DOCTOR OF PHILOSOPHY

Omic-enabled Dectection and Studies of Phytophthora capsici in the Field

Mcleod, Rory

Award date:
2020

Omics-enabled Detection and Studies of *Phytophthora capsici* in the Field

Rory A. McLeod

Doctor of Philosophy

School of Life Science, University of Dundee

May 2020

University of Dundee

Table of contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of Tables

Chapter 1

Chapter 2

List of Abbreviations

BAK1	Brassinosteroid Insensitive 1-Associated Kinase
BLASTp/n	Basic Local Alignment Search Tool (protein/nucleotide)
CAA	Carboxylic Acid Amine
CDS	Coding Sequence
CM334	Coriollo de Morales 334 Pepper
CSV	Comma Separated Value
DEFRA	Department of Environment, Food and Rural Affairs
DNA	Deoxyribonucleic Acid
DPI	Days Post Infection
ELISA	Enzyme Linked Immunosorbent Assay
EHM	Extrahaustorial Matrix

ETI	Effector Triggered Immunity
ETS	Effector Triggered Susceptibility
EWINDO	East West Seed Indonesia
EZ	Enza Zaden, the Netherlands
GAPDH	Glyceraldehyde-3-Phosphate Dehydrogenase
GFF	General Feature Format
HR	Hypersensitive Response
HMW	High Molecular Weight
ITS	Internal Transcribed Spacer Region
JGI	Joint Genome Institute
JHI	James Hutton Institute
LFT	Lateral Flow Immunoassay
LOH	Loss of Heterozygosity
MAMP	Microbe Associated Molecular Pattern
MCL	Markov Clustering
MUX	Multiplexer
NCBI	National Centre for Biotechnology Information
NGS	Next Generation Sequencing
NLS	Nuclear Localisation Signal
OEDs	Omics-Enabled Detection Pipeline
ONT	Oxford Nanopore Technology
PAMP	Pathogen Associated Molecular Pattern
PCR	Polymerase Chain Reaction
PDP	Primer Diagnostic Pipeline
PenSeq	Pathogen Enrichment Sequencing
POI	Position of Interest
PR	Pathogenesis Related
PRR	Pathogen Recognition Receptor
PTI	PAMP Triggered Immunity
QTL	Quantitative Trait Loci
RIL	Recombinant Inbred Line

ROC	Receiver Operating Characteristic
RLP	Receptor-like Kinase
RLP	Receptor-like Protein
RNA	Ribonucleic Acid
RNAseq	Sequencing of Ribonucleic Acid
ROS	Reactive Oxygen Species
SNP	Single Nucleotide Polymorphism
SSR	Short Sequence Repeats
UV	Ultraviolet
ZMW	Zero Mode Waveguides

Acknowledgements

First of all, I would like to thank Dr Edgar Huitema and Dr Leighton Pritchard for their continual guidance, support and advice throughout the PhD project.

Edgar, thank you for the opportunity to study and carry out this PhD project in your lab. I am grateful for your help and support over the last 4 years, both academically and personally. Thank you for inspiring me during meetings and drop-ins when times have been tough. Before starting this PhD, computational biology was completely new to me. Leighton, thank you for your patience with me and showing me the bioinformatic ropes. I appreciate the new skillset I have learned (and still learning) and hope to use them in my future career.

Thank you to Dr Petra Bleeker, Dr Walter Verweij, Dr Richard Feron and the plant pathology team at Enza Zaden. Also, thanks to Ramadhani Safitri and the plant pathology team at EWINDO. A special thanks to Cepy, I loved travelling around Indonesia with you, visiting all the farmers and of course, finding *P. capsici*! Thank you for your help and I hope we will meet again.

The Huitema lab has been a constant support group during the 4 years. I'd like to thank each and every one of them. In particular, I'd like to thanks Chris Cornelisse, without your impressive bioinformatic skillset, OEDs would not be what it is today. Thank you! Michael, Victor, Jasmine, Graeme, Tiago, Gaetan, and Natalja, thank you all for helping me in the lab (mostly with molecular biology techniques!). It has been a privilege to work with you. I would also like to thank Dr Piers Hemsley and Professor David Horn as my thesis committee for their input and advice on the project.

My friends and family have helped me extensively throughout the project. Mum, Dad and Kirsty, thank you for your guidance and for always believing in me. Chris, Rob, Gregor and Elliot, thank you for all being there to take my mind off of work and reminding myself to chill out. And thank you to Susan, Bill and Molly for your support over the last 10 years.

But most of all, to my wife Bethany, I cannot thank you enough for your continual support throughout the PhD. Your love and positive attitude has encouraged me to keep going and achieve my goal. I look forward to spending the future with you, wherever life takes us!

Declarations

The results presented in this PhD thesis are from investigations conducted by myself. Any work that is not my own is clearly identified with the appropriate references and publications. I hereby declare that I am the author of this work and it has not been previously accepted for any higher degree.

Rory A. McLeod

We certify that Rory McLeod has fulfilled the relevant ordinance and regulations of the University Court and is qualified to submit this thesis for the degree of Doctor of Philosophy.

Dr. Edgar Huitema, School of Life Science, University of Dundee

Dr. Leighton Pritchard, Information and Computational Sciences, The James Hutton Institute

Summary

Phytophthora capsici is a devastating pathogen capable of infecting important crops worldwide. Disease in the field can lead to eradication of crops and large financial losses if left untreated. Current diagnosis of the pathogen from the field is time consuming, difficult and requires highly trained specialists to handle and process samples. A more efficient diagnosis method is needed to ensure farmers can effectively maintain and manage their crops.

Working with recently isolated *P. capsici* strains from the field is beneficial for both phytopathologists and plant breeders to identify the mechanisms used by the pathogen to cause infection and to develop resistant, commercial crops. To bridge the gap between laboratory and field knowledge, successful diagnosis and isolation of the pathogen are necessary.

A genome of a single isolate of *P. capsici* has been sequenced and is publicly available. I have used this genome – as well as sequencing three other field isolates using three different sequencing technologies – to use a genomics approach to address the diagnostic issues that are currently faced.

Here, I have employed two bioinformatic pipelines to aid the diagnosis effort to diagnose *P. capsici* from the field. The first (OEDs), designs diagnostic primers that are species-specific; and the second (PDP), designs primers that can discriminate within the species, resulting in isolate-specific primers.

I conclude that both pipelines can design discriminatory diagnostic primers, but more sequence data and validation are required to substantiate these claims. These pipelines show great promise for diagnosing eukaryote plant pathogens found in the field.

Chapter 1: General Introduction

Plant pathogens have devastated crops from the dawn of domesticated farming (Stukenbrock and McDonald, 2008). Numbers of plant pathogenic organisms have increased over time with the appearance of novel crop varieties and improved detection methods (Tewari, 2018). Today, multi-billion dollar losses are incurred each year by plant pathogens that infect and cause disease on important food crops worldwide (Tewari, 2018). With the human population predicted to rise to more than 9 billion by 2050, providing enough food by sustainable farming methods will be of extreme importance (Quentin Grafton, 2015). Therefore, it is essential that disease outbreaks are efficiently managed to prevent largescale crop losses and famine. In order to supply enough food, existing farmlands and fields need to be utilised more efficiently to minimise the risk of potential pathogenic outbreaks.

Plants have co-evolved a symbiotic relationship with soil microbes in order to survive (Selosse and Rousset, 2011). Symbiosis has been key for the transferal of important compounds through the root system for millions of years. For example, nitrogen and phosphorus uptake can be aided by rhizobia bacteria and mycorrhizal fungi respectivley, to convert it into a form that the plants can utilise (Morgan, 2013). However, not all microbial interactions are beneficial to the host. Various bacteria, viruses, fungi and oomycetes can cause disease in plants. As a result of these interactions, plants are under a fluctuating selection pressure. The pathogen also has to adapt to overcome the host immune system in order to survive and proliferate, resulting in a host-pathgen arms race.

1.1 Oomycetes: An important group of animal and plant pathogens

Oomycetes are eukaryotic organisms that grow using a filamentous, hyphal network and have the ability to form water and airborne spores (Jiang and Tyler, 2012). Oomycetes were long considered to be fungi until the mid-20th century (1967) as molecular and morphological techniques improved (McCarthy and Fitzpatrick, 2017, Arx, 1967). These methods and further studies have classified oomycetes as to being more closely related to brown algae and diatoms in the Stramenopiles phylum, rather than fungi.

The Stramenopile kingdom contains algae and the oomycetes. While several oomycete species play key ecological roles in decomposing organic matter, others have pathogenic life styles, invading animal and plant species (Lamour et al., 2007, Kamoun, 2003). For example, *Saprolegnia parasitica* infects fresh water fish, where it completes the majority of its lifecycle. *Pythium insidiosum* is an important pathogen of mammals causing pythiosis dogs, horses and immunocompromised people (van West, 2006, De Cock et al., 1987, Gaastra et al., 2010).

The majority (>60%) of the 800 known oomycete species are capable of causing disease in plants (Restrepo et al., 2014, Thines and Kamoun, 2010). Over 100 of these species belong to the *Pythium* genus (Kamoun, 2003). An aggressive *Pythium* species that causes multiple diseases on plants is *P. aphanidermatum*. This pathogen can cause root rot, seedling rot, damping-off cottony blight and stalk rot on a wide range of hosts including soybean, cucumber and pepper (Johnstone et al., 2005, Postma, 2009, Rosso, 2008). Current *Pythium* research is focussed on finding a suitable and effective biological control management strategy (Parveen, 2015). However, the most notable oomycete plant pathogens belong to the *Phytophthora* genus.

1.2 *Phytophthora*: Notable oomycete pathogens

Members of the *Phytophthora* genus rank amongst the most devastating oomycete pathogens. The most recent phylogenetic analysis within the *Phytophthora* genus, suggests that 142 known and 43 provisionally named *Phytophthora* species have been identified and characterised worldwide to date (Yang et al., 2017). *Phytophthora* spp. can infect the majority of dicots including a variety of important crops and several monocots (Lamour et al., 2007, Lamour et al., 2012b, Kamoun, 2003). The host range for many *Phytophthora* species is diverse and can result in a variety of diseases. For example, *Phytophthora ramorum* strains can cause sudden oak death on oak trees and also leaf blight on woody ornamentals (Grunwald et al., 2008). There is currently, no effective treatment for *P. ramorum* disease in the field with resistance to fungicides reported (Rizzo et al., 2005). One management method used is the removal of infected and surrounding hosts to prevent further spread of the disease. The most notable *Phytophthora* species is *P. infestans. Phytophthora infestans* was the pathogen that was responsible for causing the infamous potato famine in the mid-19th Century, where millions of Irish and mainland European humans died (Ristaino, 2002). Other examples of *Phytophthora species* include *P. sojae* which infects soybean causing root rot (Tyler, 2007), *P. kernoviae* causes necrotic lesions on trees and wild flowers (Brasier et al., 2005) and *P. cactorum* infects an extremely wide range of hosts including trees, wild flowers and fruit crops causing collar rot and root rot (Hantula, 2000, Plich, 1979).

1.3 *Phytophthora* reproduction and infection cycle

All *Phytophthora* species have a hemi-biotrophic lifecycle, having two distinct infection stages – biotrophy then necrotrophy (Irwin et al., 1997). The biotrophic stage occurs during the initial stages of infection where the pathogen survives on living tissue

without any disease symptoms (Lamour et al., 2012b). The necrotrophic phase is where the infected cells are killed and the pathogen proliferates (Lamour et al., 2012b). Many species within the *Phytophthora* genus can reproduce sexually, asexually or both. Sexual reproduction in *Phytophthora* requires two mating types, A1 and A2 (Lamour et al., 2012b). When both mating types are in close proximity, male and female gametangia are formed (antheridia and oogonia respectively). Thick walled oospores are formed after fertilisation and can survive in the soil and in harsh conditions for long periods of time. When conditions are favourable, the oospores germinate producing sporangia.

In the absence of the two mating types, or when only one is present, *Phytophthora* reproduce asexually. The sporangia can infect hosts directly, or when dislodged by rain water or irrigation, release ~30 motile zoospores. Zoospores are asexual motile spores that are chemotactically attracted to plant tissue (Babadoost, 2004). Examples of chemo-attractants are sugars, amino acids, alcohols and phenolic compounds (Hardham, 2001). Two flagella aid zoospores to actively swim towards the intended host. An anterior flagellum pulls the zoospore and a posterior flagellum acts as a rudder (Hardham, 2001). Glycoproteins are used by the zoospore to physically adhere to the host surface (Hardham, 2007).

When the zoospores adhere to the host, the flagella are removed and the appearance of a cell wall develops around the zoospore, creating a cyst (Hardham, 2001). These cysts produce a germ tube and then differentiate into appresoria structures that can penetrate the epidermal layer or produce new sporangia on the surface in favourable conditions (Figure 1.1.A) (Grenville-Briggs et al., 2005, Lamour et al., 2012b). The resulting sporangia can release zoospores, or infect directly, thus completing the infection cycle.

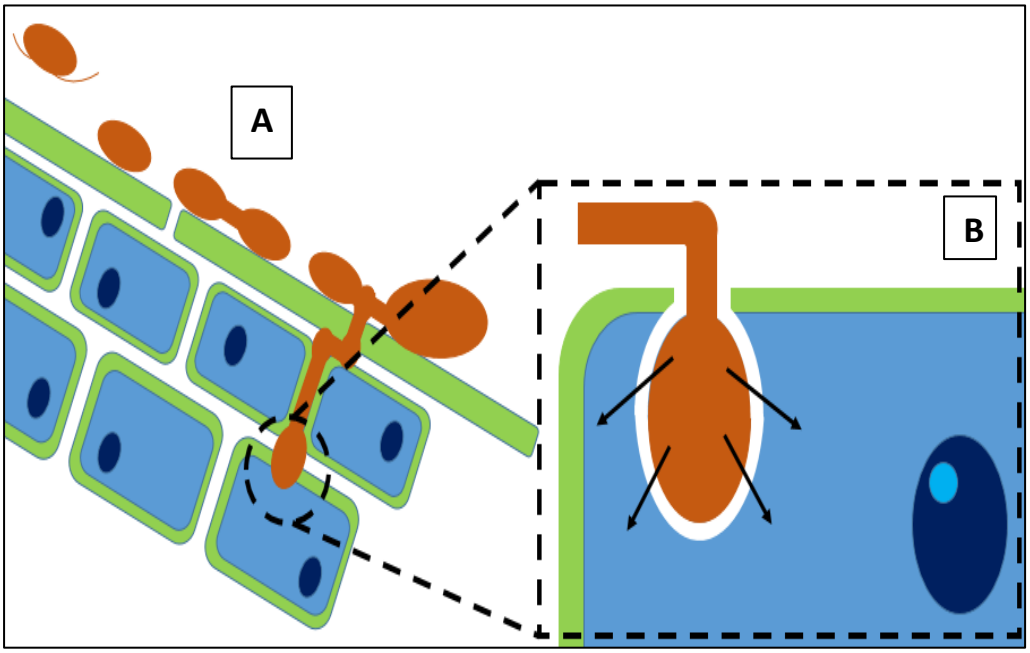


Figure 1.1: *Phytophthora* infection development: (A) Initial host contact via zoospore to eventual development of sporangium on the surface and haustoria penetrating the host cell. (B) In-depth look at haustoria penetrating host cell. Arrows represent the transfer of effectors from the pathogen.

1.4 The host response to *Phytophthora* infection

Plants continually interact with pathogenic microbes in or around their environment and require the ability to recognise potential threats and combat them when under attack to prevent infection. This complex signalling network is regulated by multiple phytohormones including jasmonic acid and salicylic acid (Denance et al., 2013). Host membrane bound pathogen recognition receptors (PRRs) detect the presence of pathogenic organisms by microbe or pathogen-associated molecular patterns (MAMPs/PAMPs) such as flagellin in bacteria, β-Glucans in fungi and transglutaminase in oomycetes (including *Phytophthora*) (Nurnberger et al., 2004, Dodds and Rathjen, 2010, Jones and Dangl, 2006). Once the pathogen has been detected by the host, or host cells have been damaged, signals are sent to initiate the first line of host immunity, PAMP triggered immunity (PTI). PTI events include the

production of reactive oxygen species (ROS), ion fluxes, transcriptional reprogramming and MAP-kinase activation in order to inhibit or slow down the pathogen (Chaparro-Garcia et al., 2015). Receptor like proteins/kinases (RLP/RLK) regulate PTI. For example, The RLK, BAK1 (brassinosteroid insensitive 1-associated kinase 1) modulates PTI in *Arabidopsis thaliana* and *Nicotiana benthamiana* during *P. infestans* infection. Silencing of BAK1 in these hosts saw an increase in susceptibility (Chaparro-Garcia A, 2011). Many plant pathogens including *Phytophthora* spp. can overcome PTI by attacking the PTI signalling pathways or by synthesising their own phytohormones or toxins (Ma and Ma, 2016). Once PTI is overcome, the pathogen can proliferate.

1.5 Effector proteins – The pathogen's arsenal for causing infection

Once the epidermal layer of the host has been breached by *Phytophthora*, mycelium grows intercellularly into the mesophyll layer in the form of hypha. Hyphae protrude into the host cells where haustoria are produced (Figure 1.1.B) (Avrova et al., 2008). Thought to be secreted at the haustorial interface, effectors are pathogen proteins that manipulate host biology to enable pathogen proliferation (Wang et al., 2018). Amongst other things, these effector proteins can inhibit or suppress the plant immune response resulting in effector triggered susceptibility. However, in response to the detection and recognition of these effectors by host resistance proteins (R proteins) the second layer of plant immunity – effector triggered immunity (ETI) – is initiated often resulting in programmed cell death, or a hypersensitive response (HR).

A number of models are described for the perception of pathogen effectors; the gene-for-gene hypothesis, where R proteins or nucleotide-biding site, leucine rich repeat (NB-LRR) proteins directly or indirectly target the effector (Chisholm et al., 2006,

Dodds and Rathjen, 2010). The guard hypothesis describes where R proteins monitor the targets of effectors and, if perturbed, trigger the host immune response (van der Hoorn and Kamoun, 2008). In *Arabidopsis*, the phosphorylation of a plasma-membrane bound protein, RIN4, by AvrRpm1 and AvrB (two unrelated bacterial effectors) is thought to activate the NB-LRR protein, RPM1 (Jones and Dangl, 2006).

The decoy model postulates that some effectors may have multiple targets within the host, but some targets act as decoys. Once the effector binds to the decoy, if an R protein is present, this will result in an immune response, but if there is no R protein present, the pathogen does not proliferate, nor does it initiate an immune response as it is trapped in a recognition event (van der Hoorn and Kamoun, 2008). The tomato R protein, Cf-2 targets the fungal *Cladosporium fulvum* effector, Avr2 – a protease inhibitor. A host protease, RCR3, acts as a decoy for the effector by trapping it into a recognition event (Shabab et al., 2008, van der Hoorn and Kamoun, 2008). If Cf-2 is present, a cell death immune response is triggered (Shabab et al., 2008).

However, pathogens can respond by secreting bodyguard decoy effectors to protect the virulent effectors. For instance, during *P. sojae* infection on soybean, a secreted effector protein in the apoplast, xyloglucan-specific endoglucanase (*Ps*XEG1) is inhibited by a host secreted protein, gluconase inhibitor protein 1 (GmGIP1) (Ma et al., 2017). In response, the pathogen secretes a *Ps*XEG1 paralog, *Ps*XLP1 that mimics *Ps*XEG1 but has no known enzymatic activity (Paulus and van der Hoorn, 2018). This bodyguard decoy is thought to protect the effector and its function to interact with the host cell wall during infection.

The study of oomycete effector proteins is a prominent topic in current plant pathology research. Functional genomics and bioinformatic studies have identified a large array

of effector proteins from multiple pathogens including *Phytophthora* species (Stam et al., 2013). There are two distinct classes of effectors – apoplastic and cytoplasmic (Kamoun, 2006). Both classes can lead to effector triggered susceptibility (ETS) by modifying host signalling and host cellular processes if the plant immune system is bypassed (Howden and Huitema, 2012).

1.5.1 Apoplastic effectors

Apoplastic effectors are secreted from the pathogen into the extracellular volume between host cells. The host detects these "foreign molecules" and triggers an immune response (Doehlemann and Hemetsberger, 2013). Host pathogenesis related (PR) enzymes such as proteases, glucanases and chitinases are secreted which have been shown to inhibit the function of effectors (Jashni et al., 2015). For example, EPI1 and EPI10 are kazal-like serine protease inhibitors secreted from *P. infestans*. These effectors bind to the PR serine protease, P69B in the apoplast to prevent the

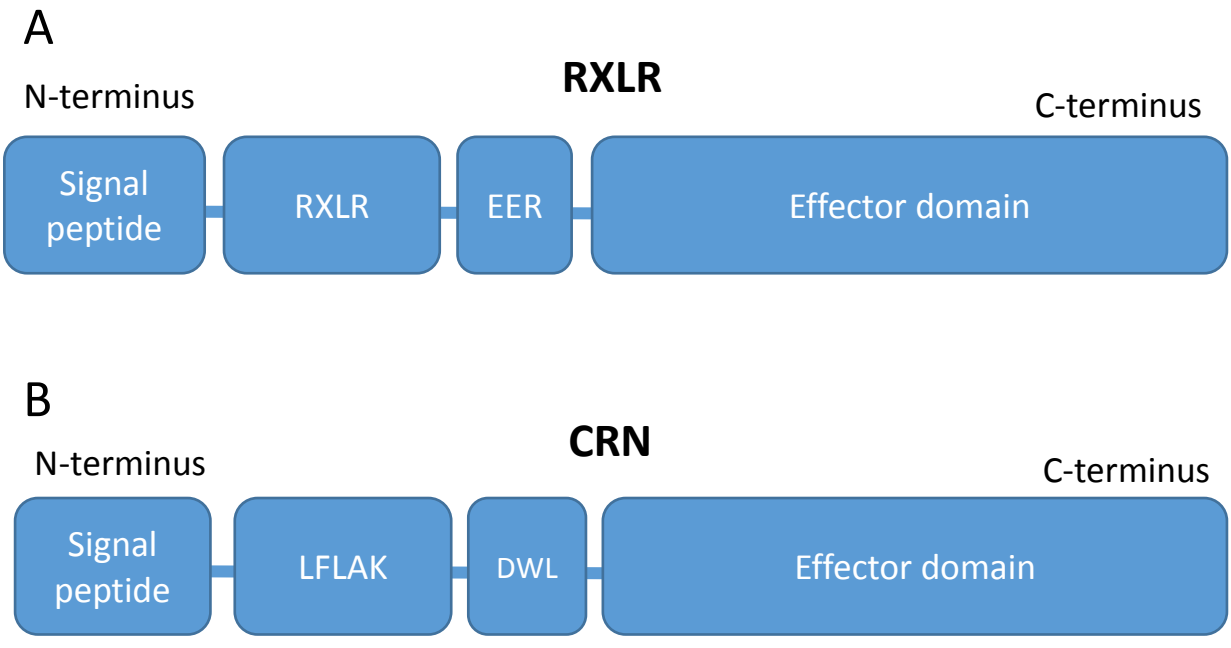


Figure 1.2: Modular structures of cytoplasmic effectors, RXLR (A) and CRN (B). Both require a signal peptide for secretion from the pathogen, conserved motifs for translocation and an effector domain at the C-terminus.

degradation of pathogenic proteins (Tian et al., 2005, Tian et al., 2004, Tornero et al., 1997).

1.5.2 Cytoplasmic effectors

Cytoplasmic effectors are translocated across the haustorial interface into host cells by a signal peptide and conserved translocation sequence at the N-terminus (Wang et al., 2018). The effector domain at the C-terminus is thought to be responsible for pathogenicity. *Phytophthora* cytoplasmic effectors target specific cellular components/mechanisms within the host cell (Boevink et al., 2016, Mafurah et al., 2015, Stam et al., 2013). There are two distinct classes of cytoplasmic effectors in the *Phytophthora* genus, the RXLRs and the Crinklers (CRNs) (Figure 1.2).

1.5.3 The RXLR protein family

The RXLRs are a family of effectors secreted during the early stages of *Phytophthora* infection at the haustorial interface. The RXLRs are named after the conserved Arg-X-Leu-Arg motif (where X is any amino acid) following from the signal peptide at the N-terminus (Figure 1.2.A) (Lamour et al., 2012b, Stam et al., 2013). The pathogenicity of the RXLRs is due to the genetically diverse effector domain at the C-terminus. RXLRs have been identified in all *Phytophthora* species sequenced so far (Haas et al., 2009). A study in the oomycete pathogen *Hyaloperonospera arabidopsis* showed that RXLRs can localise in different cellular compartments (Caillaud et al., 2012). The RXLRs are known to alter host cellular signalling and suppress PTI within the host cell (Birch, 2008). Although thought to be translocated at the haustorial interface, the mechanism(s) of delivery into the host cell are not fully understood (Wang et al., 2018).

The RXLR effectors can promote infection as HR inducers (Stam, 2013). RXLR interaction with host targets can suppress an immune response. The *P. infestans*

RXLR effector, Pi04314 is upregulated during the biotrophic stage of infection and interacts with three host proteins – phosphatase 1 catalytic (PPc1) isoforms. The effector re-localises the host proteins from the nucleolus to the nucleoplasm during infection. This mitigates the transcriptional response for salicylic acid and jasmonic acid which in turn, supresses the host immune response (Boevink et al., 2016).

The *P. infestans* Avr3a effector suppresses host immunity by targeting the host ubiquitin E3 ligase CMPG1 which is required for cell death (Bos et al., 2010). The R protein, R3a, from the host detects Avr3a in hosts which triggers ETI (Armstrong et al., 2005). Silencing Avr3a also prevents *P. infestans* infection (Bos et al., 2010). *Phytophthora sojae* and *P. capsici* contain AVR3a homologues (Boutemy et al., 2011).

RXLRs are differentially expressed at different developmental and infection stages during *Phytophthora capsici* infection (Jupe et al., 2013). RXLRs can be classified into four groups in regards to time of gene expression. Type I and III are highly expressed at the start of infection, where type I has low expression when cysts germinate and type III has high expression. Type II and IV are highly expressed during biotrophy (8-24 hours after infection) where type II are downregulated during necrotrophy (24-72 hours after infection) and type IV are upregulated during sporulation stage (>72 hours after infection) (Jupe *et al.* 2013).

1.5.4 The CRN protein family

First identified in *Phytophthora* by Torto *et al* (2003) , another family of effectors that are thought to get secreted into the cytoplasm during infection are known as Crinklers (CRNs) – named after the CRinkling and Necrosis phenotype observed when the proteins are expressed in plant tissue. CRN effectors are characterised by conserved LFLAK and DWL domains containing LXLFLAK and HVLVVVP motifs respectively at

the N-terminus (Figure 1.2.B) (Jupe et al., 2013, Stam et al., 2013). Similarly to the RXLRs, the CRNs contain a signal peptide at the N-terminus that is required for translocation into the host cell, although nothing is known about how this is done (Schornack et al., 2010). However, we do know that all CRN effectors target the nucleus during infection but not all cause cell death, confirming that cell death is not always a phenotype to confirm CRN activity (Schornack et al., 2010, Stam et al., 2013).

The C-terminus of the *P. infestans* CRN effector, CRN8, shows sequence similarities to serine and threonine kinases which shows catalytic activity *in planta* (van Damme et al., 2012). When localised to the host nucleus with the aid of a nuclear localisation signal (NLS), CRN8 was shown to enhance virulence and promote cell death (van Damme et al., 2012).

Similarly to CRN8, Mafurah *et al.* (2015) outlined another CRN effector during *P. capsici* infection that induces cell death. The PcCRN4 effector was also observed to localise to the nucleus in host plants in order to function. However, it is not clear how CRN effectors, including PcCRN4 localise to the nucleus. Possible suggestions are either by an NLS (CRN8) or by an import factor such as importin-α (Schornack et al., 2010).

1.6 *Phytophthora capsici*: A devastating multi-crop pathogen

Phytophthora capsici was first described in 1922 by Leon H. Leonian, after the pathogen had caused devastating damage to a chilli pepper crop in New Mexico four years previously (Leonian, 1922). The pathogen has now been identified across the globe, mostly in temperate or tropical conditions such as South America and East Asia (Sun et al., 2008, Gobena et al., 2012). *Phytophthora capsici* can infect any area of the crop at any stage of growth (Sun et al., 2008).

Phytophthora capsici is a filamentous, broad host pathogen, predominantly infecting solanaceous and cucurbit crops in fields around the world (Dunn et al., 2014, Gobena et al., 2012, Granke et al., 2012). Occasionally, other solanaceous, cucurbit and leguminosae crops such as the solanaceous tomato (Kreutzer, 1946) and eggplant (Bodine, 1935) the Legumes snap beans (McGrath et al., 2011) and lima beans (Davidson et al., 2002) and the cucurbit melon (Tompkins, 1937) have also been susceptible to *P. capsici* in a field environment. However, under laboratory conditions, *P. capsici* can infect a wider range of hosts including carrots, turnips, cauliflower and peaches (Granke et al., 2012).

With the emergence of next generation sequencing (NGS) techniques, omics-enabled studies are becoming more common in order to study the genetic make-up and mechanisms of *Phytophthora* spp. (Govers and Gijzen, 2006, Pang et al., 2015). According to the National Centre for Biotechnology Information (NCBI), there are currently 27 *Phytophthora* species with sequenced genomes (Table 1.1).

The sequenced *P. capsici* isolate LT1534 that is currently used as the reference, was produced from mating two strains – LT51 ((A1) isolated from cucumber from Michigan) and LT263 ((A2) isolated from pumpkin in Tennessee) (Lamour et al., 2012a). The offspring were backcrossed twice with LT263 (A2) resulting in the inbred line LT1534 (A2) which was sequenced and is regularly used in laboratory studies.

Table 1.1 The publicly accessible *Phytophthora* genomes from the NCBI.

Species	**Citation of genome publication**
Phytophthora agathidica	(Studholme et al., 2016)
Phytophthora cactorum	(Armitage et al., 2018, Grenville-Briggs et al., 2017)
Phytophthora cambivora	(Feau et al., 2016)
Phytophthora capsici	(Lamour et al., 2012a)
Phytophthora cinnamomi	(Studholme et al., 2016, Longmuir et al., 2017)
Phytophthora citricola	(Srivastava, Unpublished)
Phytophthora colocasiae	(Vetukuri et al., 2018a)
Phytophthora cryptogea	(Feau et al., 2016)
Phytophthora fragariae	(Gao et al., 2015, Tabima et al., 2017)
Phytophthora infestans	(Haas et al., 2009)
Phytophthora kernoviae	(Sambles et al., 2015, Studholme et al., 2016, Studholme et al., 2019)
Phytophthora lateralis	(Feau et al., 2016, Quinn et al., 2013)
Phytophthora litchii	(Ye et al., 2016)
Phytophthora megakarya	(Ali et al., 2017)
Phytophthora multivora	(Studholme et al., 2016)
Phytophthora nicotianae	(Liu et al., 2016)
Phytophthora palmivora	(Ali et al., 2017)
Phytophthora parasitica	(Shan and Hardham, 2004)
Phytophthora pinifola	(Feau et al., 2016)
Phytophthora pisi	(Hosseini, Unpublished)
Phytophthora plurivora	(Vetukuri et al., 2018b)
Phytophthora pluvialis	(Studholme et al., 2016)
Phytophthora ramorum	(Tyler et al., 2006)
Phytophthora rubi	(Tabima et al., 2017)
Phytophthora sojae	(Tyler et al., 2006)
Phytophthora taxon totara	(Studholme et al., 2016)
Phytophthora x alni	(Feau et al., 2016)

1.7 Field knowledge

The majority of fundamental *Phytophthora* research has currently been studied under controlled, laboratory conditions. Understanding the molecular mechanisms that *Phytophthora* spp. use to initiate and prolong infection on hosts is important in the ever evolving arms-race between pathogen and host. With an increased knowledge on a molecular level, the likelihood of developing resistant hosts or an efficient treatment increases. However, there is relatively little known about *Phytophthora* biology and the interactions with hosts in a field environment.

Phytophthora capsici is known to cause damping-off, leaf blight, wilting and root, stem and fruit rot in susceptible hosts in a field environment (Granke et al., 2012). Disease symptoms are also known to vary on host choice (also within cultivars) and depend on the local *P. capsici* population - some isolates are also more virulent that others (Kim, 1992, Granke et al., 2012, Silvar et al., 2006).

Some isolates, including the *P. capsici* reference isolate, have never been responsible for any outbreaks in a field environment. An important part of bridging the gap between lab and field knowledge is researching isolates that have been responsible for field outbreaks rather than genetically altered and lab derived strains. In order to work with these isolates, they first have to be identified in the field and isolated for further research.

1.8 Plant pathogen diagnostics

Efficient pathogen diagnosis is vital for farmers and crop researchers worldwide. However, there is a significant amount of questions that pathologists need to think about when diagnosing disease. For example:

i) What part of the plant is affected? Is it root derived? Is it localised to a particular area?

ii) Are other plants or other surrounding plant species affected? What is the scale of the problem?

iii) Are there any outstanding abiotic factors to consider? Including drought, irrigation, soil pH, soil nutrients, chemical deposits (herbicides or pesticides), air pollution, temperature, extreme weather.

iv) Is there any phenotypic evidence that would suggest which pathogen may be the causal agent? Distinctive markings/lesions on infected tissues can suggest which group of organisms may be causing disease. For example, mosaic patterns on leaves would suggest a viral pathogen whereas spore formation on surface tissue would indicate a fungal or oomycete pathogen.

It is crucial that plant pathologists can take all of these factors into consideration before arriving at a final conclusion as to what pathogen may be responsible.

1.9 Current molecular diagnostic methods

There is a wide range of diagnostic tools for multiple plant pathogens that are available to farmers. Some methods include, enzyme-linked immunosorbent assay (ELISA) tests, DNA amplification, microscopy, double stranded RNA analysis (Putnam, 1995). More recently, on-site DNA sequencing has become a reality with portable Nanopore technology (Chalupowicz et al., 2019). The diagnostic method of choice needs to be reliable, accessible and workable in order for adoption in an efficient management strategy. There is often a trade off in diagnostic tools between sensitivity (the ability to identify all incidences of the pathogen) and specificity (the ability to identify particular

incidences of a pathogen). Receiver operating characteristic (ROC) curves can compare sensitivity and specificity across a range of values to determine the optimal cut off values of specificity and sensitivity for clinical diagnostic tests (Florkowski, 2008). For instance, a test with high specificity and low sensitivity will likely predict more positive results for a disease resulting in many false negatives, whereas high sensitivity and low specificity will result in many false positives but fewer false negatives (Figure 1.3). The accuracy of a diagnostic test is often shown as the area under the ROC curve. A higher accuracy results in the values predicted in the test to be true.

