sites that were consistently scored in technical replicates (W1733-NSW, PH185-France and CIN-9-Dominican Republic) with no missing value and at least 10x coverage. The upper and lower five per cent depth quantiles were removed for each sample. Indels, non biallelic loci and non-genotyped loci were removed. *Phytophthora sojae* (P7076) was used as an outgroup in the analysis. *P. sojae* reads were mapped to the *P. cinnamomi* reference genome using the above described procedure to call variants. Only the variants that were scored in *P. sojae* and *P. cinnamomi* isolates were retained for the final analysis. A total of 299 high quality SNPs were retained for further data analysis.

3.3.3 Mating types and A1:A2 ratio

Identification of mating type was done by pairing an isolate of unknown mating type of *P. cinnamomi* with an isolate of known mating type. Oospores were observed to confirm the mating type. We tested the hypothesis of 1:1 mating type ratio for each regional population in R using two sided binom.test function in R.

3.3.4 Genetic diversity and population structure

Populations were defined by regions as follows: Taiwan, Vietnam, Australia, Africa, North America and South America. Note that sampling in certain continents was only conducted on a limited basis: South America (Chile), North America (USA and Dominican Republic). Vietnam and Taiwan are of particular interest and these populations were thus not merged into an Asia sample to allow study of differentiation and migration among these populations. Genotypic diversity was assessed by calculating the number of multilocus genotypes (MLGs). MLG is defined as a unique combination of single nucleotide polymorphisms (SNPs) in this case. We assessed the number of MLGs using the R package poppr v.2.5.0 (Kamvar et al. 2014). MLGs were merged in to a larger group called multilocus lineage using average neighbor algorithm (Kamvar, Brooks, and Grünwald 2015). A greater number of MLGs, private MLGs and high MLGs diversity is expected for a sexually reproducing population compare to clonally reproducing population. A rarefaction curve was drawn to assess the MLGs diversity using the R package vegan v.2.4-2 (Oksanen et al. 2016) to correct for uneven sample size.

Population structure was inferred using software ADMIXTURE v.1.3 (Alexander et al. 2009). ADMIXTURE models the probability of genotypes for different ancestor groups similar to software STRUCTURE (Pritchard et al. 2000) which is highly used for microsatellite data.

Vcf data was first converted to plink format data to allow to be used in ADMIXTURE software (Purcell et al. 2007). A total of 2 to 10 clusters (K) were evaluated. Cross validation errors were calculated to infer the right number of clusters. ADMIXTURE was also run for each population individually following the above method.

3.3.5 Phylogenetic analyses and ancestral population probability

A maximum likelihood (ML) phylogenetic tree was generated using RAxML v.8.2.11(Stamatakis 2006, 2014) to determine relationships among samples. Genotypes were extracted using vcfR v.1.6.0 (Knaus and Grünwald 2017) and coded as a 3 character multistate (0/0 = 0; 0/1 or 1/0 = 1 and 1/1=2). The ASC-MULTIGAMMA model with "lewis correction" was used to correct for ascertainment bias resulting due to exclusion of invariable positions (Lewis 2001). *P. sojae* was used as an outgroup to root the tree. A total of 1,000 bootstrap replicates were performed and a tree was plotted using the R package ape v.5.0 (Paradis et al. 2004; Paradis 2010). The ML tree was used to infer the ancestral state probabilities for the geographic origin of nodes using the ace function implemented in the package ape. Geographic locations were mapped to nodes using the R package phytools v.0.6-44 (Revell 2012). A similar approach was used to infer that ancestral bats were of laryngeal echolocation type and that ancestral angiosperms flowers were bisexual (Sauquet et al. 2017; Thiagavel et al. 2018).

3.3.6 Inferring the mode of reproduction

The index of association (I_A) is a measure of linkage disequilibrium and can be used to infer the mode of reproduction in a population. I_A was calculated on clone corrected data for Taiwanese, Vietnamese and Australian population using the R package *poppr* (Kamvar et al.,

2015, 2014). First we simulated the distribution of I_A under clonal (90 % linked SNPs), partially clonal (50 % linked SNPs) and sexual (10 % linked SNPs) modes of reproduction with *adegenet* v.2.1.1(Jombart & Ahmed, 2011) and compared our observed I_A with simulated I_A (Tabima, Coffey, Zazada, & Grünwald, 2018). I_A was calculated on 100 randomly sampled SNPs with 100 replicates.

3.3.7 Phylogeographic analyses

Phylogeographic migration analysis was performed using coalescent approaches in the software Migrate-n v.4.2.14 (Beerli 1998, 2009; Beerli and Felsenstein 2001). We determined the likely migration pathways among the 4 populations that had moderate to high genetic diversity: Taiwan, Vietnam, Australia and Africa. Various models of divergence and migration were tested using randomly selected 50, 100 and 299 SNPs for these four populations. Two independent runs were performed with 5 and 10 randomly sampled isolates per population. The Vcf file was converted to the Migrate-n hapmap format file using the function vcfR2migrate implemented in the package vcfR (Knaus and Grünwald 2017). Migrate-n was run for 10,000, 20,000, 50,000, and 100,000 generations with a step size of 100. The first 10% of the analysis were discarded as burn-in. Heating was turned on with four chains set to 1.0, 1.2, 3.0, and 1000000.0. The models were evaluated using Bayesfactor.

3.4 Results

3.4.1 Two pan global lineages, intermediate diversity in Australia and South Africa, and highest diversity in Asia

We obtained a total of 165 multilocus genotypes (MLGs) from a total of 201 *P. cinnamomi* isolates sampled worldwide. We collapsed MLGs based on the technical replicates and a genetic distance cutoff of 0.06 which resulted in 28 MLGs. This critical step removes variation within technical replications and provides variation at the level of a clone. Isolates from Asia (Vietnam and Taiwan) had the highest number of MLGs followed by Australia and Africa (Fig 3.2A, Fig 3.3A). Private MLGs were observed predominantly for Asian samples with the exception of one private MLG observed in South Africa. Rarefaction curves showed highest diversity for Asian isolates and lowest diversity for South American (Chile) isolates (Fig 3.3B). None of the other populations had private MLGs. South Africa and Australia populations were nearly identical. Two pan global, clonal lineages of A2 mating type were commonly observed for populations outside of Asia (either mostly black or pink clusters; Fig 3.2B).

Population structure assigned by ADMIXTURE analysis at K = 5 supported the presence of two dominant pan global, clonal lineages (black and pink) and admixed populations in Taiwan and Vietnam. The Australian and African populations were intermediate and each were assigned predominantly to 3 clusters (black, pink, and red; Fig 3.2B).

3.4.2 Phylogenetic analyses and ancestral state reconstruction suggests Taiwan as a candidate center of origin

To infer relationships between *P. cinnamomi* populations we constructed a maximum likelihood tree with 1,000 bootstrap replicates and "Lewis" correction to correct for ascertainment bias using RAxML. Our ML tree corroborates the previous observation of existence of two panglobal, clonal lineages and high diversity in Taiwanese and Vietnamese populations (Fig 3.4). Taiwanese isolates were observed at the root of the tree. Isolates from Australia and South Africa clustered together indicating genetic similarity. The ML tree was used to infer the likelihood of the probability of nodes by geographic origin. Ancestral state reconstruction for geographic origin of populations supports Taiwan as the ancestral population (Fig 3.5).

3.4.3 Mating types and mode of reproduction

More frequent occurrence of the A2 mating type compared to the A1 mating type has been reported previously. We observed both A1 and A2 mating types in Asia, South Africa and Australia. However, populations from Asia (Taiwan and Vietnam) and Africa met our expectation of a 1:1 mating type ratio consistent with the indication of sexual reproduction. Occurrence of the A1 mating type outside of Asia was rare (Table 3.1). Populations from Taiwan and Vietnam indicated a semi clonal mode of reproduction based on simulations of linkage evaluated using index of association (Fig 3.6). This is based on testing the observed index of association to the simulated distributions under a clonal (90 % linkage), partially clonal (50 % linkage) and sexual (10 % linkage) mode of reproduction.

3.4.4 Phylogeographic analysis suggests Taiwan as ancestral population

We simulated scenarios for multiple models of population divergence and migration using the coalescent approaches implemented in migrate-n. The model with the highest probability varied between two different models depending on the number of individuals, number of SNPs and number of MCMC runs (Fig 3.7). The difference between these two models is only in the emergence of either the African or Australian population outside of Asia. The first model suggests Australia was the first population emerging from Asia whereas the other model suggests Africa. Regardless of which population first emerged from Asia, both models suggest Taiwan as an ancestral population and bidirectional gene flow between Africa and Australia.

3.5 Discussion

Crandall and Gravatt (1967) first proposed Southeast Asia as the center of origin of *P. cinnamomi*. This was based on the fact that native vegetation in the region was resistance to disease. Our analyses on global populations of *P. cinnamomi* strongly support the idea put forwarded by Crandall and Gravatt 1967. High genetic diversity, presence of both mating types in a 1:1 ratio and ancestral population estimation suggests Taiwan as the likely center of origin for the pathogen. No admixture was observed for isolates except from Asia suggesting clonal population in these regions which is consistent with the hypothesis of *P. cinnamomi* being introduced from Asia to rest of the world. Low genotypic diversity in *P. cinnamomi* population was also observed as speculated previously (Eggers et al. 2012; Engelbrecht et al. 2017; Beaulieu et al. 2017).

Population structure analysis of *P. cinnamomi* from Australia, South Africa and the United States has shown populations to be genetically less diverse and provided support for

clonal reproduction. Our analyses confirm the previous works regarding their diversity and mode of reproduction. Predominance of single mating type (A2) further supports that the idea of clonal reproduction in those regions. Both mating types were detected from the African population (South Africa) which provides an opportunity for sexual reproduction however no cross over isolates have been described in the region. Linde et al. (1999) suggested the presence of both mating types in South Africa with the possibility of subtle sexual reproduction. Genotypes from South Africa resemble those from Australia. This could be possibly due to introduction from the same source or migration between these populations. Our migration analyses do suggest bidirectional gene flow between African and Australian populations. Genetic similarity between these population has also been reported (Linde et al. 1999).

One of the recurring themes in *P. cinnamomi* population structure is the emergence of two A2 lineages and a rare A1 mating type tending to form their own groups. Phylogenetic analysis based on ML tree adds support to genetic diversity in Asia and the existence of two A2 clades. Presence of two different A2 clonal lineages has been reported by multiple studies in Australia, South Africa and the United States (Dobrowolski et al. 2003; Pagliaccia et al. 2013; Eggers et al. 2012; Engelbrecht et al. 2017; Beaulieu et al. 2017; Linde et al. 1999, 1997). Dobrowolski et al. (2003) identified three clonal lineages of *P. cinnamomi* from Australia which includes two A2 lineages and one A1 lineage. However A1 isolates from Taiwan and Vietnam did not form a group to support the hypothesis of reproductive isolation.

Our samples of *P. cinnamomi* from Asia are limited to the two nations of Taiwan and Vietnam. Inclusion of samples from Indonesia and Malaysia will improve our understanding of diversity in a wider geographic region. Similarly Papua New Guinea has been also described to have a diverse population of *P. cinnamomi* and is also suggested as center of diversity (Arentz

and Simpson 1986). Recently a few more genomes of *P. cinnamomi* have been sequenced and made available, use of these new genomes to map our short reads could identify novel variants (Longmuir et al. 2017; Studholme et al. 2015).

In short, we surmise that Taiwan is the center of origin of *P. cinnamomi* with the highest diversity in Southeast Asia. However, for the rest of the world, two dominant clonal lineages of A2 mating types and one lineage of an A1 mating type was identified. The A1 mating type is rare in other parts of the world. The presence of A1 in Taiwan and Vietnam provides an opportunity for sexual reproduction probably leading to aggressive genotypes.

Table 3.1 Number of A1 and A2 mating type isolate by population. A1/A2 significance was tested using a binomial distribution.

Population	**Number of isolates**	**A1**	**A2**	**A1/A2**	**p-value**
Asia	68	33	35	0.95	0.90
Australia	37	2	27	0.08	< 0.05
Africa	11	5	5	1	1.00
Europe	45	0	45	0	< 0.05
North America	26	0	7	0	0.01
South America	14	0	14	0	< 0.05

Figure 3.1 Impact of the invasive plant pathogen *P. cinnamomi* on native plant ecosystems in Australia. Left: Heath landscape, showing Phytophthora dieback in *a Eucalyptus* stand in the mid-ground, in the Stirling Range, Western Australia. Right: Sign warning about the risk of moving the invasive pathogen *Phytophthora cinnamomi* in Western Australia via infested soil to 'help stop the rot'. (Photo by: Gnangarra, Wikimedia).

Figure 3.2 Global distribution and admixture population structure of *P. cinnamomi*. (A) Distribution of *P. cinnamomi* multilocus genotypes. Sampled countries are colored in the world map. Each color in the pie charts represents a multilocus genotype. (B) ADMIXTURE plot of *P. cinnamomi* isolates by populations at K = 5. Each bar represents an isolate with proportional ancestry assigned to one of each K populations. The two panglobal lineages are shown in black and pink.

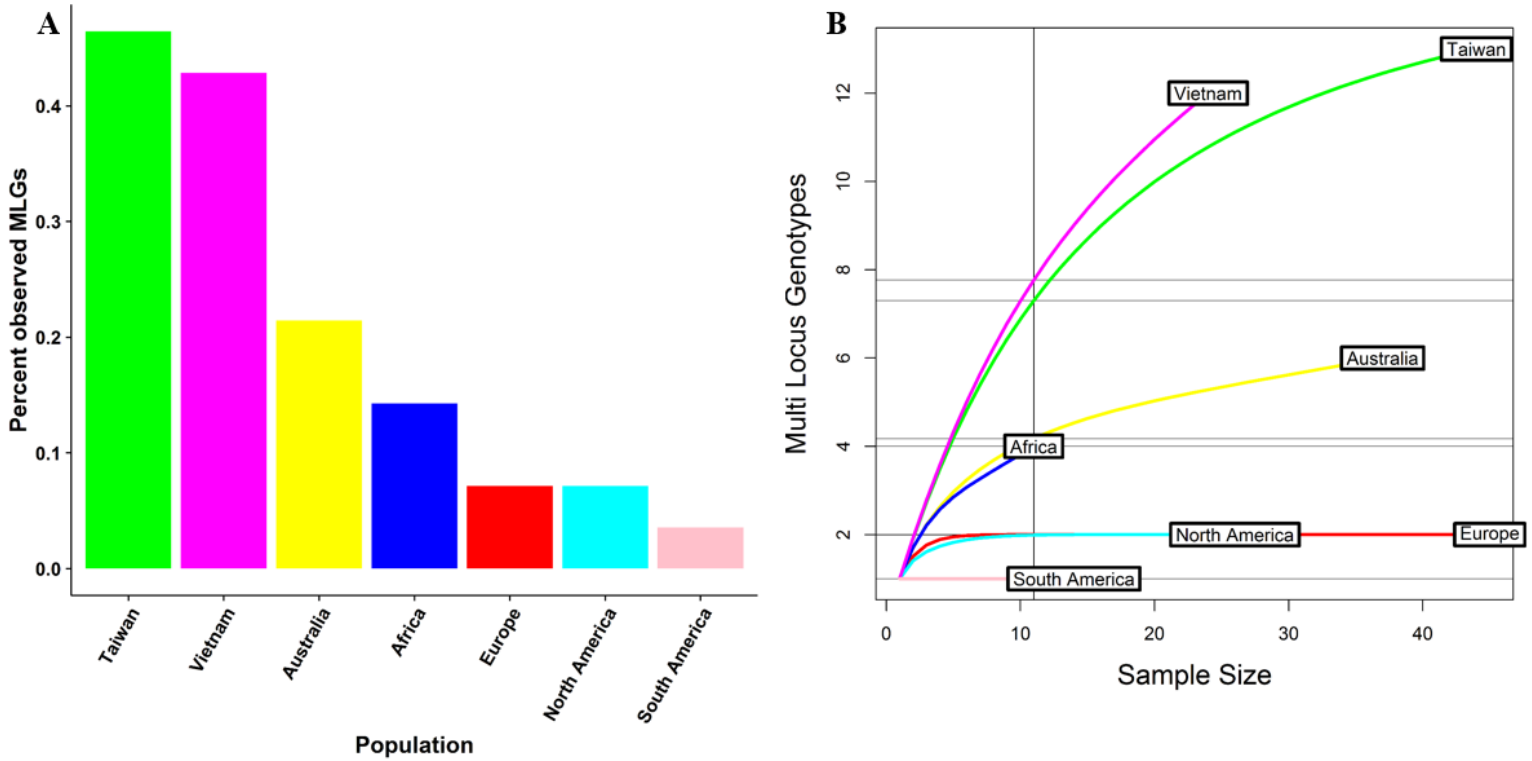

Figure 3.3 Global genetic diversity observed in *P. cinnamomi* by populations. (A) Proportion of observed multilocus genotypes (MLG) by regional population using a bitwise cutoff distance of 0.06. (B) Rarefaction curves by regional population.

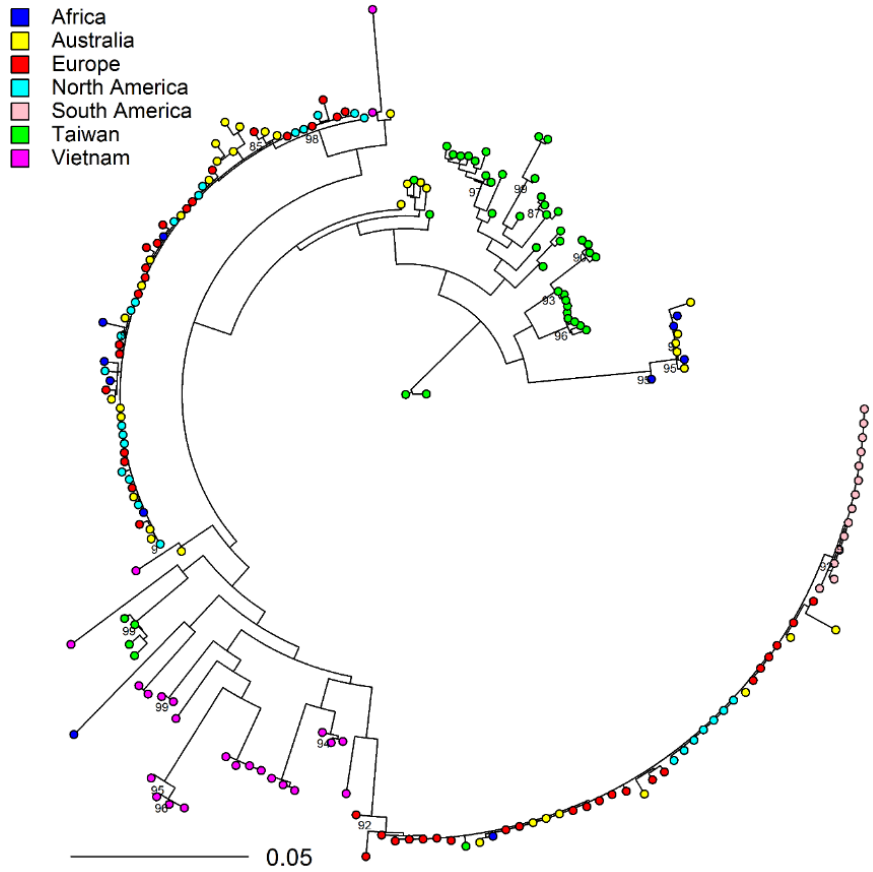

Figure 3.4 Maximum likelihood tree of *P. cinnamomi* by regional populations. The tree was inferred using 299 concatenated single nucleotide polymorphism (SNP) with 1000 bootstrap replicates using RAxML version 8.2.11. ASC_MULTIGAMMA substitution model with "Lewis correction" was used to correct for ascertainment bias. Number at node indicates bootstrap support. Tree was rooted using *P. sojae* as outgroup.