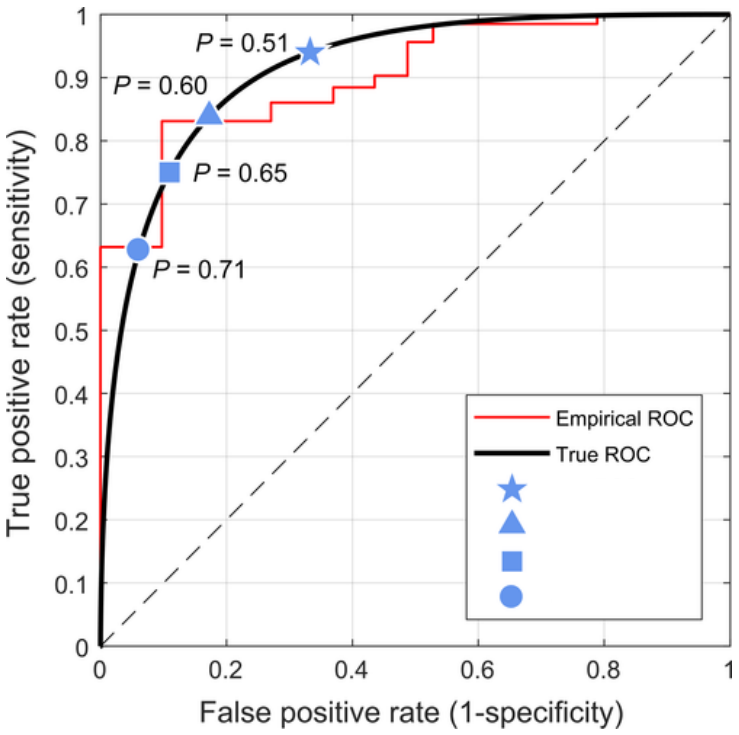


Figure 1.3. An example of a ROC curve adapted from Skelsey (2017) showing the viability of inoculum in fields encompassing *Phytophthora* infection. There are two lines shown - empirical (data) in red and true (binormal) in black. The four blue markers are probability values ranging from less strict to most strict. In this instance, the best performing cut off with the highest sums of sensitivity and specificity was the triangle (8.0 and 9.1 respectively). The star had a high sensitivity (>8.0) but low specificity (<9.1). The other two (circle and square) have low sensitivity (<8.0) and high specificity (>9.1).

However, diagnostic tools are not always specific and results can sometimes be misleading due to false positive or false negative readings. A number of devices are on the market to diagnose multiple *Phytophthora* species. The Department of Environment, Food and Rural Affairs (DEFRA) and the Forestry Commission in the United Kingdom are currently using portable, lateral flow immunoassay (LFT) devices from Pocket Diagnostics® (Kox et al., 2007). Similar to pregnancy tests, these handheld tests give a positive or negative result dependent on the presence of a particular analyte or antibody in the sample. Whilst it is unclear whether the Pocket Diagnostic® LFT can identify *P. capsici*, studies have shown that further diagnosis is needed for an accurate diagnosis, especially on emerging outbreaks (Kox et al., 2007). Pocket Diagnostics® are suitable for prescreening – rejecting true negatives – but for confirmation of *Phytophthora* and species identification, further tests are required.

The current method to diagnose *P. capsici* disease from the field involves isolating the pathogen from infected samples onto media, extracting DNA and using PCR to identify a particular region that is presumed to be species-specific. This method is not only time consuming and laborious but is very prone to contamination and requires a well trained laboratory technician to minimise these errors. The PCR primers are not as discriminatory as previously thought due to the amplification of closely related species (Chapter 2).

There is still a desired need for an efficient *P. capsici* diagnostic tool from both farmers and researchers. Farmers will benefit from a timely diagnosis by having more time to manage their crops to minimise further infection and reduce further crop losses. Researchers and plant pathologists will also have a robust method of diagnosing multiple pathogens without the need to isolate and grow the pathogen, saving time, resources and money.

1.10 Management of disease

There are three management strategies that farmers can use to prevent spread of *P. capsici*: crop rotation, effectual irrigation and application of chemical compounds.

Crop rotation is a common method used to try and prevent infection year after year by growing different crops annually on the same site – usually followed by a non-host resistant crop. However, *P. capsici* can survive many years in the form of thick walled oospores and chlamydospores in the soil and can infect a range of hosts. Therefore, rotating the crops is not always effective.

Effective irrigation can reduce the spread of *P. capsici* infection. As *Phytophthora* zoospores are chemotactically attracted to plants in water, watering closer to the ground will reduce the splash distance therefore reducing the chance of transporting zoospores. Also, planting crops underneath cover such as a poly tunnel will protect against splashback from rain water. However, this is not always possible for large scale farms, especially in developing countries.

For many years, chemical compounds have been used to treat our crops to eradicate pests and diseases in order to maintain high yields. *Phytophthora* outbreaks in the past have been treated with many different chemicals including the phenylamide fungicides, metalaxyl and mefenoxam.

Metalaxyl was used intensively to control a range of oomycete pathogens including *Peronospora tabacina*, *Bremia lactucae* and *Phytophthora* spp. (Parra and Ristaino, 2001). The phenylamide fungicides target polymerase complex 1 in the pathogen which inhibits the biosynthesis of ribosomal RNA (rRNA) (Childers et al., 2015).

However, metalaxyl resistant *Phytophthora* spp. were quickly detected due to the extensive use and the rapid selection of the oomycetes. This prompted the introduction

of mefenoxam to manage *Phytophthora* outbreaks (Parra and Ristaino, 2001). Mefenoxam is a more active isoform of metalaxyl at inhibiting rRNA biosynthesis. Mefenoxam resistant *P. capsici* isolates emerged during the 1980s and the use of the fungicide reduced subsequently (Childers et al., 2015). The resistance was due to a mutation in the large subunit of RNA polymerase 1 in the pathogen (Randall et al., 2014).

Pyrimorph – a novel carboxylic acid amine (CAA) based fungicide – has been shown to control *P. capsici* outbreaks and is now patented and used in China and the United States of America (Pang et al., 2014). Pyrimorph works by inhibiting cell wall biosynthesis and energy production.

Despite there being no signs of resistance to Pyrimorph in the field as yet, a number of resistant isolates of *P. capsici* have been grown in the lab (Pang et al., 2013). However, as there are currently no signs of resistance to Pyrimorph in the field, perhaps *Phytophthora* fungicide research should focus on CAA based rather than phenylamide based fungicides. CAA based fungicides inhibit both the cell wall synthesis and energy production of many plant pathogens.

Common practice used in the field to prevent *P. capsici* infection rely on a combination of these methods. Hausbeck (2004) recommends multiple control strategies including planting hosts in a well-drained field, in raised beds (if possible), watering conservatively, removing rotten fruit, frequently applying fungicides (a combination of different types throughout production if possible) and not watering before harvesting.

1.11 Resistance

Due to the consistent insensitivity observed with multiple fungicides against *P. capsici*, the optimal strategy for managing the disease is further research into developing resistant crops (Barchenger et al., 2018). High levels of *P. capsici* resistance have been noted in two crop varieties – Coriollo de Morales 334 pepper line (CM334) and LA407 tomato accession (Foster and Hausbeck, 2010). Despite not being commercial lines of pepper and tomato, crop breeders are working with these particular lines to generate resistant cultivars that farmers and consumers can grow and enjoy. Collaborations with pathogen researchers concentrating on field isolates will help focus efforts that are relevant to current outbreaks. Working with *P. capsici* isolates taken directly from the field will give breeders a better idea of which varieties to develop, saving resources and time.

1.12 Thesis outline

This thesis uses a genomic approach to design, develop and diagnose *P. capsici* isolates from the field. Using a combination of whole genome sequencing tools and computational biology, species and isolate-specific markers were designed for a range of *P. capsici* isolates for a PCR based diagnosis method.

In chapter 2, I will discuss the sequencing methodologies and techniques used to obtain draft genomes to use for functional genomic studies. I used next generation, Illumina MiSeq to sequence three field isolates. Third generation sequencing technology to improve the *P. capsici* reference genome using the MinION (Oxford Nanopore Technology). Due to multiple technical issues, PacBio was also used to generate long read data for the reference isolate, as well as two other *P. capsici* field isolates.

The development and testing of two bioinformatic pipelines will be discussed in chapters 2 and 3. Chapter 2 will focus on OEDs – a pipeline that produces species-specific primers and chapter 3 will focus on PDP – a previously published pipeline that produces isolate-specific primers. Both chapters will explain how the pipelines work and the wet lab PCR testing of the generated primers with *P. capsici* isolates and other *Phytophthora* species.

Chapter 5 focuses on two field trips. The first to Indonesia, where I visited farms across the country in search of *P. capsici* infected pepper plants. Samples were taken back to the lab for isolation, where only 1/20 samples were confirmed to be *P. capsici*. This backs-up claims from plant pathologists that *P. capsici* is difficult to isolate from infected material. The single Indonesian isolate was used in a large phenotypic infection assay with three other *P. capsici* isolates during my second trip – to Enza Zaden in the Netherlands.

The phenotype assay was performed on a recombinant inbred line (RIL) of pepper plants from a cross between a resistant line (CM334) and a susceptible line (Maor). Results indicated that the recently isolated *P. capsici* isolates (including the Indonesian isolate), were far more aggressive in causing disease symptoms in the majority of the RIL population compared to the commonly used *P. capsici* lab isolates (including the reference isolate). This prompts questions for future *P. capsici* research as to what isolates should we focus on for genomic and molecular studies.

Finally, the last chapter will draw all conclusions of this project together and highlight the important points from previous chapters. Future research is also discussed in context of this project and to bridge the gap between lab and field knowledge in regards to *P. capsici* biology.

Understanding the biology of the pathogen, the host and their interaction will help to manage and possibly prevent further outbreaks. Identifying how pathogens target hosts and cause disease in hosts may indicate how to prevent or suppress progression. Likewise, extensive studying of the host immune system during infection may indicate what proteins/biochemical pathways are involved that can possibly be manipulated in a way to prevent or suppress infection progression.

Chapter 2: A Bioinformatic Pipeline to Produce Species-specific Primers for a PCR based Diagnosis Method – Omics-Enabled Detection pipeline (OEDs)

2.1 Introduction

Phytophthora capsici is a devastating pathogen with a broad host range that negatively impacts the production of important food crops. Indeed, *P. capsici*-incited losses, along with costs incurred by chemical control, continue to pose a grave challenge to farmers. To successfully combat *P. capsici* (and pathogens in general), crop production practices, limiting pathogen spread, and early intervention are critical. Therefore, the early detection and diagnosis form an essential requirement to prevent significant crop losses and further dissemination. Current methods of diagnosing *P. capsici* in the field are time-consuming, difficult and require trained specialists to handle and process samples. Most farmers thus have limited access to information that describe crop health.

In this chapter, I present a bioinformatic pipeline that I and others have developed. This pipeline takes (publically) available genome sequences for a range of pathogens (within the *Phytophthora* genus) and identifies regions of genomes that are unique to a given pathogen. By using sequences from multiple isolates within a species, the pipeline can automatically design species-specific primers for validation and deployment in the lab and field respectively.

There has been more than a twofold increase in the number of newly discovered *Phytophthora* spp. over the last decade (Yang et al., 2017). In 1996, there were 58

known *Phytophthora* species within the genus (Erwin, 1996). Currently, more than 150 *Phytophthora* species are defined, a number that is expected to grow in the future (Yang et al., 2017). This increasing number of (closely related) species within *Phytophthora* has rendered the classification and identification on morphological and structural features more difficult. In 1963, *Phytophthora* spp. were categorised into 6 taxonomic groups (Waterhouse, 1963). More recently, there have been a number of phylogenetic analyses on *Phytophthora* on molecular and sequence data which have ordered the genus into 10 clades (Figure 2.1) (Cooke et al., 2000, Blair et al., 2008, Martin et al., 2014, Yang et al., 2017, McCarthy and Fitzpatrick, 2017).

Phytophthora capsici resides in clade 2 as defined by Waterhouse (1963) with *P. tropicalis*, *P. mexicana*, *P. glovera* and *P. plurivora* (Figure 2). *P. tropicalis* infects woody and perennial crops and does not infect pepper whereas *P. capsici* infects a large range of vegetable crops, including pepper (Enzenbacher et al., 2015). Despite their distinct host ranges, *P. capsici* and *P. tropicali*s were originally thought to be the same species due to similar morphological structure of the sporangia. However in 2001, they were declared separate species (Aragaki and Uchida, 2001).

There are currently 110 genome assemblies for 26 species of *Phytophthora* that are available from the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/genome/?term=*Phytophthora*). Four of the 26 *Phytophthora* species are in clade 2 – *P. multi*vora, *P. plurivora*, *P. colocasiae* and *P. capsici*. However, there is no genome sequence for *P. tropicalis*.

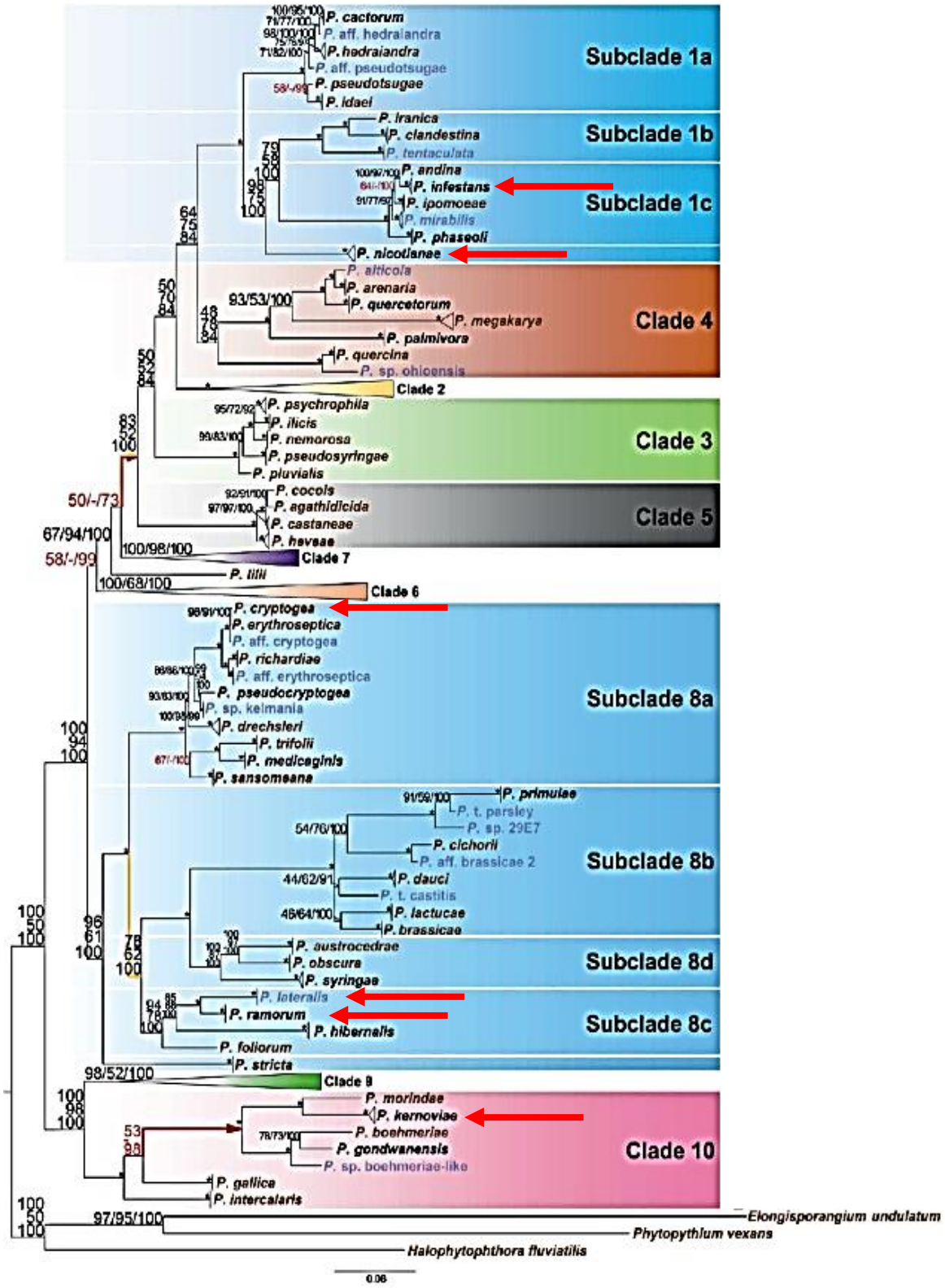


Figure 2.1. *Phytophthora* phylogenetic tree from Yang et al. (2017) based on concatenated sequences of seven nuclear genetic markers. Topology and branch lengths of maximum analysis are shown. Red arrows indicate the species that were used in the designing of OEDs. (*P. parasitica* (clade 1) not shown). Figure 2.2 shows the *Phytophthora* species used for designing OEDs in Clade 2. The remaining 4 species are found in clade 7: *P. cinnamomi*, *P. cambivora*, *P. rubi*, *P. fragariae* and *P. sojae*.

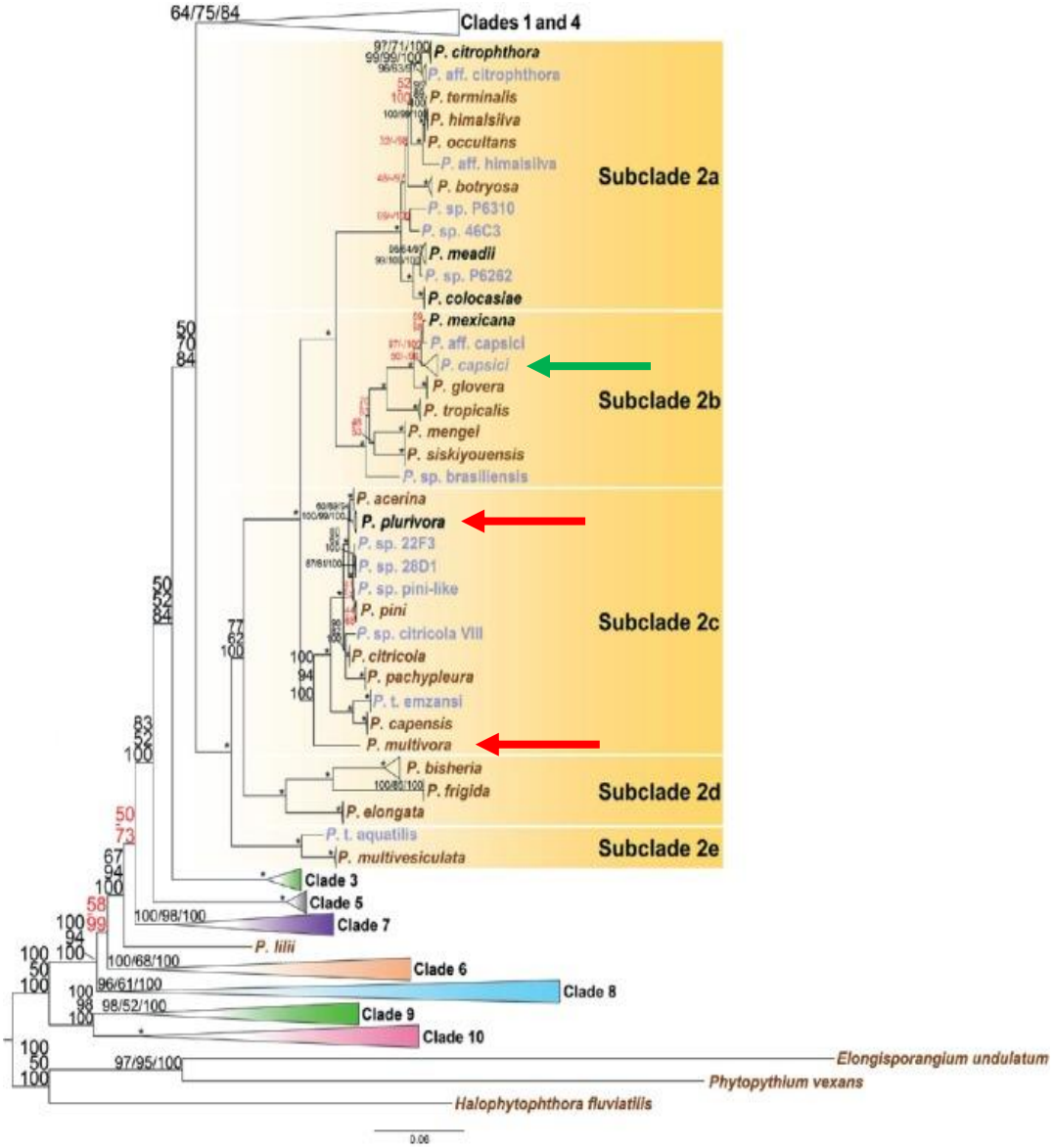


Figure 2.2 Clade 2 from the *Phytophthora* phylogenetic tree from Yang et al. (2017). The green arrow indicates the location of *P. capsici* in clade 2b. Red arrows indicate species used in the designing of OEDs.

2.1.1 Current diagnostic procedure for identifying *P. capsici*

A common method of diagnosing *P. capsici* involves the use of PCR to identify the internal transcribed spacer 1 regions (ITS1). These regions were thought to be conserved amongst species and are species-specific. However, two different species can share the same ITS1 sequence and also, multiple variants of ITS1 sequences can

be found in a single species (Martin et al., 2012). A single set of primers would be ideal for a species-specific diagnosis, but the amplified region would have to be present in all isolates, accounting for sequence variety between isolates, and not found in off-target species. For a species-specific primer set to be used to diagnose *P. capsici* isolates, the candidate primer sets need to be tested with multiple target isolates and off-target species to confirm the hypothesis. Currently, using ITS1 primers would entail further amplicon sequencing to confirm or deny a positive diagnosis. This is inconvenient, especially if used for field applications. For the majority of *Phytophthora* species, there are very few markers for diagnosing on a species level (Kong et al., 2003, O'Brien, 2008, Bhat, 2010).

There are a range of computational pipelines for primer design to diagnose various pathogens from the field (Pritchard et al., 2012, Rodrigues-Luiz et al., 2017, Giordano et al., 2018, Dreier, 2019a). Although similar in output, the methods to arrive at species-specificity differ slightly. Dreier et al (2019) identifies and selects core gene regions that are specific to certain species, then designs primers within those regions. In contrast, Rodrigues-Luiz *et al* (2017) designed taxon-specific primers by targeting either microsatellite markers (short sequence repeats – SSRs) or orthologue genes.

Traditional plant pathogen diagnostic methods include morphological and serotyping, and are not sufficient at discriminating between closely related species and sub-species (Martin et al., 2012). The use of modern computational tools has increased due to the rise in access to available genomic data and computational programs, and the decrease in cost of whole genome sequencing. These methods have made it easier to discriminate and distinguish between closely related species and taxa. Also, for pathogens that have been challenging to diagnose in the past, computational tools have proven to be effective (Giordano et al., 2018).

This chapter focusses on the construction and use of a bioinformatic pipeline that identifies multiple species-specific markers that can be used to detect *P. capsici*. As *P. capsici* has a wide host range and poses a threat to multiple farmers worldwide, it is imperative that a more effective diagnosis method is used.

2.2 Methods

2.2.1 OEDs pipeline construction

The Omics-Enabled Detection pipeline (OEDs) is a modular pipeline used to generate species-specific PCR primers written in Python (Figure 2.3). A target *P. capsici* reference genome with three sequences of *P. capsici* isolates (AD84, Q108 and Y006) were used to construct OEDs. The output consists of a list of predicted primers that amplify regions in all isolates of *P. capsici* and are species-specific. Four genomes from other *Phytophthora* species were used as an off-target data set to discriminate between species: *P. cinnamomi*, *P. fragariae*, *P. infestans* and *P. ramorum*. Each section of the pipeline was tested manually before combining all steps in one Python script that can be run from the command line.

2.2.2 Quality control

The first step in OEDs is quality control (Figure 2.3.A). The sequenced reads and genome data are analysed to see if they are in the appropriate file format (FASTA). The read quality is reported using FastQC (Andrews, 2010). After quality control, the pipeline forks into two routes – mapping and assembling.

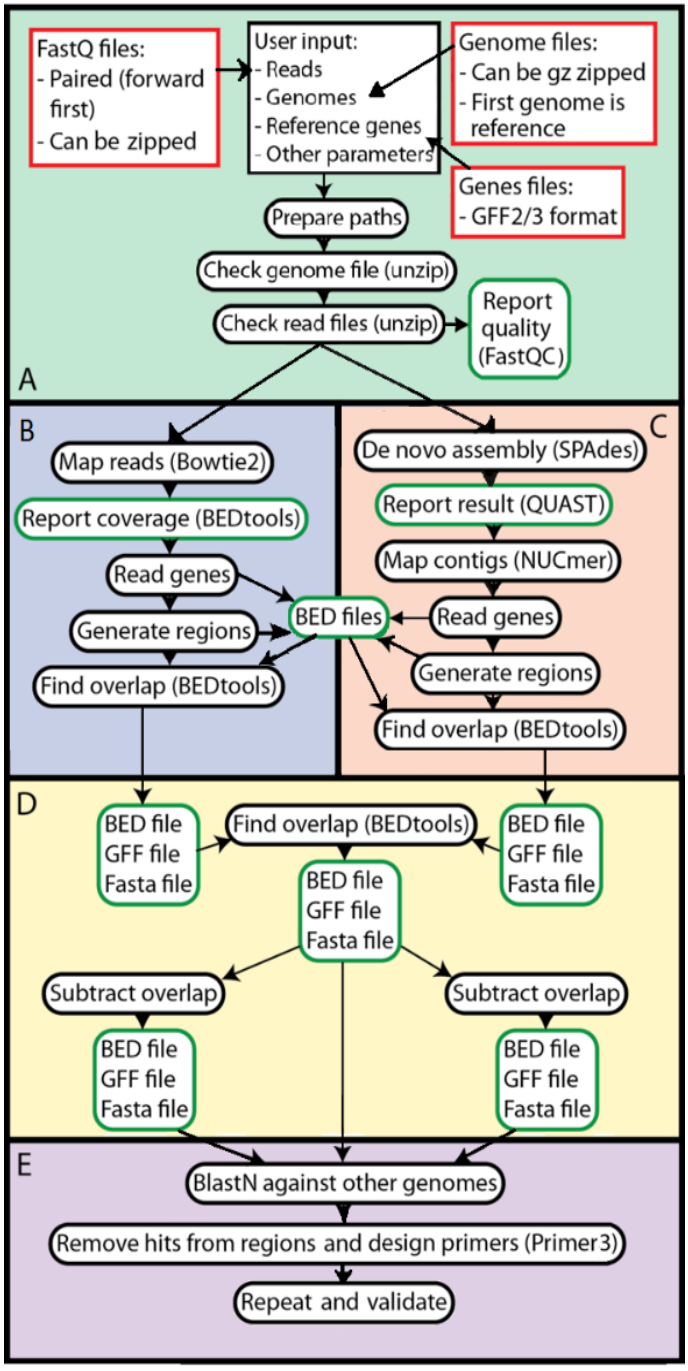


Figure 2.3. OEDs flow diagram. (A): Input files and quality control. (B): Mapping. (C): *de novo* assembly, (D): Comparing mapping and assembly (E): Blast filter against other species, primer design and validate. (Red rectangles contain input files, black rectangles contain the user input required, black ovals show steps in the pipeline and green circles indicate output.) OEDs is in a private repository on GitHub at https://github.com/Rory-McLeod/OEDS. The source code is available in Appendix 1.1

2.2.3 Mapping reads to the reference genome

After the quality control stage – reported using FastQC – the reads of each isolate are mapped to the reference genome using Bowtie2 (Langmead, 2010) with default settings (Figure 2.3.B). Coverage per nucleotide is reported using SAMtools (Li et al., 2009). Regions of the reference genome that are within gene locations are identified with a GFF file. Sequences that are within gene positions and also have a minimum mapping depth of 12 and have a minimum length of 28bp are saved in BED files for each target isolate. The sequences are then aligned using Bedtools (Quinlan and Hall, 2010) with default settings to find conserved regions between isolates. Conserved regions were saved as the "*mapping*" dataset.

2.2.4 *De novo* assembly of field isolate reads

The reads of each field isolate were also assembled *de novo* using SPAdes (3.1.1) (Nurk et al., 2013) with default settings (Figure 2.3.C). The quality of each assembly was reported using Quast (Gurevich et al., 2013) to observe the N50, genome fraction % and number of contigs. The resulting contigs were mapped to the reference genome using NUCmer (Delcher et al., 2002). The sequences were assessed with Bedtools with default settings to find conserved regions between isolates. Conserved regions were saved as the "*de novo*" dataset.

2.2.5 Comparison of mapping and *de novo* assembly

Both datasets – "*mapping*" and "*de novo*" – were aligned using BEDtools to analyse the output of both methods. BEDtools was used to identify common regions in the two datasets whilst also identifying the regions that were unique to each method. Three new datasets were generated (Figure 2.3.D) (Figure 2.4).

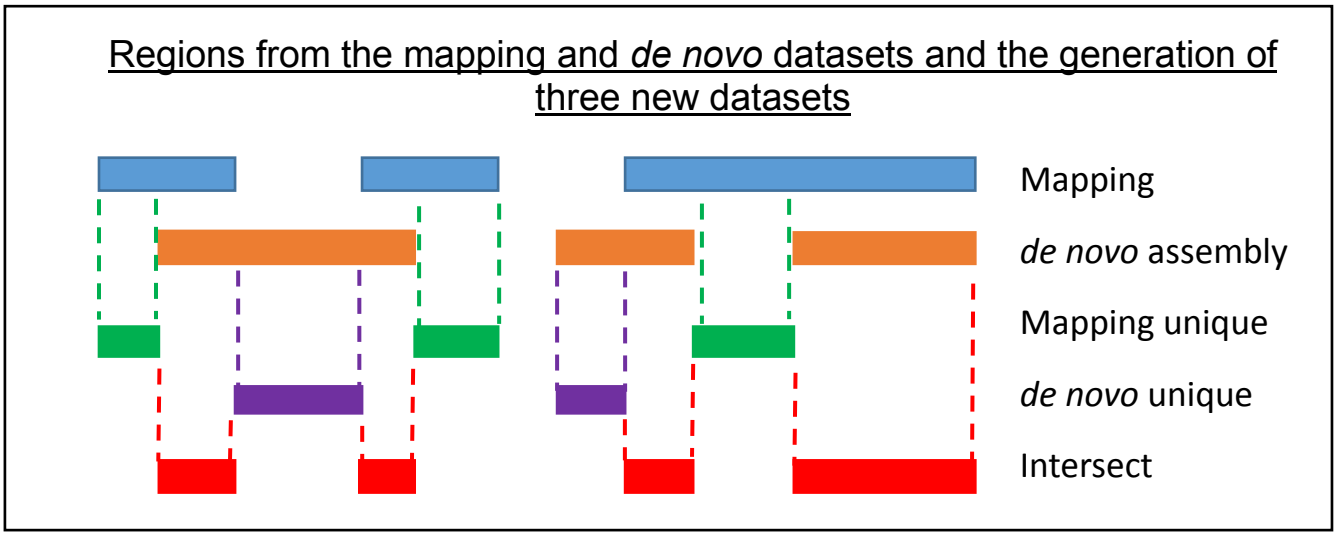


Figure 2.4. The generation of 3 further datasets using BEDtools. The "*mapping unique*" (green) contains the regions identified only by the "*mapping*" dataset and not found in the "*de novo*" dataset. The "*de novo unique*" (purple) dataset contains regions identified only by the "*de novo*" dataset and not found in "*mapping*". The "*intersect*" (red) dataset contains regions found by both "*mapping*" and "*de novo*" datasets.

2.2.6 Blastn against other *Phytophthora* spp.

The penultimate stage in OEDs is a local Blastn (2.2.28) search with default settings (E value: 10.0) (Figure 2.3.E). All of the generated regions from each of the five datasets are the "query" and four off-target *Phytophthora* species genomes were the "subject" (Table 2.1). If the Blastn search returned a hit on default Blastn settings with another *Phytophthora* species, the region is not specific to *P. capsici*, therefore, the region is removed from the dataset. The Blastn results were saved in CSV files.

Table 2.1. Off-target *Phytophthora* genomes used in the OEDs pipeline to generate *P. capsici* specific conserved regions. Publicly available from the NCBI.

Off-target species	Clade	Genome size (Mbp)	Published
Phytophthora cinnamomi	7	77.9	(Longmuir et al., 2017, Studholme et al., 2016)
Phytophthora fragariae	7	73.7	(Gao et al., 2015, Tabima et al., 2017)
Phytophthora infestans	1	228	(Haas et al., 2009)
Phytophthora ramorum	8	65	(Tyler et al., 2006)

2.2.7 Primer design

The final stage in OEDs is primer design using Primer3 (2.3.0) (Figure 2.3.E). Parameters were adjusted to obtain amplicons from species-specific markers that are between 200-500bp in length and have minimum and maximum primer lengths at 15bp and 21bp respectively. Default thermodynamic parameters were used.

2.2.8 Validation of primers using *in silico* and PCRs

The validation of species-specific primers occurs out-with the pipeline with computational (*in silico*) and wet lab PCRs. The *in silico* PCRs were performed using Primersearch (Rice et al., 2000). All primer sets predicted from OEDs for each dataset (mapping, mapping unique, intersect, *de novo*, *de novo* unique) were subject to *in silico* PCRs using Primersearch with 14 off-target *Phytophthora* species (Table 2.2). Primers that showed amplification with other *Phytophthora* species (with a 10% mismatch allowance) were discarded before selecting primers to use in wet lab PCRs.

Table 2.2. 14 *Phytophthora* species used as off-target genomes for the *in-silico* PCRs to reject OEDs primer sets that showed non-specific amplification. Publicly available from the NCBI.

Off-target species	Clade	Assembly size (Mbp)	Published
Phytophthora cinnamomi	7	77.9	(Longmuir et al., 2017, Studholme et al., 2016)
Phytophthora fragariae	7	73.7	(Gao et al., 2015, Tabima et al., 2017)
Phytophthora infestans	1	228	(Haas et al., 2009)
Phytophthora cambivora	7	230.6	(Feau et al., 2016)
Phytophthora cryptogea	8	63.8	(Feau et al., 2016)
Phytophthora kernoviae	10	38.1	(Sambles et al., 2015, Studholme et al., 2016)
Phytophthora lateralis	8	52.4	(Feau et al., 2016, Quinn et al., 2013)
Phytophthora multivora	2	40.1	(Studholme et al., 2016)
Phytophthora nicotianae	1	69-80	(Liu et al., 2016)
Phytophthora parasitica	1	95.5	(Shan and Hardham, 2004)
Phytophthora plurivora	2	41	(Vetukuri et al., 2018b)
Phytophthora rubi	7	74	(Tabima et al., 2017)
Phytophthora sojae	7	95	(Tyler et al., 2006)
Phytophthora ramorum	8	65	(Tyler et al., 2006)

DNA from five isolates of *P. capsici* (four field isolates and one reference isolate), and three other *Phytophthora spp*. (*P. tropicalis*, *P. nicotianae* and *P. katsurae*) was extracted using the DNeasy Plant MiniKit (Qiagen, Valencia, CA) from mycelia grown on pea broth media for 5 days in the dark at 25 °C. A subset of 15 primers generated from OEDs were chosen for wet lab validation (three from "*mapping*", three from "*de novo*", three from "*mapping unique*", three from "*de novo unique*" and three from the

"*intersect*"). PCR set up was as follows; 94 °C for 30 seconds, then a 35x cycle of 94 °C, 56 °C and 72 °C for 10 seconds, 45 seconds and 1 minute 4 seconds respectively. The PCR set up concludes with 75 °C for 5 minutes. The PCR product was run on a 1% agarose electrophoresis gel with Sybr safe DNA stain and viewed under UV light. Bands present were sequenced and cross-referenced with the expected sequence for product confirmation.

2.2.9 Testing the current *P. capsici* diagnostic primers

A wet lab PCR was set up as above for three *P. capsici* isolates (AD84, LT1534 and LT6535) and 2 other *Phytophthora* species – *P. tropicalis* and *P. cryptogea* – using 2 ITS1 primer sets (Silvar, 2005). Both primer sets amplify the same region at different lengths 452bp and 595bp. The forward primer is the same for both primer sets but the reverse primers differ.

2.3 Results

2.3.1 OEDs generates species-specific primers

Species-specific primers were generated using OEDs. All primers generated from OEDs amplified all *P. capsici* isolates in PCR testing (including one isolate not involved in the designing of OEDs – AP154, discussed further in Chapter 5). The primers were also tested for non-specific amplification with four off-target *Phytophthora* species (including the closely related species *P. tropicalis*). Eight of the 15 primers tested in PCRs amplified *P. tropicalis*. No other signs of amplification in off-target *Phytophthora* species was observed.

2.3.2 Mapping – high alignment rate and low coverage depth

Between 73-80% of the paired end reads for each field isolate mapped to the reference genome using Bowtie2 (Table 2.3).

Table 2.3. The alignment rate for the reads of three field isolates of *P. capsici* when mapped to the *P. capsici* reference genome as part of OEDs.

***P. capsici* Isolate**	**Alignment rate (%)**
AD84	74.28
Q108	83.39
Y006	76.91

Despite the alignment rate, the average coverage depth was far from the theoretical coverage depth. Theoretical coverage depth is the number of times a base is read during sequencing and is calculated by:

$$\text{Theoretical cov depth} = no.of\ reads \times \left(\frac{avg\ read\ length}{assembly\ length}\right)$$

The theoretical coverage depth for AD84, Q108 and Y006 was 46.1, 36.0 and 43.8 respectively. However, actual coverage depth was reported as 12.5, 12.1 and 15.5 for AD84, Q108 and Y006 respectively.

There were 35,659 regions in the "*mapping*" dataset that were all considered to be conserved between all 4 *P. capsici* isolates (Table 2.4).

Table 2.4. The number of regions, Blast hits against off-target *Phytophthora* species and primer sets for each data set throughout the process of OEDs.

	Mapping	**Mapping unique**	***De novo***	***De novo* unique**	**Intersect**
Conserved regions	35,659	2,810	52,404	52,352	33,792
Number of Blast hits	633,906	7,629	737,601	147,722	626,509
Number of regions removed after Blast	7,776	291	19,582	15,184	7,535
Number of regions after Blast – for primer design	27,883	2,519	32,822	37,168	26,257
Generated primer sets	10,376	301	107,379	114,165	10,095

2.3.3 *De novo* assembly

The "*de novo*" dataset resulted in 52,404 regions that were conserved amongst all four *P. capsici* isolates (Table 2.4). The quality of the SPAdes assemblies for all three field isolates of *P. capsici* was reported using Quast (Table 2.5). When compared to the reference genome (10,760 contigs and 397,000 NG50), the three field isolates had more contigs and were considerably shorter as confirmed by the NG50 – defined as the length of the shortest contig at 50% of the total genome length. Also, only around 50% of the reference genome is covered by contigs for any of the three field isolates.

Table 2.5. The SPAdes *de novo* assembly Quast report for three field isolates (Illumina data).

	AD84	**Q108**	**Y006**
# contigs	23,636	27,271	17,439
Total length (>=0bp)	40,803,527	44,026,057	53,272,093
Total length (>=1000bp)	29,633,252	31,608,204	46,814,684
N_{50} (bp)	2,042	1,864	5,389
NG_{50} (bp)	835	974	3,639
Genome fraction (%)	46.594	56.561	62.263

2.3.4 Comparison of mapping and *de novo* assembly datasets and Intersect dataset

The output of the "*mapping*" and the "*de novo*" datasets were analysed using BEDtools where a further three datasets were generated: the "*intersect*" dataset, the "*mapping unique*" dataset and the "*de novo unique*" dataset (Figure 2.3). The number of regions identified in each of these data sets are reported in Table 2.4.

2.3.5 Blastn search and filtering

The local Blastn search with default settings returned a list of regions that were found in *Phytophthora* species other than *P. capsici*, therefore not *P. capsici* specific. These regions were discarded from the dataset. This resulted in 20% (7,776) and 37% (19,582) of regions removed from the "*mapping*" and "*de novo*" datasets respectively. 291 and 15,184 regions were removed from the unique datasets ("*mapping unique*" and "*de novo unique*" respectively) with 7,535 regions removed from the "*intersect*" dataset.

2.3.6 Primer design and *in silico* PCR

Primers were generated from sequences that passed the Blastn screen using Primer3 (Table 2.4). This resulted in 10,376 primers for the "*mapping*" dataset, 107,379 primers for the "*de novo*" dataset, 301 primers for the "*mapping unique*" dataset, 114,165 primers for the "*de novo unique*" dataset and 10,095 primers for the "*intersect*" dataset. All of these primers are predicted to be specific to *P. capsici*.

Table 2.6. The number of primers after the *in silico* PCR using Primersearch

	Mapping	**Mapping unique**	***De novo***	***De novo* unique**	**Intersect**
No. of primers before *in silico* PCR	10,376	301	107,379	114,165	10,095
No. of primers after *in silico* PCR	8,300	236	78,386	82,198	8,076

The primers generated from OEDs were tested for cross-hybridisation in off-target genomes (Table 2.2). Between 20 and 30% of primers were removed from the dataset after performing *in silico* PCRs using Primersearch with a 10% mismatch rate (Table 2.6).

2.3.7 PCR validation

A subset of 15 predicted *P. capsici* specific primers from OEDs were randomly selected for PCR (Table 2.7). Five primers were randomly selected from three datasets – "*mapping unique*", "*de novo unique*" and "*intersect*". The primers were

tested for *P. capsici* specificity with *P. capsici* isolates and other *Phytophthora* species (Figure 2.5). All primers generated from OEDs amplified all isolates of *P. capsici* (excluding one – Primer set 8). Primer set 8 did not amplify AP154 in one of the two PCRs. AP154 was not included in the development of OEDs. This isolate was recently isolated from the field (after OEDs had been developed). Eight of the *P. capsici* specific primers also amplified *P. tropicalis*. This included primer set no. 5 (POI15705) from the "*intersect*" dataset which presented mixed results showing a band for *P. tropicalis* in one PCR and no band for *P. tropicalis* in another (data not shown). No amplification was seen for any other *Phytophthora* species.

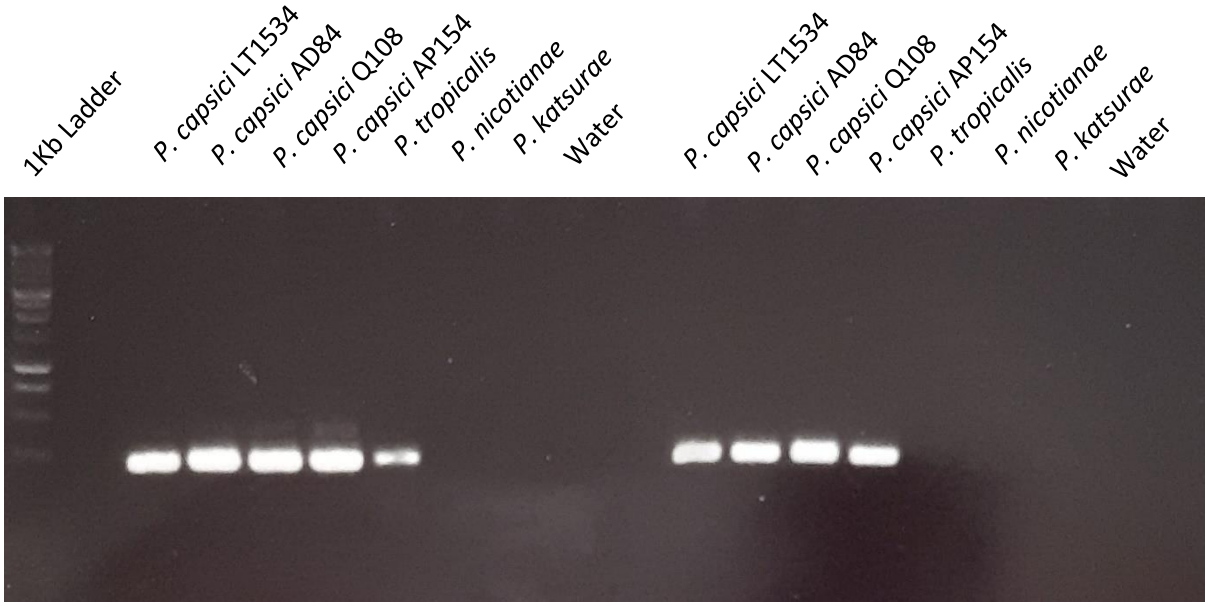


Figure 2.5. Validation of OEDs PCR primer sets 1 (left – POI17144) and 2 (right – POI22580) both at the expected size (220bp).

Table 2.7. The results from the PCR validation of OEDs primers. Green indicates a band present in the electrophoresis analysis, red indicates no band detected. (-) indicates that the sample (DNA) was not used in the PCR set up.

Primer no.	Oligo Name	Dataset	Test	*P. capsici* isolates					*P. tropicalis*	*P. nicotianae*	*P. katsurae*	*P. cambivora*	Water
				LT1534	AD84	Q108	Y006	AP154					
1	POI17144	Mapping Unique	1				-					-	
			2				-			-	-	-	
2	POI22580	Mapping Unique	1				-					-	
			2				-			-	-	-	
3	POI18174	Mapping Unique	1				-					-	
			2				-			-	-	-	
4	POI13258	Intersect	1				-					-	
			2				-			-	-	-	
5	POI15705	Intersect	1				-					-	
			2				-					-	
6	POI12334	Intersect	1				-					-	
			2				-					-	
7	POI23939	*De novo* Unique	1				-					-	
			2				-					-	
8	POI36383	*De novo* Unique	1				-					-	
			2				-					-	
9	POI49051	*De novo* Unique	1				-					-	
			2							-	-		
10	POI45474	*De novo* Unique	1				-					-	
			2							-	-		
11	POI26381	*De novo* Unique	1				-					-	
			2							-	-		
12	POI629	Mapping Unique	1							-	-		
13	POI19169	Mapping Unique	1							-	-		
14	POI50088	Mapping Unique	1							-	-		
15	POI11540	Intersect	1							-	-		

2.3.8 ITS1 primers identify *P. tropicalis* as well as *P. capsici*

The previously published, “species-specific” ITS1 primers that are currently used for diagnosing *P. capsici* cannot distinguish between *P. capsici* and *P. tropicalis*. The electrophoresis gel shows amplification in all three isolates of *P. capsici* and also *P. tropicalis* (Figure 2.6).

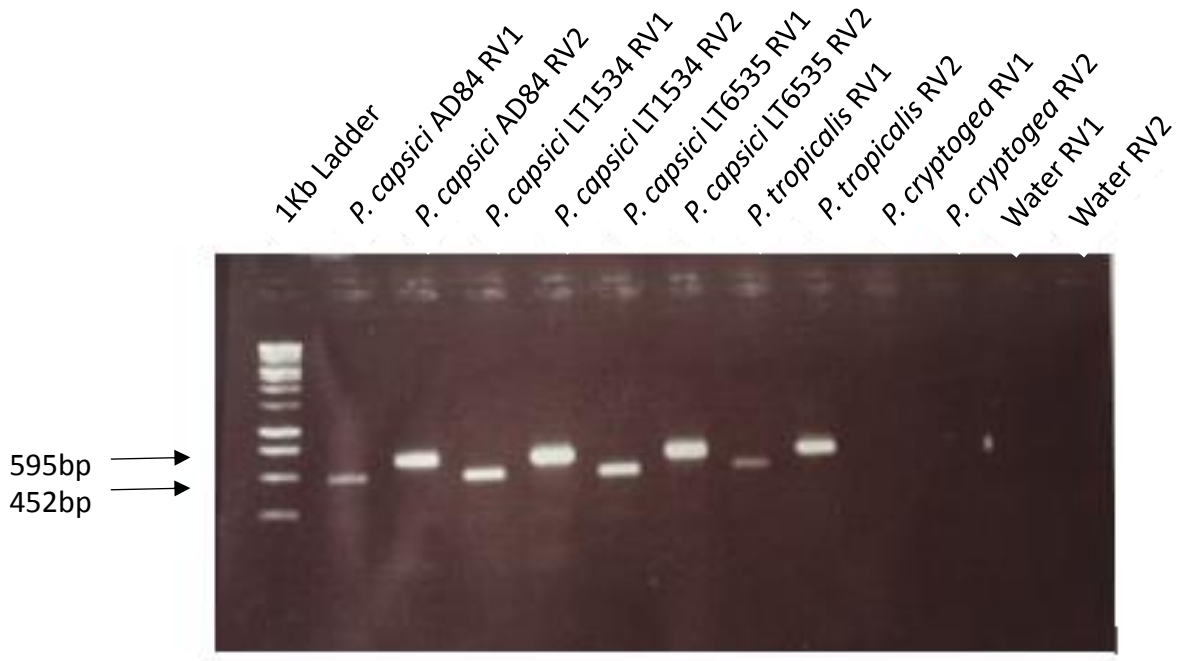


Figure 6. Previously published *P. capsici* specific primers with *P. capsici*, *P. tropicalis* and *P. cryptogea* isolates. Lanes 2-7: *P. capsici* genomic DNA from three isolates (AD84, LT1534 and LT6535). Lane 8-9: *P. tropicalis*. Lanes 10-11: *P. cryptogea.* Lanes 12-13: Water. Each sample has the same forward primer, whereas there are two different reverse primers: RV1 and RV2 resulting in expected sizes of 452bp and 595bp respectively.

2.4 Discussion

I present a bioinformatics pipeline that is suited to the fast and efficient design of primers, able to diagnose Eukaryote pathogens. The OEDs pipeline can take (publicly) available genome sequences along with Illumina sequence datasets to identify genomic loci, suited for diagnostic PCRs. I built and tested the OEDs pipeline for the diagnosis of *Phytophthora capsici* in the field. *P. capsici* is notoriously difficult to diagnose in the field and only few diagnostic primer pairs are currently available. OEDs that can play an important role in pathogen diagnostic research. The wet lab PCR validation of the species-specific primers show that OEDs is successful in identifying *P. capsici* and discriminating between other *Phytophthora* species. All of the primer sets that I have tested in PCRs amplify all *P. capsici* isolates despite

around half of them also amplifying *P. tropicalis*. Identifying multiple marker regions that can be used for diagnosing the pathogen is beneficial due to not having to be continuously reliant on just one or two markers. Having multiple targets can also be beneficial due to the ever-evolving nature of the pathogen. As the pathogen evolves, some markers may not be as reliable for species diagnostics due to sequence changes – having multiple targets resolves this issue. Furthermore, as new isolates are continuously being diagnosed and sequenced, these novel sequences can be incorporated into OEDs to develop marker regions that are relevant and specific to up to date field conditions.

2.4.1 Poor mapping and assembly – downstream analysis problems?

It is likely that the results from mapping and *de novo* steps for the three *P. capsici* isolates may not be optimal. This may be due to the issues experienced during the sequencing of the three field isolates (discussed more in Chapter 4). Despite this, OEDs still produced a large number of candidate, species-specific, diagnostic primer sets. This shows great promise for OEDs despite issues upstream. However, the specificity and sensitivity of OEDs can be improved by firstly incorporating improved sequence data from *P. capsici* isolates and more off-target genomes – especially sequence data from closely related species (*P. tropicalis*). Secondly, the validation would need to involve more predicted primers tested with more isolates of *P. capsici* and more off-target species to indicate how specific and sensitive the primers are.

2.4.2 Mapping *P. capsici* field isolate reads to the reference genome

The sequenced field isolates were mapped to the reference genome to identify regions that were conserved between them. The alignment rate for the field isolate reads was between 73-80%. However, there was a large difference between the

theoretical and actual read depth coverage for all three field isolates. Chapter 4 shows us that only around 50% of the reference genome is covered by each of the field isolates (assembly analysis). This is not ideal for performing comparative genomics (searching for conserved regions in all isolates). We are effectively not searching half of the genome. To rectify this issue, sufficient sequence data is required to have enough genome coverage and read depth to design conserved regions throughout the genome. This gives us the best chance at finding more conserved regions to be used in the downstream analysis.

2.4.3 *De novo* assembly for three field isolates of *P. capsici*

Assembling the three field isolates using SPAdes resulted in a low NG50 value and only 50% of the reads mapping to the genome. Similarly to the mapping stage, these results are far from ideal for designing species-specific markers for multiple isolates as there is an increasing likelihood of missing sequence data and therefore not obtaining markers that would be present and conserved amongst all *P. capsici* isolates. To resolve this issue, similar to the mapping (above), more sequence data from each isolate is required to obtain assemblies of decent length and sufficient read depth.

2.4.4 Comparison of mapping and *de novo* datasets

The mapping and *de novo* assembly stages resulted in 35,659 and 52,404 regions respectively that were larger than 28 bp and had a read depth of >12 (Table 2.4). This may suggest that the *de novo* method was more successful than the mapping method as more regions were identified. Despite the "*de novo*" data set obtaining more regions than the mapping data set, 37% of the regions were removed after having a Blastn hit with another *Phytophthora spp*.. Only 20% of the "*mapping*" regions were removed. The number of candidate regions in both data sets were

similar after the Blast stage (27,883 “*mapping*” and 32,822 “*de novo*”). A similar percentage of primers from each dataset were discarded after the *in silico* PCR. Between 20 and 30% of the primers were removed after *in silico* PCR using Primersearch.

2.4.5 The “*intersect*” dataset has the greatest confidence of a species-specific marker

The “*intersect*” dataset was generated after mapping and the *de novo* assembly. Regions in this dataset (33,792) were present in both “*mapping*” and “*de novo*” datasets. 22% of these regions were discarded for having a hit during the Blastn stage. Detection in both datasets and no Blastn hits from other *Phytophthora* spp. is promising for producing species-specific markers. The 10,095 primer sets generated from the “*intersect*” dataset have greater confidence than primers generated from one of the other datasets alone.

2.4.6 “Mapping unique” and “*de novo* unique” datasets

The two “*unique*” datasets are regions that are found only in that particular dataset (e.g. regions that are in “*mapping unique*” are not found in the “*de novo*” dataset and visa-versa). These two datasets are contrasting from each other in regards to number of regions, Blastn hits and primers produced. The “*mapping unique*” dataset produced 2,810 regions, 10% of which were removed resulting in 301 primer sets generated from Primer3. The “*de novo unique*” dataset generated 52,352 regions, 29% of which were removed resulting in 114,165 primer sets.

2.4.7 Identifying isolates unknown to OEDs in PCRs

The 15 primer sets were all tested at least twice in wet lab PCRs. Introducing AP154 – the *P. capsici* isolate which has yet to be sequenced and was not included in

OEDs development – to the wet lab PCR validation further proved how successful OEDs is at developing species-specific primers. As OEDs was developed using the four *P. capsici* isolates included in the wet lab validation, it was not surprising that the all designed primer sets amplified all isolates. However, AP154 was an untested isolate and no sequence data was available. To observe species-specific amplification in all but one wet lab PCRs, shows that OEDs is successful at designing species-specific primers that can detect unknown/novel isolates in a given species.

2.4.8 The lack of *P. tropicalis* sequence data results in OEDS produces false positive results.

Despite the lack of *P. tropicalis* sequence data available, OEDs has proven to generate primers that can differentiate between the two closely related species (8/15 primer sets) (Table 2.7). However, the PCR images (Figure 2.5) show that there is still an inconsistency in the candidate primers being truly species-specific when *P. tropicalis* is also amplified. To correct this, sequence data from *P. tropicalis* is needed. A *P. tropicalis* genome incorporated into OEDs would greatly improve the primers generated to be truly species-specific. In an ideal situation, OEDs would benefit from genome sequences from all *Phytophthora* spp., other oomycetes and further fungal genomes. A major pitfall in this “ideal situation” is the lack of computational storage/power needed to run OEDs with considerably more genomes.

2.4.9 The current *P. capsici* diagnostic primers are not fit for purpose

It is clear that the previously published primers that amplify the ITS1 region are not suitable for *P. capsici* diagnosis. Firstly, the existing *P. capsici* ITS1 diagnostic primers also amplify *P. tropicalis* (Figure 2.6). Species-specific primers should only be able to amplify a single species. Secondly, as isolates of *P. capsici* continuously

evolve, having a single region to determine a diagnosis is problematic. OEDs solves this problem by generating multiple primer sets in multiple regions, for example, 82,198 potential primer sets in 37,168 regions ("*de novo unique*" dataset). Despite the lack of *P. tropicalis* sequence data, OEDs generates primers that can distinguish between the two closely related species, whilst also detecting all *P. capsici* isolates. A recommendation for future *P. capsici* diagnostic tests is that the existing primers should be replaced with primers generated from OEDs.

Similarly to *P. capsici*, *Pythium* and fungal species can also infect pepper. To increase the confidence in OEDs in a real life setting would be to incorporate genome sequences from these species to rule them out.

The majority of current *Phytophthora* species-specific PCR diagnostic markers are reliant on very few target sequences. This pipeline has the ability to generate multiple candidate species-specific primers for a single species diagnosis.

OEDs has the ability to design species-specific primers for diagnosing a pathogen of interest. OEDs requires a set of isolates from said species, a reference genome and genomes from closely related species. OEDs has been successful in regards to identifying *P. capsici* and discriminating between other *Phytophthora* species. However, sequencing and incorporating a *P. tropicalis* genome into the pipeline would reduce the number of false positive primer results during PCR validation. Despite the lack of sequence data, a number of primer sets are still able to discriminate between the two species.

Although OEDs has been developed to produce markers for *P. capsici* specificity, the pipeline is not restricted to this pathogen. As long as the input criteria are met, OEDs can theoretically be used to diagnose other pathogens for species-specificity.

Chapter 3: Primer Diagnostic Pipeline – Developing discriminatory primers to differentiate *Phytophthora capsici* isolates

3.1 Introduction

Phytophthora capsici is a devastating pathogen with a broad host range that is prevalent worldwide in hot, tropical conditions. Sexual reproduction of *P. capsici* requires two mating types; A1 and A2 (Lamour et al., 2012b). If both mating types are present and in close proximity, the resulting progeny will contain genetic material from both parents, increasing genetic diversity. In order to compete and evade the host's immune system, genetic diversity is beneficial in the ever-evolving arms race between host and pathogen.

Many phenotypic and genotypic studies on *P. capsici* populations have provided contrasting results in *P. capsici* population biology around the world. Clonal lineages of *P. capsici* have been identified in South America and Northern China and have the ability to survive for long periods (Sun et al., 2008, Hurtado-Gonzales et al., 2008, Hu et al., 2013b, Gobena, 2011). However, the population structure in the USA and South Africa has been described as dynamic with a high level of genetic diversity (Gobena et al., 2012, Lamour and Hausbeck, 2001, Meitz, 2010). Some isolates have been shown to be genetically diverse after infecting different hosts (Silvar et al., 2006). A deeper understanding of *P. capsici* genetic variation within a designated region would help to develop better management and control strategies during an outbreak. A

diagnostic tool that could identify particular isolates of *P. capsici* would be beneficial to farmers in order that the outbreak can be efficiently controlled.

3.1.1 Computational Pipelines as Diagnostic Tools

The emergence of low-cost next generation sequencing (NGS) has enabled affordable sequencing of whole genomes that can be utilised by computational tools for genomic analysis to help diagnose pathogens more quickly and more precisely than before. Although developing bioinformatic tools may require expertise in computational biology, most tools are designed to be user friendly and easy to use by wet lab biologists. There are multiple computational pipelines that are used for diagnosing human and plant pathogens (Rodrigues-Luiz et al., 2017, Giordano et al., 2018, Dreier, 2019b).

A common method marker-based diagnostic for *P. capsici* is dependent on only one genetic location, the internal transcribed spacer region 1 (ITS1). This locus is sub optimal for species diagnosis. For example, it is known that the ITS1 region cannot differentiate between closely related species, including *P. capsici* and *P. tropicalis* (Chapter 2) and also *P. fragariae* and *P. rubi* (Martin et al., 2012). Also, if there is any genetic variation within this region between isolates, misdiagnosis is likely. The pipeline, described in Chapter 2, OEDs, presents a computational pipeline that generates multiple target sites for a PCR-based species-specific diagnoses. This chapter focuses on detecting particular isolates within the *P. capsici* species by generating isolate-specific primers from a range of different *P. capsici* isolates.

3.1.2 Primer Diagnostic Pipeline – Isolate-specific diagnoses

Pritchard et al. (2012) initially published the Find Differential Primers pipeline by designing diagnostic primer sets to identify the isolate of *E. coli* during an outbreak in

cucumber plants in Germany in 2011 which caused 4000 individual infections and 53 deaths across Europe. The dataset included 11 draft genomes of the target isolate (O104:H4) and 69 publically available chromosomal or plasmid sequences of other *E. coli* isolates as a negative sequence set. Individual generated primer sets showed 100% sensitivity with low false positive rates (9%-22%). However, using a combination of two primer sets showed 100% sensitivity and 100% specificity. Using the same mentality for the reasoning behind constructing the pipeline, we decided to employ it for the ultimate goal of diagnosing different isolates of *P. capsici*.

Primer Diagnostic Pipeline (PDP) is a modular pipeline derived from the Find Differential Primers pipeline (available on GitHub at https://github.com/widdowquinn/find_differential_primers). The published pipeline has seven steps – validate input config file, concatenate sequences, identify features, predict primer locations, cross amplification, BLAST screen and classify. PDP has the ability to produce diagnostic primer candidate specific at species, sub-species and isolate level (depending on the dataset). The sub-species-specific primers can group particular isolates together whilst distinguishing them from other isolates. This may be due to changes in the sequence where there is a clear difference between isolates in the same species. For example, *P. capsici* isolates in Asia, may not share the same characteristics or sequence similarity with *P. capsici* isolates in South America.

The challenges facing farmers today is the lack of an efficient *P. capsici* diagnostic tool to correctly identify the pathogen causing disease on crops. It is important for species-specific diagnostic tests to identify all isolates in a defined species – this is covered in Chapter 2. This chapter discusses a second computational pipeline that can be used to design isolate-specific primers for a PCR based diagnostic method. Identifying the particular causative isolate or isolates, will give valuable information for

P. capsici population biology within the region to see how the species is evolving and how best to control the outbreak.

3.2 Methods

3.2.1 Input sequence used during the pipeline

PDP takes genome sequences as input data. Four draft genomes of four *P. capsici* isolates were used in this pipeline as targets (isolates we are designing diagnostic primers for). The four isolates include the published reference *P. capsici* isolate (LT1534) and three field isolates (AD84, Q108 and Y006 – discussed in Chapter 4) (Lamour et al., 2012a). Fourteen genomes of *Phytophthora* species were also included in the pipeline as off-targets (not *P. capsici*) to be able to prevent designing non-specific diagnostic primer sets (Table 3.1).

3.2.2 Pipeline workflow

The first stage of the pipeline is to produce and validate a config file (Figure 3.1). The config file contains information about the *P. capsici* isolates (the target isolates). The first column in the file contains the name of the isolate. The second column contains the designated categories assigned to each isolate. These categories are classes that one or more of the isolates belong to. For instance AD84, Q108 and Y006 are all field isolates, therefore, they all have the “field” category, whereas the laboratory isolate, LT1534, has the “lab” category. The pipeline can design primers that can be category specific, therefore designing primers that can amplify only field isolates, or only lab isolates. Finally, the third column contains the filesystem path to the directory holding the isolate sequence data.

Table 3.1. Off-target *Phytophthora* genomes used in PDP pipeline to design *P. capsici*-specific diagnostic primer sets.

Off-target species	**Clade**	**Genome size (Mbp)**	**Published**
Phytophthora cinnamomi	7	77.9	(Longmuir et al., 2017, Studholme et al., 2016)
Phytophthora fragariae	7	73.7	(Gao et al., 2015, Tabima et al., 2017)
Phytophthora infestans	1	228	(Haas et al., 2009)
Phytophthora cambivora	7	230.6	(Feau et al., 2016)
Phytophthora cryptogea	8	63.8	(Feau et al., 2016)
Phytophthora kernoviae	10	38.1	(Sambles et al., 2015, Studholme et al., 2016, Studholme et al., 2019)
Phytophthora lateralis	8	52.4	(Feau et al., 2016, Quinn et al., 2013)
Phytophthora multivora	2	40.1	(Studholme et al., 2016)
Phytophthora nicotianae	1	69-80	(Liu et al., 2016)
Phytophthora parasitica	1	95.5	(Shan and Hardham, 2004)
Phytophthora plurivora	2	41	(Vetukuri et al., 2018b)
Phytophthora rubi	7	74	(Tabima et al., 2017)
Phytophthora sojae	7	95	(Tyler et al., 2006)
Phytophthora ramorum	8	65	(Tyler et al., 2006)

```
# Config file for all P. capsiciseq input
# Name      Categories  Computational_Path_to_Genome_Sequence
AD84  Pcap,field,AD84   location/of/genome/sequence/AD84_genome.fasta   -
Q108  Pcap,field,Q108   location/of/genome/sequence/Q108_genome.fasta   -
Y006  Pcap,field,Y006   location/of/genome/sequence/Y006_genome.fasta   -
LT1534      Pcap,lab,LT1534   location/of/genome/sequence/LT1534_genome.fasta -
```

Figure 3.1. An example of an input config file which contains the necessary information required for PDP to run and discriminate between isolates. The file contains the name of the isolate, a list of categories the isolate belongs to and the location of the genome sequence.

The second stage of the pipeline, “fix_sequences”, concatenates the genomes into one consecutive nucleotide sequence – a pseudo chromosome – for each isolate in the config file (to enable diagnostic primers from fragmented draft genomes). Regions containing deletions and ambiguity symbols are replaced with N to ensure only ATCGNs are present. This stage also creates a json file with the information in the config file with the addition of the location of the new, “fixed” concatenated sequence file.

Primers are designed using ePrimer3 that conform to default thermodynamic parameters including GC content and melting temperature (Rice et al., 2000). The primers can also be designed in any desired regions either by the inclusion of a GFF file or from Prodigal output (for bacterial genomes).

The primer sets are filtered by the removal of duplicate primers and a Blastn screen. By using specific thermodynamic settings to design primers, it is likely that identical primer sets will be created for individual isolates. The duplicates are removed to reduce the computational load in the downstream analysis. The primers are then screened against a Blastn database containing off-target (other *Phytophthora*)

genome sequences (Table 3.1). This step is optional, but recommended. If primer sets amplify regions in any of these off-target genomes, they are removed from the dataset as they are non-specific to *P. capsici*.

The primer sets that do not show any non-specific amplification are further analysed to identify potential cross-hybridisation with the other target input genomes using Primersearch and are only retained if they are predicted to amplify the desired product sizes. Primers that cross-hybridise with one or more isolates are assigned a label for each genome that it is predicted to amplify. After all the Primersearch results are

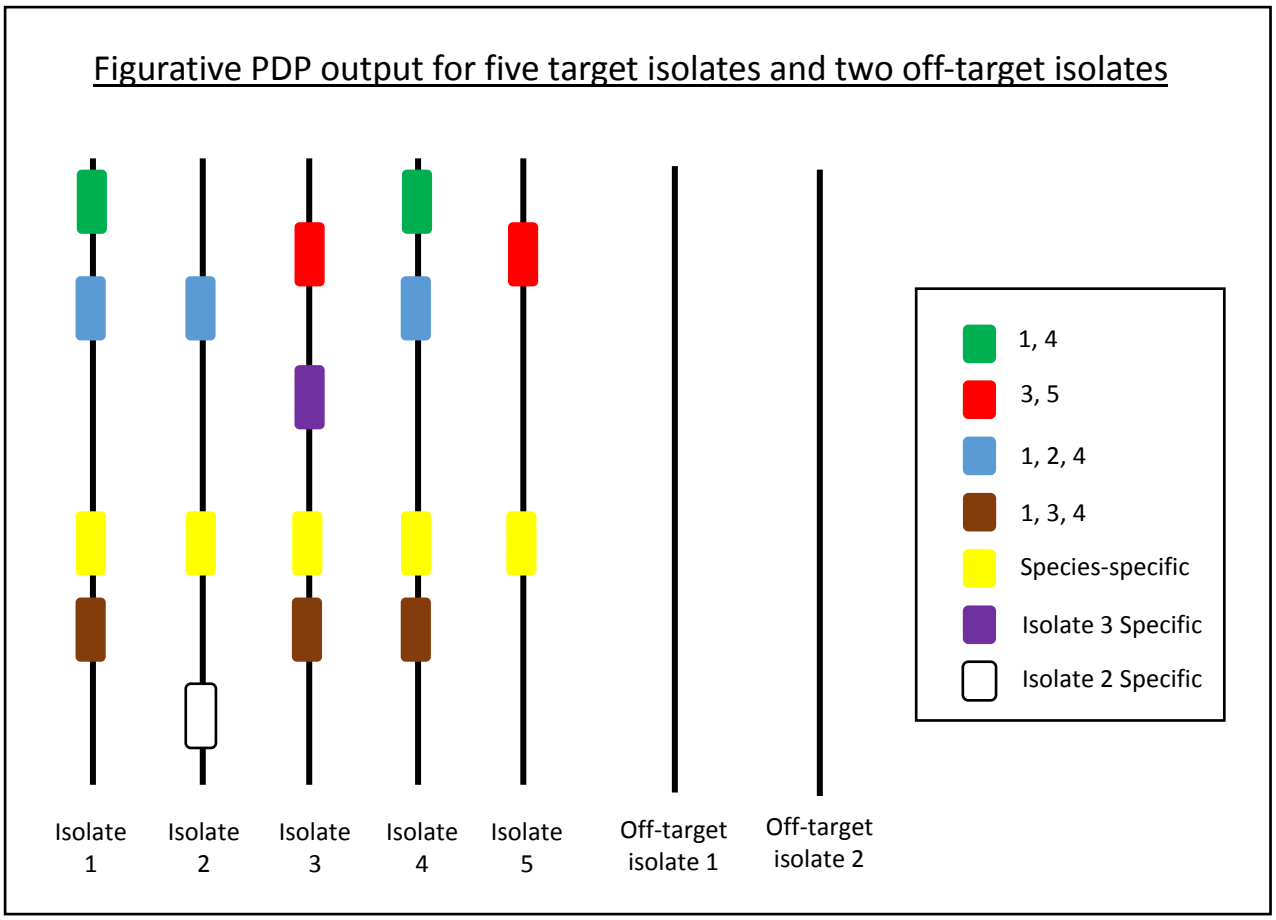


Figure 3.2. An example of a figurative representation of the PDP output from the classify step. Amplicon genomic locations across 5 different isolates of a target organism are highlighted on the left in different colours. The colour denotes the category that the primer set (amplicon) belongs. For example, Isolates 3 and 5 may be field isolates and 1, 2 and 4 are lab isolates. The red group would be classified as field isolate specific primer sets and the blue group would be classified as lab isolate specific primer sets. This is all assuming that these primers are not present in any other off-target organism.

processed, the genomes associated with each primer set are compared to the genomes associated with each category in the config file. If the two sets match, the primer set is considered specific for that category (Figure 3.2).

3.2.3 Incorporating an effector-coding sequence dataset into PDP to identify primers in effector-coding regions

To generate primers that amplify only from known effector-coding regions, a FASTA file containing the sequences of known *P. capsici* effector genes was also incorporated into the pipeline as a 5th target isolate. A category was added to all target isolates in the config file indicating that primers generated in the 5th isolate (effector sequences), would be in effector-coding regions. For instance if a primer set was identified in both AD84 and the effector-coding dataset, we could state that the primer is likely to identify AD84 in an effector-coding region. Therefore, the other *P. capsici* isolates in the config file that did not contain that particular primer set, do not have the effector, as it would have been identified during the cross-hybridisation step.