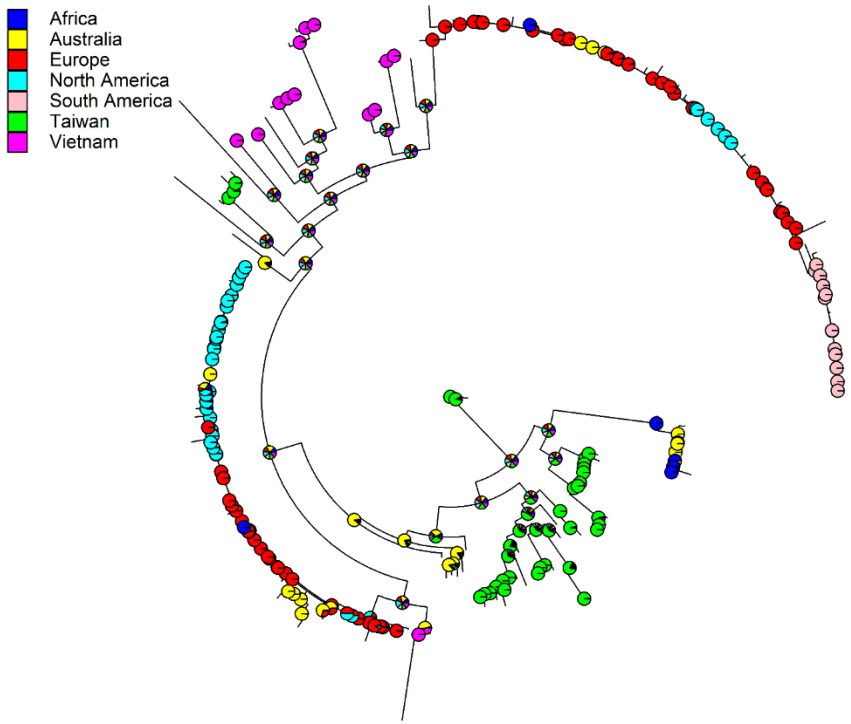

Figure 3.5 Maximum likelihood probability of nodes by geographic origin. Ancestral population estimation at each node using a ML tree. Pie chart represents the likelihood of the node being assigned to each original geographic population.

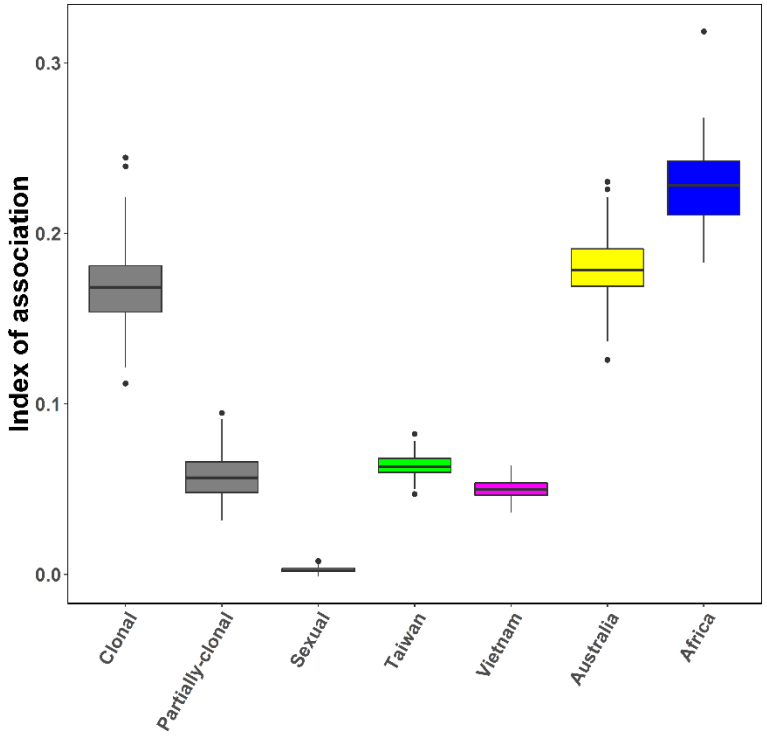

Figure 3.6 Measurement of linkage disequilibrium using the index of association (I_A). The first three box plots in grey represent the simulated I_A under clonal, partially-clonal and sexual reproduction scenarios. The next four box plots represents the observed I_A for Taiwan, Vietnam, Australia and Africa population. The Taiwanese and Vietnamese populations suggest a partially clonal mode of reproduction, whereas the Australian and African populations suggest clonal reproduction.

Figure 3.7 Phylogeographic scenarios with the highest likelihood based on coalescent analysis in migrate-n. Various combinations of divergence and migration models were tested using migrate-n. Above two represents the models with highest likelihood based on Bayes factor. In both models, Taiwan is inferred as the ancestral population and bi-directional migration between Australia and Africa which might have led to the genetic similarity between these regions.

Chapter 4

Evolution of RxLR genes in the Irish famine pathogen *Phytophthora infestans* and its close relatives

Shankar K. Shakya, Amirhossein Bahramisharif, Howard S. Judelson, Brian J. Knaus, Laura E. Rose and Niklaus J. Grünwald

Target journal: MPMI

4.1 Abstract

The Irish potato famine pathogen *Phytophthora infestans* is considered to be one of the most devastating plant pathogens. This economically destructive pathogen can rapidly evolve new virulent genotypes because of a large effective population size, sexual recombination, changes in ploidy and gene duplication events. The genome of *P. infestans* harbors hundreds of effectors that contribute to virulence and pathogenicity on a host plant. One of the main families of effector proteins is characterized by the presence of an RxLR amino acid domain, which is required for translocation into the host cell. However, little is known about the conservation and evolution of these effectors. Therefore, to gain greater insight into the mechanisms involved in plant-pathogen interactions, the molecular evolution of avirulence (Avr) genes, core RxLR genes and RxLR genes with a nuclear localization signal (NLS) motifs from three *Phytophthora* clade 1c species were evaluated. Our analysis of the presence and absence in RxLR genes across this group shows that *Phytophthora ipomoeae* and *P. mirabilis* have fewer RxLR homologs compared to *P. infestans,* which is to be expected given the ascertainment bias of mapping all sequences against the *P. infestans* reference genome. Furthermore, all three clade 1c species have only two well characterized Avr genes (*Avrvnt1* and *AvrSmira1)* in common. In general, nucleotide diversity was higher in the RxLR genes compared to core orthologous genes. In addition, some of the substitutions are predicted to alter the secondary structure of the protein which may affect protein function. The RxLR effectors also showed different expression patterns during infection of *Solanum* species. Finally, 11 genes showed evidence of recent diversifying selection. A deeper understanding of these effectors could provide efficient and novel strategies for long-term disease management and developing pathogen-resistant crops through effector-mediated molecular breeding.

4.2 Introduction

Plants and oomycete pathogens interact via a chain of molecular events. Most plants constitutively express innate immune receptors, also called pathogen recognition receptors or PRRs (Cooke et al., 2012; Jones & Dangl, 2006; Katagiri & Tsuda, 2010). At the time of attack, these receptors are able to recognize pathogen/microbe associated molecular patterns or PAMPs/MAMPs that trigger immune response called PAMP-triggered immunity (PTI) (Cooke et al., 2012; Jones & Dangl, 2006; Katagiri & Tsuda, 2010). In turn, many plant pathogens including oomycetes have evolved to modulate PTI by secreting effector proteins that are required for pathogenesis (Tyler et al., 2006). However, effector molecules can be recognized by specific plant resistance (R) proteins. The interaction between effector proteins and their cognate R proteins may lead to an immune response called "Effector Triggered Immunity" (ETI) which may result in a hypersensitive response (HR) at the site of infection. Effectors that lead to ETI in the host are referred to as avirulence (Avr) proteins. Over time, evolutionary changes in Avr genes allow pathogens to evade recognition by their hosts R-proteins. These changes can include loss of Avr genes, the acquisition of a premature stop codon leading to a truncated protein or amino acid changes resulting in the loss of recognition by the cognate R gene and/or loss of expression of the Avr gene (Armstrong et al., 2005; Gilroy et al., 2011; van Poppel et al., 2008).

Oomycete effectors can be divided into two main categories: apoplastic and cytoplasmic effectors. Apoplastic effectors act outside the plant cells (intercellular space) and cytoplasmic effectors function inside plant cells following translocation through the haustorium (Schornack et al., 2009; Whisson et al., 2007). One prominent class of cytoplasmic effector proteins is characterized by a conserved RxLR motif in the N-terminal region (Schornack et al., 2009).

Although, over 500 RxLR genes have been reported from the genome of *P. infestans* strain T30-4 (Haas et al., 2009), only a few genes have been shown to encode Avr activity (Bos et al., 2010; Champouret et al., 2009; Du et al., 2018; Gilroy et al., 2011; Oh et al., 2009; Rietman et al., 2012; van Poppel et al., 2008). The Avr genes have been a primary focus in the development of late blight resistant potatoes (Jo et al., 2014). Due to their important role in pathogenicity and virulence, RxLR genes have attracted the attention of many plant pathologists. While it is known that the RxLR motif is necessary for translocation inside the host cell (Schornack et al., 2009), the evolutionary history and diversification of this class of proteins has not been elucidated. Furthermore, there is no information on the presence or diversity of RxLR genes in the sister species *P. ipomoeae* and *P. mirabilis.*

The Irish famine pathogen, *Phytophthora infestans* (Mont.) de Bary causes late blight and results in serious yield losses on potato and tomato (Agrios, 2005; Fry et al., 2015), making this pathogen a formidable threat to two important crops. This pathogen belongs to the phylum Oomycota and interacts with plants in two sequential phases: first as a biotroph and then as a necrotroph (Whisson et al., 2007). Under suitable conditions, *P. infestans* is capable of destroying entire fields of crops in a short period of time, due to a rapid asexual life cycle (Fry, 2008; Fry et al., 2015).

P. infestans is closely related to *Phytophthora ipomoeae* and *P. mirabilis* within clade 1c of the genus *Phytophthora* (Flier et al. 2002; Galindo-A and Hohl 1985). These species are known to have coexisted in central Mexico at their center of origin (Flier et al., 2002; Goss et al., 2014; Grünwald & Flier, 2005). Unlike *P. infestans, P. ipomoeae* and *P. mirabilis* have not been reported to cause blight symptoms on *Solanum* species. *P. ipomoeae* has been reported to cause

leaf blight on *Ipomoea longipedunculata* L. (morning glory), while *P. mirabilis* has been reported to damage foliage of *Mirabilis jalapa* L. (four o'clock) (Flier et al., 2002; Goodwin & Fry, 1994). Comparative genomic analysis with clade 1c taxa can provide novel insights into the ancestral states and evolutionary history of *P. infestans*.

The overall goal of this study was to determine the patterns of molecular evolution of RxLR genes in clade 1c of *Phytophthora* to gain better insight into the mechanisms involved in plant-pathogen interactions. The aims of this study were to (1) identify a core set of RxLR effector genes conserved within *Phytophthora* clade 1c, (2) identify RxLR genes with a nuclear localization signal motif, (3) determine the putative impact of genetic variation on protein function in terms of secondary structure, (4) investigate genes with signatures of diversifying selection and (5) evaluate patterns of gene expression of *P. infestans* in cultivated and wild *Solanum* species during infection.

4.3 Materials and Methods

4.3.1 Data collection and sequencing

Raw genome sequence data of 12 strains of *P. infestans* and one strain of *P. mirabilis* were downloaded from GenBank (Knaus et al. 2019; Table 4.1). In addition, 10 isolates of *P. infestans*, four isolates of *P. ipomoeae* from Mexico and three isolates of *P. mirabilis* from Mexico were *de novo* sequenced using the Illumina Hiseq3000 technology (San Diego, CA, USA).

4.3.2 Reads mapping

Read libraries were mapped to the *P. infestans* reference genome (T30-4) using the Burrows-Wheeler Aligner (BWA) v0.7.13 with default settings (Haas et al. 2009; Li and Durbin 2009). Sequence coverage was evaluated by displaying the number of reference bases that are covered by mapped sequencing reads at various depths. To determine the rate of evolution, whole genome sequence data of 30 isolates were used for single-nucleotide polymorphism (SNP) discovery and genotype calling using GATK HaplotypeCaller v. 4.0.11.0 (DePristo et al., 2011). Genotype calling and phased haplotype calling were performed using BEAGLE v4.1 under default settings (B. L. Browning & Browning, 2013; S. R. Browning & Browning, 2007). The gene sequences were reconstructed using the function "vcfR2DNAbin" implemented in the "vcfR v.1.6.0" package (B. J. Knaus & Grünwald, 2017).

4.3.3 Breadth of Coverage and presence/absence polymorphism

Breadth of Coverage (BOC) was used as a metric for gene presence/absence polymorphisms (Cooke et al., 2012; Yoshida et al., 2013). BOC was defined as the proportion of positions sequenced in an isolate, at least once, compared to the reference genome. First, the coverage for each gene was calculated and divided by the length of gene sequence to get a BOC value. A BOC cutoff value of 0.90 was used to call RxLR gene presence/absence polymorphisms, given that at this threshold the numbers of RxLR genes in *P. ipomoeae* and *P. mirabilis* were comparable to other *Phytophthora* species (Haas et al., 2009).

4.3.4 Haplotype diversity

We determined haplotype diversity on validated Avr gene sequences. The number of observed haplotypes were calculated for each Avr gene by species using the R package "pegas v.0.10" (Paradis, 2010). Haplotype counts were used to calculate haplotype diversity for each Avr gene (Nei, 1978).

4.3.5 Nucleotide diversity

Nucleotide diversity was calculated for 218 RxLR genes that were present in all three clade 1c *Phytophthora* species at a BOC cutoff value of 0.90. This value was further compared to the nucleotide diversity value of 218 randomly selected core orthologous genes. Nucleotide diversity was determined using the R package "pegas v.0.10" (Paradis, 2010). To determine if the distribution of nucleotide diversity differed between RxLR and core orthologous genes, a non-parametric Kruskal-Wallis test was conducted in R (R Core Team 2014). The value of nucleotide diversity was also evaluated for *P. infestans* for the 343 RxLR genes that showed presence/absence polymorphism and for 220 RxLR genes conserved in all three *Phytophthora* species. Nucleotide diversity was also calculated for synonymous sites and nonsynonymous sites separately using DnaSP v6 (Rozas et al., 2017).

4.3.6 Nuclear localization signal prediction

To identify a nuclear localization signal (NLS) motif in the RxLR genes, NLStradamus was used (Ba, Pogoutse, Provart, & Moses, 2009). NLStradamus uses a simple hidden Markov model to predict novel NLSs in proteins. Putative RxLR pseudogenes detected by Haas et al. (2009) were excluded from the NLS detection.

4.3.7 Protein secondary structure

To determine the putative impact of genetic variation on protein function, comparative methods were used. Predicted protein sequence alignments of Avr genes and genes with NLS motifs were inferred using the Geneious plugin ClustalW 2.1. Secondary protein structures were predicted using EMBOSS 6.5.7 by using the Garnier Osguthorpe Robson algorithm (Garnier, Osguthorpe, & Robson, 1978; Rice, Longden, & Bleasby, 2000). The differences in predicted secondary structures were compared and the underlying causative nucleotide substitutions were identified.

4.3.8 Gene expression

To investigate regulation of gene expression of five RxLR-NLS genes during infection with *P. infestans* D12-2, gene-specific primers were designed using the Geneious plugin Primer 3 2.3.7 (Rozen & Skaletsky, 1999) (Table 4.2). Using reverse transcription polymerase chain reaction (RT-PCR), the transcript abundance of a subset of RxLR genes, as well as the cytochrome *c* oxidase subunit II (*COX2*) was determined for *Solanum lycopersicum* and *S. pimpinellifolium* at six time points (0 hpi, 6 hpi, 24 hpi, 48 hpi, 72 hpi, 96 hpi) following inoculation with *P. infestans* strain D12-2. The RT-PCR reactions were carried out in a total volume of 20 µl consisting of Green GoTaq® Flexi Buffer, 2U GoTaq® Flexi DNA Polymerase (Promega, Fitchburg, WI, USA), 1.25mM $MgCl_2$, 0.1 mM dNTPs and 0.2 mM of each primer. The T100™ Thermal Cycler (Bio-Rad, Hercules, CA, USA) was used for amplification of *P. infestans* with the following PCR conditions: initial denaturation at 95°C for 3 min, 34 cycles of 95°C for 30 s, annealing for 30 s at 60°C, extension at 72°C for 1.30 min, and final extension at

72°C for 5 min. PCR products were run on a 1% agarose gel and DNA was stained using Midori Green Advance (Nippon Genetics Europe, Dueren, Germany).

4.4 Results

4.4.1 Sequencing and mapping

The majority of short sequencing reads from *P. infestans*, *P. ipomoeae* and *P. mirabilis* could be aligned to the *P. infestans* T30-4 reference genome. Under a BOC cutoff value of 90%, 220 RxLR genes were detected in samples from all three species of *Phytophthora*. As expected, higher BOC cutoff values resulted in a lower number of genes retained. For example, at a BOC cutoff of 0.95, only 177 core effector genes were detected in all three species of clade 1c *Phytophthora*. The number of core RxLR effectors detected at a BOC cutoff value of 0.90 in each species were: 333 for *P. ipomoeae,* 353 for *P. mirabilis and* 397 for *P. infestans.*

4.4.2 RxLR genes in *Phytophthora* species

The coverage of RxLR genes within each genome relative to the *P. infestans* reference genome (T30-4) was evaluated. Both isolates of the US-23 clonal lineage, and isolates FL2009P4 (538 RxLR) and BL2009P4 (537 RxLR) had the highest number of RxLR genes shared with T30-4 followed by P13527 (Ecuador) and FP-GCC (US-11 lineage) which had 529 RxLR genes. The genomes of individuals of *P. infestans* from the sexual Mexican populations showed lower RxLR gene content compared to the genomes from other clonal individuals of *P. infestans*. The coverage of RxLR genes was the lowest for *P. mirabilis* and *P. ipomoeae* compared to *P. infestans* which is to be expected given the higher genetic divergence of these taxa relative to the T30-4 reference genome. For *P. mirabilis* the number of retained RxLR genes

ranged from 381 to 414 whereas *P. ipomoeae* had RxLR genes ranging from 364 to 381 (Fig 4.1).

4.4.3 A subset of RxLR genes encode a nuclear localization signal

To determine the potential contribution of the RxLR genes in cell death induction, the presence of a nuclear localization signal (NLS) was investigated. The presence of a NLS motif was predicted in the C-terminus of 47 RxLR genes. Seventeen of the RxLR genes with a predicted NLS motif are reported to belong to avirulence gene families. These genes include the members of the *Avr2* family (*PITG_06071*, *PITG_06077*, *PITG_07499*, *PITG_08278*, *PITG_19617*, and *PITG_20025*) and the members of the *Avrblb2* family (*PITG_04081*, *PITG_04085*, *PITG_04086*, *PITG_04090*, *PITG_04097*, *PITG_09632*, *PITG_15972*, *PITG_18683*, *PITG_20300*, *PITG_20301*, and *PITG_20303*).