The effector sequences were extracted from the existing reference genome, LT1534, which is also used in the PDP runs. Therefore, if a primer set is found in one isolate (eg AD84) and the effector-coding dataset, it is safe to say that is also found in the LT1534 dataset.

The PDP pipeline was run multiple times to design different number of primers for each target isolate. Initially, 100 primers for each target isolate were generated – this was also part of a pipeline test run. The pipeline was run subsequently to generate 1,000 and 10,000 primer sets for each of the five target isolates. The pipeline was run three times with three different primer set numbers to obtain primers that were isolate-specific.

3.2.4 PCR validation

DNA from five isolates of *P. capsici* (four target isolates and one untested isolate) and three other *Phytophthora* species (*P. tropicalis*, *P. nicotianae* and *P. katsurae*) were used to confirm species and isolate-specific primer sets generated from PDP in wet lab PCRs. PCR was performed as described in Chapter 2.

3.2.5 Validating species-specific effectors for species-specificity

Groups of species-specific effectors were identified by Thilliez (2016). Effectors were deemed to be *P. capsici* species-specific when no hits were returned after using Blastp with a cut off value of 1e-5 against a database containing *P. infestans* effector proteins (Thilliez, 2016). Primers were designed for 11 of these effectors using ePrimer3 with default thermodynamic settings and wet lab PCRs were set up as described in chapter 2 (Rice et al., 2000). The PCRs were performed as above.

3.3 Results

3.3.1 10,000 Primer sets/isolate generated enough data to classify primers into pre-determined categories

PDP was run on three occasions generating 100, 1,000 and 10,000 primers for each isolate (including the effector-coding dataset) in each run. Running PDP with 100 and 1,000 primers for each isolate did not create any isolate-specific primers. Therefore, the results will focus on the 10,000 primer data set as this run generated enough primer sets to be able to classify into groups and to generate isolate-specific predicted primers.

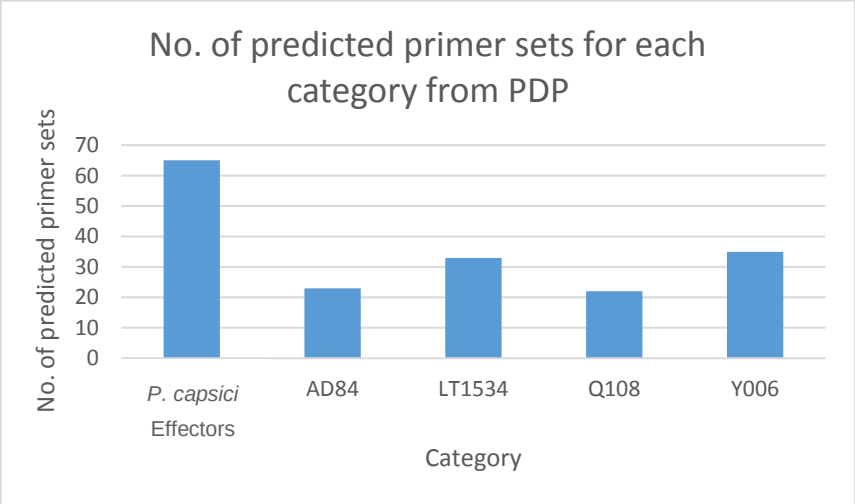
Designing 10,000 primer sets for each target isolate resulted in species-specific and isolate-specific predicted primers (Figure 3.3). There are also no primers designed that

were field specific (from AD84, Q108 and Y006). However, there were lab specific primers (as only one isolate was from the lab, therefore lab and isolate-specific.

3.3.2 Predicted species-specific PDP primers

PDP generated 65 primer sets that were predicted to be species-specific (found in all isolates and the effector-coding dataset) (Figure 3.3). As these primers are all found in the effector-coding dataset as well as the four isolates, we can predict that the primers amplify sequences in all isolates that may be in effector gene locations, indicating that these effectors may be present in these unannotated isolates.

Figure 3.3. PDP results. The number of PDP predicted primer sets that are identified for each category from the config input file after the “classify” step.



3.3.3 Predicted isolate-specific PDP primers

PDP generated predicted *P. capsici* isolate-specific primers. From the 10,000 primers generated for each target isolate, 22-35 primer sets were classified as isolate-specific for each of the four target isolates (Figure 3.3). The effector dataset did not contain any isolate-specific primer sets as all effector sequences are derived from LT1534.

3.3.4 Consistency and validation PCRs confirmed species-specificity

Five predicted species-specific primer sets that are in effector-coding regions were selected randomly and tested in wet lab PCRs with four *P. capsici* isolates and three other *Phytophthora* species. Three of the four isolates used in the PDP run were tested (AP154 was not – it had not been isolated from the field). Primer sets were deemed to be species-specific if there was a visible band on the electrophoresis gel after electrophoresis at the expected size. Two of the five primer sets amplified all *P. capsici* isolates and no other *Phytophthora* species (P3 and P4) (Figure 3.4). Two primer sets amplified all *P. capsici* and *P. tropicalis* isolates (P1 and P2) and one primer did not amplify any *Phytophthora* species (P5) (Table 3.2).

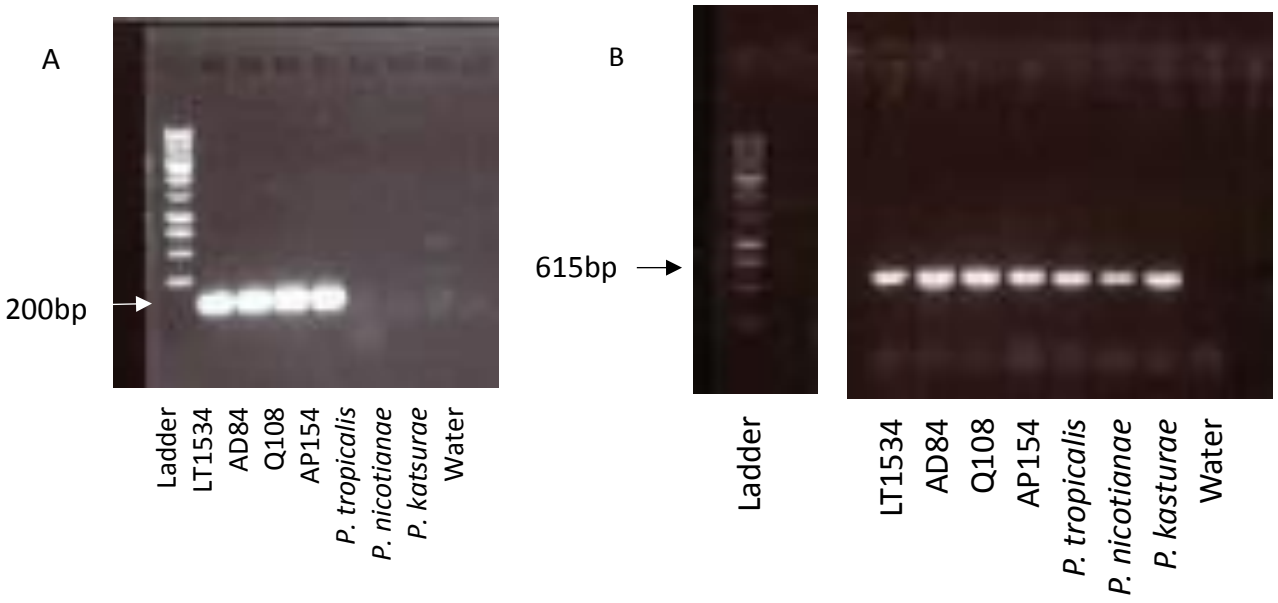


Figure 3.4. A) Electrophoresis gel image of the wet lab validation of species-specific primer – primer set 4 (P4) from PDP showing an expected size band of 200bp. B). Electrophoresis gel showing the positive control for the presence of DNA (PcRPL-18 primers) showing an expected size of 615bp.

Table 3.2. The results from the PCRs of the species-specific predicted primer sets that amplify regions within effector locations. Green with "✓" indicates a positive result with a product on a gel at the correct size. Red indicates no amplification. White indicates that the DNA was not used to test the primer set. Water was included as a negative control and pcRPL-18 was included as a positive control (not in table – see Figure 3.4).

Primer number	Oligo name	Fw/Rv	Sequence	*P. capsici* isolates					*P. tropicalis*	*P. nicotianae*	*P. castanae*	H_20
				LT1534	AD84	Q108	Y006	AP154				
P1	FPc201_CRN83_198	Forward	CCAACAGAACCCTCACCTTT	✓	✓	✓	✓	✓	✓			
	RPc201_CRN83_198	Reverse	CGCAATCCAGACAGTGTACC									
P2	FPc295_RxLR411	Forward	TGGTCGTCTTCGTAACTGGA	✓	✓	✓	✓	✓	✓			
	RPc295_RxLR411	Reverse	ATTGGAGGACTGGGCAATAG									
P3	FPc646_RxLR184	Forward	GGCTTGAACGGGTTGTTAAT	✓	✓	✓	✓					
	RPc646_RxLR184	Reverse	GCCATCGCACTGATAGTTGT									
P4	FPc925_CRN85_180	Forward	TGTACACGGGATGAAATTGG	✓	✓	✓	✓	✓				
	RPc925_CRN85_180	Reverse	TGGGAGCAGTAGGAAAGGAT									
P5	FPc937_CRN26_246	Forward	AGAATTCTCGTGGGCTGTTT									
	RPc937_CRN26_246	Reverse	AAAGCGAATGCTTCCAAATC									

3.3.5 Consistency and validation PCRs with isolate-specific primers did not confirm isolate-specificity for the majority of primers tested

The five predicted isolate-specific primer sets for each *P. capsici* target isolate were tested against their originating genomes (consistency tests). Although predicted to be isolate-specific, PCRs indicated that the majority of predicted isolate-specific diagnostic primer sets were either not isolate-specific, presenting amplification of all *P. capsici* isolates and/or *P. tropicalis*, or showing no signs of amplification for any *Phytophthora* isolates (including the target isolates). Of the 20 primer sets tested, only one was confirmed to be isolate-specific (LT1534 P3) (Figure 3.5). LT1534 P2 has a more intense band with LT1534 compared to the other *P. capsici* isolates, but amplification was also observed for all other isolates tested (not shown).

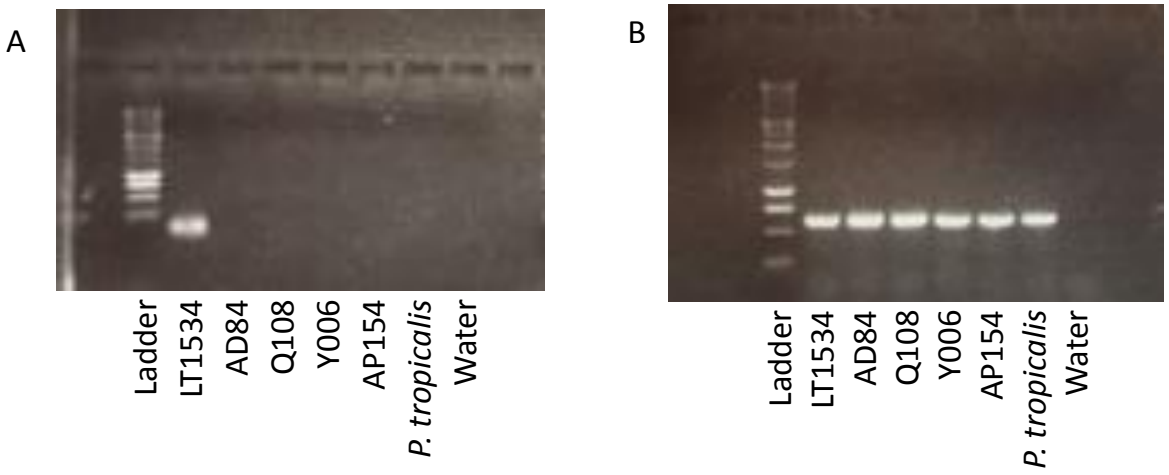


Figure 3.5. A). Electrophoresis gel of the LT1534 isolate-specific primer set (P3) tested with 5 *P. capsici* isolates, 1 *P. tropicalis* isolate and water as a negative control. B) Electrophoresis gel showing the positive control for the presence of DNA (PcRPL-18 primers).

Table 3.3. The results from the PCRs of the MCL generated species-specific effector dataset. Green with "✓" indicates a positive result with a product on a gel at the correct size. Red indicates no amplification. Water was included as a negative control. White indicates that the isolate was not used. LT1534 was the positive control.

Primer number	Oligo name	Fw/Rv	Sequence	*P. capsici* isolates					*P.tropicalis*	*P.nicotianae*	*P.castanae*	H_20
				LT1534	AD84	Q108	Y006	AP154				
P1	FPcRxLR019_GT	Forward	CTGGCCTTGGTTGAAGAGAG	✓	✓	✓	✓		✓			
	RPcRxLR019_GT	Reverse	TTACCCCGTCAAGGATGAAC									
P2	FPcRxLR011_GT	Forward	CGTGACGAAGACGATTACCC	✓	✓	✓	✓		✓			
	RPcRxLR011_GT	Reverse	CGAGACAGCCGTAACCAGAT									
P3	FPcRxLR015_GT	Forward	GCGAATGTTCTCGTTCTTCC	✓	✓	✓		✓	✓			
	RPcRxLR015_GT	Reverse	GAACTCGCCCAAGCTCATAC									
P4	FPcRxLR010_GT	Forward	AGACAGCCTCAACCAGATCG	✓	✓	✓		✓	✓			
	RPcRxLR010_GT	Reverse	GCACGAGAGTATCGCAACTG									
P5	FPcRxLR282_GT	Forward	CCGTCAATGATGAACGTGAC	✓	✓	✓		✓	✓			
	RPcRxLR282_GT	Reverse	CTGGCCTTGGTTGAAGAGAG									
P6	FPcRxLR138_GT	Forward	AAAACGATTCCTGCGAAGTG	✓	✓	✓		✓				
	RPcRxLR138_GT	Reverse	ACGGGTACGCACCTCTAGC									
P7	FPcRxLR030_GT	Forward	GCCGCGTACTCGTTGTAGAG	✓	✓	✓		✓	✓			
	RPcRxLR030_GT	Reverse	AATCCGTCACTGAGGGTCAA									
P8	FPcRxLR298_GT	Forward	TACGGGTATGCATTGTGAGC	✓	✓			✓				
	RPcRxLR298_GT	Reverse	AAAACGATTCCTGCGAAGTG									
P9	FPcRxLR489_GT	Forward	ATTTCTCCTTGCCGCCTAAC	✓	✓			✓				
	RPcRxLR489_GT	Reverse	CGAAAGTGACAGCAGCAAAC									
P10	FPcRxLR467_GT	Forward	ATCTGCTGCCCCATTGTAGA	✓	✓	✓		✓	✓			
	RPcRxLR467_GT	Reverse	CCCTGTCGAGACCATTGAGT									
P11	FPcRxLR457_GT	Forward	ACTGCTCTTTTGGCAACCAC	✓		✓		✓				
	RPcRxLR457_GT	Reverse	CTTTTGAGGACGGCTGAAAG									

3.3.6 Confirming species-specificity from MCL dataset

PCR was used to establish *in* vitro species-specificity of the 11 predicted species-specific effectors from the MCL analysis (Thilliez, 2016) (Table 3.3). Only one of the effectors was found to be *P. capsici* specific, where there was amplification for all *P. capsici* isolates but no amplification in any off-target species (P6 – PcRXLR_138). A further three effectors were also *P. capsici* specific although not amplifying all *P. capsici* isolates (all three failed to amplify one isolate of *P. capsici*). The remaining seven primer sets amplified all *P. capsici* isolates and *P. tropicalis*, but no other off-target *Phytophthora* species.

3.4 Discussion

3.4.1 Designing isolate-specific primers using PDP

The PDP run generating 10,000 primers resulted in between 22-35 isolate-specific primers for each of the four *P. capsici* isolates (Figure 3.3). These were predicted to amplify that particular isolate and no other *P. capsici* isolate, or off-target species. However, wet lab PCRs using the predicted isolate-specific primers with target and off target isolates/species, resulted in results that contradicted the computational output.

There are several possible explanations for the failure to amplify isolate-specific regions with these primers. Missing sequence data from draft genomes from the target organisms. If sequence data is missing from the target organisms, the pipeline will identify a region to be isolate-specific if not found in any other target isolate, resulting in a false positive primer set. If there are mistakes in the sequencing, or assembling of the target genomes, PDP will identify regions that are not a true representative of the genetic makeup of the organism. To rectify this issue, sequences of target

genomes that are involved in the PDP process, would need to have good quality to minimise these potential errors.

PDP designs primers using ePrimer3 with default thermodynamic parameters. When testing these primers in wet lab PCRs, in theory, these should be isolate-specific. However, there may be slight amplification in other isolates that would not have been identified during the design process. This results in PDP designing false positive results (a positive result but not the target isolate). True isolate-specific primers should not amplify any other isolate, therefore specificity of PDP needs to be addressed. To rectify this issue, the thermodynamic parameters could be relaxed in order to pick up any signs of amplification in off-target species or other isolates. However, you could argue to tune the thermodynamic parameters to maximise discriminatory ability but this is reliant on good quality genome data.

3.4.2 Species-specificity – Identifying species-specific effectors

The pipeline was not able to incorporate a GFF file containing the locations of effector genes on the target reference isolate. The tool was not able to read the file as a GFF format. As an alternative strategy, effector sequences were instead incorporated into the pipeline as if they were another target isolate genome – but an isolate containing only effector gene sequences. Due to the effector sequences being derived from the reference target isolate, there could be no “effector specific” primers, as primers capable of amplifying an effector sequence would be identified in at least two isolates. However, 65 predicted primer sets were identified that amplified only all target isolates and the “effector isolate”. These primers are predicted to be species-specific. Five of these predicted species-specific primer sets were selected for PCR validation, with five isolates of *P. capsici* and three isolates from other *Phytophthora* species. The PCR results indicate that two of five primer sets are *P. capsici* specific, with another

two amplifying only *P. capsici* isolates and *P. tropicalis*. One primer set failed to amplify any of the isolates.

This is a promising result as PDP had no sequence data from *P. tropicalis* but was still able to differentiate between the two closely related species. If *P. tropicalis* was sequenced and a high-quality genome included as an off-target genome, we should expect more candidate primers to be species-specific. It was also encouraging to observe that the predicted species-specific primers in effector regions amplified isolate AP154 – an isolate recently isolated from the field with no sequence data available. Similarly to OEDs (Chapter 2), more validation of these primers is required with other *P. capsici* and *P. tropicalis* isolates (not in this study) from around the world to be more confident that these primers are truly species-specific.

One bottleneck for designing primers that are within effector regions, is that the known effectors are all from the reference isolate. As the reference isolate has been regularly sub-cultured on media and not in a natural, field environment, the effector repertoire of this isolate may not contain effectors that are present in field isolates. Therefore, we may be missing out on effectors with diagnostic capability. A possible step to overcome this would be to annotate the three field isolate genomes to identify effectors, including RXLRs and CRNs, then run the pipeline with these regions as targeted regions by the inclusion of a GFF file with the locations on the genomes of the effector genes. This would be interesting to explore in other oomycetes and plant pathogens.

The published pipeline from Pritchard et al. (2012) was successful in generating primer sets that were 100% sensitive and 100% specific to the *E. coli* outbreak isolate and detecting *Dickeya* species (Pritchard, 2013). Employing the pipeline to generate *P.*

capsici isolate-specific primers required sequence data from target and off-target species. Currently, there is only one published *P. capsici* genome sequence – a laboratory isolate derived from back-crossing two field isolates. For PDP to generate isolate-specificity, more sequenced isolates of *P. capsici* were required (discussed further in Chapter 4). The inclusion of three other sequenced genomes of *P. capsici* isolates gave us the foundations to be able to discriminate between isolates by generating isolate-specific primers.

3.4.3 Validation of primers in MCL predicted effectors

The majority of the effectors predicted to be *P. capsici* specific from the MCL analysis also amplified the closely-related species, *P. tropicalis*. Incorporating protein sequence data from *P. tropicalis* during the Blastp search would likely discard all of these shared effectors. Despite the protein data deriving only from one target isolate (LT1534 – the published reference genome), amplification from other *P. capsici* isolates was seen in most cases. Three effectors were not identified in all *P. capsici* isolates. This may indicate a loss or change in effector repertoire notifying the genetic diversity amongst different isolates. On the other hand, the primer binding sites may have mutated but the effector sequence is maintained. More validation of these primer sets is required.

3.4.4 Enhancements to PDP – Emulating Primersearch

A proposed enhancement for PDP is to emulate Primersearch by using Bowtie2. The cross-hybridisation step takes up the majority of the computational time used while running PDP (Pritchard et al., 2012). An effective way to reduce the time taken to run the pipeline could be to emulate the role of Primersearch in the cross-hybridisation step using Bowtie2.

Primersearch searches for the presence a primer pair (in the correct orientation and with the correct separation on the target) in a target genome. This is a serial search, one primer set at a time, and can take a considerable amount of time with large datasets. Bowtie2 is an alignment tool that aligns short reads to longer sequences, such as draft or reference genomes. Bowtie2 works by indexing the reference genome, meaning that genomes are cut up into small fragments and locations are "tagged" for reference. This can make PDP more efficient by not having to compare against the complete genome, but instead, smaller fragments.

To emulate the Primersearch step, a possible approach would be to use Bowtie2 to map all the predicted primer sequences to the target genome as if they were paired-end reads, end-to-end and reporting all alignment locations. SAM/BAM output would contain the mapping locations including the insert size (amplicon). The desired amplicon size could then be searched for.

3.4.5 Future work – Compare output with OEDs' and ultimately use in the field

An interesting next step would be to compare the output of OEDs and the output of PDP. As both pipelines are capable of generating species-specific primers by different methods, there may be overlap in that the same regions/primer sets are identified by both pipelines. Both pipelines are capable of generating species-specific primers within and outwith CDS regions. However, the species-specific primers that I have discussed in this chapter from PDP are only within effector-coding regions due to the inclusion of the "effector isolate" dataset. When the GFF file issue has been resolved, it would be interesting to observe and compare the outputs of both pipelines. Would there be any overlap in similar or same regions? Would one pipeline dominate in terms of numbers of primers generated? What are the success rates of both pipelines in

diagnosing target isolates? This would be tested by running both pipelines with the same input data (draft genomes instead of read data in OEDs) and incorporating the same GFF file. *In silico* and wet lab PCRs would help validate the sensitivity and specificity of the pipelines.

The ultimate goal of PDP is to design discriminatory primers that are capable of distinguishing between species, isolates within a species or subgroups within a species and to use these to diagnose *P. capsici* isolates in the field. To meet this goal for *P. capsici*, more work is required. More sequence data from target (*P. capsici*) and off-target species are required to achieve effective sequence targeting and filtering steps to minimise false positive primer sets. The primer sets also have to be thoroughly validated in the lab with target isolates that were not included in the pipeline. What is also important to consider is the application of these primers in the field. Are the primers able to diagnose from infected material, or does the pathogen still require to be isolated and grown in the lab as is done currently. Ideally, diagnosis should occur on site, so isolation of the pathogen in the lab should be advised against if there is a suitable, effective alternative diagnosis method available. However, comparing the two growth methods in regards to diagnostic ability would be an interesting next step.

Chapter 4: Sequencing the *P. capsici* reference genome and multiple field isolates using Illumina, MinION and PacBio technologies

4.1 Introduction – the importance of genome sequences

Availability of a high-quality reference genome is highly desirable to researchers. Genomes allow evolutionary, functional and comparative analysis to be performed on a genome-wide scale (Michelmore, 2000). The ability to sequence and analyse full genomes is fundamental in genomics. In plant pathology, genetic studies are important in understanding the mechanisms that underpin the infection process. Genomics have accelerated the discovery of genes required for pathogenesis, virulence and avirulence in pathogens. For example 84 full length CRN effector proteins were identified from the *P. capsici* reference genome by using a computational pipeline approach (Stam et al., 2013). The identification of novel pathogen effectors and R genes in crops from genomic studies has driven recent plant pathology research in the lab (Stam et al., 2013, Bos et al., 2003, Neupane et al., 2018). Performing genomic studies from field infections would give researchers an up-to-date snapshot as to what is happening in the fields, rather than the lab.

4.1.1 Short read sequencers – first, second and generation technologies

Emerging in 1977, the Sanger sequencing method has been the most commonly used method for whole genome sequencing for three decades (Sanger et al., 1977, Kulski, 2016). Also known as chain terminating, the Sanger method involves fluorescently tagged nucleotides with a DNA primer, polymerase and a template strand (Figure 4.1).

The DNA strand is generated until a fluorescently tagged nucleotide is added which ceases extension and emits light. These signals can be arranged in size order determining the sequence (Sanger et al., 1977). However, this process was, and still is, expensive and is impractical for larger sequencing projects.

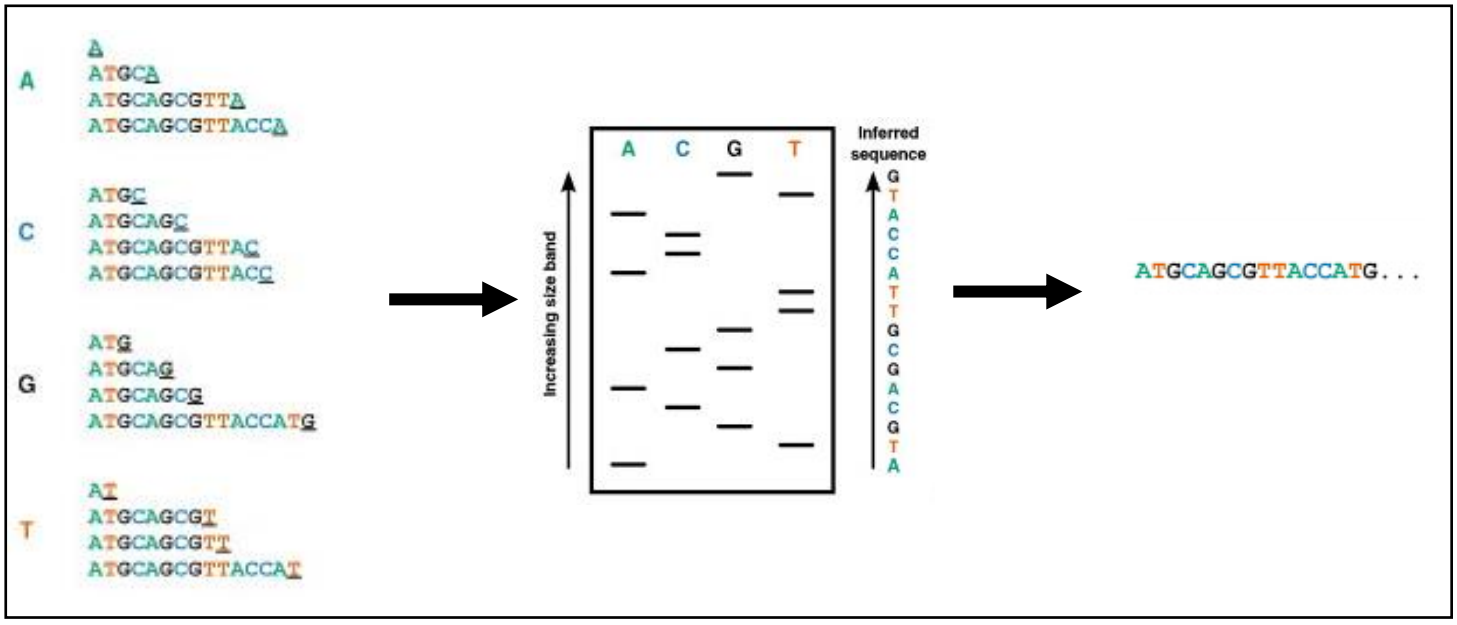


Figure 4.1. The Sanger sequencing method (chain termination). Left) The addition of fluorescent nucleotides (A, T, G and C) prevents further sequencing. Centre) observing the fragment sizes by electrophoresis. Right) The fully sequenced fragment. Image adapted from (Heather and Chain, 2016)

The development of second generation technologies including 454 (Roche) and SOLiD (Applied Biosystems/Life Technologies) were prominent in the 2000s. The 454 method uses a sequence-by-synthesis approach, similar to Sanger, but with a number of differences. Light emitted from luciferase was used to identify nucleotide changes instead of fluorescently tagging them (Heather and Chain, 2016). Also, 454 method produced millions more reads than the Sanger method. The sequencing by oligonucleotide ligation and detection (SOLiD) method uses a sequence-by-ligation

approach (Heather and Chain, 2016). DNA ligase is used to ligate nucleotide probes to a DNA template. When a probe ligates to the template, light is emitted and the sequence can be read (Pereira, 2015).

By the mid-2000s, next generation sequencing (NGS) methods were introduced to the sequencing community, offering higher throughput at a significantly reduced cost (Heather and Chain, 2016). Illumina sequencing offers millions of high-quality reads at a significantly lower cost compared to Sanger sequencing, using a method known as sequencing-by-synthesis. Although both Illumina and Sanger systems use fluorescent signals to determine base changes, both systems are different. Illumina platforms such as the MiSeq and HiSeq, sequence millions of reads at a time by continually recording the fluorescence emitted from each nucleotide added on an array containing millions of fragments, whereas Sanger systems sequence one fragment at a time as described above (Sanger et al., 1977, Mardis, 2008). Two different oligo fragments are attached to the Illumina flowcell. Adapters are added to the ends of the DNA fragment to be sequenced. One of the adapters hybridises to one of the oligos on the flowcell surface. The sequencing of the DNA fragment can be read from recording the order in which coloured light is emitted from the addition of a new nucleotide to the template strand. Millions of these sequencing events occurs simultaneously across the surface of the flowcell, creating millions of reads. As sequencing costs reduced significantly and the quantity of sequence data increased in the late 2000s/early 2010s, this gave rise to an increase in whole genome construction publications (Ossowski et al., 2008, Schatz et al., 2010, Potato Genome Sequencing et al., 2011, Tomato Genome, 2012).

4.1.2 Third generation technology – long read sequencers

Between 2014-2015, the price of genome sequencing reduced further. Biotechnology companies such as Oxford Nanopore Technologies (ONT) and Pacific Bioscience (PacBio) have developed novel sequencing methods (third generation sequencing) that can generate ultra-long reads.

4.1.3 The MinION portable sequencing device

The MinION from Oxford Nanopore technologies is a third generation, real time, portable sequencing device capable of generating long reads from high molecular weight (HMW) DNA. The sequencing is performed on a detachable flowcell which contains a sensor chip array where the DNA is sequenced. The array is made up of 2048 wells embedded on an electrically charged membrane. During a sequence run, there are 512 pores available for sequencing at any time split into 4 multiplexer MUX) stages. DNA is sequenced by passing a strand through a pore in the charged membrane. The passage of bases generates a sequence-specific signature which is interpreted and converted to a readable sequence (Figure 4.2). As this technology is yet in its infancy, the basecall error rates are relatively high and have less coverage depth than Illumina.

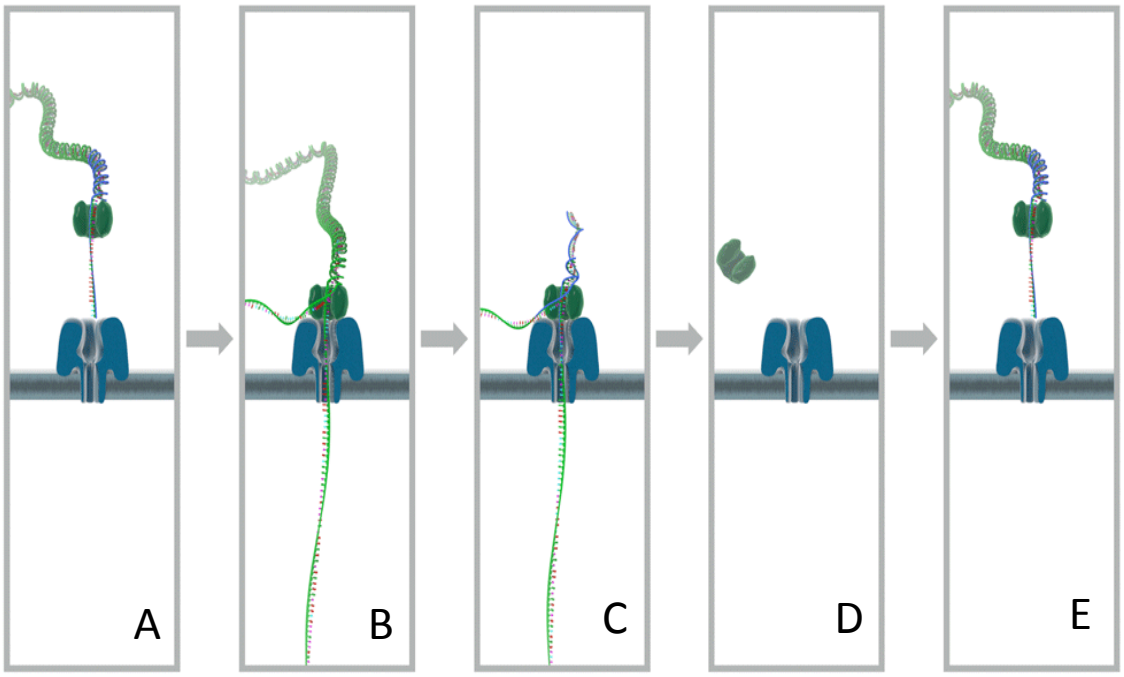


Figure 4.2. Sequencing DNA using nanopore technology using the MinION. A: The DNA is attached to a motor protein that binds to a pore protein embedded on an electrically charged membrane. B and C: The DNA is unzipped by the motor protein and one strand is passed through the pore protein. As the strand passes through the pore, the electrical current is altered. Changes in current (known as "events") can be interpreted and translated into nucleotide base pairs/dyads thus, delivering a readable sequence. D and E: After the DNA has been passed through, the motor protein detaches and waits the arrival of another DNA strand to be sequenced.

Often incorporating Illumina short reads, MinION sequence data has been used for many recent genome constructions and re-assemblies including bacteria, yeast, viruses and animals (Loman et al., 2015, Jansen et al., 2017, Karl et al., 2017, Hoenen, 2016, Liem et al., 2017, Istace et al., 2017). Like Illumina, the MinION is not restricted to genome assembly projects, the technology has also been used to diagnose emerging pathogens (Quick et al., 2015, Quick et al., 2016). The portability of the MinION has been beneficial for researchers in many fields to use the technology outwith a laboratory setting. Sequencing on-site and in real time, can increase the prospect of obtaining a correct diagnosis much faster than previous methods. Also, the MinION can be used on-site which enables rapid genomic diagnostics of emerging

diseases in resource-poor locations (Hayashida et al., 2019, Hoenen et al., 2016). Despite the increasing number of publications of genome assemblies using the MinION, as far as we are aware, the device has not been used to sequence any oomycete genomes (including *Phytophthora* spp.).

4.1.4 PacBio Sequencing

Similarly to the MinION device from ONT, PacBio also generates long reads from HMW DNA. PacBio uses Single Molecule, Real-Time (SMRT) technology which involves a SMRT cell containing 150,000 – 1,000,000 bottomless wells (depending on which SMRT platform is used). These wells are known as nanoscale observation chambers (Zero Mode Waveguides (ZMW)). During library preparation, two hairpin adapters are added to each end of the double stranded DNA fragment, resulting in circularisation – this is termed a SMRTbell construct. The library is loaded onto the SMRT cell where a primer and polymerase anneal to one of the SMRTbell adapters in each well. In an ideal sequencing run, every well would be utilised with a single SMRTbell construct. The sequencing reaction occurs within each ZMW, where a small light signal is recorded with every addition of a fluorescently tagged nucleotide in real time. Sequencing of a circularised SMRTbell generates a linear DNA fragment.

PacBio sequencing, like all sequencing methods, is error prone. The error rate of PacBio sequencing is around 13% and is likely caused by low sequencing depth which can result in base calling and mapping errors (Kulski, 2016). As PacBio sequences in a circular motion, longer SMRTbell constructs may result in less overlapping sequence passes, resulting in a higher chance of single pass error and less depth coverage (Ardui et al., 2018). However, sequencing short length SMRTbell constructs results in lots of replicates, known as sub reads. These subreads can be combined together to obtain a highly accurate consensus sequence taking into account the depth of

coverage. The length of these reads are shorter but more accurate than longer SMRTbell constructs. Sequencing longer SMRTbell fragments can result in read lengths of up to 10Kb.

There are many advantages of using third generation technologies including the low cost per genome, the ability to sequence repetitive genomic regions, the ability to sequence highly homologous genomic regions, the ability to identify structural variants, and the ability to use alongside high quality, short read data to obtain hybrid assemblies. MinION and PacBio technologies have allowed researchers to sequence and reconstruct genomes more easily, evidenced by an increasing number of published genome sequences that have been genereated with third-genereation technologies. (Chalupowicz et al., 2019, Loman et al., 2015).

Genome construction projects are now using a combination of long read technology (MinION and PacBio) and short read technology (Illumina) to improve existing genomes and construct novel genomes using a hybrid assembly approach (Karl et al., 2017, Kranz et al., 2017, Gulvik et al., 2019). This approach benefits from the long reads that can span longer distances – sometimes repeat rich or homologous sections – and the high quality from the short reads to boost the confidence of the consensus.

4.1.5 The *Phytophthora capsici* reference genome

To date, there is currently one publicly available *Phytophthora capsici* sequence from a single isolate (Lamour et al., 2012a). This isolate (LT1534) was produced by mating two field isolates; LT263 and LT51. The offspring were backcrossed twice with LT263 to generate LT1534. The isolate was sequenced using 454 and Sanger sequencing at 454 Inc. (Branford, CT) and the Joint Genome Institute (JGI, Walnut Creek CA) respectively. The sequenced reads were assembled using Arachne (v.20071016) and

passed through Rebuilder and SquashOverlaps to merge assembled sequences together. The 64Mbp reference genome is comprised of 917 scaffolds and has an N50 of 706kbp (50% of the assembly can be produced with contigs of this length or longer) with a genetic map of 54Mbp (84% assembled genome length).

The *P. capsici* genome is made up of multiple scaffolds and has many gaps and N regions, and the process of creating a genetic map was further complicated by a high number of heterozygous polymorphisms (1 in every 100bp) (Hu, 2019). Also, regions within the genomes in both parent and progenies were observed to switch to one of the parental haplotypes (known as loss of heterozygosity (LOH)). LOH is common within the *Phytophthora* genus and is thought to be responsible for genetic diversity between species (Hu et al., 2013a, Hu et al., 2013b, Shrestha et al., 2014, Turner et al., 2017). *Phytophthora* species are known to have large repetitive regions across the genome (*P. infestans* – 74%, *P. sojae* – 39%) (Haas et al., 2009). However, compared to other species, only 19% of the *P. capsici* genome is made up of repeated regions. Similarly to other *Phytophthora* species, the majority (85%) of the repetitive regions are retrotransposons with long terminal repeat sections. These issues, combined with the second generation sequencing techniques used, where the reads were much shorter than the repetitive regions makes it harder to assemble, were responsible for the length of time the project took to complete (7 years) (Lamour et al., 2012a).

In this chapter, I will describe the process and analysis of using multiple sequencing methods and technologies to sequence five *P. capsici* isolates (LT1534, AD84, Q108, Y006 and AP154). Our aims included constructing three draft genomes of known *P. capsici* field isolates to use with our two computational pipelines. Another aim was to sequence *P. capsici* isolates using the MinION/PacBio technology to (i) attempt to improve the existing *P. capsici* reference genome (LT1534) and (ii) sequence *P.*

capsici field isolates of interest (two aggressive field isolates – one of which was newly isolated from the field – Chapter 5).

4.2 Methods

4.2.1 *Phytophthora capsici* growth conditions

All *P. capsici* isolates were grown on V8 media for 5 days in the dark at 25°C. Four, 1cm chunks of V8 were cut from each plate and placed in a petri dish with pea broth solution for 5 days in the dark at 25°C to form mycelial mats. The V8 chunks were cut from the mycelia and discarded. The mycelia was washed and dried at room temperature for an hour before flash freezing with liquid nitrogen and stored at -70°C.

4.2.2 DNA extraction method for Illumina sequencing

DNA was extracted from three field isolates of *P. capsici* (AD84, Q108 and Y006) using a chloroform/phenol method for Illumina sequencing. For each isolate, mycelia were ground to a fine powder with liquid nitrogen and 400 μl of CTAB 3X buffer. After inversion for 30 minutes at 65°C, one volume of 1:1 phenol:chloroform was added and inverted for a further 10 minutes. The supernatant was collected after samples were centrifuged at 13,000 rpm for 30 minutes. The supernatant was precipitated in one volume of cold isopropanol and 0.1 volume of sodium acetate and stored at -20°C overnight.

Samples were centrifuged for 30 minutes at 13,000 rpm and the pellets were washed twice with 400 μl 70% ethanol and centrifuged again at 13,000 rpm for 5 minutes after each wash. The ethanol was discarded and the samples were left to dry at room temperature for 2 hours. The pellets were re-suspended in sterile distilled water, 2 μl of RNAse was added and incubated at 37°C for 1 hour. The samples were then

cleaned using magnetic beads (Agencourt) and concentration was measured with Qubit then sent to Enza Zaden (EZ) for re-sequencing.

4.2.3 Illumina Sequencing – three *P. capsici* isolates

The three *P. capsici* field isolates were re-sequenced using the Illumina Miseq platform (V3 2x300bp) at EZ. The bioinformatic analysis was performed at the James Hutton Institute (JHI). Reads were trimmed and assembled using multiple combinations of trimming tools and assemblers (Figure 4.3). Quast was used to report the quality of each assembly. The combination of tools that had optimal assembly quality for all three isolates was used to construct the draft genomes.

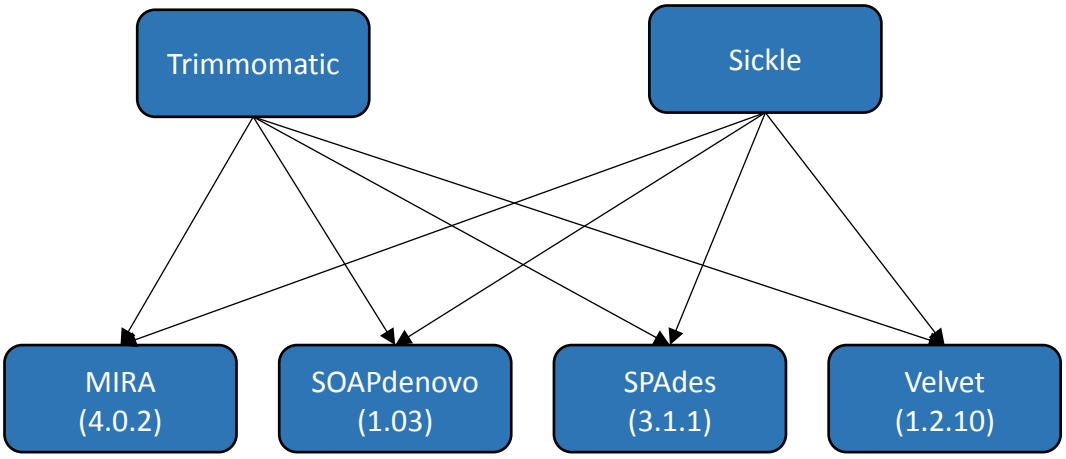


Figure 4.3. A flow diagram showing the combinations used for trimming and assembling *P. capsici* sequence data from EZ. Top line are the two trimming tools used, Trimmomatic (Bolger et al., 2014), Sickle (Joshi, 2011) and the bottom line are the four assemblers used, MIRA (Chevreux, 1999), SOAPdenovo (Li et al., 2010), SPAdes (Nurk et al., 2013) and Velvet (Zerbino and Birney, 2008) .

4.2.4 DNA extraction methods for MinION sequencing runs 1-2

Phytophthora capsici mycelia were ground into a fine powder using a mortar and pestle with liquid nitrogen and re-suspended in four volumes of DNA extraction buffer (100nM tris-HCL pH 8, 70nM EDTA pH 8, 2% (w/v) SLS, 2% (w/v) 2-mercaptoethanol and 100µg/ml proteinase K). After incubation a 55°C for 1 hour with several inversions, the sample was cooled to 27°C and 800µl chloroform/isoamyl alcohol (24:1, v/v) was

added. The samples were vortexed and centrifuged for 10 minutes at 10,000 rpm and the supernatant was collected. The chloroform/isoamyl alcohol step was repeated. DNA was precipitated by transferring the supernatant to an Eppendorf tube with 1ml isopropanol. After inverting five times, and stored at -20°C for 15 minutes, a DNA pellet was formed by centrifugation at 10,000 rpm for 10 minutes. The supernatant was discarded and the pellet was washed with 1ml 70% ethanol then 1 ml absolute ethanol. After air drying the pellet, the samples were re-suspended in 100µl TE buffer (10mM Tris/HCl, pH 8 and 1mM EDTA) at 55°C. The DNA quality was assessed by electrophoresis with a high molecular weight ladder and Nanodrop analysis.

4.2.5 DNA extraction methods for MinION sequencing runs 3-7 (Blaxter Lab Method)

Phytophthora capsici mycelia was ground into a fine powder using a mortar and pestle with liquid nitrogen. 600μl Qiagen Cell Lysis buffer and 10ul Proteinase K (20mg/ml) was added to ~100mg of mycelia powder for each sample. After mixing by rotation and gently spun down in a microfuge, samples were stored at 56°C overnight with periodical mixing by inversion. 4μl RNAse (100mg/ml) was added and gently mixed by inversion and gently spun down then incubated at 37°C for 1 hour. 200μl Qiagen Protein Precipitation Buffer was added before gently mixing by inversion, gently spun down and incubating for 10 minutes on ice. Samples were centrifuged for 31.5 minutes at 16,600G at 4°C. The supernatant was collected and 600ul isopropanol was added before mixing by inversion, gently spun down then stored at -20°C for 10 minutes. Samples were centrifuged at 4°C for 31.5 minutes at 16600G. The supernatant was removed and 600μl of 70% ethanol was added to resuspend the pellet. The samples were mixed by inversion, gently spun down and stored in ice for 10 minutes. Samples were centrifuged again at 4°C for 31.5 minutes at 16,600G. The supernatant was

discarded and the pellet was left to dry at room temperature for ~ 1 hour. The pellets were resuspended in 20μl EB buffer solution and stored at 4°C overnight before permanent storage at -20°C. The DNA quality was assessed by electrophoresis with a high molecular weight ladder and Nanodrop analysis.

4.2.6 MinION sequencing – the *P. capsici* reference isolate

All (seven) sequence runs followed the SQK-RAD003 rapid 1D sequencing protocol with FLO-MIN107 flow cells. The initial MinION sequence runs (1-3) had a DNA concentration of input of 500ng (as stated in official protocol). Subsequent MinION sequencing runs (runs 4-7) had increased concentrations of DNA of 700ng. MinKNOW (1.4.2) was used on an Intel core i7 processor running Ubuntu 16.04 LTS for sequencing. Basecalling was performed using Albacore. Nanoplot was used to assess the quality of the reads before *de novo* assembly using the long read assembler, Canu (Koren et al., 2017). Assembly quality was reported with Poretools (Loman and Quinlan, 2014).

4.2.7 Pacific Biosciences Sequencing – three *P. capsici* isolates

DNA from three *P. capsici* isolates (LT1534, AD84 and AP154) was extracted using the DNA extraction method mentioned above (adapted Blaxter lab protocol, University of Edinburgh (previously used for DNA extraction of cyanobacteria and nematodes)).

DNA quality and quantity were measured using Nanodrop and Qubit. The three DNA samples were sent to Novogene (Hong Kong), an external sequencing company, to carry out the library preparation and PacBio sequencing using the PacBio Sequel platform. Sequenced read data was corrected, trimmed and assembled using Canu (1.9) (Koren et al., 2017). Assembly quality was reported with Quast (Gurevich et al., 2013).

4.3 Results

4.3.1 Sequencing three field isolates at EZ using the Illumina MiSeq

Three *P. capsici* field isolates (AD84, Q108 and Y006) were sequenced at EZ using a MiSeq Illumina platform (Table 4.1). The average read length was considerably shorter than expected, although the number of reads was high (9.2-11.9 million reads per isolate). Theoretical coverage assuming a genome size of 64Mbp was 18.75x. Although extremely low, downstream analysis was pursued.

Table 4.1. The number and average length of sequencing three *P. capsici* isolates using Illumina MiSeq at EZ.

***P. capsici* isolate**	**No. of reads**	**Average read length (bp)**
AD84	11,925,908	91.1
Q108	9,287,596	101.8
Y006	11,215,904	116.2

Reads were trimmed and assembled using a combination of different tools available on the JHI computer cluster (Figure 4.3). Two trimming tools, Trimmomatic (Bolger et al., 2014) and Sickle (Joshi, 2011), and four assemblers, SPAdes (Nurk et al., 2013), MIRA (Chevreux, 1999), Velvet (Zerbino and Birney, 2008) and SOAPdenovo (Li et al., 2010). Assemblies produced by each approach were assessed for quality using Quast. The output from Quast is shown in supplementary file 1.

The Trimmomatic and SPAdes combination had higher N50 values, higher genome fraction % coverage (when mapped to the *P. capsici* reference genome) and a lower number of contigs with a similar total length compared to all other combinations (supplementary file 1). The graphical output shows that SPAdes had longer and a fewer number of contigs (Figure 4.4). The three assembled genomes were used in subsequent experiments and studies, including the development of OEDs (Chapter 2) and PDP pipelines (Chapter 3).

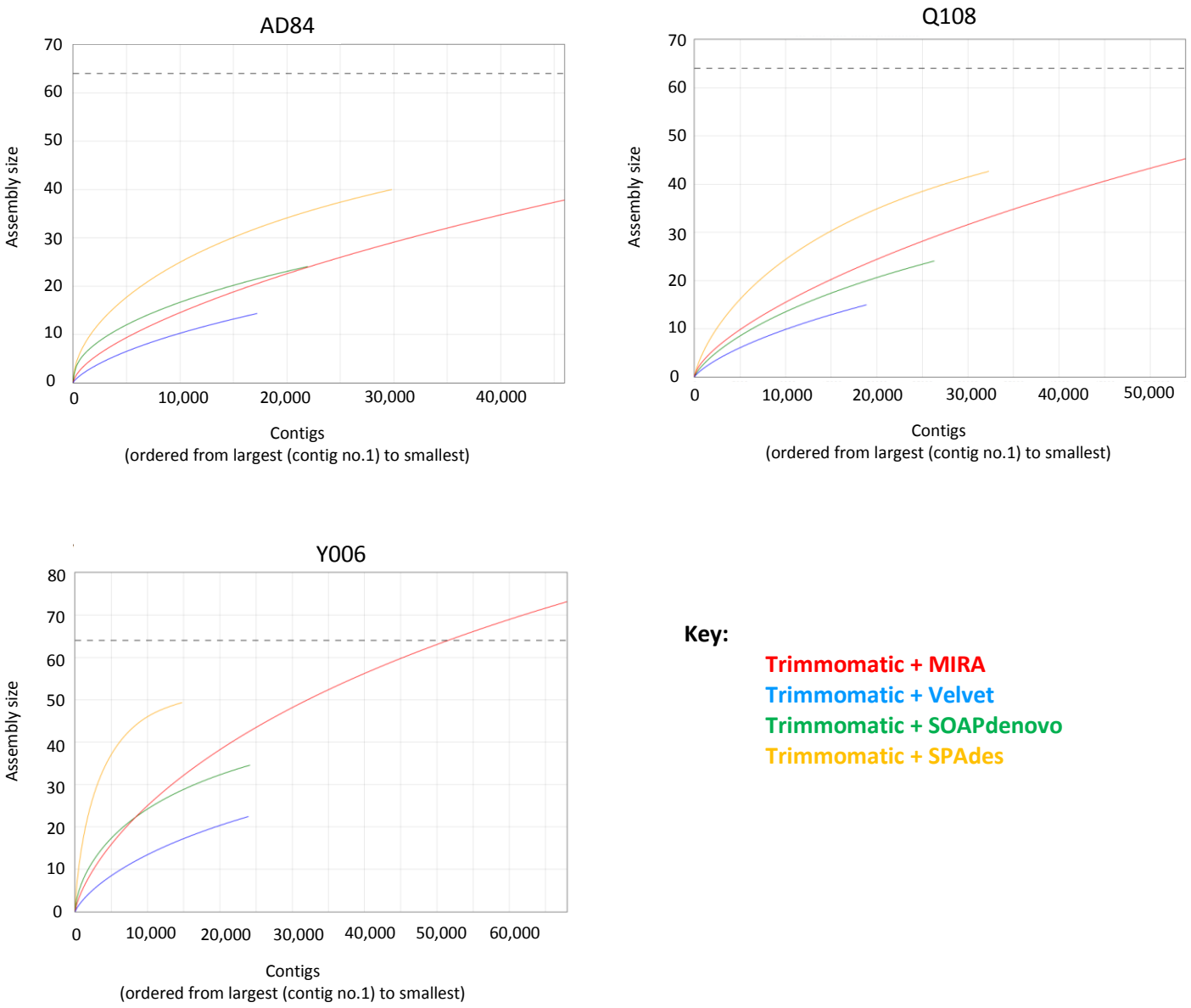


Figure 4.4. Graphical output from QUAST for each isolate showing the different assembly methods from Trimmomatic trimmed reads. Number of contigs and the length of contigs are shown on the x and y axis respectively. The different assemblers are represented by different colours, as shown in the key. The segmented line indicates the size of the reference genome (Mbp).

4.3.2 Sequencing the reference isolate using the MinION

Seven MinION flowcells were used to sequence the reference *P. capsici* isolate, LT1534. The methodology was adapted after each flowcell to try and increase read count, length and quality. The results from each run improved over time after adapting the protocol (Table 4.2). For instance, the mean read length and N50 both increased after each run. Several issues prevented us from obtaining the expected read count, read length and quality including the physical process of loading the sample onto the flowcell without the introduction of air bubbles and also the number of viable pores on the flowcell that were available for sequencing.

Table 4.2. Nanoplot statistics for the first four sequencing runs of *P. capsici* with the MinION. (Run 1 contained lambda DNA from the flowcell test run).

Sequencing run	**Flowcell number**	**DNA extraction method**	**Total reads**	**Total base pairs**	**Mean length**	**min**	**median**	**max**	**N50**
P. capsici LT1534 Run 1 (mostly lambda DNA)	**1**	Isoamyl-alcohol	959	2,525,924	2633.91	253	974	40,295	9154
P. capsici LT1534 Run 2	**2**	Isoamyl-alcohol	46 327	58,549,791	1263.84	118	930	14,346	1570
P. capsici LT1534 Run 3	**2**	Blaxter	739	1,123,156	1519.83	295	1120	8,461	2075
P. capsici LT1534 Run 4	**3**	Blaxter	18 517	31,869,889	1728	158	1278	14,290	2405

Reads from the *P. capsici* sequencing runs 1-6 were compiled together and assembled *de novo* using Canu and assessed with Quast (Figure 4.5 and supplementary file 2). With an assembly error rate of 10% (not aligning sequences with >10% differences), a total of 2,346 contigs were produced from the compiled sequence dataset from MinION runs 1-6, with an N50 of 17,940, NG50 of 5,154 covering 43.72% of the *P. capsici* genome. Read distribution was shown for all read data used in the assembly (MinION runs 1-6) in the Canu output (Figure 4.6).

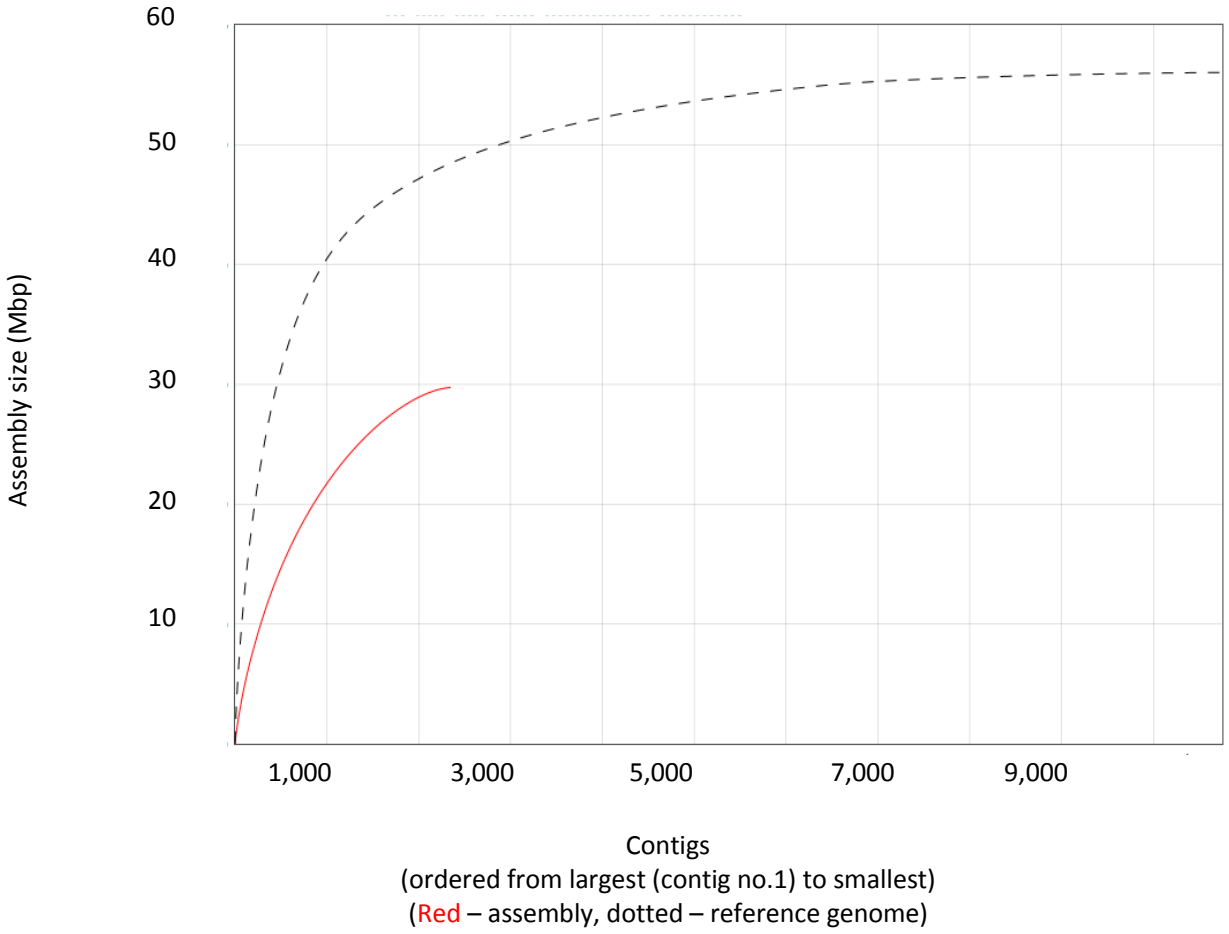


Figure 4.5. Graphical output from Quast for the assembly of the combination of read data from runs 1-6 of the MinION with the *P. capsici* reference isolate LT1534

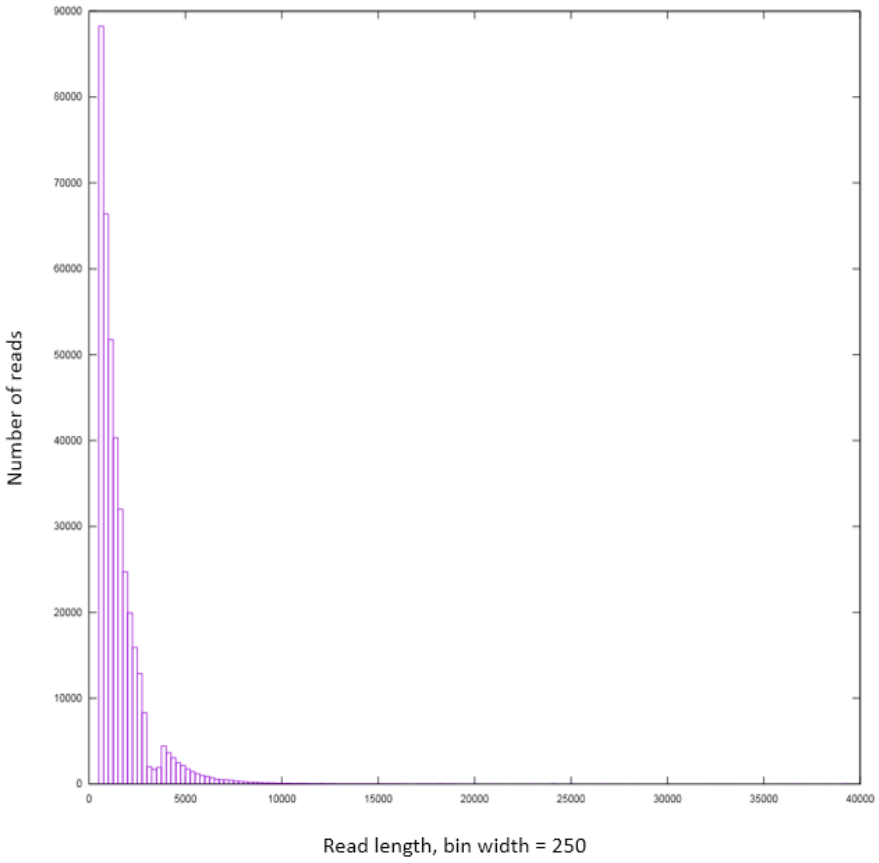


Figure 4.6. The read length distribution of *P. capsici* isolate, LT1534 after compiling MinION sequencing data from runs 1-6. Graph generated by FastQC.

4.3.3 Sequencing multiple *P. capsici* isolates with PacBio

The DNA of three *P. capsici* isolates (LT1534, AD84 and AP154) was extracted using the Blaxter lab method (section 4.2.5) and sent for PacBio sequencing at Novogene, HK. The quality and fragment sizes of the DNA was assessed before the library preparation by electrophoresis on a 1% agarose gel (Figure 4.7).

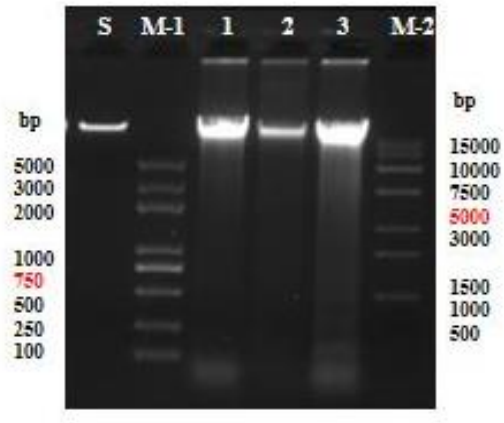


Figure 4.7. The electrophoresis gel with HMW DNA of three *P.* capsici isolates run by Novogene (HK) for quality control before PacBio sequencing library preparation S: standard sample (50ng), M-1: Trans 2k plus DNA ladder, M-2: Trans 15k DNA ladder, 1: AD84, 2: LT1534, 3: AP154. 0.5ul loaded for each sample

The quality control results indicated slight RNA contamination and moderate precipitation in all samples. Purification of all three samples was required and undertaken before sequencing on the PacBio Sequel platform. Read statistics from PacBio sequencing showed that the two field isolates had similar average read lengths, with a slightly lower average for the reference isolate (LT1534) (Table 4.3). This pattern was seen in other statistics, with the two field isolates (AD84 and AP154) having a higher number of reads, total number of bases and higher N50 than the reference isolate. The read distributions of the two field isolates are similar to that of the reference isolate (Figure 4.8).

Table 4.3. Read statistics from PacBio sequencing of three *P. capsici* isolates.

Sample	**Reads bases (G)**	**Reads number**	**Average reads length**	**N50**
AD84	12.634	1470079	8594	10938
LT1534	8.588	1253599	6850	7682
AP154	11.99	1390899	8619	11058

The reads were converted into fastq file format from bam files using Bam2Fastq, then corrected, trimmed and assembled using Canu with default settings.

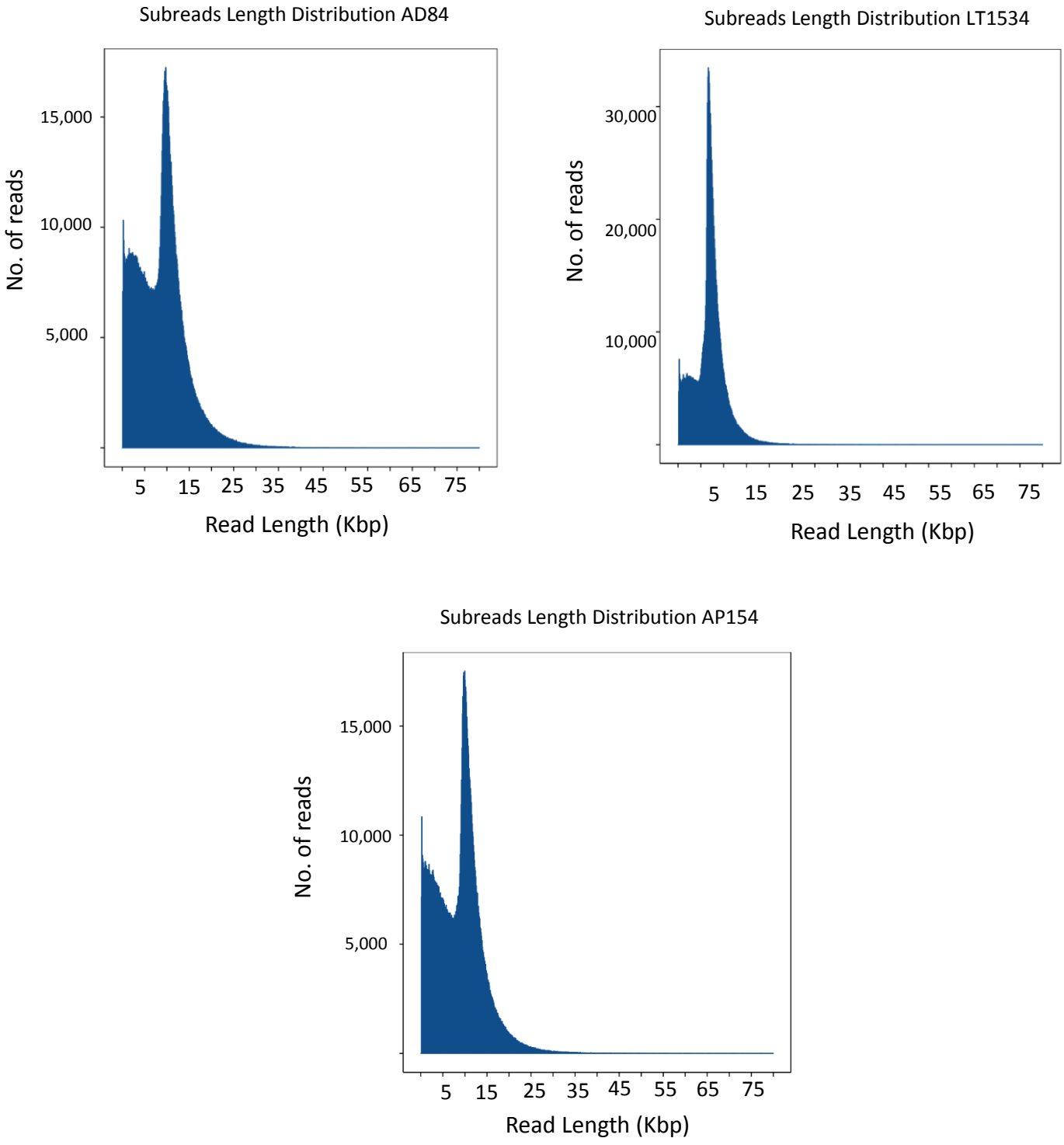


Figure 4.8. The distribution of read length from PacBio sequencing for three *P. capsici* isolates (AD84, LT1534, AP154) sequenced at Novogene (HK)

The graphical output from Quast reflects similar features to the distribution (Figure 4.9 and Figure 4.8). The statistical output from Quast is found in supplementary file 3. Assembling the two field isolates (AD84 and AP154) resulted in a similar number of contigs, N50 and NG50 (50% of the reference genome can be assembled using contigs of this length or longer) (Table 4.4). All three assemblies were extremely large. The reference genome has a total length of 56,034,254bp whereas all three PacBio assemblies were around 3x in length.

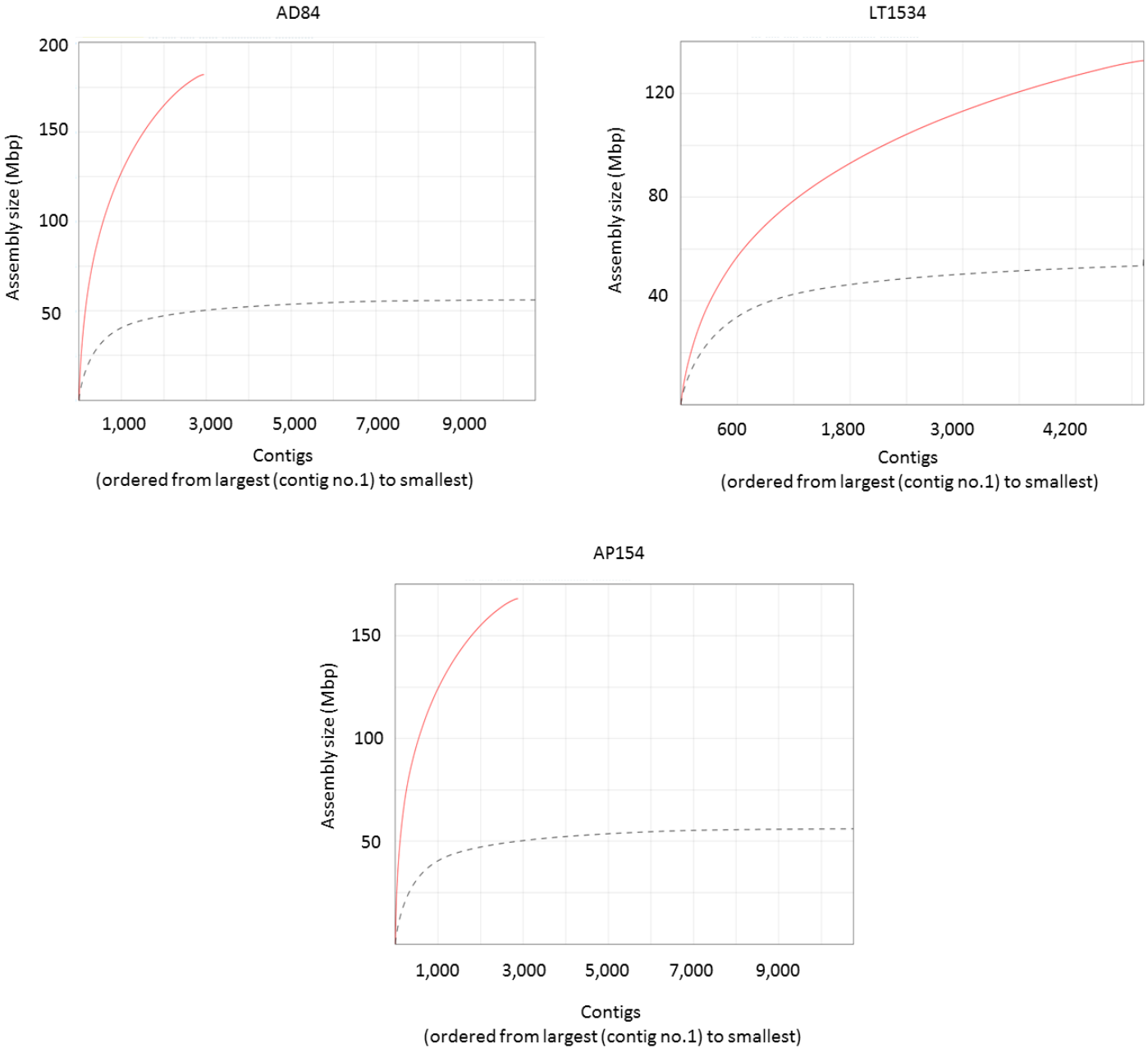


Figure 4.9. Graphical Output for analysing the assembly quality of PacBio data for the three *P. capsici* isolates (red) with the *P. capsici* reference genome (segmented line) for comparison.