4.4.4 Presence/absence variation in avirulence genes

The genes displaying presence/absence variation (343 genes) were distributed across 106 RxLR gene families out of 252 RxLR families originally described in Haas et al. (2009). The largest number of genes with presence/absence variation belonged to RxLR family 1 (42/85), followed by RxLR family 5 (21/22), RxLR family 9 (16/17) and RxLR family 7 (15/18). RxLR family 1 does not contain any known Avr genes. However, the RxLR families 5 and 7 contain *Avrblb2* and *Avr2* genes and their paralogs.

The pattern of presence/absence variation was further investigated for previously characterized Avr genes (Fig 4.2). In total, two Avr genes, including *AvrSmira1* (*PITG_07550*), *Avrvnt1* (*PITG_16294*) and one recently described *Avr-1* like gene (*PITG_06432*) were found in

all the isolates of *P. infestans*, *P. ipomoeae*, and *P. mirabilis*. The *Avr1* gene (*PITG_16663*) was missing from most isolates of *P. infestans*, but was present in the isolates of clonal lineages US-1 and US-22. This gene was not detected in *P. mirabilis*, but was detected in *P. ipomoeae*. The *Avr2* gene was found in all samples, except 3 isolates of *P. mirabilis*. The *Avr3a* gene (*PITG_14371)* which is an important pathogenicity determinant in *P. infestans* was present in all the isolates of *P. infestans*, regardless of their geographic origin. However, none of the isolates of *P. ipomoeae* and *P. mirabilis* had the *Avr3a* gene.

The *Avr4* gene (*PITG_07387)* has been shown to have frameshift mutations resulting in truncated proteins which are not recognized by the corresponding *R*-gene (*R4* resistance) (van Poppel et al., 2008). The value of BOC for the *Avr4* gene was lower than or close to 0.90 in all the isolates of *P. ipomoeae*, *P. mirabilis* and *P. infestans* from Mexico, including isolate 06_3928A (blue13 strain). The *Avr4* gene was absent in one of the isolates belonging to the US-22 lineage (IN20009T1) and one of the Mexican isolates of *P. infestans* (Pic97335/inf3).

The *Avrblb1* gene (*PITG_21388)* was present in all the isolates of *P. infestans* and one isolate of *P. mirabilis*. However, the *Avrblb2* gene (*PITG_20300)* had an unusual pattern of presence/absence polymorphism. The *Avrblb2* gene was present in all the *P. mirabilis* isolates, but it was absent in *P. ipomoeae*. In *P. infestans*, *Avrblb2* had a BOC value lower than 0.90 for the isolates belonging to US-8 (1/2), US-22 (3/3), US-24 (1/1) and Mexico (3/5). Together, these results indicate that some Avr genes are core to clade 1c, some are core to *P. infestans* and that a proportion of Avr genes across the populations show presence/absence variation, suggesting that some Avr genes can be isolate- or species-specific.

4.4.5 Nucleotide diversity is higher at RxLR genes compared to core genes

Nucleotide diversity for RxLR genes was compared to a set of core orthologous genes. No sequence variation was detected at two (PITG_14294, PITG_23199) out of 220 core RxLR effectors. Therefore, nucleotide diversity was calculated for the remaining 218 RxLR genes that were present in all isolates of *P. infestans*, *P. ipomoeae* and *P. mirabilis*. The nucleotide diversity of the RxLR genes was compared to that of 218 randomly selected core orthologous genes. A Kruskal-Wallis test showed statistically significant difference between these groups of genes. Across all three species, nucleotide diversity was significantly higher for the set of conserved RxLR genes compared to randomly selected core orthologous genes ($P < 0.001$) (Fig 4.3A). Diversity was the lowest at both RxLR and core orthologous genes in the samples from *P. ipomoeae*. Nucleotide diversity was not significantly different between the set of conserved RxLR genes and those which showed presence/absence polymorphism (Fig 4.3B).

4.4.6 Nucleotide diversity is lower in sexual populations

The Mexican population of *P. infestans* is the only population in our sample known to reproduce sexually. The other samples included are reported to reproduce asexually and are therefore classified as clonal. From a population genetics perspective, sexually reproducing populations are predicted to maintain higher levels of genetic diversity than asexual populations. We tested whether this was true for the RxLR genes in our sample. We found that genetic diversity was significantly lower in the sexually reproducing population from Mexico compared to a pooled sample of all clonal lineages (p-value 0.00008). Therefore, although genetic variation in a clonally reproducing population is expected to be lower than that of a sexually reproducing

population, substantial genetic variation can be maintained if multiple independent clonal lineages coexist (and are pooled), as seems to be the case here.

4.4.7 Population genetic analyses

We evaluated if there was evidence of selection operating on RxLR genes. A total of 47 genes were identified that had sufficient coverage for population genetic analyses (Fig 4.2). We first evaluated the ratio of nonsynonymous substitutions per nonsynonymous sites (π_a) to synonymous substitutions per synonymous sites (π_s). This ratio (π_a / π_s) can be used as an indicator of selective constraint within species (Nei and Li 1979). A ratio less than one indicates negative or purifying selection, a ratio of one is expected if the gene is evolving neutrally and a ratio greater than one implies balancing selection. Eleven genes (*PITG_04090*, *PITG_04266*, *PITG_06308*, *PITG_09837*, *PITG_12276*, *PITG_14371*, *PITG_14986*, *PITG_18683*, *PITG_20300*, *PITG_22978* and *PITG_22990*) surveyed had a ratio of π_a / π_s greater than unity (Fig 4.4). This may indicate that some of the amino acid variants within these genes are favored by natural selection.

Another method to detect natural selection is Tajima's D allowing for comparison within species (Tajima 1989). A significant positive Tajima's D value is consistent with the action of balancing selection, while a significant negative value is consistent with the action of recent positive selection. Tajima's D was calculated on the set of conserved RxLR genes using complete gene sequence and only effector domain or C-terminal domain. When compared within species, none of the conserved RxLR genes were found to be under positive selection either with whole gene sequence or with C-terminal sequence for all three species. Few conserved RxLR genes were found to be under balancing selection. *P. infestans* had 28 RxLR genes under balancing

selection, whereas *P. mirabilis* had 11 and *P. ipomoeae* had 6, using the complete gene sequence. None of the genes were shared across all three *Phytophthora* species for balancing selection. However, three RxLR genes (*PITG_06375*, *PITG_13072* and *PITG_22925*) were shared between *P. infestans* and *P. mirabilis* and only one RxLR gene (*PITG_09732*) was shared between *P. mirabilis* and *P. ipomoeae*. A selection test using only the C-terminal domain resulted in 20, 9 and 6 genes under balancing selection for *P. infestans*, *P. mirabilis* and *P. ipomoeae,* respectively. This resulted in 10 new RxLR genes that were found to be under balancing selection using the C-terminal domain, but not with the whole gene sequence.

4.4.8 Variation in Avr genes and selection analysis

Well characterized Avr genes from *P. infestans* were further studied for their nucleotide diversity π, number of synonymous and nonsynonymous substitutions and selection. Haplotype diversity for all Avr genes in *P. infestans* was greater than 0.5 except for *Avr2* (Table 4.3). The *Avr2* gene had only one nonsynonymous substitutions (N31K) which resulted in two haplotypes and low haplotype diversity. Haplotype diversity was highest for *Avrblb1* in *P. infestans* (0.87). The *Avrblb1* and *Avrblb2* genes had the highest number of haplotypes (13) followed by *AvrSmira2* with 12 haplotypes (Table 4.3). The *AvrSmira1* gene had the highest number of haplotypes for *P. mirabilis* and *P. ipomoeae* isolates. The number of synonymous and nonsynonymous nucleotide substitutions varied a lot between species. Within *P. infestans*, a maximum of 15 nonsynonymous substitutions were observed for *AvrSmira2* followed by *AvrSmira1* (13) and *Avrblb1* (12). *AvrSmira2* also had the highest number of synonymous changes (9) followed by *AvrSmira1* (6) and *Avrblb1* (5) in *P. infestans*. No synonymous changes were observed for *Avr4* in *P. infestans*. *P. mirabilis* had nonsynonymous changes ranging from 1

to 49. *AvrSmira1* had the maximum of 49 nonsynonymous and 26 synonymous substitutions in *P. mirabilis* compared to the T30-4 reference. *Avrblb1* and *AvrSmira2* also had more than or equal to 15 nonsynonymous changes in *P. mirabilis*. In contrast *Avrvnt1* in *P. ipomoeae* had maximum of 22 nonsynonymous and 9 synonymous substitutions. Another Avr gene (*AvrSmira1*) in *P. ipomoeae* had 16 nonsynonymous and 8 synonymous changes. In general, most nonsynonymous substitutions were occurred in C-terminal domain.

Tajima's D was also interpreted for the Avr genes. In this test, the average number of pairwise differences is compared with the number of segregating sites. C-terminal domain or the functional domain of gene sequence is known to harbor more mutations compared to the other regions. Thus, Tajima's D statistic was calculated for both the full-length gene sequence and the C-terminal region. Interestingly, only one Avr gene was found under balancing selection in *P. infestans* (Table 4. 4). *AvrSmira1* (PITG_07550) was shown to be under balancing selection, suggesting the existence of multiple alleles in the *P. infestans* population.

4.4.9 Robustness of protein secondary structure to mutation

To predict mutations that could alter the role of the effector protein secondary structure elements were examined for their ability to tolerate mutation. In the RxLR genes, alpha helices and beta strands accumulated more mutations than coils. In general, more SNPs occurred in the alpha helix regions. Furthermore, the alpha helices seem to accept more synonymous and nonsynonymous mutations and they are more robust to mutations than other regions. Interestingly, one Avr gene with an NLS motif (*PITG_18683*), two Avr (*PITG_07550*, *PITG_21388*) and five NLS-RxLR genes (*PITG_12276*, *PITG_15038*, *PITG_15225*,

PITG_22798, PITG_23193) showed a conversion of secondary structure elements in the effector domain that was induced by a single-site mutation. (Fig 4.5).

4.4.10 Expression analysis

To investigate the regulation of gene expression, five RxLR genes including *PITG_12276*, *PITG_15038*, *PITG_19994*, *PITG_22978* and *PITG_22990* were chosen. All of the selected genes had an NLS motif in which three of them have undergone diversifying selection (*PITG_12276, PITG_22978* and *PITG_22990*). The NLS-RxLR gene *PITG_15038* showed high rate of secondary structure changes due to mutations, whereas *PITG_19994* had no variants. To determine the expression pattern of these genes during infection, cDNA reverse transcribed from six time points were used to conduct RT-PCR. Across all the selected genes, the RxLR genes *PITG_22798* and *PITG_15038* were upregulated at early stages of infection (24hpi and 48 hpi) in both *S. lycopersicum* and *S. pimpinellifolium* (Fig 4.6). Furthermore, RT-PCR revealed that the RxLR genes *PITG_ 19994* and *PITG_12276* were also upregulated but only at the necrotrophic phase of infection, whereas *PITG_22990* was not induced in any stages of infection (Fig 4.6).

4.5 Discussion

In this study, we conducted comprehensive comparative genomic analyses and investigated the molecular evolution, the presence/absence polymorphism and diversity of RxLR in *Phytophthora* clade 1c species. Loss or mutation in RxLR genes can evade recognition and facilitate infection. This study showed that *P. ipomoeae* and *P. mirabilis* had fewer RxLR genes compared to *P. infestans* Most of the Avr genes were detected in *P. infestans*, except *Avr1*,

suggesting their importance in pathogenesis. The RxLR genes were also shown to harbor more diversity compared to core orthologous genes, as speculated and were shown to be under balancing selection of multiple haplotypes.

Oomycete effectors are secreted at different sites of the host (Kamoun, 2006). The known RxLR effectors that have Avr activities have been reported to be translocated inside the plant cell (Schornack et al., 2009). This study identified 47 RxLR encoding genes that are predicted to have NLS motifs in their C-termini and were therefore, predicted to move to the nucleus and possibly activate or manipulate host gene expression. The NLS-RxLR effector *PITG_22279* has been previously reported to target the host nucleus and cause cell death in an agroinfiltration assay using *Nicotiana benthamiana* (Wang et al., 2017). Interestingly, the study by Wang et al. (2017) showed that intentional changes to the NLS region of *PITG_22279* disabled localization of the protein in the nucleus which may have prevented the induction of cell death.

The RxLR genes products can be recognized by large arsenals of immune receptors called resistance (R) proteins that activate host defense. This interaction is called the "gene-for-gene" response and has been inferred in many plant-pathogen interactions, independent of a past history of breeding. Gene-for-gene refers to the observation that specific genotypes of the host are resistant to specific genotypes of the pathogen. However, the effectors are believed to have undergone evolutionary forces (Hogenhout, Van der Hoorn, Terauchi, & Kamoun, 2009; Kamoun, 2007). In fact, high rates of sequence polymorphisms and amino acid changes have been observed in effector proteins, because they rapidly evolve under diversifying selection (Liu et al., 2004; Win et al., 2007). The current study identified 11 RxLR genes with signatures of

diversifying selection. This knowledge can provide insight into genetic engineering and can greatly facilitate the search for novel resistance genes.

The discovery of novel resistance genes has been challenging for many years. In *P. infestans*, nucleotide substitutions and gene expression polymorphism have been described as possible mechanisms for the loss of Avr genes in order to overcome resistance genes (Bos et al., 2010; Gilroy et al., 2011; van Poppel et al., 2008). The *Avr1* gene was missing from most of the *P. infestans* isolates, but it was present in the US-1 and US- 22 lineages. The loss of *Avr1* from *P. infestans* in the modern samples, except US-22, might be due to R1 gene deployment. A study by Yoshida et al. (2013) has shown that the *Avr1* gene has been lost in the modern isolates of *P. infestans* , except one isolate belonging to the US-22 lineage. The current study confirmed that the three tested US-22 isolates have an intact *Avr1* gene with no sequence variation. Absence of *Avr1* in the modern isolates, except US-22 may suggest that the US-1 lineage could be a possible parent in the emergence of the US-22 lineage. Another important Avr gene, *Avr3a* was present in all the isolates of *P. infestans,* but absent in *P. ipomoeae* and *P. mirabilis*. This suggests that the *Avr3a* gene in *P. infestans* is associated with pathogenesis. The *Avr3a* gene contains two forms: *Avr3a*$K^{80}I^{103}$ and Avr3$E^{80}M^{103}$. The *Avr3aKI* forms can induce resistance in the presence of R3a gene, whereas the *Avr3a*EM forms are pathogenic on plant harboring R3a gene. The *Avr3aEM* form has been shown to evolve under diversifying selection to avoid resistance (Bos et al., 2006; Martin et al., 2013). Our observation of KI and EM variants is consistent with previous reports (Cooke et al., 2012; Martin et al., 2013; J. Yin et al., 2017). Only the isolates of the US-1 lineage and 1306 were observed to be homozygous KI. The rest of the isolates were either EM homozygous or EM/KI heterozygous. This EM homozygous or EM/KI heterozygous isolates have shown to be virulent on plant with *R3a* gene (Cooke et al., 2012). Recently, the *P. infestans*

isolates from China have been shown to be EM homozygous (Yin et al. 2017) with only three nonsynonymous positions. In our study, the *Avr3a* gene from diverse samples showed more segregating sites than previous reports (J. Yin et al., 2017; Yoshida et al., 2013). Furthermore, the current study showed that BOC value for *Avr4* was less than 100 in most isolates, because the *Avr4* gene was probably truncated. Early stop codons leading to truncated protein have been described as a mechanism to overcome *R4* gene mediated resistance (van Poppel et al., 2008). The *Avrblb1* and *Avrblb2* genes showed an increase in the number of nucleotide haplotypes in *P. infestans* compared to other Avr genes.

High levels of genetic variation in RxLR genes play an important role in the evolution of a population. Genetic variation can be caused by a single nucleotide mutation (Raffaele & Kamoun, 2012). Furthermore, a single mutation can alter the structure of a protein which may result in loss of biological functions (Lodish et al., 2000). In this study, several RxLR genes showed changes in protein secondary structure that was caused by a single mutation. This is important because mutations that lead to pathogenesis have been reported to change the secondary structure of the protein (Abrusán & Marsh, 2016). Also, it was notable that mutations mostly occurred on the alpha helices, suggesting that this secondary structural element can tolerate more variation (SNPs) than other elements. Previous studies on human genome similarly showed that the alpha helices are more robust to SNPs (Abrusán & Marsh, 2016).

The oomycete RxLR effectors are known to have different expression patterns during infection of the host. Several RxLR effectors with Avr activities, including *Avr3a, Avr1d, Avr4* and *Avrblb* have been reported to be expressed early during infection (Haas et al., 2009; W. Yin et al., 2013). This study showed that the RxLR effectors *PITG_15038* and *PITG_22798* were

upregulated during the biotrophic phase of infection (24 hpi) by *P. infestans* in both cultivated and wild tomatoes. Similar to this study, Wang et al. (2017) found that the effector gene *PITG_22798* was expressed during early stages of potato infection. On the other hand, a few effector genes such as *PiNPP1.1* has been reported to be upregulated later during infection of tomato. Likewise, in the current study, two of the RxLR genes *PITG_12276* and *PITG_19994* showed expression at the necrotrophic phase of the infection in both *Solanum* species.

In conclusion, this study showed that the *P. infestans* isolates have retained the RxLR genes including Avr genes, suggesting their importance in host-pathogen interactions. In contrast, *P. ipomoeae* and *P. mirabilis* had significantly fewer RxLR and Avr genes. Furthermore, 220 core RxLR effectors were identified to be conserved among three clade 1c species. These core RxLR effectors can be used as key targets in developing late blight resistant crops. This study also showed that diversifying selection has operated on 11 RxLR genes. Knowing the evolutionary history gives us insights into how quickly a pathogen can evolve novel infection strategies. This knowledge, in turn, can help to shape plant breeding programs for long-term crop protection.

Table 4.1 Isolates of *P. infestans, P. mirabilis* and *P. ipomoeae* used in the study.