Table 4.4. Selected assembly statistics from Quast for the assembly of three *P. capsici* isolates from PacBio sequence data.

Sample	No. of contigs	No. of contigs >50,000 bp	N50 (without reference)	NG50 (with reference genome)	Genome fraction (%)	Total assembled length
AD84	2944	1049	91,189	303,497	84.52	182,098,825
LT1534	4923	575	37,651	105,190	88.26	132,701,540
AP154	2879	817	95,655	427,269	77.43	168,076,253

4.4 Discussion

Sequencing technologies have adapted and been developed extensively over the last 50 years. Advances in sequencing technologies has enabled researchers to sequence and study multiple genomes at a relatively low cost. Sequencing plant pathogen genomes has allowed researchers to study the genetics of the pathogen and also the mechanisms that are involved in causing diseases on host plants. A deeper understanding of the mechanisms that underpin plant pathogen infection is crucial to prevent crop losses, develop resistant crops and provide sufficient crops for the increasing population.

Three different sequencing technologies have been used in this chapter: Illumina, ONT and PacBio. Illumina sequencing generates high quality, short reads and there are an abundance of computational tools and online advice available. However, long read sequence technology is less mature and not as routinely available as short read data and, therefore, there are not as many computational tools or assistance on data analysis. As long read technology advances and more studies are published, more

tools and advice are becoming available making it easier to progress with third generation sequencing data. Hybrid assembly approaches using both short and long read sequencers is becoming more popular for generating or improving genome sequences.

4.4.1 Illumina sequencing at Enza Zaden (EZ)

The sequenced reads from Illumina MiSeq at EZ, were shorter than expected. To extract the best possible assemblies, the reads were trimmed and assembled using a range of different tools to find the combination that provided the optimal results. From analysing the statistical and graphical output from Quast for all assembly combinations, Trimmomatic and SPAdes had consistent results for all three isolates and also had superior stats compared to the other combination of tools in regards to % genome coverage, number of contigs and N50. However, sequence data from each of the three isolates only cover around 50% of the reference genome. With each assembly covering only half of the reference genome, it was clear that all three assemblies were poor. Despite the quality of the assemblies, they were used in further experiments, including OEDs and PDP development as input data (chapter 2 and 3 respectively).

The sequencing produced very short reads even before trimming. This may have been due to the DNA extraction protocol used. The fragments of DNA may have been excessively sheared and therefore resulting in shorted than expected reads. To rectify this, assessment of DNA fragment length and quality could be extensively reviewed before sequencing. For example, electrophoresis - running the DNA on an agarose gel to observe any degradation and average fragment length, or using the Qubit to assess for any impurities in the sample.

4.4.2 MinION – A promising tool for future genome sequencing projects

Although my work with the MinION sequencer improved after every run, the flowcell never produced the expected read lengths or numbers for an efficient or acceptable reference genome assembly. The quality and fragment length of the DNA used in runs 1 and 2 was not good enough (the flowcell used for run 1 was also used to test the flowcell and loading procedure with lambda DNA. Inefficient washing of the flowcell resulted in residual DNA from the previous run). This was down to the extraction method used where the DNA was excessively sheared with the vortex rigorous inverting. Therefore, for subsequent runs, the Blaxter method (section 4.2.5) was used. This step does not use the vortex or any rigorous shaking of the samples, reducing unnecessary shearing. Further improvements were made to the protocol for subsequent runs, such as pipetting slower, cutting off pipette tips to make them wider and using an increased DNA concentration for loading onto the flowcell. Even after combining reads from runs the first six runs and assembling as a single genome with Canu, only 43% of the *P. capsici* reference genome was covered (Table 4.2). As the output for each run gradually improved (increased average read length and N50 etc.), there was evidence that the changes implemented to the protocol had positive impacts but not enough to generate enough data for a hybrid genome assembly.

It became clear that sequencing the *P. capsici* reference isolate with the MinION was a challenging process. Obtaining high throughput, good quality, and ultra-long reads from the MinION is heavily reliant on HMW DNA. It is widely accepted amongst the Nanopore community (not so much in the published literature) that the DNA extraction and preparation has to be meticulously handled throughout the protocol. For example, to obtain expected reads lengths of >1000bp, vortexes and shakers cannot be used as these shear the DNA into smaller fragments. In addition, the MinION can detect the

slightest presence of contaminants in the DNA sample. Chloroform and phenolic compounds used during extraction protocols can wreak havoc on the sequencing array, reducing the number of pores available for DNA molecule uptake and therefore reducing sequence output.

The technique of loading the library sample was also intricate and required technical know-how and sufficient practice. The action of slowly pipetting the sample into the flowcell can easily introduce air bubbles into the liquid channels on the flowcell if not performed carefully and correctly. Introduction of air bubbles into the array during the loading of the sample onto the flowcell and damage the pores, rendering them dysfunctional. In regards to the technique of loading the sample for sequencing, there is little guidance available online and we found it to be very much a trial and error method. The action of loading the sample and minimising the introduction of air bubbles was practiced on used flowcells to help practice and gain confidence in the required technique.

Furthermore, we found that the majority of the active pores (“ready for sequencing”) had “died”/become unavailable for sequencing after the first 24 hours during a 48 hour sequence run. In our experience, running the sequence run after 24 hours resulted in an increase in short reads which can skew read distribution graphs and output statistics. Therefore, sequencing runs 4-7 were stopped after 24 hours. Flowcells are designed to be re-used by washing the previous library with a designated wash kit. However, I found that after washing the array, the number of pores available for sequencing was low and therefore, was not viable for a second sequencing run forcing me to use fresh flowcells for each sequencing run.

The continuous use of sequencing runs on fresh flowcells and library preparation reagents was not financially viable for this project. The MinION has many benefits and has great potential for future use in diagnostics and genome sequencing but unfortunately, I was unable to take advantage of them in this project. A hybrid assembly approach using MinION data was not attempted due to the small number and short length of the MinION reads. I decided to pursue an alternative sequencing method to guarantee quantity and quality data to work with.

4.4.3 PacBio Sequencing – Improved sequence data

Similarly to MinION sequencing, PacBio sequencing requires HMW DNA of high quality. In this case the *P. capsici* DNA was sent to an external company for library preparation and sequencing (Novogene, HK) by experienced lab technicians who routinely use PacBio sequence technology. HMW DNA of three isolates of *P. capsici* were sequenced using PacBio (LT1534, AD84 and AP154), compared to one *P. capsici* isolate attempted with MinION sequencing (LT1534).

The three isolates sent for PacBio sequencing required purification before library preparation. The quality control performed by Novogene reported that all samples had signs of RNA contamination and precipitation.

The two field isolates (AD84 and AP154) produced better assemblies (more reads, longer reads, fewer contigs, higher N50 and NG50) than the reference isolate. The read length distributions for all three isolates were far greater than the MinION data, with more and longer reads. Interestingly, but not surprisingly, the sequenced reference isolate had the highest percentage genome coverage when aligned to the reference genome, compared to the two field isolates. Each assembly for the three isolates generated around 3x the reference genome length. This was unexpected and

may be rectified by incorporating Illumina reads to enhance the sequence depth coverage and boost confidence in the consensus sequence for each assembly. I may have enough sequence data to improve the existing reference genome. In-depth QC of the three assemblies may indicate if this is the case. It may be of interest to highlight the areas from other isolates that were not aligned to the reference genome to identify potential genes that may explain the aggressiveness of these isolates in the field (possible effector genes).

4.4.4 Future work

I am currently in the process of further assessing the three PacBio assemblies using BUSCO (Benchmarking Universal Single-Copy Orthologs) (Waterhouse et al., 2017), then aiming to annotate them with a view to identify effector proteins. A hybrid genome assembly for the *P. capsici* reference isolate would be a desirable next step. The long reads generated from the third-generation sequencing platforms can span long and repetitive regions where the short reads can align to and boost the confidence of the consensus sequence by extensive depth coverage. From comparing the outputs from the two, third-generation sequence methods, PacBio and Illumina reads would be the optimal choice for a chance to improve the existing *P. capsici* reference genome.

Annotating the two PacBio sequenced, field isolates would be of interest to *P. capsici* researchers and industries. Obtaining annotated genomes from multiple *P. capsici* isolates could give researchers valuable information of present pathogen biology in the field. This may include newly identified effector genes that are not found or present in the reference. *P. capsici* researchers may want to direct their focus to studying *P. capsici* genomic studies using annotated field isolates rather than the current laboratory reference isolate in the future. Studying the genome of an isolate that was

recently isolated from the field will give researchers an up-to-date picture as to what is occurring in the fields.

Chapter 5: Collaborative efforts to bridge the *Phytophthora capsici* knowledge gap between the laboratory and the field

5.1 Introduction

We are facing a food security crisis due to climate change, unsustainable practices, a growing population and disease pressures (Fujimori et al., 2019, Bommarco et al., 2013). To meet future demands, there is an urgent need for high yielding crops that allow sustainable food production in diverse and changing environments. Researchers and breeders must therefore rapidly respond to changing demands while maintaining focus on the development of successful field traits. Plant pathogens such as *P. capsici* are a threat to global crop production and industries are continually working to overcome this. Before vital research can be performed on these pathogens to understand the biology and molecular mechanisms that underpin the disease process, initial research into identification and diagnosis is required.

5.1.1 Common industrial isolation and diagnostic procedure of *P. capsici*

Phytophthora capsici can infect a large number of crops from different plant families. It poses a threat to many different crops around the world. There is currently no fool-proof, efficient diagnostic tool to identify *P. capsici* from infected field material. Current best practice is to first isolate the pathogen from infected material in a laboratory, and sequence the ITS1 region (discussed in chapter 2). Isolating *P. capsici* from infected

material is notoriously difficult (Quesada-Ocampo et al., 2009, Manohara, 2004). The most commonly employed method is to use selective antibiotic media (pimaricin, ampicillin, rifampicin, penta-chloronitrobenzene (PCNB)) (PARP) (Jeffers, 1986), but this is prone to fungal contamination which hinders the diagnostic process. After *P. capsici* is successfully isolated, the ITS1 region is sequenced to confirm the pathogen species. However, the ITS1 regions within the species can vary on a nucleotide level, therefore increasing the likelihood of a misdiagnosis. In addition, this diagnostic method relies heavily on experienced pathologists. There is a clear need for a highly sensitive and specific *P. capsici* diagnostic tool that can be routinely used in the field. Importantly, such resources must be accessible and the tools workable to allow deployment in a crop setting.

5.1.2 Collaborations between Researchers and Industry

Collaborative efforts between academic and industrial researchers are seen to be advantageous for both sides and will be important in the battle against plant pathogens on important crops (Evans and Austin, 2010, Evans, 2010). The combination of the understanding of the pathogen biology from the researchers and the commercial and plant breeding knowledge from the industrial partners uniquely underpins productive efforts towards developing pathogen-resistant crops in a commercial setting.

Collaborations between industry and academics are very beneficial (Guimón, 2013). From a research aspect, we can study pathogen biology on crops used in a field environment whilst in a controlled manner (Jupe et al., 2013). Results from these experiments can focus future research and hypotheses that are currently relatable to the field, rather than in the lab. The industry can benefit by understanding the host response from infection of pathogens that are prevalent in the field. This can drive breeding efforts that are focussed on developing resistant, commercial plants to

current plant pathogen threats (Visschers et al., 2019). Both sides also benefit from the knowledge transfer of working together with experts in both fields. Ultimately, collaborations can help production of fully commercial, resistant crops that can mitigate the food security crisis.

The collaboration on this project with seed breeding company, Enza Zaden (EZ) and the University of Dundee has allowed lab-based research to move into a field-based setting. As pathologists, the majority of our research and focus is with pathogens in controlled lab settings. However, collaborating with an industry partner has enabled us to observe the *P. capsici* infection process on crops in a controlled field experiment. This has enabled us to understand what isolates may be of interest in regards to future genomic and molecular studies.

One interest of Enza Zaden (EZ) is breeding crops for resistance against pathogens (including *P. capsici*) (Garibaldi, 2004, Vos et al., 1998, Visschers et al., 2019). Plant breeders research host biology and the mechanisms that are used to prevent pathogens from causing disease. This has been complicated further for *P. capsici* susceptible hosts due to the range in disease symptoms that the pathogen can cause. For instance, foliar blight and stem blight in pepper plants involve independent evaluation from breeders which may require identification of different or multiple host resistance genes for both disease symptoms (Barchenger et al., 2018, Monroy-Barbosa, 2010).

Resistance in commercial pepper plants is continually broken by *P. capsici* outbreaks, and results in whole crop losses when left untreated. However, the non-commercial pepper variety Criollo de Morelos (CM334) is known to have a strong resistance to *P. capsici* (Ortega et al., 1991). Efforts from seed breeding companies and research

organisations (including EZ) have been on-going to breed for resistance using CM334 as a source of resistance. Despite the lack of knowledge of the mechanisms that encode resistance to *P. capsici*, several quantitative trait loci (QTL) have been identified from CM334. Highlighting and exploiting QTLs can be beneficial in the long-term aim of the breeder to deliver a commercial crop that is continually resistant to pathogen outbreaks (including *P. capsici*) in the field. This is done by using markers to track potentially useful known loci in the host line.

Plant breeders at EZ have generated a recombinant inbred line (RIL) of pepper plants derived from CM334 and the susceptible Maor variety resulting in over 200 pepper accessions. RIL populations can be a strong tool for genetic mapping, identifying QTLs of interest for plant breeders to focus on (Broman, 2005). RILs are produced by breeding two inbred lines together, then selfing or breeding with siblings to generate a “mosaic patterned” genome from the two parents. By genotyping these lines, novel QTLs of importance to developing resistant, commercial lines can be identified.

The pepper RIL population at EZ has been phenotyped after *P. capsici* infection with a single isolate (Q108). The Q108 isolate was originally isolated from the field in 2008, and the routine use and sub-culturing on media plates in the lab may have potentially reduced the rate of infection and/or altered the genetic makeup, as seen in other organisms (Ansari and Butt, 2011, Shah et al., 2007). An experiment was set up to inoculate accessions of the RIL population with multiple isolates of *P. capsici* with the following aims: to identify possible isolate-specific outbreaks within individual members of the RIL (if particular isolates can cause infection in accessions that other isolates do not); to identify resistant RIL members that may identify novel QTLs; and to determine whether any of the field isolates of *P. capsici* break host resistance.

5.1.3 RNA sequencing (RNAseq) – Genetic expression of *P. capsici* isolates during early stages of infection

In regards to *P. capsici* biology, it is important to identify the genes responsible that cause and proliferate infection. When genes are identified, studies can be focused on the exact roles they play and how they interact with the host. Jupe *et al.* (2012) showed there are distinct changes in gene expression in regards to *P. capsici* infection on tomato in a time course microarray experiment. Subsequent studies have used RNAseq to observe gene expression at various life stages of the pathogen including mycelial growth, zoospores and cysts (Chen et al., 2013)

Expression profiles of individual isolates during the early stages of infection on pepper would be advantageous to researchers. This would enable researchers to investigate possible effectors that are expressed early in the infection process and also their targets in the host. Commonly expressed genes at this particular time point as well as isolate-specific expressed genes, may be of interest to phytopathologists. Expression profiles can also give rise to the discovery of polymorphisms in genes between different isolates which may have a bearing on levels of expression and/or infection. For example, genes differentially expressed in a single isolate may indicate activities/processes specific to that isolate.

5.1.4 EWINDO – Isolating *P. capsici* isolates from Indonesian fields

East West Seed Indonesia (EWINDO) is a seed breeding company based in Indonesia. They work directly with farmers, advising on best agricultural practices and products to boost yields. The heat and humidity in Indonesia are ideal conditions for *Phytophthora capsici* to thrive. *Phytophthora capsici* is prevalent in many fields across Indonesia (especially chilli pepper) and threatens to wipe out entire crops. To help prevent this, EWINDO researchers and technicians are deployed to farms to take

samples and diagnose and advise the best method to treat the crops to prevent further losses. *Phytophthora capsici* is extremely difficult to isolate from infected material, and diagnosis at EWINDO is performed by observing the growth on media, and microscope examination of sporangia morphology. This method has a number of disadvantages: mycelial growth on plates can differ amongst isolates (Figure 5.1) and be similar to off target species such as *Pythium*; and it is an expert task to distinguish between species when diagnosing by sporangia morphology.

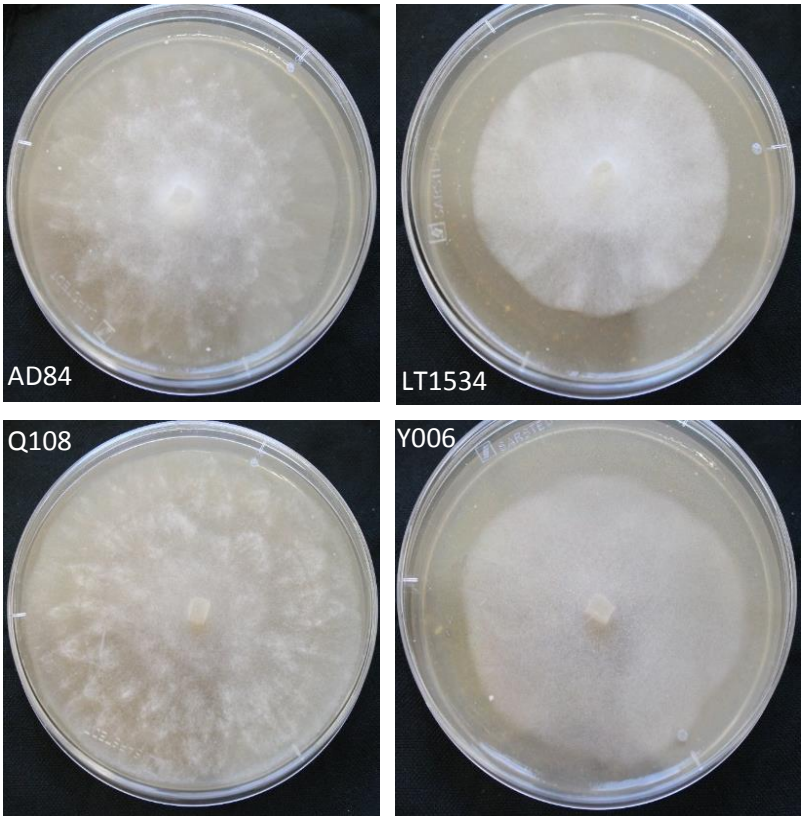


Figure 5.1. The mycelial morphology of four *P. capsici* isolates on V8 agar grown in the dark at 25°C for 5 days. The laboratory isolate (LT1534) and three field isolates (AD84, Q108 and Y006).

In this chapter, I will discuss experiments conducted at two industry companies (EZ and EWINDO) and how they are related to the research project and important for both *P. capsici* researchers and crop breeding industries. One aim of this chapter was to isolate field strains of *P. capsici* from predicted *P. capsici* infected crops. These isolates (and others) would then be used to infect the pepper RIL population at EZ. The outcomes of this experiment would benefit both pathogen researchers and plant breeders. By infecting the RIL population, differences in pathogen population (lab and field isolates) could be observed whilst the breeders would gain valuable information on potential RIL accessions that may or not show resistance to a range of *P. capsici* isolates, informing future breeding programs.

5.2 Materials and Methods

5.2.1 Sampling infected material from Indonesian farms

Sampling of infected material took place at three different locations on two Indonesian Islands (Lembang and Yogyakarta on Java Island, and Lampung on Sumatra Island) (Figure 5.2). I visited multiple pepper fields were visited at each location at three different altitudes (Lowlands, Midlands and Highlands) (Table 5.1). Infected plants were easily identified compared to healthy plants (Figure 5.3.A). Diseased plants showed wilting, lack of foliage and stunted growth. Samples were taken from suspected *P. capsici* infected plants at the site where the pathogen was expected to be active on the host. This would normally be on the stem or branches where there was a clear differentiation of dead tissue and living tissue categorised by a healthy green/brown colour, transitioning into a blackening colour change (Figure 5.3.B). Multiple stems were sampled from each infected plant (if present), wrapped up in a

paper tissue then inserted into a plastic bag for containment. This process was repeated for multiple suspected *P. capsici* infected crops. Samples were sent back to the EWINDO laboratory for isolation.

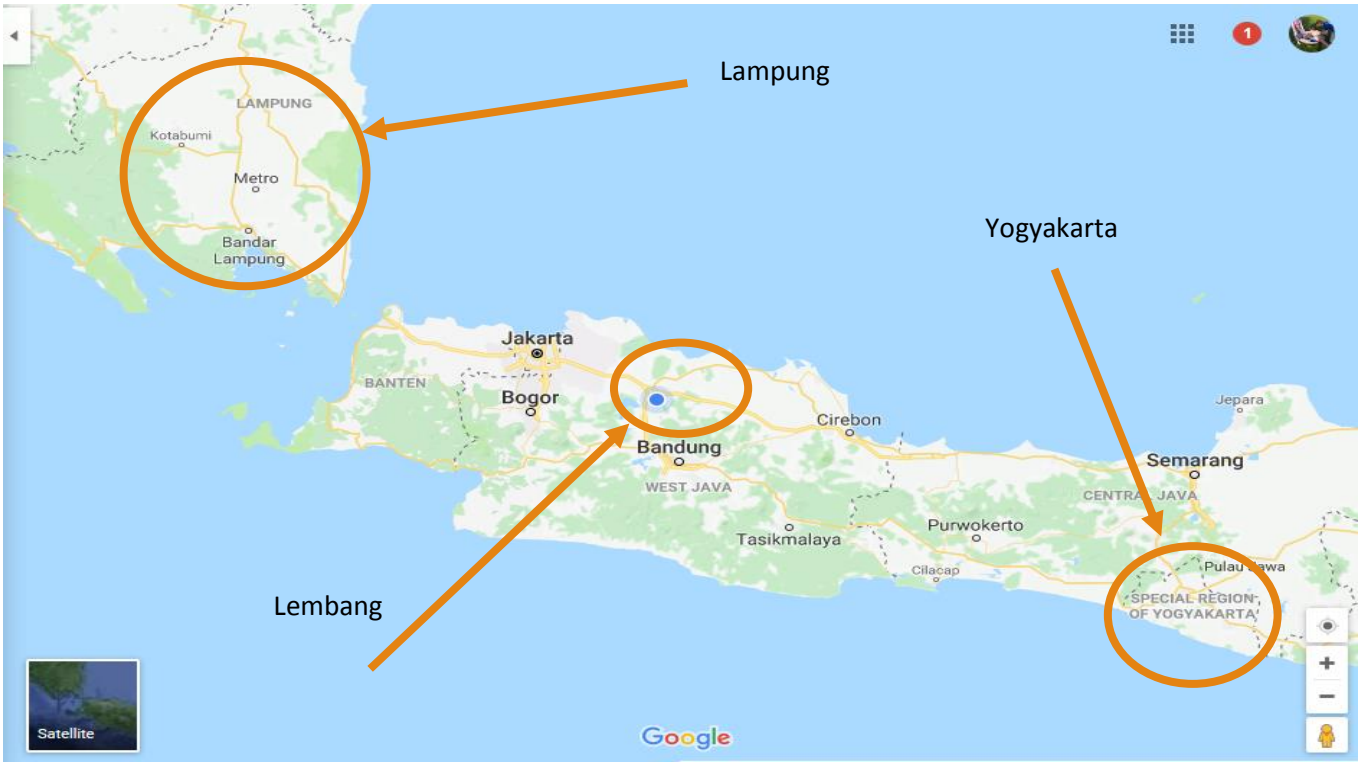


Figure 5.2. Part of Indonesia. The locations where the infected pepper plants were sampled from. Lembang and Yogyakarta on Java Island and Lampung on Sumatra Island.

Table 5.1. Descriptions of the three altitudes of farms visited around Indonesia and the average weather conditions.

Location	**Altitude (meters)**	**Average weather conditions**
Lowland	<250	Very hot and dry
Midland	>250 and <1000	Humid and hot
Highland	>1000	Cooler and wetter

Figure 5.3. Sampling infected pepper plants in Indonesia (A: Left) A suspected *P. capsici* infected pepper plant (foreground, red arrow) (Lembang). (B: Right) Close up of a stem of a suspected *P. capsici* infected pepper plant (middle stem at the green/black transition indicated by the red arrow) (Lembang)

5.2.2 Isolation of *P. capsici* from infected material

Multiple scrapings of each infected stem sample (containing the transition between living and dead tissue) were cut from the samples. Two different methods of culturing were used (Figure 5.4). One method was to insert 3-4 scrapings into media on one plate, the other was to add one piece of plant material underneath media. Two different types of media were used; water agar and standard V8. 4/5 plates were set up for each sample. After 3-4 days stored in the dark at 25°C, the plates were sub-cultured onto fresh media plates. This was done by taking a small piece of mycelial growth from the initial plate and placing it underneath a fresh plate. After 3-4 days, the process was repeated from the sub-cultured plate until the mycelia was clear of bacterial or fungal contamination.

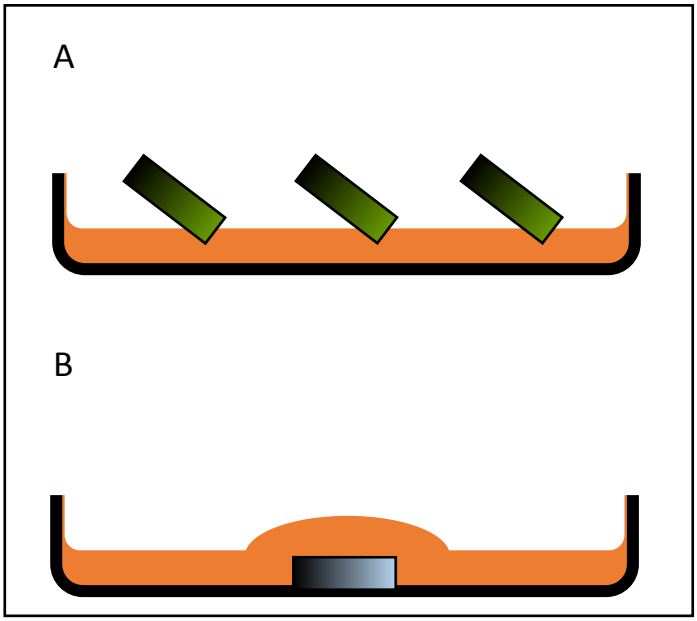


Figure 5.4. Two different methods of isolating *P. capsici* from infected scrapes on petri dishes. A: Three scrapings inserted into the media (V8). B: One infected scrape added underneath the media (V8).

When clear of contamination and sufficient growth on the surface, plates were put in the light for 2-3 days at room temperature to induce sporangial growth. Plates were flooded with 10 ml ice cold water and spread with a sterile glass rod to dislodge sporangia and viewed under the microscope for identification of *P. capsici* spores and zoospores. Successful *P. capsici* isolates were identified by PCR using species-specific primers (from PDP) in a wet lab PCR (set up described further in Chapter 2).

5.2.3 Initial experiment set up conditions

Four *P. capsici* isolates were used (AD84, Q108, LT1534 and AP154) in the large phenotyping assay with the pepper RIL population. The experiment was run in two batches, using two isolates at a time, due to limited greenhouse space. AD84 and Q108 were used in part 1 and LT1534 and AP154 were used in the part 2. Results were combined then discussed.

5.2.4 Growth of RIL Population

84 seeds were sown from each of the 200 accessions of the RIL population on rock wool plugs with vermiculite on top. After 3 weeks of growth, 40 viable plants per accession were transferred to a greenhouse where they underwent *P. capsici* infection assays. 10 plants from each line were put in a plastic container and water was added to the top of the wool plug. This was repeated for each *P. capsici* isolate – 1 pot of 10 plants for each isolate of *P. capsici*. If there were fewer than 10 plants available for each container, the plants were evenly sorted into groups.

5.2.5 *P. capsici* culturing and preparation

Four *P. capsici* isolates (AD84, Q108, LT1534 and AP154) were grown on V8 media in sterile conditions for five days in the dark at 25°C. After sufficient growth on the surface of the plate, the V8 media was cut in half and distributed between two empty petri dishes. Each half of the media was cut into smaller pieces and added to a fresh petri dish with 15 ml ice cold water to induce sporulation. These plates were incubated at 25°C in the light for 30 minutes. The water was removed and 15 ml fresh water was added. Plates were then returned to the light incubator for 24 hours. The plates were stored at 4°C for 1 hour to liberate the zoospores. The suspension was collected and the concentration was calculated and normalised to obtain 15,000 zoospores/ml inoculum.

5.2.6 *P. capsici* inoculation

2 ml of zoospore suspension inoculum was added to the water in each container and these containers were stored in the greenhouse. The phenotype of infected plants per container was recorded for each pepper accession on days 1, 5, 11 and 14. The plants were recorded as either “fully resistant” – showing no signs of infection, “lesions” – showing signs of initial infection (lesions and thinning base of stem) or “susceptible” –

wilted and collapsed. The results used to calculate the resistance percentage – the proportion of plants from each accession that showed no infection symptoms. For plants showing “lesion” symptoms (not dead, but showing infection symptoms) a correction calculation was included to the resistant percentage score:

$$\text{resistant percentage} = \left(\frac{x \times 100}{y}\right) + \left(\frac{\left(\left(z \times \frac{1}{4}\right) \times 100\right)}{y}\right)$$

where *x* is the number of resistant plants at a chosen DPI, *y* is the total number of plants in accession tested and where *z* is the number of “lesion” plants at a chosen DPI. This calculation is in common use at EZ for phenotypic infection assays.

5.2.7 RNA sequencing

Stem bases from susceptible plants (OP177) at six days post inoculation (DPI) from each *P. capsici* isolate (LT1534, AD84, Q108 and AP154) were cut for RNA sequencing. 3cm regions of stem containing healthy and dead tissue were cut with a sterile blade, added to a 1.5ml Eppendorf and stored at -80°C. Four biological reps were collected from each of the four *P. capsici* isolates. RNA extraction was performed using the Machery-Nagel RNA isolation kit. Assessment of the quality and amount of extracted RNA was carried out by running the samples on a 1% agarose gel, a PCR with primers – one amplifying *P. capsici* (effector CRN 83_198 from PDP (Chapter 3)) and a housekeeping gene in pepper (glyceraldehyde-3-phosphate dehydrogenase, GAPDH) and running the samples on the Nanodrop and BioAnalyser.

5.3 Results

5.3.1 Indonesian field visits and sampling

A total of 20 infected samples from three locations on two islands were brought to the laboratory for *P. capsici* isolation (Table 5.2).

Table 5.2. Metadata from the infected crop samples taken from Indonesian fields.

Location	Island	Altitude	Crop	Variety	Date sampled
Lembang	Java	Midland	Curly Pepper	Unknown	26/04/18
Lembang	Java	Midland	Bird Pepper	Unknown	26/04/18
Lembang	Java	Midland	Bird Pepper	Unknown	26/04/18
Lembang	Java	Highland	Big Pepper	Unknown	26/04/18
Lembang	Java	Highland	Big Pepper	Unknown	26/04/18
Lembang	Java	Highland	Big Pepper	MegaTop	26/04/18
Lembang	Java	Highland	Curly Pepper	Serambi	26/04/18
Lembang	Java	Highland	Cucumber	Roberta	26/04/18
Lembang	Java	Highland	Cucumber	Roberta	26/04/18
Yogyakarta	Java	Midland	Bird Pepper	Cempluk	30/04/18
Yogyakarta	Java	Midland	Bird Pepper	Cempluk	30/04/18
Yogyakarta	Java	Midland	Bird Pepper	Shypoon	30/04/18
Yogyakarta	Java	Midland	Bird Pepper	Unknown	30/04/18
Yogyakarta	Java	Midland	Bird Pepper	Unknown	30/04/18
Yogyakarta	Java	Highland	Curly Pepper	Unknown	30/04/18
Yogyakarta	Java	Highland	Curly Pepper	Unknown	30/04/18
Yogyakarta	Java	Highland	Curly Pepper	Unknown	30/04/18
Yogyakarta	Java	Highland	Curly Pepper	Unknown	30/04/18
Lampung	Sumatra	Lowland	Curly Pepper	Unknown	03/05/18
Lampung	Sumatra	Lowland	Curly Pepper	Unknown	03/05/18

5.3.2 Low success rate in isolating *P. capsici*

There was a low success rate from isolating *P. capsici* from suspected *P. capsici* infected samples. Samples were regularly sub-cultured underneath fresh agar to prevent contamination (Figure 5.4.B). When growth was first observed on the surface of the plate – usually after two or three days – a small fraction of media was taken and put under fresh media and allowed to grow. From the 20 suspected *P. capsici* infected samples gathered from fields across Indonesia, only one was confirmed to be *P. capsici* in the lab. Where the sporangia were observed through the microscope, the majority of the other isolates were contaminated and looked to have ascomycete spores. The single isolate was taken forward for PCR identification.

5.3.3 A single isolate of *P. capsici* was successfully isolated from infected pepper plant

PCR results with species-specific primers confirmed the suspected isolate to be *P. capsici* (Figure 3.4). The isolate was named AP154.

5.3.4 Differences in RIL infection rates from *P. capsici* isolates

The majority of the RIL population, including the positive and negative controls (susceptible Maor and resistant CM334 respectively) grew sufficiently well to provide 10 plants per accession for each *P. capsici* isolate.

Infection symptoms were observed at four time points, 1, 5, 11 and 14 days. The RIL accessions became susceptible in all *P. capsici* isolate groups as time progressed. Infection symptoms were first observed on day 5 in isolates, AD84 and AP154 and progressed to day 11 and 14. However, for isolates LT1534 and Q108, infection symptoms started to become apparent at day 11 and progressed at a slower rate than the other two isolates (Figure 5.5). More pepper accessions were susceptible to AD84

and AP154 than LT1534 and Q108. Susceptibility was also more severe in AD84 and AP154 than LT1534 and Q108, where symptoms were seen much earlier.