Species/Strain	Genotype	Accession no. †	Origin	Year	Reference
P. infestans					
1306_2	…	…	California	Unknown	Knaus et. (2016)
blue13	…	ERS226850	England	2006	Cooke et al. (2012)
P13527	…	ERS226844	Ecuador	2002	Yoshida et al. (2013)
P13626	…	ERS226845	Ecuador	2003	Yoshida et al. (2013)
NL07434	…	ERS226846	The Netherlands	2007	Yoshida et al. (2013)
DDR7602	US-1	ERS226848	Germany	1976	Yoshida et al. (2013)
LBUS5	US-1	ERS226849	South Africa	2005	Yoshida et al. (2013)
RS2009P1	US-8	ERS258000	Pennsylvania	2009	Martin et al. (2013)
US040009	US-8	…	New York	Unknown	B. Knaus et al. (2016)
FP-GCC	US-11	…	New York	Unknown	B. Knaus et al. (2016)
P10127	US-18	ERS241587	North Carolina	2002	Yoshida et al. (2013)
IN2009T1	US-22	ERS258001	Pennsylvania	2009	Martin et al. (2013)
US10006	US-22	…	Kentucky		B. Knaus et al. (2016)
P17777	US-22	ERS226847	New York	2009	Yoshida et al. (2013)
BL2009P4	US-23	ERS258002	Pennsylvania	2009	Martin et al. (2013)
FL2009P4	US-23	…	Kentucky	2009	B. Knaus et al. (2016)
ND822Pi	US-24	…	North Dakota	Unknown	B. Knaus et al. (2016)
Pic97146/inf2	…	…	Mexico	1997	Goss et al. (2014)
Pic97335/inf3	…	…	Mexico	1997	Goss et al. (2014)
Pic97750/inf5	…	…	Mexico	1997	Flier et al. (2002)
Pic97785/inf6	…	…	Mexico	1997	Goss et al. (2014)
P10650	…	ERS241584	Mexico	2004	Yoshida et al. (2013)
P. ipomoeae					
PIP-07-001/ipo1	…	…	Mexico	1999	Current study
PIP-07-003/ipo2	…	…	Mexico	1999	Current study
PIP-07-096/ipo4	…	…	Mexico	1999	Current study
PIP-07-097/ipo5	…	…	Mexico	1999	Current study
P. mirabilis					

PM-07-001/mir1	…	…	Mexico	1999	Current study
PM-07-099/mir4	…	…	Mexico	1999	Current study
PM-07-100/mir5	…	…	Mexico	1999	Current study
P7722	…	ERS241588	Mexico	1992	Yoshida et al. (2013)

† Whole-genome sequencing was conducted for the strains with no accession numbers.

Table 4.2 The universal and gene-specific primers used in reverse transcription polymerase chain reaction (RT-PCR).

Primers	Sequence 5´–3´	Reference
COX2-F	GGCAAATGGGTTTTCAAGATCC	Hudspeth, Nadler, and Hudspeth (2000)
COX2-R	CCATGATTAATACCACAAATTTCACTAC	Hudspeth et al. (2000)
12276-F	ACTCTTTGCAAGCGCCAAAG	Current study
12276-R	ATCTGTGCGACGACCCTTTT	Current study
15038-F	TCTTCTGGCCAATCCGCAAT	Current study
15038-R	CAGTCTGCATCCTCTTGGCA	Current study
19994-F	GGTGCGGATATGGTCTCCAG	Current study
19994-R	TGTCGTCTGCTGCGTTAAGT	Current study
22798-F	AACAAGTTAGCTGCGGTCGA	Current study
22798-R	GAGCTCGGATCCAGACCTTG	Current study
22990-F	GAGAGCTGGCCAAGGACTTT	Current study
22990-R	CACCTTTGGGGATGTACGCT	Current study

Table 4.3 Number of well characterized avirulence genes, nucleotide haplotypes and haplotype diversity observed in *P. infestans*, *P. mirabilis*, and *P. ipomoeae.*

	***P. infestans* (22)**[†]			***P. mirabilis* (4)**			***P. ipomoeae* (4)**		
Avr gene	Isolates[‡]	Haplotypes	HD[§]	Isolates	Haplotypes	HD	Isolates	Haplotypes	HD
Avr1	5	5	0.67	…[¶]	…	…	4	4	0.78
Avr2	22	2	0.13	1	2	1	4	2	0.43
Avr3a	22	5	0.65	…	…	…	…	…	…
Avr4	21	8	0.69	…	…	…	…	…	…
Avrblb1	22	13	0.87	1	2	1	…	…	…
Avrblb2	17	13	0.69	4	1	0	…	…	…
AvrSmira1	22	5	0.73	4	8	1	4	8	1
AvrSmira2	22	12	0.57	3	5	0.93	…	…	…
Avrvnt1	22	8	0.67	4	4	0.86	4	4	0.78

[†] Sample size is indicated in parentheses.

[‡] Number of isolates carrying the avirulence gene.

[§] Haplotype diversity.

[¶] Genes inferred as missing using BOC of less than 0.90.

Table 4.4 Analysis of Tajima's D value for the avirulence genes using whole gene sequence and C-terminal sequence for all three *Phytophthora* species.

Avr genes	**Whole gene sequence**			**C-terminal sequence**		
	Pinf_D	Pmir_D	Pipo_D	Pinf_D	Pmir_D	Pipo_D
Avr1	-0.51	…[†]	0.42	-0.56	…	0.38
Avr2	-0.38	…	0.33	…	…	0.33
Avr3a	0.82	…	…	0.59	…	…
Avr4	0.02	…	…	0.85	…	…
Avrblb1	1.08	…	…	1.33	…	…
Avrblb2	-0.25	…	…	-0.11	…	…
AvrSmira1	2.61[‡]	0.63	0.67	2.43*	1.02	0.67
AvrSmira2	0.72	-0.82	…	0.86	-0.67	…
Avrvnt1	0.83	1.79	-0.3	0.72	1.44	-0.3

[†] Tajima's D was not calculated for the genes that were missing (BOC value < 0.90) or did not have enough sequences.

[‡] Tajima's D value was significantly different from zero ($P < 0.05$).

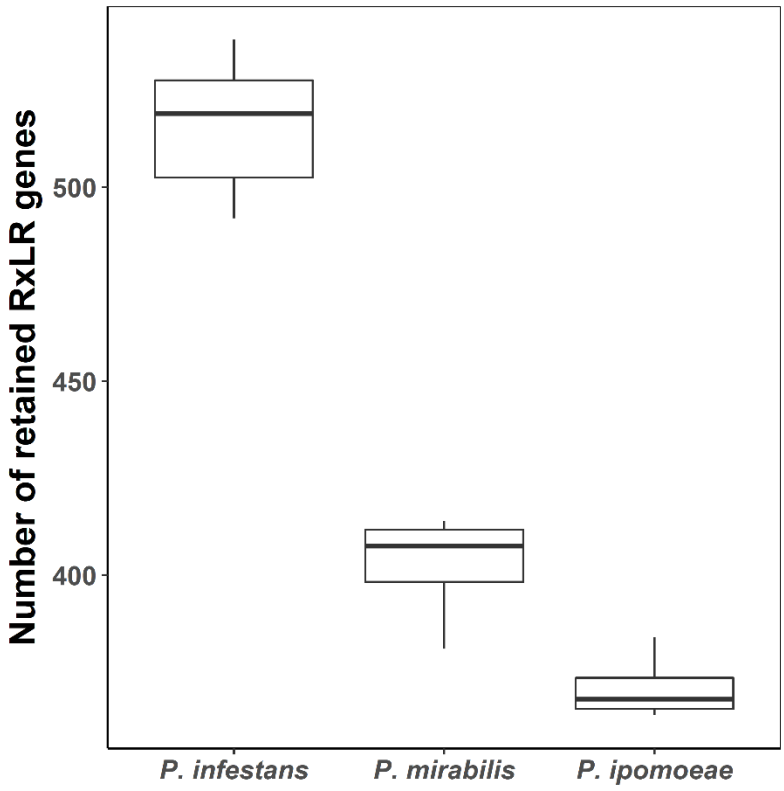

Figure 4.1 Number of RxLR genes detected in three *Phytophthora* species relative to *P. infestans* reference genome T30-4. Breadth of coverage (BOC) cutoff value was set at 0.90.

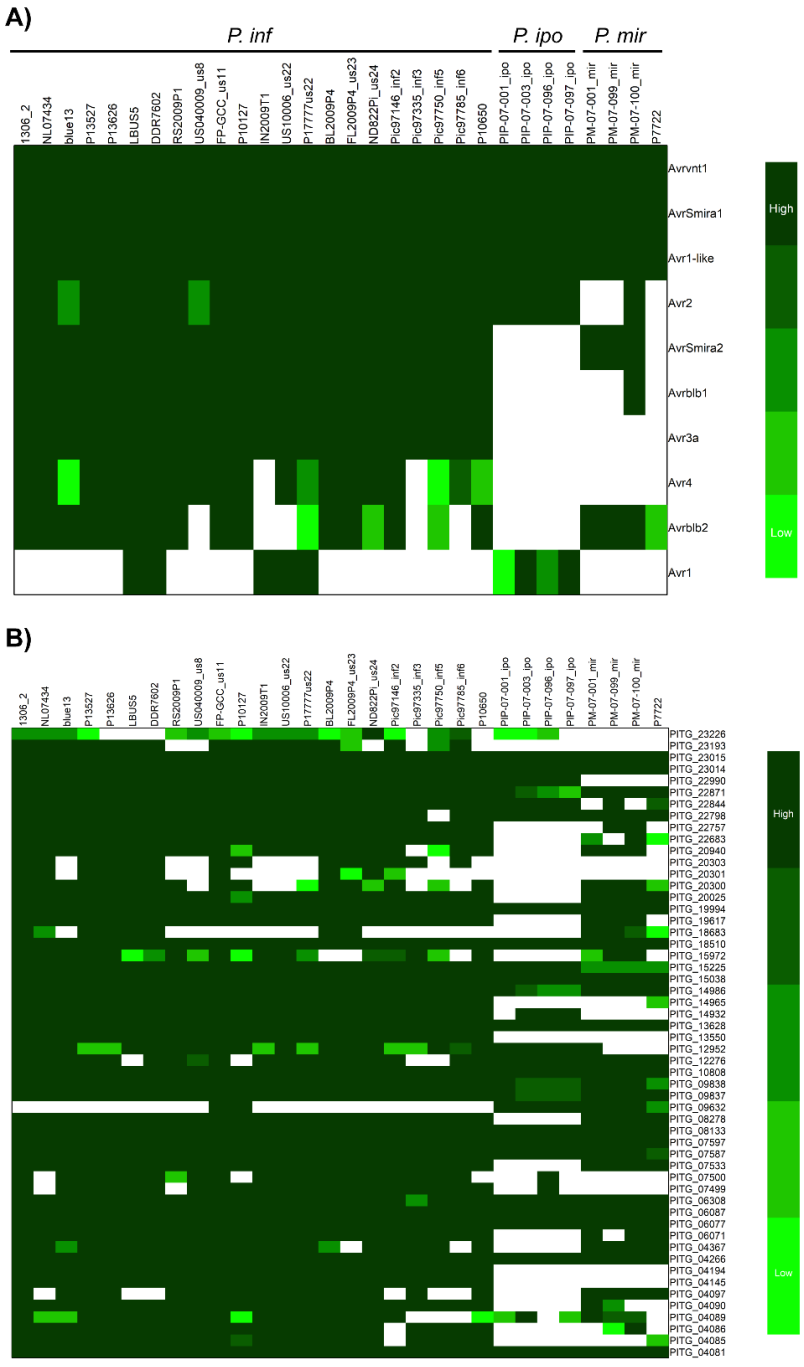

Figure 4.2 Breadth of coverage for A) 10 avirulence and B) 47 NLS-RxLR genes in *Phytophthora* clade 1c genomes. The RxLR genes are indicated in rows and *Phytophthora* genomes are shown in columns. *Phytophthora infestans*, *P. ipomoeae* and *P. mirabilis* are indicated as *P. inf*, *P. ipo* and *P. mir*, respectively. Color indicates BOC value ranging from 0.90 (light green) to 1 (dark green). BOC value of less than 0.90 is indicated in white and, thus treated as missing.

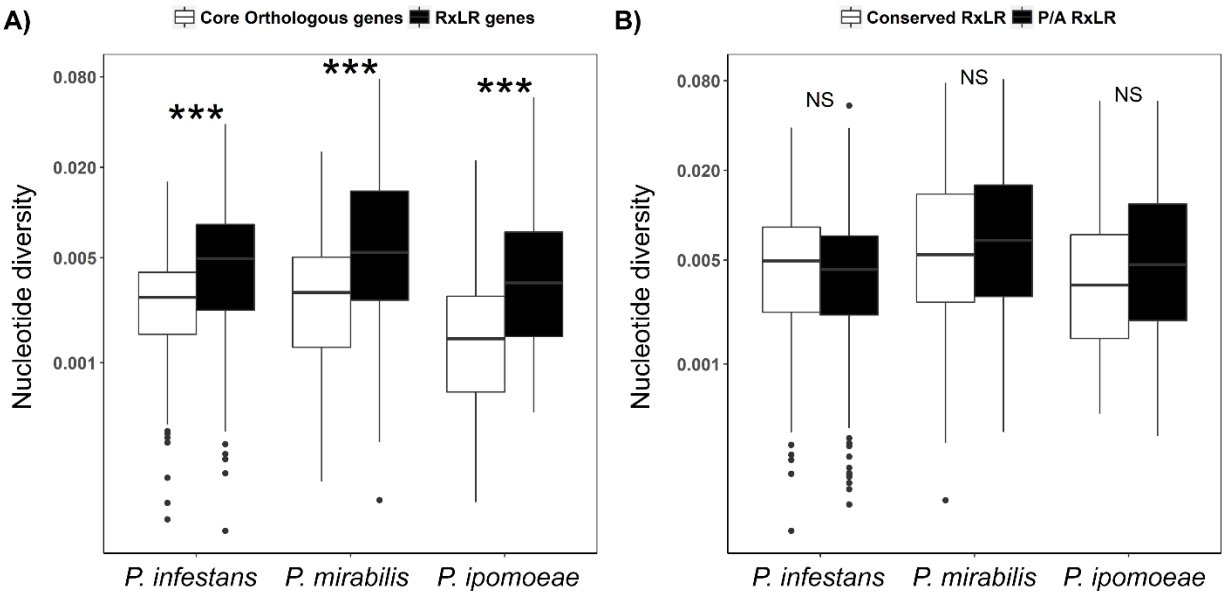

Figure 4.3 Nucleotide diversity observed at core orthologous and RxLR genes in three *Phytophthora* species. A) Nucleotide diversity of 218 conserved RxLR genes compared with randomly selected 218 core orthologous genes for three *Phytophthora* species. B) Nucleotide diversity at conserved RxLR genes compared to RxLR genes with presence/absence polymorphism. Statistical differences were tested using Kruskal-Wallis test.

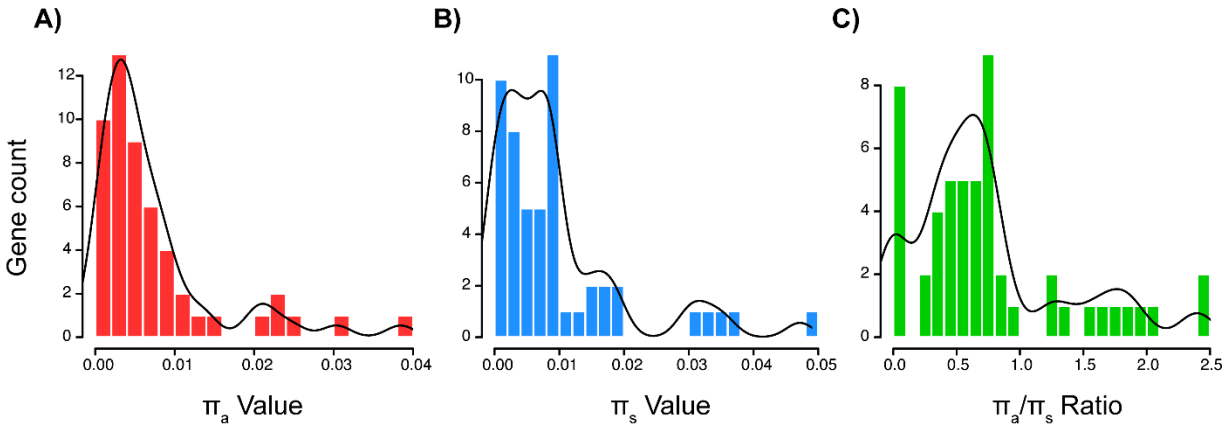

Figure 4.4 Frequency distribution of sequence variation across RxLR genes: A) nonsynonymous substitutions (π_a), B) synonymous substitutions (π_s) and C) the ratio (π_a/π_s) of nonsynonymous substitutions to synonymous substitutions for the RxLR genes. A ratio of π_a/π_s less than 1 indicates purifying selection ($\pi_a<\pi_s$), a ratio of 1 indicates neutral selection and a ratio above 1 indicates balancing selection ($\pi_a>\pi_s$).

Figure 4.5 Predicted changes in secondary structure of the NLS-RxLR gene *PITG_22798* caused by a single nucleotide polymorphism (SNP) in the C-terminal domain. The location of polymorphisms are indicated by a star. The reference genome for *Phytophthora infestans* is labeled T30. The two haplotypes for each isolate are labeled 0 and 1.

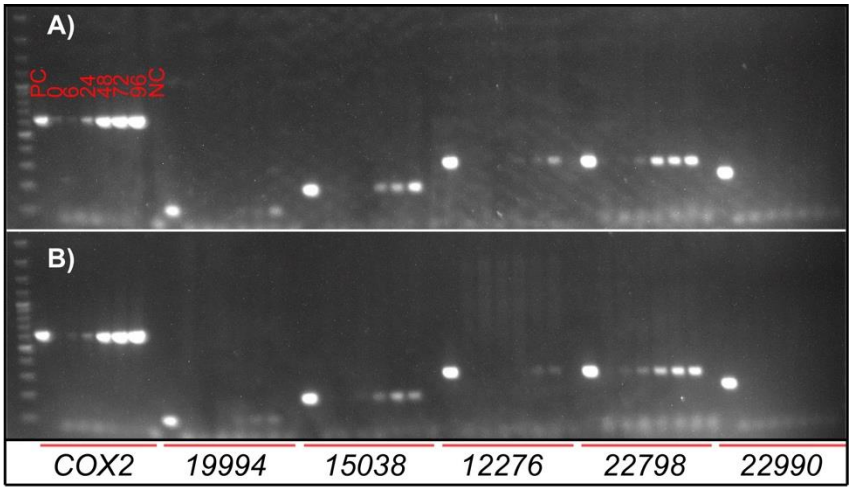

Figure 4.6 Induction of NLS-RxLR genes during infection of *P. infestans* D12-2 on *Solanum* species. Reverse transcription polymerase chain reaction (RT-PCR) was conducted for five NLS-RxLR genes, as well as the COX2 region on A) *S. lycopersicum* and B) *S. pimpinellifolium*. The plants were inoculated with *P. infestans* D12-2 and samples were taken at six time points (0 hpi, 6hpi, 24 hpi, 48 hpi, 72 hpi and 96 hpi). Samples were loaded into the gel from left to right in the following order: positive control (PC), 0 hpi, 6 hpi, 24 hpi, 48 hpi, 72 hpi, 96 hpi and negative control. A 100 bp DNA ladder (L) was used for the size of the fragments on a gel, during electrophoresis.

Chapter 5

Reference bias in *Phytophthora infestans* genomes

Shankar K. Shakya, Brian J. Knaus and Niklaus J. Grünwald

Target journal: MPMI

5.1 Abstract

Sequencing reads that fail to map against a reference genome are often ignored in whole genome sequencing studies. There is growing evidence that the reads which fail to map to a reference genome might provide insight into biological functions. Thus, we hypothesized that unmapped reads represented critical aspects of the uniqueness of a strain and could be assembled to predict strain specific genes. We first mapped the short reads from 22 *Phytophthora infestans* isolates to the *P. infestans* reference strain T30-4 using bowtie2 and filtered the unmapped reads. The number of reads that did not map to the reference strain ranged from 3-58%. We then performed *de novo* assembly with the velvet assembler using those unmapped reads for each isolate and predicted the gene content using AUGUSTUS. Assembly size ranged from 1.43 - 16 Mbp and the gene content ranged from 394-3938. Orthofinder was used to identify orthologous gene clusters along with strain specific genes. We were able to identify 62 clusters with genes present from all 22 *P. infestans* isolates and a total of 1,554 singletons. These newly identified genes had orthologs in other *Phytophthora* species which suggests that the unmapped reads are not just contaminations and could potentially be used to improve the reference genome.

5.2 Introduction

Use of whole genome sequence data has improved our understanding of the evolution of genome architecture (Lander et al. 2001; Myers et al. 2000; Bevan and Walsh 2005; Haas et al. 2009). Whole genome sequence data can be analyzed and summarized in two major ways. First, if the genome information is lacking for the organism of interest, then sequences at higher coverage can be assembled to produce a reference genome (Lander et al. 2001; Myers et al.

2000; Bevan and Walsh 2005; Haas et al. 2009). Second, if the reference genome is available then the resequencing of isolates can be done at lower coverage in order to reduce the cost and then short reads can be mapped to the reference genome to identify genomic variants. However, during the process of mapping short reads to the reference genome some percentage of reads fail to map (Whitacre et al. 2015; Weide et al. 2016; Gouin et al. 2015; Faber-Hammond and Brown 2016b; Singla 2016; Zarif Saffari 2012; Faber-Hammond and Brown 2016a). The degree of mapping percentage can vary depending on genetic relatedness of the reference strain and the re-sequenced strain. This results in a reference bias due to various factors: the reference genome is not fully assembled; variants in resequenced samples fail to map because of error divergence from the reference or indels. Reads that fail to map against the reference genome due to this bias are often neglected for various reasons. However, there is growing evidence that these unmapped reads might provide insight into novel variants and may contribute to our understanding of gene content in an individual (Whitacre et al. 2015; Weide et al. 2016; Gouin et al. 2015; Faber-Hammond and Brown 2016b; Singla 2016; Zarif Saffari 2012; Faber-Hammond and Brown 2016a; Liu et al. 2014).

Assembly of reads that fail to map against the reference genome can provide critical new insights into genome biology. Numerous papers discussing the significance of unmapped reads in different species including aphids, Zebrafish, dogs, rats, rainbow trout and humans have been published (Whitacre et al. 2015; Weide et al. 2016; Gouin et al. 2015; Faber-Hammond and Brown 2016b; Singla 2016; Faber-Hammond and Brown 2016a; Liu et al. 2014). For example, Sherman et al. (2018) has shown that the presence of more genetic sequences by deep sequencing and assembling the unmapped reads from African human population compared to the reference human genome (GRCh38) with the higher aim of creating a human pan-genome.

Similarly, assembly of unmapped reads from dogs was able to close 12,503 gaps in the canine reference genome (Holden et al. 2018). Arniella et al. (2018) identified novel genes in rainbow trout by assembling the unmapped reads which were then shown to have a role in magnetoreception. Additionally, potential contaminants were also identified but at a low frequency. All of this work points towards the common theme of reference bias, identification of novel genes and suggests the possibility of improvement of the reference genome.

The oomycete plant pathogen *Phytophthora infestans* causes potato and tomato late blight. The reference genome of *P. infestans* strain T30-4 revealed extensive repetitive regions (74 %) and a large genome size compared to other *Phytophthora* species (Haas et al. 2009). Although the reference genome has been used for numerous resequencing studies, the importance of unmapped reads have not been evaluated (Yoshida et al. 2013; Martin et al. 2013; Knaus et al. 2019; Cooke et al. 2012). Only two studies thus far have assembled unmapped reads from *P. infestans* (Martin et al. 2013; Cooke et al. 2012). Cooke et al. (2012) first assembled the unmapped reads from the 06_392A strain (aka "blue13") of *P. infestans* to produce a 2.77 Mbp *de novo* assembly and identified 6 candidate RxLR effector genes absent from the T30-4 reference strain. Another study by Martin et al. (2013) also performed the *de novo* assembly of unmapped reads of historic and modern lineages of *P. infestans* to produce a total assembly of 4-5 Mbp for each modern isolates and 110-134 Mbp assembly for each herbarium isolates from the 1800s. Only 5 novel RxLR genes were identified from modern *P. infestans* isolates and none from the herbarium samples. There are two limitations to the studies by Cooke et al. (2012) and Martin et al. (2013). First, both studies focused only on identifying RxLR genes from the *de novo* assembly of unmapped reads and ignored other gene families. Second, in both studies isolates belonging to clonally reproducing lineages were analyzed. *P. infestans* is known to

reproduce sexually in Mexico at its center of origin and clonally reproduce in most of the rest of the world (Grünwald and Flier 2005; Goss et al. 2014; Shakya et al. 2018). Given that more genetic variation is observed in sexual populations we might expect reference bias to be higher for sexually reproducing *P. infestans* populations compared to clonal lineages.

In this study, we aimed to quantify the magnitude of reference bias in sexually and clonally reproducing *P. infestans* by quantifying the percentage of reads that fail to map against the reference genome. We hypothesized that sexually reproducing individuals are genetically diverse compare to clonally reproducing isolates thus resulting in increased reference bias. We also aimed to identify genes missing from the reference T30-4 isolate and genes that are specific to each isolate. Identification of a core set of genes from the *de novo* assembly of unmapped reads can improve the reference genome in terms of gene content. Our final aim was to identify novel RxLR and CRN genes which plays role in the process of plant colonization.

5.3 Materials and Methods

5.3.1 *P. infestans* isolates, filtering unmapped reads and *de novo* assembly

Whole genome sequences of *P. infestans* isolates were downloaded from NCBI (Table 5.1) Reads from each *P. infestans* isolate were first mapped to the *P. infestans* reference genome T30-4 using bowtie2 v 2.2.3 (Langmead and Salzberg 2012). Unmapped reads were filtered from the SAM (sequence alignment map) file using samtools flag -f 4 (Li et al. 2009). The short-read assembler velvet was used to assemble the filtered reads that didn't map to the reference genome (Zerbino and Birney 2008). The velvet assembler provides flexibility in kmer length selection and allows filtering of contigs by size. Velvetoptimiser was used to identify the

optimum kmer value for assembly. Minimum contig length was set to 500bp for the assembly. Assembly statistics are presented in Table 5.2.

5.3.2 Gene prediction using *de novo* assembly

Gene prediction was done using the *ab initio* gene prediction tool, AUGUSTUS (Stanke and Morgenstern 2005). AUGUSTUS is a gene prediction tool for eukaryotes and uses a generalized hidden Markov model. AUGUSTUS software has previously defined parameters for model organisms but not for *P. infestans*. Therefore, we first trained the software using transcript data from reference *P. infestans* strain to improve gene calling. Only complete genes were called using AUGUSTUS by specifying the parameter "complete genes". The number of predicted genes are reported in Table 5.2.

5.3.3 Orthologous gene clustering

Orthofinder was used to identify homologous gene clusters (Emms and Kelly 2015). Orthofinder uses all-versus-all blast and Markov Clustering algorithm to identify orthologous gene clusters. The number of genes in each orthologous cluster and the sample specific genes are placed in separate files which makes it easier for visualization. The number of genes in each orthologous cluster and strain specific genes (singletons) were plotted using the R package *ggplot* (Wickham 2009). Gene duplication events were also identified using Orthofinder.

5.3.4 Identification of candidate RxLR and CRN genes

RxLR and CRN genes are well studied in *Phytophthora* and known to be involved in pathogenicity. We used regular expressions to search for candidate RxLR and CRN genes using

the R package *effectR* (Tabima and Grünwald 2019). Newly identified RxLR and CRN gene lengths were computed and compared against gene lengths observed in the T30-4 reference genome. Candidate RxLR and CRN genes were tested for the presence of encoded signal peptides using SignalP 5.0 (Armenteros et al. 2019). A second round of orthologous clustering was performed only using the newly identified RxLR genes to identify RxLR genes unique to each strain of *P. infestans.*

5.3.5 Homologs identification

To identify the potential homologs, the longest gene from each orthologous group was selected and grouped to produce a single fasta file of the predicted protein. The AAI profiler was used to identify homologs (Medlar et al. 2018). The AAI-profiler is a homology identification tool that computes the amino acid identity (AAI) between the query proteome against the Uniprot database. The amino acid identity cutoff value was set to 50% to reduce potential biases due to short matches. The resulting taxonomic information was visualized using the R package *metacoder* (Foster et al. 2017).

5.4 Results

5.4.1 Variation in the percentage of unmapped reads, assembly size and gene content

Large variation (3 to 58) in the percentage of unmapped reads was observed among the 22 individuals of *P. infestans* studied here. The *P. infestans* isolate from the Netherlands (NL07434) had the smallest percentage (3.61%) of unmapped reads whereas the isolate PIC97146 from Mexico had the highest percentage (58.09%) of unmapped reads. The majority of the isolates had unmapped reads less than 20%, and only 6 isolates had unmapped reads

greater than 20 % out of which 4 were from Mexico, and the other two belonged to the US-8 clonal and US-18 clonal lineages (Fig 5.1A). The assembly size using the unmapped reads ranged from 1.43 Mbp to 16 Mbp. Most of the assemblies were slightly smaller than 5 Mb; however, three isolates originating from Mexico had assembly size greater than 10 Mbp (Fig 5.1B). Contigs N50 was also highest for those three isolates from Mexico (Fig 5.1C). For majority of the isolates, the number of predicted genes were slightly less than 1000 except for three isolates from Mexico (Fig 5.1D).

5.4.2 Identification of orthologous gene cluster

Orthofinder was used to identify orthologous gene clusters. A total of 21,465 *ab initio* predicted genes were used to identify orthologous gene clusters. Out of 21,465 genes only 19,911 genes were classified into 2766 orthologous gene clusters. Orthologous clusters were sorted by the number of gene content and plotted. First orthologous gene cluster had the maximum of 218 genes. Three isolates from Mexico (PIC97785, PIC97335 and PIC97146) which had more predicted genes, had genes occurring in almost 2000 orthologous gene clusters (Fig 5.2A). Genes that could not be classified into any orthologous cluster, known as singletons, were identified for each sample and plotted (Fig 5.2B). Two isolates ND822Pi (US-24 lineage) and DDR7602 (US-1 lineage) had no singleton genes whereas isolate PIC97146 had the highest number of singleton genes followed by the 1306_2 isolate (Fig 5.2B). PIC97146 had the highest number of gene duplication events compared to other isolates.

5.4.3 Candidate RxLR and CRN genes prediction

Candidate RxLR and CRN genes were first identified based on the presence of RxLR and LFLAK motifs respectively (Fig 5.3). Isolate PIC97750 had the highest number of RxLR genes followed by isolate RS2009P1 (US-8 lineage). Three isolates (ND822Pi, FP-GCC, and PIC97335) had no RxLR genes from the assembly of unmapped reads (Fig 5.3A). Abundance of CRN genes was low compared to RxLR genes. Isolate 06_3928A (blue13) had the highest number (4) of CRN genes. Candidate RxLR and CRN genes were also tested for the presence of an encoded signal peptide which is involved in secretion of the effector. The majority of candidate CRN genes, based on the LFLAK motif, did not have aencode signal peptides (Fig 5.3B). Only four isolates (P17777, 06_3928A, PIC97785 and PIC97146) had a CRN gene with an encoded signal peptide. Newly predicted RxLR and CRN gene length distributions were compared with RxLR and CRN gene length distribution from the T30-4 reference genome (Fig 5.4). Clustering of RxLR genes into orthologous groups identified shared and unique RxLR genes from the assembly of unmapped reads. Nine isolates had at least one unique RxLR gene whereas isolate PIC97146 and PIC97750 had 3 and 2 unique RxLR genes respectively.

5.4.4 Genes from the assembly of unmapped reads have homologs in other *Phytophthora* species and bacteria

AAI-profiler was used to identify homologs which uses a query proteome to search against the Uniprot database. The majority of genes had homologs in other closely related *Phytophthora* species (Fig 5.5A). These genes were primarily annotated as uncharacterized proteins in *Phytophthora* but also included a few carbohydrate degrading enzymes and necrosis

inducing proteins (NPP). Some of the genes had homologs in bacteria primarily *Paenibacillus* species (Fig 5.5B). These included MFS and ABC transporter genes and transcriptional regulator genes (AraC, LysR, GntR, LacI). A very similar pattern was revealed with singleton genes which had homologs in *Phytophthora* and *Paenibacillus* species (Fig 5.6).

5.5 Discussion

Here, we demonstrated the phenomenon of reference bias in the late blight pathogen *P. infestans.* Reference bias is observed because the reference genome fails to capture the genomic diversity at the population level (Whitacre et al. 2015; Weide et al. 2016; Gouin et al. 2015; Faber-Hammond and Brown 2016b; Singla 2016; Zarif Saffari 2012; Faber-Hammond and Brown 2016a; Liu et al. 2014). Quantifying the percentage of reads that fail to map to the reference genome is one way of measuring the reference bias. Our work on *P. infestans* shows this bias can range from 3 to 60 percent of reads. We then assembled and predicted genes using these unmapped reads and identified genes that were not represented in the reference genome. These newly identified genes have homologs in other *Phytophthora* species and a bacterial genus, *Paenibacillus*.

Cooke et al. (2012) and Martin et al. (2013) previously assembled unmapped reads from a few isolates of *P. infestans*. An assembly size of 3-5 Mbp was observed in both studies. We observed very similar assembly sizes for most of our isolates with few exceptions. For the majority of the isolates, the total assembly size was below 8 Mbp; however, three isolates (PIC97785, PIC97335, and PIC97146) had assembly sizes larger than 10 Mbp. The majority of the isolates with an assembly size less than 8 Mbp are usually the ones which belong to clonal lineages. Assembly sizes reported by Cooke et al. (2012) and Martin et al. (2013) were based on

clonally reproducing isolates. The three samples (PIC97785, PIC97335, and PIC97146) with the highest percentage of unmapped reads and assembly sizes larger than 10 Mbp were isolated from central Mexico. Central Mexico is the center of origin of *P. infestans* where it reproduces sexually (Goss et al. 2014; Shakya et al. 2018). This high percentage of unmapped reads and assembly sizes for these three isolates is not unusual and probably represent the high genetic diversity of *P. infestans* found in central Mexico. A similar result was observed by Sherman et al. (2018) who identified almost 10 % more DNA content in the African human population compared to the reference human genome.

The number of genes predicted from the assembly of unmapped reads were slightly less than 1,000 for the majority of the isolates. Increased gene content was observed for the same three isolates which had assembly sizes greater than 10 Mbp. Orthologous and paralogous clustering of the genes revealed that increased gene duplication events were responsible for the increased gene count. Copy number variation and change in ploidy has been observed for *P. infestans* previously (Li et al. 2017; Knaus et al. 2019). Very few pathogenicity genes (RxLR and CRN) were identified from the *de novo* assembly similar to what has been reported by Cooke et al. (2012) and Martin et al. (2013). Most of the candidate RxLR genes showed the presence of and encoded signal peptide but this was not true for candidate CRN genes. Crinkler (CRN) genes are known to be difficult to identify the encoded signal peptide (Stam et al. 2013). The presence of a signal peptide, RxLR or LFLAK motif and an appropriate gene length suggests a functional pathogenicity gene.

The majority of the newly predicted genes have homologs in other *Phytophthora* species and *Paenibacillus* bacteria. The genus *Phytophthora* is known to harbor numerous pathogenicity

genes encoding RxLRs, crinklers and carbohydrate degrading enzymes (Haas et al. 2009; Adhikari et al. 2013; Stam et al. 2013; Cooke et al. 2012). The presence of these genes in the assembly of unmapped reads in not surprising given the abundance of these genes in the genome. Similarly, homologs of genes in *Paenibacillus* species could be because of horizontal gene transfer (HGT) from bacteria to *Phytophthora*. HGT has been primarily described in prokaryotes but there is evidence of horizontal gene transfer between bacteria and the oomycete pathogen, *Phytophthora* (McCarthy and Fitzpatrick 2016; Levy et al. 2018).

Here, we demonstrate the phenomenon of reference bias with the genome of the potato late blight pathogen, *Phytophthora infestans*. This bias is primarily due to the singularity of the reference genome which fails to capture the genomic diversity of a population. Assembling the reads which fail to map to the reference genome could be one of the ways to correct for this reference bias and *de novo* gene prediction can provide insight into novel gene content for each individual. For a plant pathogen like *P. infestans* which can reproduce both sexually and asexually, the reference bias varied depending on its mode of reproduction. *P. infestans* isolates originating from Mexico had a significantly greater percentage of unmapped reads compared to clonal lineages. This was expected given the high genetic diversity of isolates found in Mexico. Very few novel RxLR and CRN genes were identified from the assembly of unmapped reads suggesting that the reference *P. infestans* isolate covers a majority of RxLR and CRN genes.

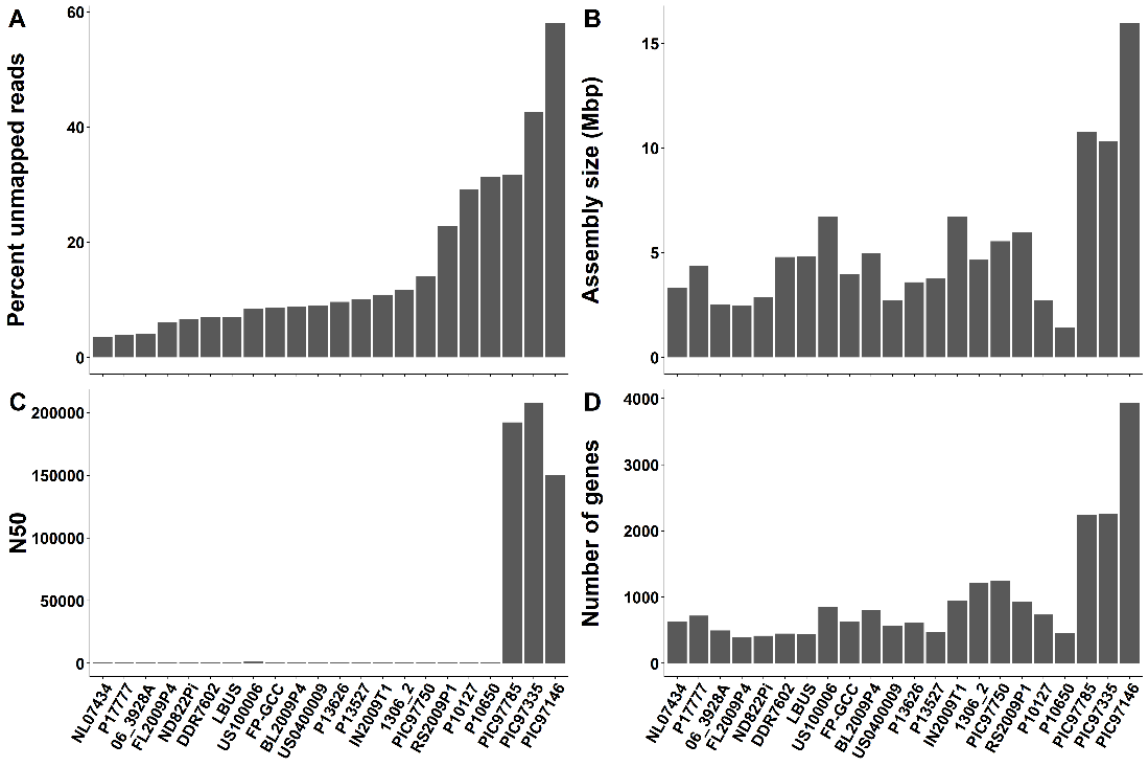

Figure 5.1. Analyses of unmapped reads (reads that fail to map against the reference genome *P. infestans* T30-4 strain) of *P. infestans* isolates. **A.** Percentage reads that failed to map to the reference genome. **B.** *de novo* assembly size using unmapped reads. **C.** Assembly N50. **D.** Number of predicted genes from *de novo* assembly of unmapped reads. X-axis in the above graph represents each isolate.