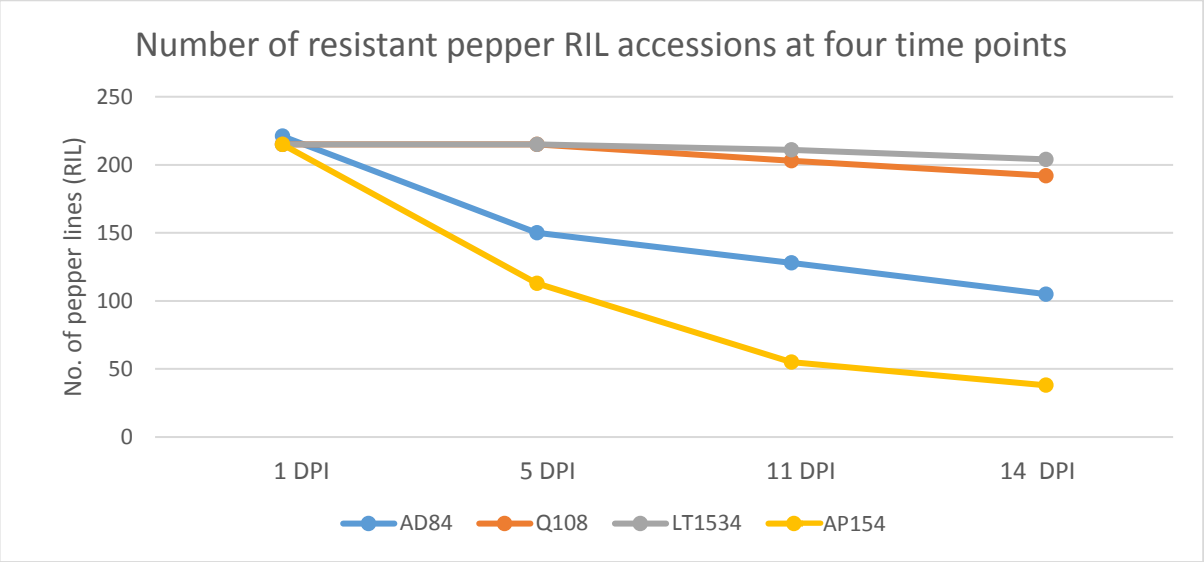


Figure 5.5. Number of resistant pepper accessions in a RIL population over the course of the infection phenotype experiment with four *P. capsici* isolates at Enza Zaden.

At day 14, the majority of LT1534 and Q108 infected RILs were resistant. Around half of the RILs infected with AD84 were susceptible, whilst AP154 infected RIL members showed most susceptibility (Figure 5.6).

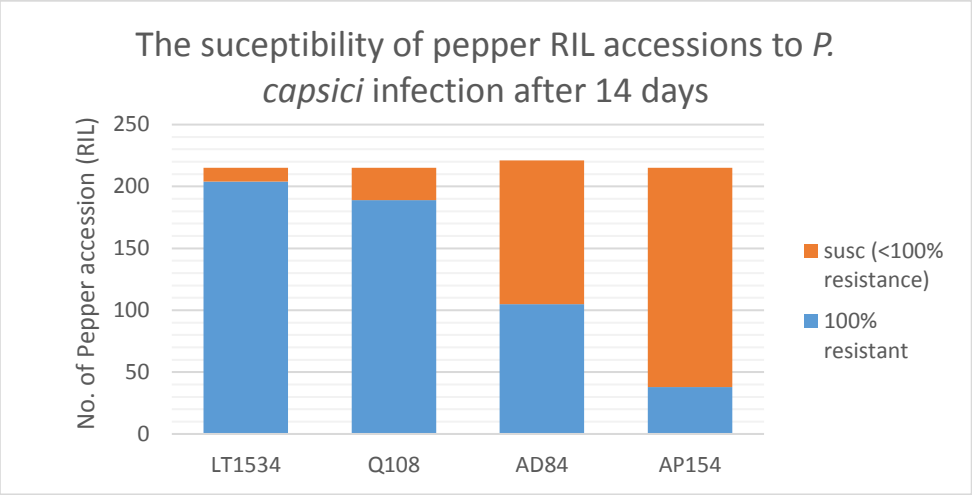


Figure 5.6. Numbers of resistant and susceptible pepper accessions after 14 DPI from *P. capsici* isolates. Blue indicates all plants in each accession showed no symptoms of infection. Orange indicates that at least one plant in each accession showed infection symptoms. Results shown are from one rep.

5.3.5 Extracting RNA for RNAseq

RNA was extracted from 16 stems (four biological replicates for four *P. capsici* isolates) from one susceptible pepper accession (OP177). The extracted RNA was found to be highly contaminated with DNA from *P. capsici* and pepper (Figure 5.7).

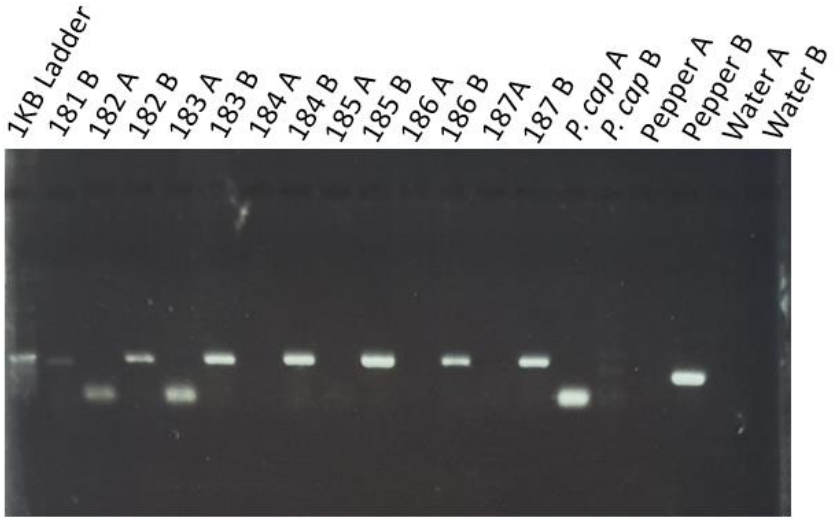


Figure 5.7. PCR of RNA samples of *P. capsici* infected pepper stems with two primer sets A: PDP generated, species-specific primer CRN 83_198 B: GAPDH (glyceraldehyde-3-phosphate dehydrogenase) pepper housekeeping gene.

5.4 Discussion

5.4.1 EWINDO – Confirming the complex technique of isolating *P. capsici* from infected material

From the 20 suspected samples of showing phenotypic evidence of *P. capsici* infection, I was able to successfully isolate a pathogen from only one sample in the laboratory on V8 media. Isolating *P. capsici* from infected plant material is a difficult procedure (Wang et al., 2009). Plants are naturally surrounded by an abundance of microbes – especially during infection. When grown on media, there is competition for growth from many different microbes. We found that the methods currently used at EWINDO to grow the pathogen on media were suboptimal compared to the sub-culturing method we currently use (placing material underneath media). It was notable that considerably more bacterial and fungal growth was seen on isolation plates when infected material protruded from the media compared to adding it underneath media. This suggests that anaerobic competition was limited in comparison to aerobic competition. However, some pathogens including *P. capsici*, are able to grow through the media and spread across the plate. My observations led to the adoption of a modified protocol for *P. capsici* isolation at EWINDO.

The majority of *P. capsici* isolation media plates in this project were observed to be contaminated with ascomycete spores, and so were discarded. A common method used to isolate *P. capsici* employs media supplemented with antibiotics to suppress growth of unwanted microbial species. One method to isolate *Phytophthora* species from field samples is to supplement media with PARP (Morita and Tojo, 2007). Interestingly, media supplemented with PARP is also used to isolate *Pythium* species from plants and soil (Jeffers, 1986). Although this is not 100% effective, it may have helped increase the likelihood of isolating *P. capsici* from more field samples.

5.4.2 *P. capsici* inoculation and phenotype infection assay

There was a clear distinction between isolates in regards to infection rate and susceptibility on hosts. LT1534 and Q108 were the less aggressive compared to the other two isolates, AD84 and AP154. Although previously described as a field isolate, Q108 has been regularly used in the lab under controlled conditions for many years. Whereas AD84 and AP154 have been recently isolated from the field and have had less time to become accustomed to growth on media. The two isolates that were seen as less aggressive (LT1534 and Q108) are however regularly grown on media. Over a long period of time, it is hypothesised that pathogens may become less virulent when continually grown on media (Ansari and Butt, 2011) where pathogens do not have to compete or cause infection. The *P. capsici* reference isolate (LT1534) was the least aggressive isolate in this study. This isolate has been the focus for many genetic and phenotypic studies in labs across the world and may be attenuated beyond a point where it is useful to represent disease progression in the laboratory for emerging and current isolates. This raises the important question for pathogen researchers and industries; should we be focussing our efforts on isolates that do not show a true representation of what is occurring in the fields? Addressing this issue is key for future research and for breeders. Whether it involves storing our isolates differently (possibly continuous infections on hosts in place of storage on media), or selecting the isolates that we study, changes are required to keep up with real life pathogen biology in the field.

Repeating the phenotype experiment with a subset of RIL accession and with more *P. capsici* isolates would be interesting for researchers and breeders. Infecting pepper RIL accessions of interest chosen by the breeders from the original experiment with an increased number of *P. capsici* isolates would be the next experiment to judge how

resistant the crops really are. In order to do this, we would need to successfully isolate *P. capsici* isolates from infected crops all over the world and infect the chosen accessions. This would not only help the breeders to identify possible resistant crops, but would give pathologists valuable information of population biology and could influence future studies.

5.4.3 Genotyping may indicate novel QTLs for resistance

Genome Wide Marker Assisted Selection (GMAS) genotyping is currently underway at EZ on the RIL pepper population. It would be interesting to observe if there are any novel QTLs associated with resistance, identified from the resistant RIL pepper accessions (especially from the AP154 dataset). Potential markers may indicate the presence of resistance genes that plant breeders might exploit to generate *P. capsici* resistant pepper varieties. Due to the large number of RIL accessions included in this study, limitations of experimental time meant that only four *P. capsici* isolates could be used in my experiment. With unlimited resources I would choose to select candidate RIL members of interest and repeat the experiment with an increased number of *P. capsici* isolates – preferably recently isolated field strains. Results may indicate or narrow the number of QTLs that may be of interest to the plant breeder.

5.4.4 RNA sequencing – A follow up experiment

A follow up experiment from the infection assay was to perform RNAseq on infected plant material to give an overview of gene expression during the early stages of infection. An attempt to extract *P. capsici* RNA from infected pepper stems resulted in very low quality and contamination of both *P. capsici* and pepper DNA. However, due to a failure to successfully retrieve RNA of sufficient amounts and quality from infected material, RNAseq was unfortunately not possible.

RNAseq on RNA obtained from infected tissue is not an easy step. Recent studies have shown that the majority of sequence data returned from RNAseq on infected tissue belongs to the host. As a small percentage of sequence data is derived from the pathogen, studies on pathogen gene expression are problematic (De Cremer et al., 2013). To mitigate the problems associated with low transcript levels, deep sequencing is often required. An attractive way around this issue is the use of target enrichment and subsequent Illumina sequencing. Recently, Pathogen enrichment sequencing (PenSeq) methodologies have been established in Dundee that could be used for gene expression profiling (Thilliez et al., 2019). Given that PenSeq has only been used with genomic DNA thus far, development and optimisation of protocols that can handle cDNA need to be developed.

Chapter 6 General Discussion

6.1 The need for a better diagnostic tool

Plant pathogens cause an estimated loss of $2000 billion US dollars per year worldwide (Tewari, 2018). This includes crop losses and longer lasting effects, such as soil requiring remediation. The ability to successfully diagnose a pathogen from the field is essential to administer an effective treatment in order to prevent or minimise crop losses. Therefore, a successful diagnosis not only saves the farmer time and money but will also ultimately help provide enough food for the ever-growing population.

Phytophthora capsici has the potential to wipe out entire crops. However, current diagnostic methods do not meet the standards required to confidently confirm the species or isolate that is responsible for a given outbreak. For example, current diagnostic primers amplifying the ITS1 region cannot discriminate between *P. capsici* and *P. tropicalis* (Chapter 2) (Silvar, 2005). Other methods including morphological determination can be laborious and often require highly experienced technicians. An efficient and successful diagnostic tool would be beneficial for both industries and farmers.

In this thesis I set out to improve on the current *P. capsici* diagnostic method by using two computational pipelines to design species and isolate-specific markers that could be used as a target for a PCR based diagnostic approach. I also set out to improve the existing *P. capsici* genome by sequencing using Illumina, MinION and PacBio sequencing technologies. The current reference genome is not of great quality with missing regions and multiple scaffolds. Also, the sequencing technology used to

construct the *P. capsici* genome make it very difficult to process large, repetitive regions compared to modern technology (MinION and PacBio).

It was clear from visiting affected farms and laboratories around the world, that developing a *P. capsici* diagnostic tool that can be used routinely with little technical ability, would be the ideal solution. In order to achieve this, the genomes of multiple *P. capsici* isolates were sequenced using a variety of different methods. After identifying isolate and species-specific genomic markers from two computational pipelines, primers are designed and diagnosis can be confirmed using PCR. This thesis explains in detail, the development, testing and validation of two computational pipelines that design isolate and species-specific genomic markers. The process and methodology of sequencing *P. capsici* isolates using a variety of different methods is also discussed.

6.2 OEDs designs species-specific primers

For general PCR based diagnosis methods, the main problem is selecting a target sequence that can differentiate between species/isolates. For diagnostics as a whole, there are very few primer sets available for successful PCR based diagnosis method. Examples of commonly used targets for PCR based diagnosis are the ITS1 region, 16S and 23S ribosomal RNA genes and known housekeeping genes (Kong et al., 2001, Dreier, 2019b). However, these targets are not always reliable for a species-specific diagnosis due to similarities with closely related species, as seen in this study. There are a number of computational pipelines available that aid in selecting different targets from genomic data for a successful diagnosis. For example, Rodrigues-Luiz et al. (2017) designed species-specific primers around single sequence repeats (SSR) that flank orthologue genes in closely related species. The first computational pipeline described in this thesis, OEDs, was constructed to design species-specific markers to

be used for a PCR based diagnosis method. The primers were designed from conserved regions identified from sequences obtained from four *P. capsici* isolate genomes (LT1534, AD84, Q108 and Y006). These regions were filtered to exclude any conservation with other, off-target species (other *Phytophthora* species), then primers were designed within candidate *P. capsici*-specific regions. *In silico* PCR discarded any primers that amplified off-target species and that were not picked up by the pipeline filtering stage. Wet lab validation confirmed the primers specificity and sensitivity.

OEDs was successful in designing a large number of species-specific primers (considerably more than what is available for *P. capsici*). The wet lab PCRs with primers generated from OEDs were successful for *P. capsici* specificity. The primers need to be tested further with more isolates of *P. capsici* and other off-target species to improve the specificity and sensitivity. However, initial results are very promising with some primers able to differentiate between *P. capsici* and *P. tropicalis* – two very closely related species – without any *P. tropicalis* sequence data. Another method to improve our primer output in terms of specificity and sensitivity is to incorporate more sequence data – both target and off-target data. More *P. capsici* isolates will refine the identified conserved regions and more off-target genomes (including *P. tropicalis*) will help reduce the number of regions that are found in multiple species – therefore, not *P. capsici* specific. OEDs has great potential to be used with other pathogens to design multiple species-specific primer sets for diagnosis.

6.3 PDP designs isolate and species-specific primers

Similarly to OEDs, PDP also generates diagnostic primer sets. PDP was adapted from Pritchard et al. (2012), where diagnostic primer sets were designed for an isolate of *E.*

coli responsible for an outbreak on cucumber. Using three draft genomes from three *P. capsici* field isolates sequenced with Illumina MiSeq and the existing *P. capsici* reference genome, we were able to generate 20-30 primer sets per isolate, predicted to be isolate-specific. PCR validation of these primer sets was not successful. This may be due to the lack of genome coverage from the input genome data. The three draft genomes of *P.* capsici had between 40-50% reference genome coverage. With sequence data missing, the pipeline may predict regions that are isolate-specific, but may be also found in a second isolate resulting in a false positive.

We ran PDP again with draft genomes from PacBio data, showing 70-80% genome coverage, to see if improved assemblies would have an effect of the number and quality of primers predicted. Primers predicted with PacBio data should be more isolate-specific than the ones from the original data (Illumina MiSeq), due to the inclusion of an increased genome coverage (more of the genome sequence available): although wet lab PCR assays validating this hypothesis are yet to be performed. It would also be interesting to identify if there are any isolate-specific primers designed from the MiSeq PDP run in the PacBio PDP output. These primers may have originally thought to be isolate specific but with the inclusion of more sequence data to improve the genome coverage, this might not be the case with the primers amplifying more than one isolate. Further analysis is required to assess the impact that inclusion of the PacBio data had on the PDP output.

6.4 Sequencing *P. capsici* – challenges and adaptations

In chapter 4, we sequenced various *P. capsici* isolates using different methodologies. Before this project, the reference isolate LT1534 was the only publicly available genome of *P. capsici*. Due to a range of issues with the reference isolate (described

further in Chapter 4), we attempted to improve the reference sequence using third generation sequencing with the MinION portable sequencer (Oxford Nanopore Technology). In theory, the long reads from the MinION would be used in conjunction with the short reads from the original publication in a hybrid assembly approach. It is becoming common practice to use existing, high quality, short reads (NGS) to map to and boost the coverage and quality of the long reads (third-generation) (Karl et al., 2017, Gulvik et al., 2019, Chalupowicz et al., 2019). However, the technical aspects of preparing, loading and running the sample using the MinION were complex and intricate with little margin for error. Although our results improved after each sequencing run, the reads generated were not of sufficient length or quantity to use in a hybrid assembly. Moreover, combining all of our read data from all six runs into one assembly remained insufficient in regards to sequence length and genome assembly quantity. This prompted us to seek alternative sequencing methods.

PacBio, similarly to the MinION, promises high throughput and long reads. Novogene (HK) undertook the library preparation and sequencing of three *P. capsici* isolates (LT1534, AD84 and AP154) from HMW DNA using the PacBio Sequel platform. Using an external company that specialises in PacBio sequencing would minimise the technical errors we experienced using the MinION device. Initial statistics from Novogene (Chapter 4) showed that the sequence data for the reference isolate (LT1534) was not as good as the two other isolates (AD84 and AP154) – both aggressive field isolates. *De novo* assembly using Canu was far better than the sequence data from the MinION and the initial Illumina MiSeq in terms of genome coverage and average length of contigs (Chapter 4). The two field isolates were similar in regard to initial read statistics and *de novo* assembly statistics. The decision was made to run a bioinformatic pipeline (PDP) with the three PacBio assemblies to

observe the difference in output between both sets of input data (Illumina MiSeq and PacBio data).

6.5 Recently isolated *P. capsici* isolates are more aggressive than regular used lab isolates

An interesting result from the large phenotyping experiment at Enza Zaden (EZ) was that the recently isolated *P. capsici* isolates were more aggressive, causing disease symptoms in pepper much faster and more efficiently than regularly used laboratory isolates. As *P. capsici* researchers, we focus much of our work on understanding the biology behind the mechanisms that underpin the infection process on host plants. This research is usually performed with the reference isolate, from which genetic data is derived. Our results (Chapter 5) showed the reference isolate performed poorly in infecting and causing disease in the majority of pepper plants in the recombinant inbred line (RIL) population. This could be due to a number of factors: do recently isolated *P. capsici* isolates contain novel or undiscovered effector proteins able to break host resistance? What is the effector repertoire of these aggressive isolates? Has the reference isolate lost the ability to infect due to lack of continual infection on hosts and regular sub-culturing on media in the lab? i.e. has the reference isolate become “lazy”? This raises an important question for *P. capsici* future research regarding the relevance of the currently studied isolate; should *P. capsici* studies focus on and be performed with an isolate that does not have the ability to cause disease in a real life environment? Future research should therefore focus on recent and more aggressive isolates rather than the existing reference, using phenotypic and sequence data from this study.

6.6 Future work

An important next step in this project is to combine the PacBio data for LT1534 with the existing genome sequence to construct a hybrid assembly to improve the current *P. capsici* reference genome. Combining the long reads and high quality short reads has proven to be beneficial for other organisms (Karl et al., 2017, Gulvik et al., 2019). However, phenotyping results carried out at EZ, showing that the reference isolate was poorly performing infection compared to more recently isolated *P. capsici* isolates, suggest that the former may not be best suited to pursue as the standard “reference”. A possible approach would be to choose an isolate that currently causes disease in the field as a representative “reference” such as AD84 (Isolated in China in 2015) which has proven to be aggressive and which we also have Illumina and PacBio sequence data. A hybrid assembly using both sets of sequence data would give researchers a genomic bases for further studies on a recently isolated field isolate capable of causing infection as opposed to the current reference isolate (LT1534).

Although both pipelines produce diagnostic primers, the methodology differs between them. PDP was previously published and designs primers for multiple genomes of interest, then filters according to similarities with other isolates included. On the other hand, OEDs creates genomes from read data using mapping and *de novo* assembly, then identifies species-specific regions before designing primers within those regions. Comparing the outputs from both pipelines would be interesting: is there any overlap between the outputs in terms of genomic regions selected? Are those regions within gene rich or gene sparse locations? Are these genes isolate/species-specific (possibly effectors?

To make the output of the pipelines more sensitive and specific, a number of factors could be considered. First, increasing the number of *P. capsici* isolate genomes included in the pipelines will ensure true differentiation between isolates in isolate-specific primers. Secondly, including many off-target species (including other *Phytophthora* species and oomycetes) will ensure the primers are truly species-specific, amplifying only the target organisms. It was clear from the wet lab validation of both OEDs and PDP that some primers, thought to be *P. capsici* specific, amplified *P. tropicalis*. Therefore, it is highly recommended that off-target species should include closely related ones, such as *P. tropicalis* for future pipeline runs. Incorporating this sequence data will help differentiate between the two species during the pipeline rather than producing false positive primer sets. Finally, wet lab PCR validation should be carried out to test the generated primers against both target and off-target species. The validation should also be repeated to warrant robust diagnostic primer sets.

6.7 Concluding Remarks

As current methods of diagnosing *P. capsici* outbreaks in the field are time-consuming, expensive, reliant on expertise and overall not efficient, a different diagnostic tool was required to help prevent the further spread of the disease. I have developed and validated one computational tool that can rapidly design diagnostic primers for (emerging) eukaryotic pathogens. I have also utilised a previously published pipeline that can design discriminatory primers within a species. Although additional sequence data and further wet lab validation with other *P. capsici* and off target organisms are required, initial results are promising for the identification of this disease in the field.

Appendix

1.1 Source code for OEDs (Chapter 2)

```
#!/usr/bin/env python
"""
Todo:
- add cmd to reference genome('s)
- add check for GFF file
"""

import optparse
from VisualisationTools import VisualisationTools
import Assemblers
import Main
import ReadAligner
import threading
import PrimerDesign
from Blast import Blast
import copy
import NUCmer
import sys
import os
import inspect


def MapperPreRun():
    if len(worker.fastQFileList) > 2:
        fastQPairs = len(worker.fastQFileList) - 1
        i = 0
        while i < fastQPairs:
            mapper = ReadAligner.Bowtie2(worker.fastQFileList[i],
worker.fastQFileList[i + 1],
                                         worker.refGenomeList[0],
options.output_filepath)
            worker.mapperClass.append(mapper)
            mapper.start()
            i += 2
    else:
        mapper = ReadAligner.Bowtie2(worker.fastQFileList[0],
worker.fastQFileList[1],
                                     worker.refGenomeList[0],
options.output_filepath)
        worker.mapperClass.append(mapper)
        mapper.start()
    for mapper in worker.mapperClass:
        mapper.join()
    for mapper in worker.mapperClass:
        bamWorker = ReadAligner.BamTools(mapper.samFile, mapper.referenceDB)
        bamWorker.start()
        worker.bamClass.append(bamWorker)

    for bamWorker in worker.bamClass:
        bamWorker.join()

    for bamWorker in worker.bamClass:
        visualisationTool = VisualisationTools(bamWorker.samFile)
        visualisationTool.start()
        worker.visualisationClass.append(visualisationTool)

    for visualisationTool in worker.visualisationClass:
```

```
        visualisationTool.join()

    mapperPrimer = PrimerDesign.PrimerDesignByMapping()
    for visualisationTool in worker.visualisationClass:
        mapperPrimer.generateCoords(visualisationTool.depthPerPos)
    return PrimerDesign.PrimerDesign.runIntersect(mapperPrimer.coordsFile,
"/MapperPoI.gff")


def DeNovoPreRun():
    if len(worker.fastQFileList) > 2:
        fastQPairs = len(worker.fastQFileList) - 1
        i = 0
        while i < fastQPairs:
            assembler = Assemblers.Spades(worker.fastQFileList[i],
worker.fastQFileList[i + 1], options.output_filepath)
            worker.assemblerClass.append(assembler)
            assembler.start()
            i += 2
    else:
        assembler = Assemblers.Spades(worker.fastQFileList[0],
worker.fastQFileList[1], options.output_filepath)
        worker.assemblerClass.append(assembler)
        assembler.start()
    contigs = ""
    for assembler in worker.assemblerClass:
        assembler.join()
        contigs += assembler.outputDir + " "
    thread = threading.Thread(Assemblers.Assemblers.quast(contigs))
    thread.start()
    Main.Main.threadList.append(thread)
    contigs = contigs.split(" ")
    nucmerList = list()
    for contig in contigs:
        if len(contig) > 0:
            nucmerRun = NUCmer.NUCmerRun(contig)
            nucmerRun.start()
            nucmerList.append(nucmerRun)
            Main.Main.Contigs.append(contig)
    for nucmerRun in nucmerList:
        nucmerRun.join()
    denovoPrimer = PrimerDesign.PrimerDesignByDenovo()
    for contig in contigs:
        if len(contig) > 0:
            denovoPrimer.readCoords(contig)
    return PrimerDesign.PrimerDesign.runIntersect(denovoPrimer.coordsFile,
"/denovoPoI.gff")


def DeNovoContig(contigs):
    nucmerList = list()
    for contig in contigs:
        if len(contig) > 0:
            nucmerRun = NUCmer.NUCmerRun(contig)
            nucmerRun.start()
            nucmerList.append(nucmerRun)
    for nucmerRun in nucmerList:
        nucmerRun.join()
    denovoPrimer = PrimerDesign.PrimerDesignByDenovo()
    for nucmerRun in nucmerList:
        contig = nucmerRun.contigs
        if len(nucmerRun.contigs) > 0:
            denovoPrimer.readCoords(contig)
```

```
            Main.Main.Contigs.append(contig)
    return PrimerDesign.PrimerDesign.runIntersect(denovoPrimer.coordsFile,
"/denovoPoI.gff")

Main.Main.logger.info("Welcome to OEDS! please enjoy this piece of software, and
RTFM!")
externTools = ["/mnt/apps/primer3-2.3.0/src/primer3_core", "/mnt/apps/SPAdes-3.1.1-
Linux/bin/spades.py", "/mnt/apps/quast/quast-2.3/quast.py"]
if
os.path.isfile(os.path.dirname(os.path.abspath(inspect.getfile(inspect.currentframe
()))))+"/AdditionalScripts/.config.conf"):
    configFile =
open(os.path.dirname(os.path.abspath(inspect.getfile(inspect.currentframe()))))+"/Ad
ditionalScripts/.config.conf")
    for i, line in enumerate(configFile):
        if "DEFAULT" in line:
            Main.Main.ExternTool.insert(i, externTools[i])
        else:
            Main.Main.ExternTool.insert(i, line.rstrip())
else:
    Main.Main.ExternTool = externTools
    Main.Main.logger.warning("The system might experience problems. Please run
config.sh before running OEDS to make sure the depencencies are installed
correctly.")

for item in Main.Main.ExternTool:
    print item
fastQFileList = []
refGenomeList = []

parser = optparse.OptionParser()
parser.add_option('-o', '--output',
                  dest="output_filepath",
                  default="workDir",
                  help="Output directory for all the subfiles like sam, bam and
fasta files"
                  )
parser.add_option('-r', '--resultOutput',
                  dest="result_filepath",
                  default="resultDir",
                  help="Result directory for all the results like coverage plots
and primer outputs"
                  )
parser.add_option('-Q', '--fastQ',
                  dest="fastQFile",
                  action="append",
                  help="All the fastQ files (if pairwise data, input first the
first file, and then the second file)"
                  )
parser.add_option('-g', '--Genome',
                  dest="Genomes",
                  action="append",
                  help="Reference genomes. each of this genome will be used in the
program, only fasta format!"
                  )
parser.add_option('-G', '--GFFFile',
                  dest="gffFile",
                  help="GFF file. Make sure the chromosomes/scaffold have the same
name as the reference genome!")
parser.add_option('-R', '--refGenome',
                  dest="refGenome",
                  help="Reference genome file. The reference genome should be as
closely related to the read data"
```

```
                    " as possible"
                )
parser.add_option('-C', '--contigs',
                dest="contigs",
                action="append",
                help="Full address to the contigs from a de novo assembly"
                )

options, args = parser.parse_args()

if not options.refGenome:
    print "No reference genome added, please use -R or read the github readme for
more information"
    sys.exit(0)

Main.Main.makeDirectory(options.output_filepath)
Main.Main.makeDirectory(options.result_filepath)
if options.gffFile:
    Main.Main.gffFile = options.gffFile
else:
    Main.Main.gffFile = None
Main.Main.workDir = options.output_filepath
Main.Main.resultDir = options.result_filepath
Main.Main.threadList = list()
worker = Main.Main()
worker.openRefGenomes(options.Genomes, options.refGenome)
worker.openFastQFiles(options.fastQFile)
threadList = list()
for fastQFile in worker.fastQFileList:
    workLine = "fastqc " + Main.Main.fastQAdd + fastQFile + " -o " +
Main.Main.resultDir + " -q --noextract"
    thread = threading.Thread(Main.Main.execute(workLine, "Generating fastQC
reports in the background"))
    thread.start()
    Main.Main.threadList.append(thread)

preRun = list()
methodList = list()
thread = threading.Thread(methodList.append(MapperPreRun()))
thread.start()
preRun.append(thread)
if options.contigs:
    Main.Main.externContig = True
    thread = threading.Thread(methodList.append(DeNovoContig(options.contigs)))
else:
    thread = threading.Thread(methodList.append(DeNovoPreRun()))
thread.start()
preRun.append(thread)
for thread in preRun:
    thread.join()

methodList.append(PrimerDesign.PrimerDesign.runMethodIntersect(methodList,
Main.Main.workDir+"/Intersect.gff"))
PrimerDesign.PrimerDesign.removeDuplicate(methodList[2])
methodList.append(PrimerDesign.PrimerDesign.runMethodSubstract([methodList[0],
methodList[2]], Main.Main.workDir +

"/MapperUnique.gff"))
methodList.append(PrimerDesign.PrimerDesign.runMethodSubstract([methodList[1],
methodList[2]], Main.Main.workDir +

"/DenovoUnique.gff"))
```

```
genome = PrimerDesign.PrimerDesign.readRefGenome(Main.Main.genomeAdd +
Main.Main.refGenomeList[0])
blastList = list()
for PipMethod in methodList:
    fastaFile = PipMethod.rstrip()[:-3]+"fa"
    PrimerDesign.PrimerDesign.saveFasta(fastaFile,
PrimerDesign.PrimerDesign.readGFF(PipMethod, genome))
    blastList.append(fastaFile)

otherGenomes = copy.copy(Main.Main.refGenomeList)
del otherGenomes[0]
blastThread = list()
allAlias = ""
for genome in otherGenomes:
   Blast.makeDatabase(str(genome), Main.Main.workDir)
   allAlias += Main.Main.workDir + "/" + genome + " "
allAlias = allAlias.rstrip()
Blast.aliasTool(allAlias, Main.Main.workDir)
for blastItem in blastList:
    blastResult = Blast(blastItem, otherGenomes)
    blastResult.start()
    blastThread.append(blastResult)

for blastItem in blastThread:
    blastItem.join()
genome = PrimerDesign.PrimerDesign.readRefGenome(Main.Main.genomeAdd +
Main.Main.refGenomeList[0])
for contig in Main.Main.Contigs:
    genome =
PrimerDesign.PrimerDesign.getSNPtoN(PrimerDesign.PrimerDesign.readSNPDen(contig+".s
nps", ""), genome)
for file in os.listdir(Main.Main.workDir):
    if file.endswith(".vcf"):
        genome =
PrimerDesign.PrimerDesign.getSNPtoN(PrimerDesign.PrimerDesign.readSNPMap(Main.Main.
workDir+"/"+file, ""), genome)
for k, blastItem in enumerate(blastList):
    item = blastItem.rstrip()[:-2]
    thread = threading.Thread(PrimerDesign.PrimerDesign.generatePrimer3Input(
        item+"primSets", PrimerDesign.PrimerDesign.readGFF(item+"unique.gff",
genome)))
    thread.start()
    Main.Main.threadList.append(thread)

for thread in Main.Main.threadList:
    thread.join()

Main.Main.logger.info("Thank you for using OEDS. For further validation and
additional functions, see AdditionalScripts.")
```

References

ALI, S. S., SHAO, J., LARY, D. J., KRONMILLER, B., SHEN, D., STREM, M. D., AMOAKO-ATTAH, I., AKROFI, A. Y., BEGOUDE, B. A., TEN HOOPEN, G. M., COULIBALY, K., KEBE, B. I., MELNICK, R. L., GUILTINAN, M. J., TYLER, B. M., MEINHARDT, L. W. & BAILEY, B. A. 2017. *Phytophthora megakarya* and *P. palmivora*, closely related causal agents of cacao black pod rot, underwent increases in genome sizes and gene numbers by different mechanisms. *Genome Biol Evol*.

ANDREWS, S. 2010. FastQC. a quality control tool for high-throughput sequence data.

ANSARI, M. A. & BUTT, T. M. 2011. Effects of successive subculturing on stability, virulence, conidial yield, germination and shelf-life of entomopathogenic fungi. *J Appl Microbiol,* 110**,** 1460-9.

ARAGAKI, M. & UCHIDA, J. Y. 2001. Morphological distinctions between *Phytophthora capsici* and *P. tropicalis sp nov. Mycologia,* 93**,** 137-145.

ARDUI, S., AMEUR, A., VERMEESCH, J. R. & HESTAND, M. S. 2018. Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. *Nucleic Acids Res,* 46**,** 2159-2168.

ARMITAGE, A. D., LYSOE, E., NELLIST, C. F., LEWIS, L. A., CANO, L. M., HARRISON, R. J. & BRURBERG, M. B. 2018. Bioinformatic characterisation of the effector repertoire of the strawberry pathogen *Phytophthora cactorum*. *PLoS One,* 13**,** e0202305.

ARMSTRONG, M. R., WHISSON, S. C., PRITCHARD, L., BOS, J. I., VENTER, E., AVROVA, A. O., REHMANY, A. P., BOHME, 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. & BIRCH, P. R. 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**,** 7766-71.

ARX, J. A. 1967. *Pilzkunde,* Lehre, J. Cramer.

AVROVA, A. O., BOEVINK, P. C., YOUNG, V., GRENVILLE-BRIGGS, L. J., VAN WEST, P., BIRCH, P. R. & WHISSON, S. C. 2008. A novel *Phytophthora infestans* haustorium-specific membrane protein is required for infection of potato. *Cell Microbiol,* 10**,** 2271-84.

BABADOOST, M. 2004. *Phytophthora* blight: a serious threat to cucrbit industries. *The American Phytopathological Society.* St. Paul, Minnesota.