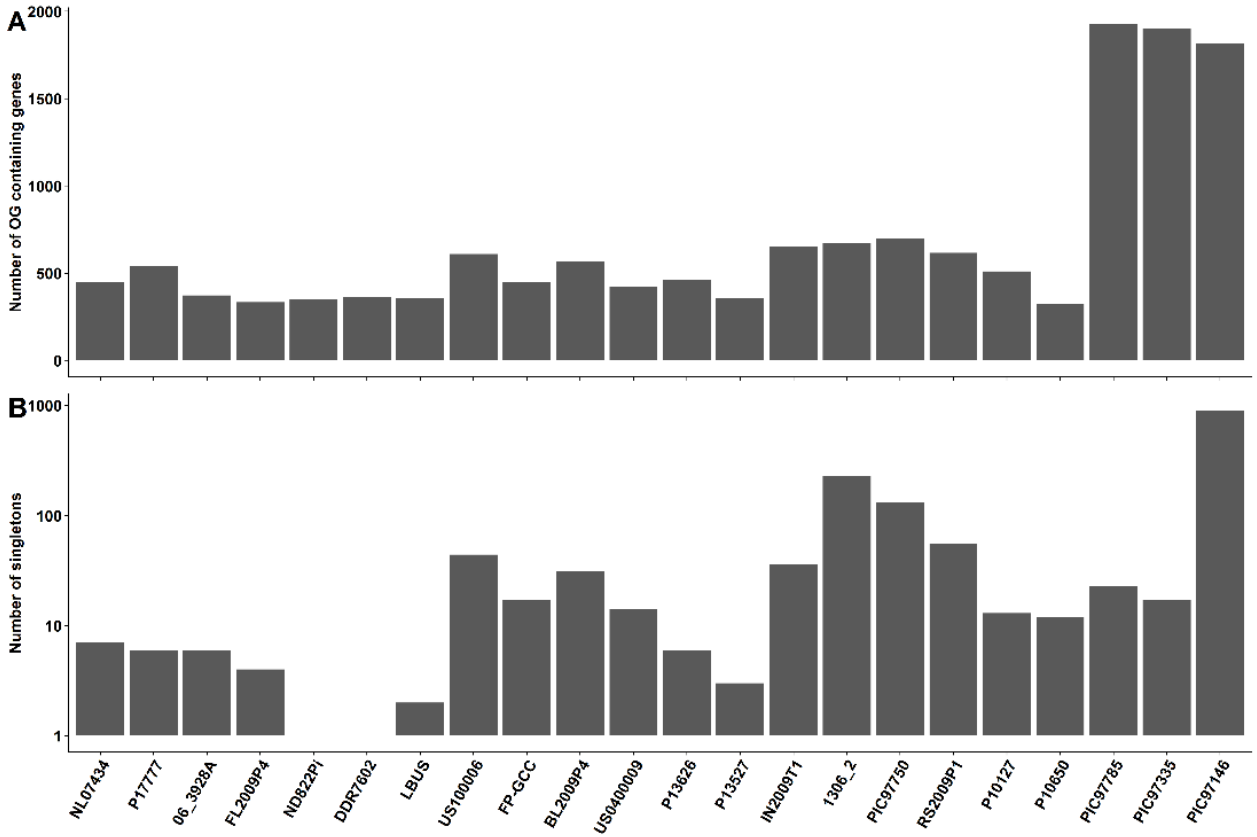

Figure 5.2. Orthologous gene clusters and number of singleton genes. **A.** Number of orthologous groups containing genes from each *P. infestans* isolate. **B.** Number of singletons (genes found only in that strain).

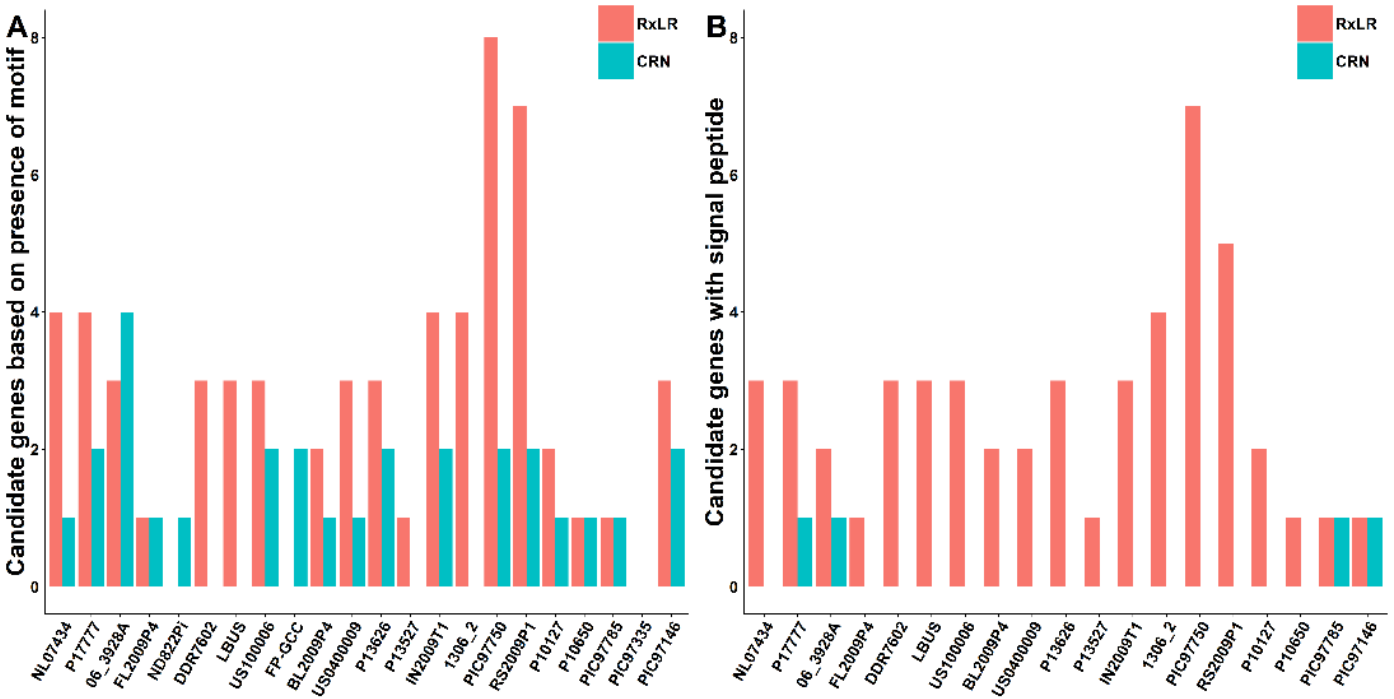

Figure 5.3 Predicted candidate RxLR (red) and CRN (blue) genes based on presence of the RxLR or LFLAK motif respectively (A) and also presence of signal peptide (B). Signal peptide was predicted using SignalP (Armenteros et al. 2019).

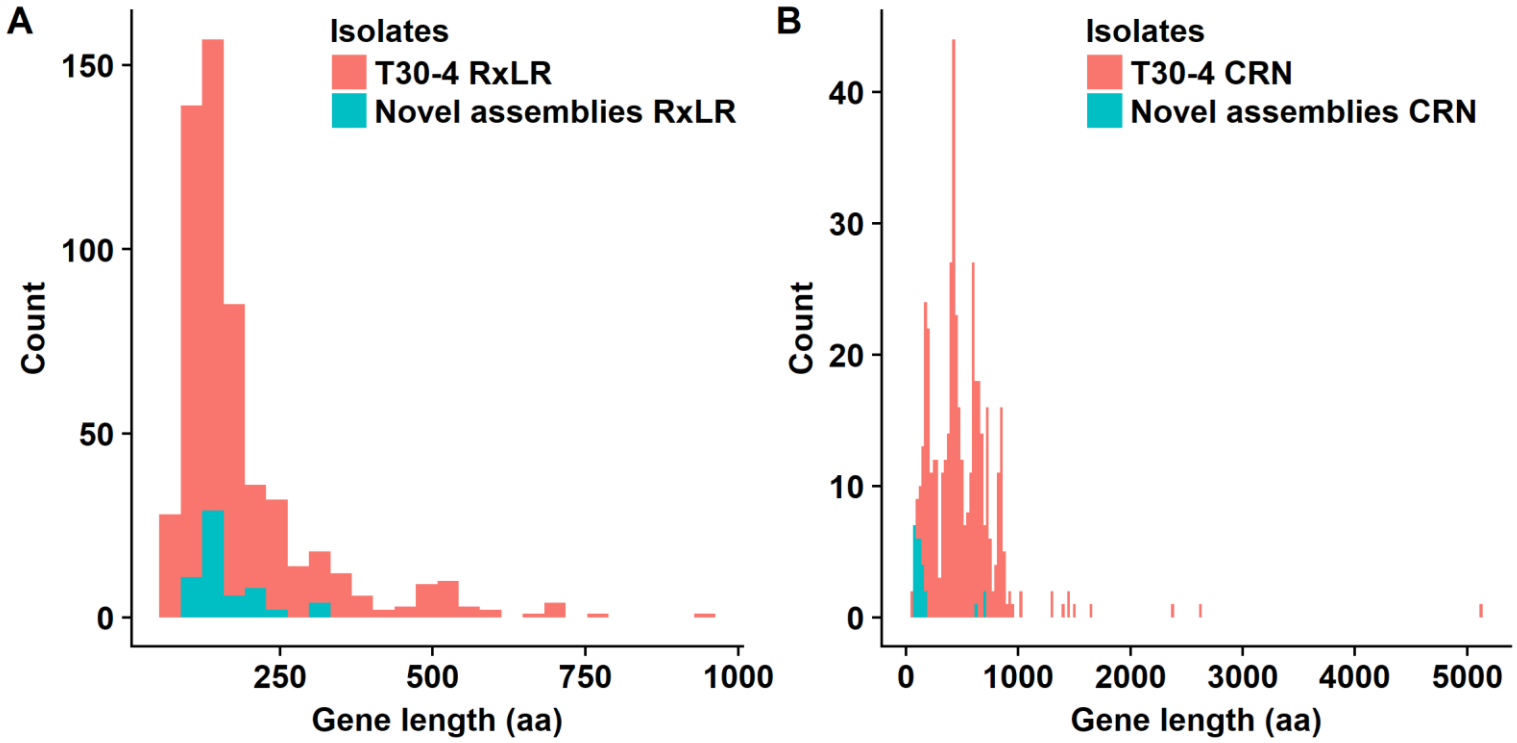

Figure 5.4 Distribution of amino acid length of predicted (A) RxLR and (B) CRN genes and comparison with the *P. infestans* reference strain T30-4 (red).

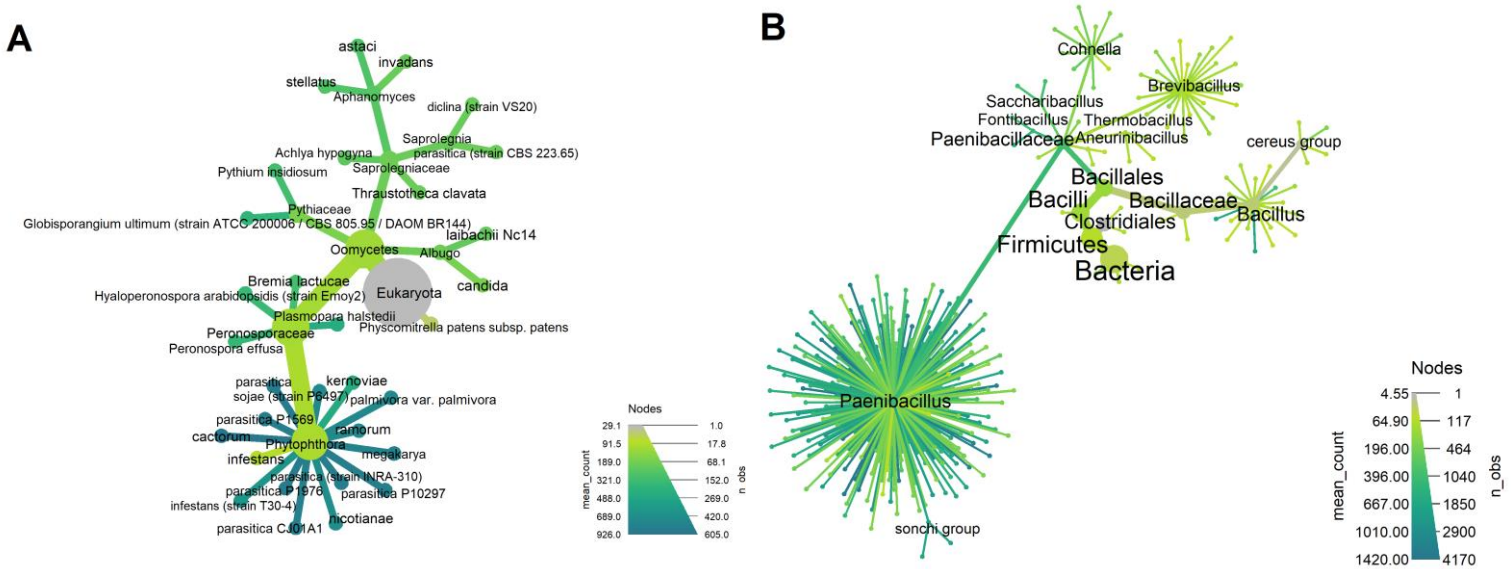

Figure 5.5 Genes identified from the assembly of unmapped reads have homologs in (A) *Phytophthora* and (B) *Paenibacillus* species.

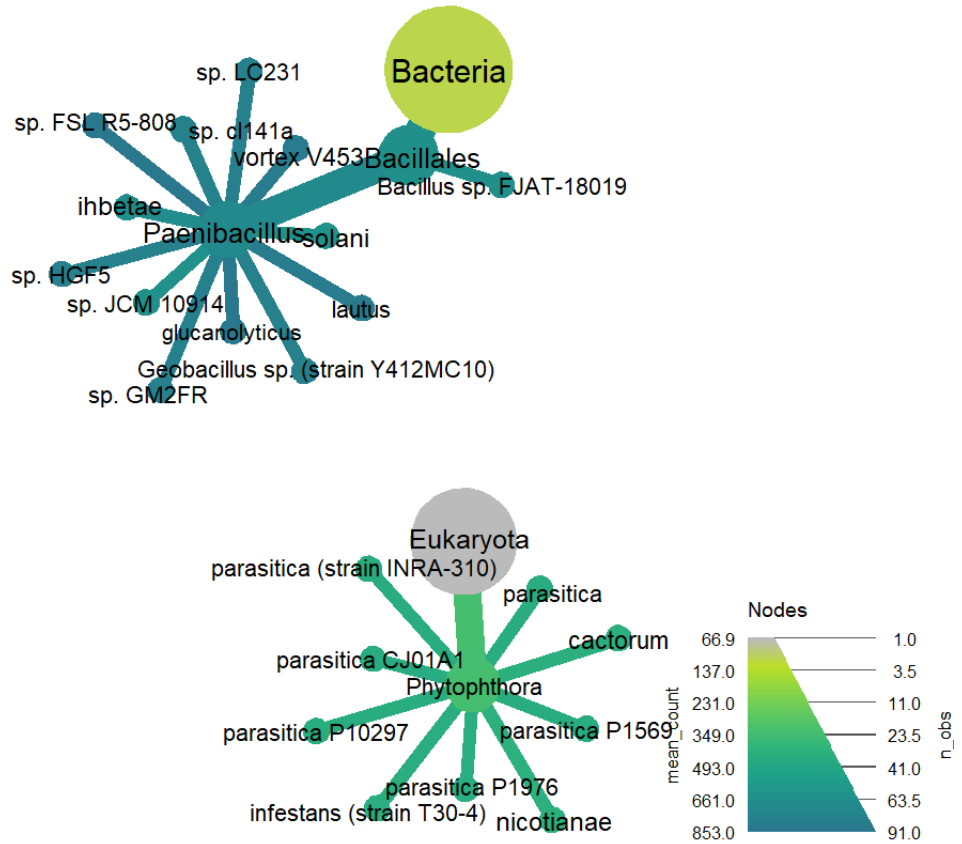

Figure 5.6 Singleton genes identified from the assembly of unmapped reads have homologs in *Phytophthora* and *Paenibacillus* species.

Table 5.1 Isolates of *P. infestans* used in the study.

Isolate	**Genotype**	**Accession no.**	**Origin**	**Year**	**Reference**
1306_2	…	…	California	Unknown	
blue13	…	ERS226850	England	2006	Cooke et al. (2012)
P13527	…	ERS226844	Ecuador	2002	Yoshida et al. (2013)
P13626	…	ERS226845	Ecuador	2003	Yoshida et al. (2013)
NL07434	…	ERS226846	The Netherlands	2007	Yoshida et al. (2013)
DDR7602	US-1	ERS226848	Germany	1976	Yoshida et al. (2013)
LBUS5	US-1	ERS226849	South Africa	2005	Yoshida et al. (2013)
RS2009P1	US-8	ERS258000	Pennsylvania	2009	Martin et al. (2013)
US040009	US-8	PRJNA542680	New York	Unknown	Knaus et al. (2019)
FP-GCC	US-11	PRJNA542680	New York	Unknown	Knaus et al. (2019)
P10127	US-18	ERS241587	North Carolina	2002	Yoshida et al. (2013)
IN2009T1	US-22	ERS258001	Pennsylvania	2009	Martin et al. (2013)
US10006	US-22	PRJNA542680	Kentucky		Knaus et al. (2019)
P17777	US-22	ERS226847	New York	2009	Yoshida et al. (2013)
BL2009P4	US-23	ERS258002	Pennsylvania	2009	Martin et al. (2013)
FL2009P4	US-23	PRJNA542680	Kentucky	2009	Knaus et al. (2019)
ND822Pi	US-24	PRJNA542680	North Dakota	Unknown	Knaus et al. (2019)
Pic97146/inf2	…	PRJNA542680	Mexico	1997	Knaus et al. (2019)
Pic97335/inf3	…	PRJNA542680	Mexico	1997	Knaus et al. (2019)
Pic97750/inf5	…	PRJNA542680	Mexico	1997	Knaus et al. (2019)
Pic97785/inf6	…	PRJNA542680	Mexico	1997	Knaus et al. (2019)
P10650	…	ERS241584	Mexico	2004	Yoshida et al. (2013)

Table 5.2 Summary table of assembly statistics for *P. infestans* reads that didn't map to the reference strain.

Isolate	Genotype	% unmapped reads	kmer	Assembly size (Mbp)	Number of contigs	N50	Largest contig (bp)	Gene count
1306_2		11.72	35	4.65	5483	817	9701	1216
06_3928A		4.15	49	2.52	2649	980	8708	498
P13527		10.03	49	3.75	4310	855	7097	474
P13626		9.64	47	3.59	3859	934	8173	611
NL07434		3.61	53	3.31	3117	1124	28792	632
DDR7602	US-1	6.93	49	4.79	5430	872	8615	448
LBUS	US-1	6.93	49	4.80	5441	869	8170	435
RS2009P1	US-8	22.83	51	5.98	6630	858	13237	934
US0400009	US-8	8.93	29	2.72	3175	836	5676	566
FP-GCC	US-11	8.71	35	3.97	4430	897	13314	629
P10127	US-18	29.16	57	2.72	3161	842	9075	735
IN2009T1	US-22	10.9	51	6.72	7181	893	16960	952
US100006	US-22	8.4	29	6.71	5949	1184	24502	857
P17777	US-22	3.97	49	4.35	4234	1071	19913	720
BL2009P4	US-23	8.8	51	4.97	5529	860	13429	808
FL2009P4	US-23	6.05	29	2.45	2781	865	8482	394
ND822Pi	US-24	6.7	31	2.86	3056	949	10113	417
PIC97146		58.09	69	15.97	2123	150474	667578	3938
PIC97335		42.58	75	10.31	2520	207930	1481946	2254
PIC97750		14.12	55	5.553	6525	811	12823	1245
PIC97785		31.66	75	10.75	2776	192180	1485270	2240
P10650		31.3	57	1.430	1762	789	5051	462

Chapter 6

Summary

This dissertation improves our limited understanding of the phylogeography and evolution of genomes of *Phytophthora* species. Usually, populations at their center of origin are characterized by increased genetic diversity due to the presence of both mating types resulting in sexual reproduction, and evolution of phylogenetically closely related species (Grünwald and Flier 2005; Goss et al. 2014; O'Hanlon et al. 2018). This has been well documented for the late blight pathogen *P. infestans* which is thought to have originated in central Mexico. Two other closely related species, *P. mirabilis* and *P. ipomoeae* also evolved in central Mexico supporting this hypothesis. *P. mirabilis* and *P. ipomoeae* are pathogens on *Mirabilis jalapa* and *Ipomoeae longipedunculata*, respectively (Flier et al. 2002; Raffaele et al. 2010a). This host specificity is thought to be associated with the evolution of specific pathogenicity factors (Raffaele et al. 2010a). We performed a comparative genomic study for these three closely related *Phytophthora* species to study RxLR gene conservation. Similarly, genome and transcriptome features like recombination, copy number variation and gene expression play an important role in determining virulence, pathogenicity and host specificity (Knaus et al. 2019; Raffaele et al. 2010b; Li et al. 2017). Thus, it is very important to understand the changes occurring in these genomes that will lead to the identification of critical genes. One such example is trying to identify the core important effector genes because they are involved in virulence and in avirulence given the presence of corresponding resistance genes. Identification of conserved avirulence genes is the first step towards screening for disease resistance genes.

Phytophthora species have caused disease losses for more than 150 years (Fry et al. 2015). One of the reasons for the emergence and re-emergence of these pathogens lies in the plasticity of their genomes (Haas et al. 2009; Raffaele et al. 2010a). Sexual reproduction in *Phytophthora* species can be from frequent to rare depending on species and geography (Goss et al. 2014; Grünwald et al. 2012; Arentz 2017). In the case of *P. cinnamomi*, our work supported a sexual mode of reproduction and high genetic diversity in Asia suggesting Asia as its center of origin as speculated previously, whereas in the rest of the world we observed two dominant panglobal lineages of A2 mating types. Migration of the pathogen out of its center of origin has led to severe yield loss in field crops and also threatened biodiversity in the past whether it was *P. infestans* out of central Mexico or *P. cinnamomi* out of Asia (Goss et al. 2014; Engelbrecht et al. 2017).

Disease management can be improved with resistance. However, breeding for resistance requires knowledge of conserved, secreted effectors with known avirulence phenotypes (Armstrong et al. 2005; Du et al. 2015). There are roughly 300-500 RxLR effectors in each *Phytophthora* species, but not all of them are functional (Haas et al. 2009; Tyler et al. 2006). Which effectors are important and which ones are dispensable without losing virulence are important questions to answer. Here we searched for core effectors within closely related clade 1c *Phytophthora* species. *P. infestans* has an unusually high number of RxLR genes compared to other *Phytophthora* species. Our approach of mapping short reads from three *Phytophthora* species to the *P. infestans* reference genome resulted in ascertainment bias in gene calling. This bias due to the reference genome was addressed by assembling and calling genes using the reads that failed to map to the *P. infestans* reference genome.

To summarize, organisms with a dual mode of reproduction (sexual and asexual) such as *Phytophthora* have been recognized as a problem for more than a century now (Fry et al. 2015). The broad host range of some *Phytophthora* species has created substantial negative impacts including losses in crop yield and ecosystem biodiversity. We want to believe that our understanding of the diversity and evolution of the genes in *Phytophthora* genomes has improved a lot substantially. The thesis presented here focuses on improving our understanding of the signatures of a population at its center of origin and outside. Invasion and establishment of certain pathogen genotypes from a diverse population can lead to unprecedented problems. In such cases proper quarantine is extremely helpful to keep pathogens out of certain areas. Similarly, knowledge about the diversity of the pathogenicity genes (RxLR and CRN) is key to the search and deployment of resistance genes. Certain effector genes are indispensable in the genome and therefore these effector genes could be targeted to breed for disease resistance. Our current approach uses short sequencing reads mapped to a reference genome to call SNPs to reconstruct allelic gene sequences. The limitation of this approach is that this fails to detect genes that are absent from the reference genome or that map poorly. To overcome this issue, unmapped reads could be assembled to predict novel genes. This work could be potentially expanded to investigate the variation in other pathogenicity genes (other than RxLR and CRN) in *Phytophthora* species and perform gene expression studies to identify core pathogenicity genes that are constitutively expressed.

Bibliography

Adhikari, B. N., Hamilton, J. P., Zerillo, M. M., Tisserat, N., Lévesque, C. A., and Buell, C. R. 2013. Comparative genomics reveals insight into virulence strategies of plant pathogenic oomycetes. PLoS ONE. 8

Alexander, D. H., Novembre, J., and Lange, K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19:1655–1664

Arentz, F. 2017. *Phytophthora cinnamomi* A1: An ancient resident of New Guinea and Australia of Gondwanan origin? For. Pathol. 47:e12342

Arentz, F., and Simpson, J. A. 1986. Distribution of *Phytophthora cinnamomi* in Papua New Guinea and notes on its origin. Trans. Br. Mycol. Soc. 87:289–295

Armenteros, J. J. A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., Heijne, G. von, and Nielsen, H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat. Biotechnol. 37:420

Armstrong, M. R., Whisson, S. C., Pritchard, L., Bos, J. I. B., Venter, E., Avrova, A. O., Rehmany, A. P., Böhme, U., Brooks, K., Cherevach, I., Hamlin, N., White, B., Fraser, A., Lord, A., Quail, M. A., Churcher, C., Hall, N., Berriman, M., Huang, S., Kamoun, S., Beynon, J. L., and Birch, P. R. J. 2005. An ancestral oomycete locus contains late blight avirulence gene Avr3a, encoding a protein that is recognized in the host cytoplasm. Proc. Natl. Acad. Sci. U. S. A. 102:7766–7771

Arniella, M. B., Fitak, R. R., and Johnsen, S. 2018. Unmapped sequencing reads identify additional candidate genes linked to magnetoreception in rainbow trout. Environ. Biol. Fishes. 101:711–721

Balci, Y., Balci, S., Eggers, J., MacDonald, W. L., Juzwik, J., Long, R. P., and Gottschalk, K. W. 2007. *Phytophthora* spp. associated with forest soils in eastern and north-central U.S. oak ecosystems. Plant Dis. 91:705–710

Beaulieu, J., Ford, B., and Balci, Y. 2017. Genotypic diversity of *Phytophthora cinnamomi* and *P. plurivora* in Maryland's nurseries and Mid-Atlantic forests. Phytopathology. 107:769–776

Beerli, P. 1998. Estimation of migration rates and population sizes in geographically structured populations. NATO ASI Ser. LIFE Sci. 306:39–54

Beerli, P. 2009. How to use MIGRATE or why are Markov chain Monte Carlo programs difficult to use? Pages 42–79 in: Population Genetics for Animal Conservation, G. Bertorelle, M.W. Bruford, H.C. Hauffe, A. Rizzoli, and C. Vernesi, eds. Cambridge University Press, Cambridge.

Beerli, P., and Felsenstein, J. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. Proc. Natl. Acad. Sci. 98:4563–4568

Bevan, M., and Walsh, S. 2005. The *Arabidopsis* genome: a foundation for plant research. Genome Res. 15:1632–1642

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 Genet. Biol. FG B. 45:266–277

Bos, J. I. B., Armstrong, M. R., Gilroy, E. M., Boevink, P. C., Hein, I., Taylor, R. M., Zhendong, T., Engelhardt, S., Vetukuri, R. R., Harrower, B., Dixelius, C., Bryan, G., Sadanandom, A., Whisson, S. C., Kamoun, S., and Birch, P. R. J. 2010. *Phytophthora infestans* effector AVR3a is essential for virulence and manipulates plant immunity by stabilizing host E3 ligase CMPG1. Proc. Natl. Acad. Sci. U. S. A. 107:9909–9914

Bos, J. I. B., Kanneganti, T.-D., Young, C., Cakir, C., Huitema, E., Win, J., Armstrong, M. R., Birch, P. R. J., and Kamoun, S. 2006. The C-terminal half of *Phytophthora infestans* RXLR effector AVR3a is sufficient to trigger R3a-mediated hypersensitivity and suppress INF1-induced cell death in *Nicotiana benthamiana*. Plant J. Cell Mol. Biol. 48:165–176

Cahill, D. M., Rookes, J. E., Wilson, B. A., Gibson, L., and McDougall, K. L. 2008. *Phytophthora cinnamomi* and Australia's biodiversity: impacts, predictions and progress towards control. Aust. J. Bot. 56:279–310

Chang, T. T., Yang, W. W., and Wang, W. Y. 1996. Use of random amplified polymorphic DNA markers for the detection of genetic variation in *Phytophthora cinnamomi* in Taiwan. Bot. Bull. Acad. Sin. 37:165–171

Cooke, D. E., Drenth, A., Duncan, J. M., Wagels, G., and Brasier, C. M. 2000. A molecular phylogeny of *Phytophthora* and related oomycetes. Fungal Genet. Biol. 30:17–32

Cooke, D. E. L., Cano, L. M., Raffaele, S., Bain, R. A., Cooke, L. R., Etherington, G. J., Deahl, K. L., Farrer, R. A., Gilroy, E. M., Goss, E. M., Grünwald, N. J., Hein, I., MacLean, D., McNicol, J. W., Randall, E., Oliva, R. F., Pel, M. A., Shaw, D. S., Squires, J. N., Taylor, M. C., Vleeshouwers, V. G. A. A., Birch, P. R. J., Lees, A. K., and Kamoun, S. 2012. Genome analyses of an aggressive and invasive lineage of the Irish potato famine pathogen. PLoS Pathog. 8:e1002940

Crandall, Bowen. S., and Gravatt, G. F. 1967. The distribution of *Phytophthora cinnamomi*. Ceiba. 13:43–53

Dell, B., and Malajczuk, N. 1989. Jarrah dieback — A disease caused by *Phytophthora cinnamomi*. Pages 67–87 in: The Jarrah Forest: A complex mediterranean ecosystem, Geobotany. B. Dell, J.J. Havel, and N. Malajczuk, eds. Springer Netherlands, Dordrecht.

Dobrowolski, M. P., Tommerup, I. C., Shearer, B. L., and O'Brien, P. A. 2003. Three clonal lineages of *Phytophthora cinnamomi* in Australia revealed by microsatellites. Phytopathology. 93:695–704

Du, Y., Mpina, M. H., Birch, P. R. J., Bouwmeester, K., and Govers, F. 2015. *Phytophthora infestans* RXLR Effector *AVR1* Interacts with Exocyst Component Sec5 to Manipulate Plant Immunity. Plant Physiol. 169:1975–1990

Eggers, J. E., Balci, Y., and MacDonald, W. L. 2012. Variation among *Phytophthora cinnamomi* isolates from oak forest soils in the eastern United States. Plant Dis. 96:1608–1614

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., and Mitchell, S. E. 2011. A robust, simple Genotyping-by-Sequencing (GBS) approach for high diversity species. PLOS ONE. 6:e19379

Emms, D. M., and Kelly, S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16:157

Engelbrecht, J., Duong, T. A., and Berg, N. v d. 2017. New microsatellite markers for population studies of *Phytophthora cinnamomi*, an important global pathogen. Sci. Rep. 7:17631

Erwin, D. C., and Ribeiro, O. K. 1996. Phytophthora diseases worldwide. Phytophthora Dis. Worldw.

Erwin, D. C., and Ribeiro, O. K. 2005. *Phytophthora Diseases Worldwide*. APS Press.

Faber-Hammond, J. J., and Brown, K. H. 2016a. Anchored pseudo-de novo assembly of human genomes identifies extensive sequence variation from unmapped sequence reads. Hum. Genet. 135:727–740

Faber-Hammond, J. J., and Brown, K. H. 2016b. Pseudo-de novo assembly and analysis of unmapped genome sequence reads in wild Zebrafish reveal novel gene content. Zebrafish. 13:95–102

Flier, W. G., Grünwald, N. J., Kroon, L. P. N. M., Sturbaum, A. K., van den Bosch, T. B. M., Garay-Serrano, E., Lozoya-Saldaña, H., Fry, W. E., and Turkensteen, L. J. 2003. The population structure of *Phytophthora infestans* from the Toluca valley of Central Mexico suggests genetic differentiation between populations from cultivated potato and wild *Solanum* spp. Phytopathology. 93:382–390

Flier, W. G., Grünwald, N. J., Kroon, L. P. N. M., Van Den Bosch, T. B. M., Garay-Serrano, E., Lozoya-Saldaña, H., Bonants, P. J. M., and Turkensteen, L. J. 2002. *Phytophthora ipomoeae sp. nov*., a new homothallic species causing leaf blight on *Ipomoea longipedunculata* in the Toluca Valley of central Mexico. Mycol. Res. 106:848–856

Foster, Z. S. L., Sharpton, T. J., and Grünwald, N. J. 2017. Metacoder: An R package for visualization and manipulation of community taxonomic diversity data. PLOS Comput. Biol. 13:e1005404

Fry, W. E., Birch, P. R. J., Judelson, H. S., Grünwald, N. J., Danies, G., Everts, K. L., Gevens, A. J., Gugino, B. K., Johnson, D. A., Johnson, S. B., McGrath, M. T., Myers, K. L., Ristaino, J. B., Roberts, P. D., Secor, G., and Smart, C. D. 2015. Five reasons to consider *Phytophthora infestans* a reemerging pathogen. Phytopathology. 105:966–981

Galindo-A, J., and Hohl, H. R. 1985. *Phytophthora mirabilis*, a new species of *Phytophthora*. Sydowia. 38:87–96

Goodwin, S. B. 1995. Use of cellulose-acetate electrophoresis for rapid identification of allozyme genotypes of *Phytophthora infestans*. Plant Dis. 79:1181

Goodwin, S. B., Drenth, A., and Fry, W. E. 1992. Cloning and genetic analyses of two highly polymorphic, moderately repetitive nuclear DNAs from *Phytophthora infestans*. Curr. Genet. 22:107–115

Goss, E. M., Tabima, J. F., Cooke, D. E. L., Restrepo, S., Fry, W. E., Forbes, G. A., Fieland, V. J., Cardenas, M., and Grünwald, N. J. 2014. The Irish potato famine pathogen *Phytophthora infestans* originated in central Mexico rather than the Andes. Proc. Natl. Acad. Sci. 111:8791–8796

Gouin, A., Legeai, F., Nouhaud, P., Whibley, A., Simon, J. C., and Lemaitre, C. 2015. Whole-genome re-sequencing of non-model organisms: lessons from unmapped reads. Heredity. 114:494–501

Gouin, A., Nouhaud, P., Legeai, F., Rizk, G., Simon, J.-C., and Lemaitre, C. 2013. Whole genome re-sequencing: lessons from unmapped reads. in: Journées Ouvertes Biologie Informatique Mathématiques,

Grünwald, N. J., and Flier, W. G. 2005. The biology of *Phytophthora infestans* at its center of origin. Annu. Rev. Phytopathol. 43:171–190

Grünwald, N. J., Flier, W. G., Sturbaum, A. K., Garay-Serrano, E., van den Bosch, T. B. M., Smart, C. D., Matuszak, J. M., Lozoya-Saldaña, H., Turkensteen, L. J., and Fry, W. E. 2001. Population structure of *Phytophthora infestans* in the Toluca valley region of central Mexico. Phytopathology. 91:882–890

Grünwald, N. J., Garbelotto, M., Goss, E. M., Heungens, K., and Prospero, S. 2012. Emergence of the sudden oak death pathogen *Phytophthora ramorum*. Trends Microbiol. 20:131–138

Grünwald, N. J., McDonald, B. A., and Milgroom, M. G. 2016. Population genomics of fungal and oomycete pathogens. Annu. Rev. Phytopathol. 54:323–346

Haas, B. J., Kamoun, S., Zody, M. C., Jiang, R. H. Y., Handsaker, R. E., Cano, L. M., Grabherr, M., Kodira, C. D., Raffaele, S., Torto-Alalibo, T., Bozkurt, T. O., Ah-Fong, A. M. V., Alvarado, L., Anderson, V. L., Armstrong, M. R., Avrova, A., Baxter, L., Beynon, J., Boevink, P. C., Bollmann, S. R., Bos, J. I. B., Bulone, V., Cai, G., Cakir, C., Carrington, J. C., Chawner, M., Conti, L., Costanzo, S., Ewan, R., Fahlgren, N., Fischbach, M. A., Fugelstad, J., Gilroy, E. M., Gnerre, S., Green, P. J., Grenville-Briggs, L. J., Griffith, J., Grünwald, N. J., Horn, K., Horner, N. R., Hu, C.-H., Huitema, E., Jeong, D.-H., Jones, A. M. E., Jones, J. D. G., Jones, R. W., Karlsson, E. K., Kunjeti, S. G., Lamour, K., Liu, Z., Ma, L., MacLean, D., Chibucos, M. C., McDonald, H., McWalters, J., Meijer, H. J. G., Morgan, W., Morris, P. F., Munro, C. A., O'Neill, K., Ospina-Giraldo, M., Pinzón, A., Pritchard, L., Ramsahoye, B., Ren, Q., Restrepo, S., Roy, S., Sadanandom, A., Savidor, A., Schornack, S., Schwartz, D. C., Schumann, U. D., Schwessinger, B., Seyer, L., Sharpe, T., Silvar, C., Song, J., Studholme, D. J., Sykes, S., Thines, M., van de Vondervoort, P. J. I., Phuntumart, V., Wawra, S., Weide, R., Win, J., Young, C., Zhou, S., Fry, W., Meyers, B. C., van West, P., Ristaino, J., Govers, F., Birch, P. R. J., Whisson, S. C., Judelson, H. S., and Nusbaum, C. 2009. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. Nature. 461:393–398

Hardham, A. R. 2005. *Phytophthora cinnamomi*. Mol. Plant Pathol. 6:589–604

Hardham, A. R., and Blackman, L. M. 2018. *Phytophthora cinnamomi*. Mol. Plant Pathol. 19:260–285

Haverkort, A. J., Boonekamp, P. M., Hutten, R., Jacobsen, E., Lotz, L. a. P., Kessel, G. J. T., Visser, R. G. F., and van der Vossen, E. a. G. 2008. Societal costs of late blight in potato and prospects of durable resistance through cisgenic modification. Potato Res. 51:47–57

Holden, L. A., Arumilli, M., Hytönen, M. K., Hundi, S., Salojärvi, J., Brown, K. H., and Lohi, H. 2018. Assembly and analysis of unmapped genome sequence reads reveal novel sequence and variation in dogs. Sci. Rep. 8

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

Jones, J. D. G., and Dangl, J. L. 2006. The plant immune system. Nature. 444:323–329

Kamoun, S. 2006. A catalogue of the effector secretome of plant pathogenic oomycetes. Annu. Rev. Phytopathol. 44:41–60

Kamoun, S. 2003. Molecular genetics of pathogenic oomycetes. Eukaryot. Cell. 2:191–199

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. Front. Genet. 6

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

Knaus, B. J., and Grünwald, N. J. 2017. VCFR: a package to manipulate and visualize variant call format data in R. Mol. Ecol. Resour. 17:44–53

Knaus, B. J., Tabima, J. F., Davis, C. E., Judelson, H. S., and Grünwald, N. J. 2016. Genomic analyses of dominant U.S. clonal lineages of *Phytophthora infestans* reveals a shared common ancestry for clonal lineages US11 and US18 and a lack of recently shared ancestry among all other U.S. lineages. Phytopathology. 106:1393–1403

Knaus, B. J., Tabima, J. F., Shakya, S. K., Judelson, H. S., and Grünwald, N. J. 2019. Genome-wide increased copy number is associated with emergence of super-fit clones of the Irish potato famine pathogen *Phytophthora infestans*. bioRxiv. :633701

Ko, W. H., Chang, H. S., and Su, H. J. 1978. Isolates of *Phytophthora cinnamomi* from Taiwan as evidence for an Asian origin of the species. Trans. Br. Mycol. Soc. 71:496–499

Kroon, L. P. N. M., Bakker, F. T., van den Bosch, G. B. M., Bonants, P. J. M., and Flier, W. G. 2004. Phylogenetic analysis of *Phytophthora* species based on mitochondrial and nuclear DNA sequences. Fungal Genet. Biol. 41:766–782

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim,

J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, Y., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y. J., Szustakowki, J., and International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. Nature. 409:860–921

Langmead, B., and Salzberg, S. L. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods. 9:357–359

Levy, A., Gonzalez, I. S., Mittelviefhaus, M., Clingenpeel, S., Paredes, S. H., Miao, J., Wang, K., Devescovi, G., Stillman, K., Monteiro, F., Alvarez, B. R., Lundberg, D. S., Lu, T.-Y., Lebeis, S., Jin, Z., McDonald, M., Klein, A. P., Feltcher, M. E., Rio, T. G., Grant, S. R., Doty, S. L., Ley, R. E., Zhao, B., Venturi, V., Pelletier, D. A., Vorholt, J. A., Tringe, S.

G., Woyke, T., and Dangl, J. L. 2018. Genomic features of bacterial adaptation to plants. Nat. Genet. 50:138

Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. 50:913–925

Li, H., and Durbin, R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 25:1754–1760

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinforma. Oxf. Engl. 25:2078–2079

Li, Y., Cooke, D. E. L., Jacobsen, E., and van der Lee, T. 2013. Efficient multiplex simple sequence repeat genotyping of the oomycete plant pathogen *Phytophthora infestans*. J. Microbiol. Methods. 92:316–322

Li, Y., Shen, H., Zhou, Q., Qian, K., van der Lee, T., and Huang, S. 2017. Changing ploidy as a strategy: The Irish potato famine pathogen shifts ploidy in relation to its sexuality. Mol. Plant-Microbe Interact. MPMI. 30:45–52

Linde, C., Drenth, A., Kemp, G. H. J., Wingfield, M. J., and Von Broembsen, S. L. 1997. Population structure of *Phytophthora cinnamomi* in South Africa. Phytopathology. 87:822–827

Linde, C., Drenth, A., and Wingfield, M. J. 1999. Gene and genotypic diversity of *Phytophthora cinnamomi* in South Africa and Australia revealed by DNA polymorphisms. Eur. J. Plant Pathol. 105:667–680

Liu, Y., Koyutürk, M., Maxwell, S., Xiang, M., Veigl, M., Cooper, R. S., Tayo, B. O., Li, L., LaFramboise, T., Wang, Z., Zhu, X., and Chance, M. R. 2014. Discovery of common sequences absent in the human reference genome using pooled samples from next generation sequencing. BMC Genomics. 15:685

Longmuir, A. L., Beech, P. L., and Richardson, M. F. 2017. Draft genomes of two Australian strains of the plant pathogen, *Phytophthora cinnamomi*. F1000Research. 6:1972

Martin, F. N., Blair, J. E., and Coffey, M. D. 2014. A combined mitochondrial and nuclear multilocus phylogeny of the genus Phytophthora. Fungal Genet. Biol. FG B. 66:19–32

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

Martin, M. D., Cappellini, E., Samaniego, J. A., Zepeda, M. L., Campos, P. F., Seguin-Orlando, A., Wales, N., Orlando, L., Ho, S. Y. W., Dietrich, F. S., Mieczkowski, P. A., Heitman, J., Willerslev, E., Krogh, A., Ristaino, J. B., and Gilbert, M. T. P. 2013. Reconstructing genome evolution in historic samples of the Irish potato famine pathogen. Nat. Commun. 4:2172

McCarthy, C. G. P., and Fitzpatrick, D. A. 2017. Phylogenomic reconstruction of the oomycete phylogeny derived from 37 genomes A.P. Mitchell, ed. mSphere. 2

McCarthy, C. G. P., and Fitzpatrick, D. A. 2016. Systematic search for evidence of interdomain horizontal gene transfer from prokaryotes to oomycete lineages. mSphere. 1:e00195-16

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M. A. 2010. The Genome Analysis Toolkit: A mapreduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–1303

Medlar, A. J., Törönen, P., and Holm, L. 2018. AAI-profiler: fast proteome-wide exploratory analysis reveals taxonomic identity, misclassification and contamination. Nucleic Acids Res. 46:W479–W485

Myers, E. W., Sutton, G. G., Delcher, A. L., Dew, I. M., Fasulo, D. P., Flanigan, M. J., Kravitz, S. A., Mobarry, C. M., Reinert, K. H., Remington, K. A., Anson, E. L., Bolanos, R. A., Chou, H. H., Jordan, C. M., Halpern, A. L., Lonardi, S., Beasley, E. M., Brandon, R. C., Chen, L., Dunn, P. J., Lai, Z., Liang, Y., Nusskern, D. R., Zhan, M., Zhang, Q., Zheng, X., Rubin, G. M., Adams, M. D., and Venter, J. C. 2000. A whole-genome assembly of *Drosophila*. Science. 287:2196–2204

Nei, M., and Li, W. H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. U. S. A. 76:5269–5273

O'Hanlon, S. J., Rieux, A., Farrer, R. A., Rosa, G. M., Waldman, B., Bataille, A., Kosch, T. A., Murray, K. A., Brankovics, B., Fumagalli, M., Martin, M. D., Wales, N., Alvarado-Rybak, M., Bates, K. A., Berger, L., Böll, S., Brookes, L., Clare, F., Courtois, E. A., Cunningham, A. A., Doherty-Bone, T. M., Ghosh, P., Gower, D. J., Hintz, W. E., Höglund, J., Jenkinson, T. S., Lin, C.-F., Laurila, A., Loyau, A., Martel, A., Meurling, S., Miaud, C., Minting, P., Pasmans, F., Schmeller, D. S., Schmidt, B. R., Shelton, J. M. G., Skerratt, L. F., Smith, F., Soto-Azat, C., Spagnoletti, M., Tessa, G., Toledo, L. F., Valenzuela-Sánchez, A., Verster, R., Vörös, J., Webb, R. J., Wierzbicki, C., Wombwell, E., Zamudio, K. R., Aanensen, D. M., James, T. Y., Gilbert, M. T. P., Weldon, C., Bosch, J., Balloux, F., Garner, T. W. J., and Fisher, M. C. 2018. Recent Asian origin of chytrid fungi causing global amphibian declines. Science. 360:621–627

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., and Wagner, H. 2016. *vegan: Community Ecology Package*.

Old, K., Dudzinski, M., and Bell, J. 1988. Isozyme variability in field populations of *Phytophthora cinnamomi* in Australia. Aust. J. Bot. 36:355–360

Old, K. M., Moran, G. F., and Bell, J. C. 1984. Isozyme variability among isolates of *Phytophthora cinnamomi* from Australia and Papua New Guinea. Can. J. Bot. 62:2016–2022

Pagliaccia, D., Pond, E., McKee, B., and Douhan, G. W. 2013. Population genetic structure of *Phytophthora cinnamomi* associated with avocado in California and the discovery of a potentially recent introduction of a new clonal lineage. Phytopathology. 103:91–97

Paradis, E. 2010. pegas: an R package for population genetics with an integrated-modular approach. Bioinforma. Oxf. Engl. 26:419–420

Paradis, E., Claude, J., and Strimmer, K. 2004. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics. 20:289–290

Pratt, B. H., and Heather, W. A. 1973. The origin and distribution of *Phytophthora cinnamomi* Rands in Australian native plant communities and the significance of its association with particular plant species. Aust. J. Biol. Sci. 26:559–574

Pratt, B. H., Heather, W. A., and Shepherd, C. J. 1973. Recovery of *Phytophthora cinnamomi* from native vegetation in a remote area of New South Wales. Trans. Br. Mycol. Soc. 60:197–204

Presser, J. W., and Goss, E. M. 2015. Environmental sampling reveals that *Pythium insidiosum* is ubiquitous and genetically diverse in North Central Florida. Med. Mycol. 53:674–683

Pritchard, J. K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics. 155:945–959

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., and Sham, P. C. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559–575

Raffaele, S., Farrer, R. A., Cano, L. M., Studholme, D. J., MacLean, D., Thines, M., Jiang, R. H. Y., Zody, M. C., Kunjeti, S. G., Donofrio, N. M., Meyers, B. C., Nusbaum, C., and Kamoun, S. 2010a. Genome evolution following host jumps in the Irish potato famine pathogen lineage. Science. 330:1540–1543

Raffaele, S., Win, J., Cano, L. M., and Kamoun, S. 2010b. Analyses of genome architecture and gene expression reveal novel candidate virulence factors in the secretome of *Phytophthora infestans*. BMC Genomics. 11:637

Rands, R. D. 1922. *Stripe canker of Cinnamon caused by* Phytophthora cinnamomi *n. sp*. Medeelingen van het Instituut voor Plantenziekten.

Rashidifard, M., Fourie, H., Véronneau, P.-Y., Marais, M., Daneel, M. S., and Mimee, B. 2018. Genetic diversity and phylogeny of South African *Meloidogyne* populations using genotyping by sequencing. Sci. Rep. 8:13816

Revell, L. J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3:217–223

Sauquet, H., von Balthazar, M., Magallón, S., Doyle, J. A., Endress, P. K., Bailes, E. J., Barroso de Morais, E., Bull-Hereñu, K., Carrive, L., Chartier, M., Chomicki, G., Coiro, M., Cornette, R., El Ottra, J. H. L., Epicoco, C., Foster, C. S. P., Jabbour, F., Haevermans, A., Haevermans, T., Hernández, R., Little, S. A., Löfstrand, S., Luna, J. A., Massoni, J., Nadot, S., Pamperl, S., Prieu, C., Reyes, E., dos Santos, P., Schoonderwoerd, K. M., Sontag, S., Soulebeau, A., Staedler, Y., Tschan, G. F., Wing-Sze Leung, A., and Schönenberger, J. 2017. The ancestral flower of angiosperms and its early diversification. Nat. Commun. 8

Schornack, S., van Damme, M., Bozkurt, T. O., Cano, L. M., Smoker, M., Thines, M., Gaulin, E., Kamoun, S., and Huitema, E. 2010. Ancient class of translocated oomycete effectors targets the host nucleus. Proc. Natl. Acad. Sci. U. S. A. 107:17421–17426

Shakya, S. K., Goss, E. M., Dufault, N. S., and van Bruggen, A. H. C. 2015. Potential effects of diurnal temperature oscillations on potato late blight with special reference to climate change. Phytopathology. 105:230–238

Shakya, S. K., Larsen, M. M., Cuenca-Condoy, M. M., Lozoya-Saldaña, H., and Grünwald, N. J. 2018. Variation in genetic diversity of *Phytophthora infestans* populations in Mexico from the center of origin outwards. Plant Dis. 102:1534–1540

Shearer, B. L., Crane, C. E., Barrett, S., and Cochrane, A. 2007. *Phytophthora cinnamomi* invasion, a major threatening process to conservation of flora diversity in the South-west Botanical Province of Western Australia. Aust. J. Bot. 55:225

Shepherd, C. J. 1975. *Phytophthora cinnamomi* - an ancient immigrant to Australia.

Sherman, R. M., Forman, J., Antonescu, V., Puiu, D., Daya, M., Rafaels, N., Boorgula, M. P., Chavan, S., Vergara, C., Ortega, V. E., Levin, A. M., Eng, C., Yazdanbakhsh, M., Wilson, J. G., Marrugo, J., Lange, L. A., Williams, L. K., Watson, H., Ware, L. B., Olopade, C. O., Olopade, O., Oliveira, R. R., Ober, C., Nicolae, D. L., Meyers, D. A., Mayorga, A., Knight-Madden, J., Hartert, T., Hansel, N. N., Foreman, M. G., Ford, J. G., Faruque, M. U., Dunston, G. M., Caraballo, L., Burchard, E. G., Bleecker, E. R., Araujo, M. I., Herrera-Paz, E. F., Campbell, M., Foster, C., Taub, M. A., Beaty, T. H., Ruczinski, I., Mathias, R. A., Barnes, K. C., and Salzberg, S. L. 2018. Assembly of a pan-genome from deep sequencing of 910 humans of African descent. Nat. Genet.

Singla, D. 2016. A lesson from unmapped reads in Next-Generation Sequencing data. SM J Bioinform Proteomics. 1:1007

Stam, R., Jupe, J., Howden, A. J. M., Morris, J. A., Boevink, P. C., Hedley, P. E., and Huitema, E. 2013. Identification and characterisation CRN effectors in *Phytophthora capsici* shows modularity and functional diversity. PLoS ONE. 8

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

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

Stanke, M., and Morgenstern, B. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 33:W465–W467

Studholme, D. J., McDougal, R. L., Sambles, C., Hansen, E., Hardy, G., Grant, M., Ganley, R. J., and Williams, N. M. 2015. Genome sequences of six *Phytophthora* species associated with forests in New Zealand. Genomics Data. 7:54–56

Stukenbrock, E. H., and McDonald, B. A. 2008. The origins of plant pathogens in agro-ecosystems. Annu. Rev. Phytopathol. 46:75–100

Tabima, J. F., Coffey, M. D., Zazada, I. A., and Grünwald, N. J. 2018. Populations of *Phytophthora rubi* show little differentiation and high rates of migration among states in the western United States. Mol. Plant. Microbe Interact. 31:614–622

Tabima, J. F., and Grünwald, N. J. 2019. effectR: An expandable R package to predict candidate RxLR and CRN effectors in oomycetes using motif searches. Mol. Plant. Microbe Interact.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123:585–595

Thiagavel, J., Cechetto, C., Santana, S. E., Jakobsen, L., Warrant, E. J., and Ratcliffe, J. M. 2018. Auditory opportunity and visual constraint enabled the evolution of echolocation in bats. Nat. Commun. 9

Tyler, B. M. 2007. *Phytophthora sojae*: root rot pathogen of soybean and model oomycete. Mol. Plant Pathol. 8:1–8

Tyler, B. M., Tripathy, S., Zhang, X., Dehal, P., Jiang, R. H. Y., Aerts, A., Arredondo, F. D., Baxter, L., Bensasson, D., Beynon, J. L., Chapman, J., Damasceno, C. M. B., Dorrance, A. E., Dou, D., Dickerman, A. W., Dubchak, I. L., Garbelotto, M., Gijzen, M., Gordon, S. G., Govers, F., Grunwald, N. J., Huang, W., Ivors, K. L., Jones, R. W., Kamoun, S., Krampis, K., Lamour, K. H., Lee, M.-K., McDonald, W. H., Medina, M., Meijer, H. J. G., Nordberg, E. K., Maclean, D. J., Ospina-Giraldo, M. D., Morris, P. F., Phuntumart, V., Putnam, N. H., Rash, S., Rose, J. K. C., Sakihama, Y., Salamov, A. A., Savidor, A., Scheuring, C. F., Smith, B. M., Sobral, B. W. S., Terry, A., Torto-Alalibo, T. A., Win, J., Xu, Z., Zhang, H., Grigoriev, I. V., Rokhsar, D. S., and Boore, J. L. 2006. *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science. 313:1261–1266

Wang, J., Fernández-Pavía, S. P., Larsen, M. M., Garay-Serrano, E., Gregorio-Cipriano, R., Rodríguez-Alvarado, G., Grünwald, N. J., and Goss, E. M. 2017. High levels of diversity and population structure in the potato late blight pathogen at the Mexico centre of origin. Mol. Ecol. 26:1091–1107

Weide, R. H. van der, Simonis, M., Hermsen, R., Toonen, P., Cuppen, E., and Ligt, J. de. 2016. The genomic scrapheap challenge; extracting relevant data from unmapped whole genome sequencing reads, including strain specific genomic segments, in rats. PLOS ONE. 11:e0160036

Whitacre, L. K., Tizioto, P. C., Kim, J., Sonstegard, T. S., Schroeder, S. G., Alexander, L. J., Medrano, J. F., Schnabel, R. D., Taylor, J. F., and Decker, J. E. 2015. What's in your next-generation sequence data? An exploration of unmapped DNA and RNA sequence reads from the bovine reference individual. BMC Genomics. 16

Wickham, H. 2009. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York.

Yang, X., Tyler, B. M., and Hong, C. 2017. An expanded phylogeny for the genus *Phytophthora*. IMA Fungus. 8:355–384

Yoshida, K., Schuenemann, V. J., Cano, L. M., Pais, M., Mishra, B., Sharma, R., Lanz, C., Martin, F. N., Kamoun, S., Krause, J., Thines, M., Weigel, D., and Burbano, H. A. 2013. The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. eLife. 2:e00731

Yuen, J. E., and Andersson, B. 2013. What is the evidence for sexual reproduction of *Phytophthora infestans* in Europe? Plant Pathol. 62:485–491

Zarif Saffari, A. 2012. *Exploring the diversity of unmapped reads from human deep sequencing*.

Zentmyer, G. A. 1988. Origin and distribution of four species of *Phytophthora*. Trans. Br. Mycol. Soc. 91:367–378

Zerbino, D. R., and Birney, E. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18:821–829