BARCHENGER, D. W., LAMOUR, K. H. & BOSLAND, P. W. 2018. Challenges and Strategies for Breeding Resistance in *Capsicum annuum* to the Multifarious Pathogen, *Phytophthora capsici*. *Front Plant Sci,* 9**,** 628.

BHAT, R. G. A. B., G. T. 2010. Specific Detection of *Phytophthora cactorum* in Diseased Strawberry Plants using Nested Polymerase Chain Reaction. . *Plant Pathology,* 59**,** 121-129.

BIRCH, P. R., BOEVINK, P. C., GILROY, E. M., HEIN, I., WHISSON, S. C. 2008. Oomycete RXLR Effectors: Delivery, functional redundancy and durable disease resistance. *Curr Opin Plant Biol,* 11**,** 373-379.

BLAIR, J. E., COFFEY, M. D., PARK, S. Y., GEISER, D. M. & KANG, S. 2008. A multi-locus phylogeny for *Phytophthora* utilizing markers derived from complete genome sequences. *Fungal Genet Biol,* 45**,** 266-77.

BODINE, E. W. 1935. Blight of peppers.

BOEVINK, P. C., WANG, X., MCLELLAN, H., HE, Q., NAQVI, S., ARMSTRONG, M. R., ZHANG, W., HEIN, I., GILROY, E. M., TIAN, Z. & BIRCH, P. R. 2016. A *Phytophthora infestans* RXLR effector targets plant PP1c isoforms that promote late blight disease. *Nat Commun,* 7**,** 10311.

BOLGER, A. M., LOHSE, M. & USADEL, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* 30**,** 2114-20.

BOMMARCO, R., KLEIJN, D. & POTTS, S. G. 2013. Ecological intensification: harnessing ecosystem services for food security. *Trends Ecol Evol,* 28**,** 230-8.

BOS, J. I., ARMSTRONG, M., WHISSON, S. C., TORTO, T. A., OCHWO, M., BIRCH, P. R. J. & KAMOUN, S. 2003. Intraspecific comparative genomics to identify avirulence genes from Phytophthora. *New Phytologist,* 159**,** 63-72.

BOS, J. I., 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. & BIRCH, P. R. 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**,** 9909-14.

BOUTEMY, L. S., KING, S. R., WIN, J., HUGHES, R. K., CLARKE, T. A., BLUMENSCHEIN, T. M., KAMOUN, S. & BANFIELD, M. J. 2011. Structures of *Phytophthora* RXLR effector proteins: a conserved but adaptable fold underpins functional diversity. *J Biol Chem,* 286**,** 35834-42.

BRASIER, C. M., BEALES, P. A., KIRK, S. A., DENMAN, S. & ROSE, J. 2005. *Phytophthora kernoviae sp. nov*., an invasive pathogen causing bleeding stem lesions on forest trees and foliar necrosis of ornamentals in the UK. *Mycol Res,* 109**,** 853-9.

BROMAN, K. W. 2005. The genomes of recombinant inbred lines. *Genetics,* 169**,** 1133-46.

CAILLAUD, M. C., PIQUEREZ, S. J., FABRO, G., STEINBRENNER, J., ISHAQUE, N., BEYNON, J. & JONES, J. D. 2012. Subcellular localization of the Hpa RxLR effector repertoire identifies a tonoplast-associated protein HaRxL17 that confers enhanced plant susceptibility. *Plant J,* 69**,** 252-65.

CHALUPOWICZ, L., DOMBROVSKY, A., GABA, V., LURIA, N., REUVEN, M., BEERMAN, A., LACHMAN, O., DROR, O., NISSAN, G. & MANULIS-SASSON, S. 2019. Diagnosis of plant diseases using the Nanopore sequencing platform. *Plant Pathology,* 68**,** 229-238.

CHAPARRO-GARCIA, A., SCHWIZER, S., SKLENAR, J., YOSHIDA, K., PETRE, B., BOS, J. I., SCHORNACK, S., JONES, A. M., BOZKURT, T. O. & KAMOUN, S. 2015. *Phytophthora infestans* RXLR-WY Effector AVR3a Associates with Dynamin-Related Protein 2 Required for Endocytosis of the Plant Pattern Recognition Receptor FLS2. *PLoS One,* 10**,** e0137071.

CHAPARRO-GARCIA A, W. R., GIMENEZ-IBANEZ S, FINDLAY K, COFFEY MD, ET AL. 2011. The Receptor-Like Kinase SERK3/BAK1 Is Required for Basal Resistance against the Late Blight Pathogen *Phytophthora infestans* in *Nicotiana benthamiana*. *PLOS One,* 6**,** e16608.

CHEN, X. R., XING, Y. P., LI, Y. P., TONG, Y. H. & XU, J. Y. 2013. RNA-Seq reveals infection-related gene expression changes in *Phytophthora capsici*. *PLoS One,* 8**,** e74588.

CHEVREUX, B., WETTER, T., SUHAI, S. 1999. Genome Sequence Assembly Using Trace Signals and Additional Sequence Information. *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB),* 99**,** 45-56.

CHILDERS, R., DANIES, G., MYERS, K., FEI, Z., SMALL, I. M. & FRY, W. E. 2015. Acquired Resistance to Mefenoxam in Sensitive Isolates of *Phytophthora infestans*. *Phytopathology,* 105**,** 342-9.

CHISHOLM, S. T., COAKER, G., DAY, B. & STASKAWICZ, B. J. 2006. Host-microbe interactions: shaping the evolution of the plant immune response. *Cell,* 124**,** 803-14.

COOKE, D. E., DRENTH, A., DUNCAN, J. M., WAGELS, G. & BRASIER, C. M. 2000. A molecular phylogeny of *Phytophthora* and related oomycetes. *Fungal Genet Biol,* 30**,** 17-32.

DAVIDSON, C. R., CARROLL, R. B., EVANS, T. A., MULROONEY, R. P. & KIM, S. H. 2002. First Report of *Phytophthora capsici* Infecting Lima Bean (*Phaseolus lunatus*) in the Mid-Atlantic Region. *Plant Dis,* 86**,** 1049.

DE COCK, A. W., MENDOZA, L., PADHYE, A. A., AJELLO, L. & KAUFMAN, L. 1987. *Pythium insidiosum sp. nov*., the etiologic agent of pythiosis. *J Clin Microbiol,* 25**,** 344-9.

DE CREMER, K., MATHYS, J., VOS, C., FROENICKE, L., MICHELMORE, R. W., CAMMUE, B. P. & DE CONINCK, B. 2013. RNAseq-based transcriptome analysis of *Lactuca sativa* infected by the fungal necrotroph *Botrytis cinerea*. *Plant Cell Environ,* 36**,** 1992-2007.

DELCHER, A. L., PHILLIPPY, A., CARLTON, J. & SALZBERG, S. L. 2002. Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res,* 30**,** 2478-83.

DENANCE, N., SANCHEZ-VALLET, A., GOFFNER, D. & MOLINA, A. 2013. Disease resistance or growth: the role of plant hormones in balancing immune responses and fitness costs. *Front Plant Sci,* 4**,** 155.

DODDS, P. N. & RATHJEN, J. P. 2010. Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat Rev Genet,* 11**,** 539-48.

DOEHLEMANN, G. & HEMETSBERGER, C. 2013. Apoplastic immunity and its suppression by filamentous plant pathogens. *New Phytol,* 198**,** 1001-16.

DREIER, M., BERTHOUD, H., SHANI, N., WECHSLER, D., JUNIER, P. 2019a. SpeciesPrimer: A bioinformatics pipeline dedicated to thedesign of qPCR primers for the quantification of bacterial species. *PeerJ Preprints,* 7.

DREIER, M., BERTHOUD, H., SHANI, N., WECHSLER, D., JUNIER, P. 2019b. SpeciesPrimer: A bioinformatics pipeline dedicated to thedesign of qPCR primers for the quantification of bacterialspecies. *PeerJ Preprints,* 7.

DUNN, A. R., BRUENING, S. R., GRUNWALD, N. J. & SMART, C. D. 2014. Evolution of an Experimental Population of *Phytophthora capsici* in the Field. *Phytopathology,* 104**,** 1107-17.

ENZENBACHER, T. B., NAEGELE, R. P. & HAUSBECK, M. K. 2015. Susceptibility of Greenhouse Ornamentals to *Phytophthora capsici* and *P. tropicalis*. *Plant Disease,* 99**,** 1808-1815.

ERWIN, D. C., RIBEIRO, O. K. 1996. *Phytophthora Diseases Worldwide,* St. Paul: MN, American Phytopathology Society.

EVANS, G. & AUSTIN, F. 2010. Collaborations among academia, government, and industry in the diagnostics space: barriers and some ideas for solutions. *Sci Transl Med,* 2**,** 63mr3.

EVANS, J. A. 2010. Industry Induces Academic Science to Know Less about More. *American Journal of Sociology,* 116**,** 389-452.

FEAU, N., TAYLOR, G., DALE, A. L., DHILLON, B., BILODEAU, G. J., BIROL, I., JONES, S. J. & HAMELIN, R. C. 2016. Genome sequences of six *Phytophthora* species threatening forest ecosystems. *Genom Data,* 10**,** 85-88.

FLORKOWSKI, C. M. 2008. Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests. *Clin Biochem Rev,* 29 Suppl 1**,** S83-7.

FOSTER, J. M. & HAUSBECK, M. K. 2010. Resistance of Pepper to *Phytophthora* Crown, Root, and Fruit Rot Is Affected by Isolate Virulence. *Plant Dis,* 94**,** 24-30.

FUJIMORI, S., HASEGAWA, T., KREY, V., RIAHI, K., BERTRAM, C., BODIRSKY, B. L., BOSETTI, V., CALLEN, J., DESPRES, J., DOELMAN, J., DROUET, L., EMMERLING, J., FRANK, S., FRICKO, O., HAVLIK, P., HUMPENODER, F., KOOPMAN, J. F. L., VAN MEIJL, H., OCHI, Y., POPP, A., SCHMITZ, A., TAKAHASHI, K. & VAN VUUREN, D. 2019. A multi-model assessment of food security implications of climate change mitigation. *Nature Sustainability,* 2**,** 386-396.

GAASTRA, W., LIPMAN, L. J., DE COCK, A. W., EXEL, T. K., PEGGE, R. B., SCHEURWATER, J., VILELA, R. & MENDOZA, L. 2010. *Pythium insidiosum*: an overview. *Vet Microbiol,* 146**,** 1-16.

GAO, R., CHENG, Y., WANG, Y., WANG, Y., GUO, L. & ZHANG, G. 2015. Genome Sequence of *Phytophthora fragariae var. fragariae*, a Quarantine Plant-Pathogenic Fungus. *Genome Announc,* 3.

GARIBALDI, A., GILARDI, G., & GULLINO, M. L. 2004. Varietal resistance of lettuce to *Fusarium oxysporum f. sp. lactucae.* . *Crop Protection,* 23**,** 845-851.

GIORDANO, P. R., WANG, J., VARGAS, J. M., JACOBS, J., CHILVERS, M. I. & ZENG, Q. 2018. Using a Genome-Based PCR Primer Prediction Pipeline to Develop Molecular Diagnostics for the Turfgrass Pathogen *Acidovorax avenae*. *Plant Dis,* 102**,** 2224-2232.

GOBENA, D., MCGRATH, M. T., & LAMOUR, K. 2011. Survival and spread of *Phytophthora capsici* on Long Island, New York. *Mycological Progress,* 11**,** 761-768.

GOBENA, D., ROIG, J., GALMARINI, C., HULVEY, J. & LAMOUR, K. 2012. Genetic diversity of *Phytophthora capsici* isolates from pepper and pumpkin in Argentina. *Mycologia,* 104**,** 102-7.

GOVERS, F. & GIJZEN, M. 2006. *Phytophthora* genomics: the plant destroyers' genome decoded. *Mol Plant Microbe Interact,* 19**,** 1295-301.

GRANKE, L. L., QUESADA-OCAMPO, L., LAMOUR, K. & HAUSBECK, M. K. 2012. Advances in Research on *Phytophthora capsici* on Vegetable Crops in The United States. *Plant Disease,* 96**,** 1588-1600.

GRENVILLE-BRIGGS, L. J., AVROVA, A. O., BRUCE, C. R., WILLIAMS, A., WHISSON, S. C., BIRCH, P. R. & VAN WEST, P. 2005. Elevated amino acid biosynthesis in *Phytophthora infestans* during appressorium formation and potato infection. *Fungal Genet Biol,* 42**,** 244-56.

GRENVILLE-BRIGGS, L. J., KUSHWAHA, S. K., CLEARY, M. R., WITZELL, J., SAVENKOV, E. I., WHISSON, S. C., CHAWADE, A. & VETUKURI, R. R. 2017. Draft genome of the oomycete pathogen *Phytophthora cactorum* strain LV007 isolated from European beech (*Fagus sylvatica*). *Genom Data,* 12**,** 155-156.

GRUNWALD, N. J., GOSS, E. M. & PRESS, C. M. 2008. *Phytophthora ramorum*: a pathogen with a remarkably wide host range causing sudden oak death on oaks and ramorum blight on woody ornamentals. *Mol Plant Pathol,* 9**,** 729-40.

GUIMÓN, J. 2013. Promoting University-Industry Collaborations in Developing Countries. *World Bank Policy Brief.* Washington: World Bank.

GULVIK, C. A., ARTHUR, R. A., HUMRIGHOUSE, B. W., BATRA, D., ROWE, L. A., LASKER, B. A. & MCQUISTON, J. R. 2019. Complete Genome Sequence of *Nocardia farcinica* W6977(T) Obtained by Combining Illumina and PacBio Reads. *Microbiol Resour Announc,* 8.

GUREVICH, A., SAVELIEV, V., VYAHHI, N. & TESLER, G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics,* 29**,** 1072-5.

HAAS, B. J., KAMOUN, S., ZODY, M. C., JIANG, R. H., HANDSAKER, R. E., CANO, L. M., GRABHERR, M., KODIRA, C. D., RAFFAELE, S., TORTO-ALALIBO, T., BOZKURT, T. O., AH-FONG, A. M., ALVARADO, L., ANDERSON, V. L., ARMSTRONG, M. R., AVROVA, A., BAXTER, L., BEYNON, J., BOEVINK, P. C., BOLLMANN, S. R., BOS, J. I., 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., GRUNWALD, N. J., HORN, K., HORNER, N. R., HU, C. H., HUITEMA, E., JEONG, D. H., JONES, A. M., JONES, J. D., 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., MORGAN, W., MORRIS, P. F., MUNRO, C. A., O'NEILL, K., OSPINA-GIRALDO, M., PINZON, 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., 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., WHISSON, S. C., JUDELSON, H. S. & NUSBAUM, C. 2009. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature,* 461**,** 393-8.

HANTULA, J., LIJA, A., NUORTEVA, H., PARIKKA, P., WERRES, S. 2000. Pathogenicity, morphology and genetic variation of *Phytophthora cactorum* from strawberry, apple, rhododendron, and silver birch. *Mycol Res,* 104**,** 1062-1068.

HARDHAM, A. R. 2001. The cell biology behind *Phytophthora* pathogenicity. *Australasian Plant Pathology,* 30**,** 91-98.

HARDHAM, A. R. 2007. Cell biology of plant-oomycete interactions. *Cell Microbiol,* 9**,** 31-9.

HAUSBECK, M. K., LAMOUR, K. 2004. *Phytophthora capsici* on Vegetable Crops: Research Progress and Management Challenges. *Plant Disease,* 88**,** 1292 - 1303.

HAYASHIDA, K., ORBA, Y., SEQUEIRA, P. C., SUGIMOTO, C., HALL, W. W., ESHITA, Y., SUZUKI, Y., RUNTUWENE, L., BRASIL, P., CALVET, G., RODRIGUES, C. D. S., SANTOS, C. C. D., MARES-GUIA, M. A. M., YAMAGISHI, J., FILIPPIS, A. M. B. & SAWA, H. 2019. Field diagnosis and genotyping of chikungunya virus using a dried reverse transcription loop-mediated

isothermal amplification (LAMP) assay and MinION sequencing. *PLoS Negl Trop Dis,* 13**,** e0007480.

HEATHER, J. M. & CHAIN, B. 2016. The sequence of sequencers: The history of sequencing DNA. *Genomics,* 107**,** 1-8.

HOENEN, T. 2016. Sequencing of Ebola Virus Genomes Using Nanopore Technology. *Bio Protoc,* 6.

HOENEN, T., GROSETH, A., ROSENKE, K., FISCHER, R. J., HOENEN, A., JUDSON, S. D., MARTELLARO, C., FALZARANO, D., MARZI, A., SQUIRES, R. B., WOLLENBERG, K. R., DE WIT, E., PRESCOTT, J., SAFRONETZ, D., VAN DOREMALEN, N., BUSHMAKER, T., FELDMANN, F., MCNALLY, K., BOLAY, F. K., FIELDS, B., SEALY, T., RAYFIELD, M., NICHOL, S. T., ZOON, K. C., MASSAQUOI, M., MUNSTER, V. J. & FELDMANN, H. 2016. Nanopore Sequencing as a Rapidly Deployable Ebola Outbreak Tool. *Emerg Infect Dis,* 22**,** 331-4.

HOSSEINI, S. Unpublished. *Phytophthora pisi* genome. *Unpublished*.

HOWDEN, A. J. & HUITEMA, E. 2012. Effector-triggered post-translational modifications and their role in suppression of plant immunity. *Front Plant Sci,* 3**,** 160.

HU, J., DIAO, Y., ZHOU, Y., LIN, D., BI, Y., PANG, Z., TROUT FRYXELL, R., LIU, X. & LAMOUR, K. 2013a. Loss of heterozygosity drives clonal diversity of *Phytophthora capsici* in China. *PLoS One,* 8**,** e82691.

HU, J., PANG, Z., BI, Y., SHAO, J., DIAO, Y., GUO, J., LIU, Y., LV, H., LAMOUR, K. & LIU, X. 2013b. Genetically diverse long-lived clonal lineages of *Phytophthora capsici* from pepper in Gansu, China. *Phytopathology,* 103**,** 920-6.

HU, J., SHRESTHA, S., ZHOU, Y., LIU, X., LAMOUR, K. 2019. Dynamic Extreme Aneuploidy (DEA) in the vegetable pathogen *Phytophthora capsici* sheds light on instant evolution and intractability.

HURTADO-GONZALES, O., ARAGON-CABALLERO, L., APAZA-TAPIA, W., DONAHOO, R. & LAMOUR, K. 2008. Survival and spread of *Phytophthora capsici* in Coastal Peru. *Phytopathology,* 98**,** 688-94.

IRWIN, J. A., CRAWFORD, A. & DRENTH, A. 1997. The Origins of *Phytophthora* Species Attacking Legumes in Australia. *Advances in Botanical Research,* 24**,** 431-456.

ISTACE, B., FRIEDRICH, A., D'AGATA, L., FAYE, S., PAYEN, E., BELUCHE, O., CARADEC, C., DAVIDAS, S., CRUAUD, C., LITI, G., LEMAINQUE, A., ENGELEN, S., WINCKER, P., SCHACHERER, J. & AURY, J. M. 2017. de novo assembly and population genomic survey of natural yeast isolates with the Oxford Nanopore MinION sequencer. *Gigascience,* 6**,** 1-13.

JANSEN, H. J., LIEM, M., JONG-RAADSEN, S. A., DUFOUR, S., WELTZIEN, F. A., SWINKELS, W., KOELEWIJN, A., PALSTRA, A. P., PELSTER, B., SPAINK, H. P., THILLART, G., DIRKS, R. P. & HENKEL, C. V. 2017. Rapid de novo assembly of the European eel genome from nanopore sequencing reads. *Sci Rep,* 7**,** 7213.

JASHNI, M. K., MEHRABI, R., COLLEMARE, J., MESARICH, C. H. & DE WIT, P. J. 2015. The battle in the apoplast: further insights into the roles of proteases and their inhibitors in plant-pathogen interactions. *Front Plant Sci,* 6**,** 584.

JEFFERS, S. N., MARTIN, S. B. 1986. Comparison of two media selective for Phytophthora and Pythium species. *Plant Dis,* 70**,** 1038-1043.

JIANG, R. H. & TYLER, B. M. 2012. Mechanisms and evolution of virulence in oomycetes. *Annu Rev Phytopathol,* 50**,** 295-318.

JOHNSTONE, M., CHATTERTON, S., SUTTON, J. C. & GRODZINSKI, B. 2005. Net Carbon Gain and Growth of Bell Peppers, *Capsicum annuum* 'Cubico', Following Root Infection by *Pythium aphanidermatum*. *Phytopathology,* 95**,** 354-61.

JONES, J. D. & DANGL, J. L. 2006. The plant immune system. *Nature,* 444**,** 323-9.

JOSHI, N. A., FASS, J. N. 2011. Sickle: A sliding window, adaptive, quality-based trimming tool for FastQ files [Software]. 1.33 ed.

JUPE, J., STAM, R., HOWDEN, A. J., MORRIS, J. A., ZHANG, R., HEDLEY, P. E. & HUITEMA, E. 2013. *Phytophthora capsici*-tomato interaction features dramatic shifts in gene expression associated with a hemi-biotrophic lifestyle. *Genome Biol,* 14**,** R63.

KAMOUN, S. 2003. Molecular genetics of pathogenic oomycetes. *Eukaryot Cell,* 2**,** 191-9.

KAMOUN, S. 2006. A catalogue of the effector secretome of plant pathogenic oomycetes. *Annu Rev Phytopathol,* 44**,** 41-60.

KARL, M. M., POEHLEIN, A., BENGELSDORF, F. R., DANIEL, R. & DURRE, P. 2017. Complete Genome Sequence of the Autotrophic Acetogen *Clostridium formicaceticum* DSM 92T Using Nanopore and Illumina Sequencing Data. *Genome Announc,* 5.

KIM, E. S., HWANG, B. K. 1992. Virulence of Korean Pepper cultivars of Isolates of Phytophthora capsici from Different Geographical Areas. *Plant Disease,* 76**,** 486-489.

KONG, F., JAMES, G., GORDON, S., ZELYNSKI, A. & GILBERT, G. L. 2001. Species-specific PCR for identification of common contaminant mollicutes in cell culture. *Appl Environ Microbiol,* 67**,** 3195-200.

KONG, P., HONG, C., JEFFERS, S. N. & RICHARDSON, P. A. 2003. A Species-Specific Polymerase Chain Reaction Assay for Rapid Detection of *Phytophthora nicotianae* in Irrigation Water. *Phytopathology,* 93**,** 822-31.

KOREN, S., WALENZ, B. P., BERLIN, K., MILLER, J. R., BERGMAN, N. H. & PHILLIPPY, A. M. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res,* 27**,** 722-736.

KOX, L. F., BROUWERSHAVEN, I. R., VOSSENBERG, B. T., BELD, H. E., BONANTS, P. J. & GRUYTER, J. 2007. Diagnostic Values and Utility of Immunological, Morphological, and Molecular Methods for In Planta Detection of *Phytophthora ramorum*. *Phytopathology,* 97**,** 1119-29.

KRANZ, A., VOGEL, A., DEGNER, U., KIEFLER, I., BOTT, M., USADEL, B. & POLEN, T. 2017. High precision genome sequencing of engineered *Gluconobacter oxydans* 621H by combining long nanopore and short accurate Illumina reads. *J Biotechnol*.

KREUTZER, W. A., BRYANT, L. R. 1946. Certain aspects of the epiphytology and control of tomato fruit rot caused by *Phytophthora capsici* Leonian. *Phytopathology,* 36**,** 329-339.

KULSKI, J. K. 2016. Next-Generation Sequencing - an overview of the history, tools and 'omic' approaches. *In:* KULSKI, J. K. (ed.) *Next-Generation Sequencing - Advances, Applications and Challenges.* Intech.

LAMOUR, K. H. & HAUSBECK, M. K. 2001. The Dynamics of Mefenoxam Insensitivity in a Recombining Population of *Phytophthora capsici* Characterized with Amplified Fragment Length Polymorphism Markers. *Phytopathology,* 91**,** 553-7.

LAMOUR, K. H., MUDGE, J., GOBENA, D., HURTADO-GONZALES, O. P., SCHMUTZ, J., KUO, A., MILLER, N. A., RICE, B. J., RAFFAELE, S., CANO, L. M., BHARTI, A. K., DONAHOO, R. S., FINLEY, S., HUITEMA, E., HULVEY, J., PLATT, D., SALAMOV, A., SAVIDOR, A., SHARMA, R., STAM, R., STOREY, D., THINES, M., WIN, J., HAAS, B. J., DINWIDDIE, D. L., JENKINS, J., KNIGHT, J. R., AFFOURTIT, J. P., HAN, C. S., CHERTKOV, O., LINDQUIST, E. A., DETTER, C., GRIGORIEV, I. V., KAMOUN, S. & KINGSMORE, S. F. 2012a. Genome sequencing and mapping reveal loss of heterozygosity as a mechanism for rapid adaptation in the vegetable pathogen *Phytophthora capsici*. *Mol Plant Microbe Interact,* 25**,** 1350-60.

LAMOUR, K. H., STAM, R., JUPE, J. & HUITEMA, E. 2012b. The oomycete broad-host-range pathogen Phytophthora capsici. *Mol Plant Pathol,* 13**,** 329-37.

LAMOUR, K. H., WIN, J. & KAMOUN, S. 2007. Oomycete genomics: new insights and future directions. *FEMS Microbiol Lett,* 274**,** 1-8.

LANGMEAD, B. 2010. Aligning short sequencing reads with Bowtie. *Curr Protoc Bioinformatics,* Chapter 11**,** Unit 11 7.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA PROCESSING, S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.

LI, R., ZHU, H., RUAN, J., QIAN, W., FANG, X., SHI, Z., LI, Y., LI, S., SHAN, G., KRISTIANSEN, K., LI, S., YANG, H., WANG, J. & WANG, J. 2010. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res,* 20**,** 265-72.

LIEM, M., JANSEN, H. J., DIRKS, R. P., HENKEL, C. V., VAN HEUSDEN, G. P. H., LEMMERS, R., OMER, T., SHAO, S., PUNT, P. J. & SPAINK, H. P. 2017. De novo whole-genome assembly of a wild type yeast isolate using nanopore sequencing. *F1000Res,* 6**,** 618.

LIU, H., MA, X., YU, H., FANG, D., LI, Y., WANG, X., WANG, W., DONG, Y. & XIAO, B. 2016. Genomes and virulence difference between two physiological races of *Phytophthora nicotianae*. *Gigascience,* 5**,** 3.

LOMAN, N. J., QUICK, J. & SIMPSON, J. T. 2015. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat Methods,* 12**,** 733-5.

LOMAN, N. J. & QUINLAN, A. R. 2014. Poretools: a toolkit for analyzing nanopore sequence data. *Bioinformatics,* 30**,** 3399-401.

LONGMUIR, A. L., BEECH, P. L. & RICHARDSON, M. F. 2017. Draft genomes of two Australian strains of the plant pathogen, *Phytophthora cinnamomi*. *F1000Res,* 6**,** 1972.

MA, K. W. & MA, W. 2016. Phytohormone pathways as targets of pathogens to facilitate infection. *Plant Mol Biol,* 91**,** 713-25.

MA, Z., ZHU, L., SONG, T., WANG, Y., ZHANG, Q., XIA, Y., QIU, M., LIN, Y., LI, H., KONG, L., FANG, Y., YE, W., WANG, Y., DONG, S., ZHENG, X., TYLER, B. M. & WANG, Y. 2017. A paralogous decoy protects *Phytophthora sojae* apoplastic effector PsXEG1 from a host inhibitor. *Science,* 355**,** 710-714.

MAFURAH, J. J., MA, H., ZHANG, M., XU, J., HE, F., YE, T., SHEN, D., CHEN, Y., RAJPUT, N. A. & DOU, D. 2015. A Virulence Essential CRN Effector of *Phytophthora capsici* Suppresses Host Defense and Induces Cell Death in Plant Nucleus. *PLoS One,* 10**,** e0127965.

MANOHARA, D., MULYA, K., PURWANTARA, A., WAHYUNO, D. 2004. Phytophthora capsici on Black Pepper in Indonesia. *In:* DRENTH, A., GUEST, D. I. (ed.) *Diversity and Management of Phytophthora in Southeast Asia.* Australia: ACIAR Monograph.

MARDIS, E. R. 2008. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet,* 9**,** 387-402.

MARTIN, F. N., ABAD, Z. G., BALCI, Y. & IVORS, K. 2012. Identification and Detection of *Phytophthora*: Reviewing Our Progress, Identifying Our Needs. *Plant Dis,* 96**,** 1080-1103.

MARTIN, F. N., BLAIR, J. E. & COFFEY, M. D. 2014. A combined mitochondrial and nuclear multilocus phylogeny of the genus *Phytophthora*. *Fungal Genet Biol,* 66**,** 19-32.

MCCARTHY, C. G. P. & FITZPATRICK, D. A. 2017. Phylogenomic Reconstruction of the Oomycete Phylogeny Derived from 37 Genomes. *mSphere,* 2.

MCGRATH, M. T., STRAUSS, J. & DILLARD, H. R. 2011. First Report of Phytophthora Blight Caused by *Phytophthora capsici* on Snap Bean in New York. *Plant Dis,* 95**,** 1028.

MEITZ, J. C., LINDE, C. C., THOMPSON, A., LANGENHOVEN, S., MCLEOD, A. 2010. *Phytophthora capsici* on vegetable hosts in South Africa: Distribution, host range and genetic diversity. *Australasian Plant Pathology,* 39**,** 431-439.

MICHELMORE, R. 2000. Genomic approaches to plant disease resistance. *Curr Opin Plant Biol,* 3**,** 125-31.

MONROY-BARBOSA, A., BOSLAND, P. W. 2010. A Rapid Technique for Multiple-race Disease Screening of *Phytophthora* Foliar Blight on Single *Capsicum annuum L.* Plants. *HortScience,* 45**,** 1563-1566.

MORGAN, J. B., CONNOLLY, E. L. 2013. Plant-Soil: Nutrient Uptake. *Nature Education Knowledge,* 4**,** 2.

MORITA, Y. & TOJO, M. 2007. Modifications of PARP Medium Using Fluazinam, Miconazole, and Nystatin for Detection of *Pythium* spp. in Soil. *Plant Dis,* 91**,** 1591-1599.

NEUPANE, S., ANDERSEN, E. J., NEUPANE, A. & NEPAL, M. P. 2018. Genome-Wide Identification of NBS-Encoding Resistance Genes in Sunflower (*Helianthus annuus* L.). *Genes (Basel),* 9.

NURK, S., BANKEVICH, A., ANTIPOV, D., GUREVICH, A. A., KOROBEYNIKOV, A., LAPIDUS, A., PRJIBELSKI, A. D., PYSHKIN, A., SIROTKIN, A., SIROTKIN, Y., STEPANAUSKAS, R., CLINGENPEEL, S. R., WOYKE, T., MCLEAN, J. S., LASKEN, R., TESLER, G., ALEKSEYEV, M. A. & PEVZNER, P. A.

2013. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J Comput Biol,* 20**,** 714-37.

NURNBERGER, T., BRUNNER, F., KEMMERLING, B. & PIATER, L. 2004. Innate immunity in plants and animals: striking similarities and obvious differences. *Immunol Rev,* 198**,** 249-66.

O'BRIEN, P. A. 2008. PCR Primers for Specific Detection of *Phytophthora cinnamomi*. *Australasian Plant Pathology,* 37.

ORTEGA, R. G., ESPAÑOL, C. P. & ZUECO, J. C. 1991. Genetics of Resistance to *Phytophthora capsici* in the Pepper Line 'SCM-334'. *Plant Breeding,* 107**,** 50-55.

OSSOWSKI, S., SCHNEEBERGER, K., CLARK, R. M., LANZ, C., WARTHMANN, N. & WEIGEL, D. 2008. Sequencing of natural strains of *Arabidopsis thaliana* with short reads. *Genome Res,* 18**,** 2024-33.

PANG, Z., CHEN, L., MIAO, J., WANG, Z., BULONE, V. & LIU, X. 2015. Proteomic profile of the plant-pathogenic oomycete *Phytophthora capsici* in response to the fungicide pyrimorph. *Proteomics,* 15**,** 2972-82.

PANG, Z., SHAO, J., CHEN, L., LU, X., HU, J., QIN, Z. & LIU, X. 2013. Resistance to the novel fungicide pyrimorph in *Phytophthora capsici*: risk assessment and detection of point mutations in CesA3 that confer resistance. *PLoS One,* 8**,** e56513.

PANG, Z., SHAO, J., HU, J., CHEN, L., WANG, Z., QIN, Z. & LIU, X. 2014. Competition between pyrimorph-sensitive and pyrimorph-resistant isolates of *Phytophthora capsici*. *Phytopathology,* 104**,** 269-74.

PARRA, G. & RISTAINO, J. B. 2001. Resistance to Mefenoxam and Metalaxyl Among Field Isolates of *Phytophthora capsici* Causing *Phytophthora* Blight of Bell Pepper. *Plant Dis,* 85**,** 1069-1075.

PARVEEN, T., SHARMA, K. 2015. *Pythium* diseases, control and management strategies: a review. *International Journal of Plant, Animal and Environmental Sciences,* 5**,** 224-257.

PAULUS, J. K. & VAN DER HOORN, R. A. L. 2018. Tricked or trapped-Two decoy mechanisms in host-pathogen interactions. *PLoS Pathog,* 14**,** e1006761.

PEREIRA, D. M., FERNANDES, J. C., VALENTAO, P., ANDRADE, P. B. 2015. "Omics" Technologies: Promises and Benefits for Molecular Medicine. *Principles of Translational Science in Medicine.* Elsevier Inc.

PLICH, M., RUDNICKI, R. M. 1979. Studies of the Toxins of *Phytophthora cactorum* Pathoegnic to Apple Trees. *Journal of Phytopathology,* 94**,** 270-278.

POSTMA, J., STEVENS, L. H., WIEGERS, G., L., DAVELAAR, E., NIJHUIS, E., H. 2009. Biological control of *Pythium aphanidermatum* in cucumber with a combined application of Lysobacter enzymogenes strain 3.1T8 and chitosan. *Biological Control,* 48**,** 301-309.

POTATO GENOME SEQUENCING, C., XU, X., PAN, S., CHENG, S., ZHANG, B., MU, D., NI, P., ZHANG, G., YANG, S., LI, R., WANG, J., ORJEDA, G., GUZMAN, F., TORRES, M., LOZANO, R., PONCE, O., MARTINEZ, D., DE LA CRUZ, G., CHAKRABARTI, S. K., PATIL, V. U., SKRYABIN, K. G., KUZNETSOV, B. B., RAVIN, N. V., KOLGANOVA, T. V., BELETSKY, A. V., MARDANOV, A. V., DI GENOVA, A., BOLSER, D. M., MARTIN, D. M., LI, G., YANG, Y., KUANG, H., HU, Q., XIONG, X., BISHOP, G. J., SAGREDO, B., MEJIA, N., ZAGORSKI, W., GROMADKA, R., GAWOR, J., SZCZESNY, P., HUANG, S., ZHANG, Z., LIANG, C., HE, J., LI, Y., HE, Y., XU, J., ZHANG, Y., XIE, B., DU, Y., QU, D., BONIERBALE, M., GHISLAIN, M., HERRERA MDEL, R., GIULIANO, G., PIETRELLA, M., PERROTTA, G., FACELLA, P., O'BRIEN, K., FEINGOLD, S. E., BARREIRO, L. E., MASSA, G. A., DIAMBRA, L., WHITTY, B. R., VAILLANCOURT, B., LIN, H., MASSA, A. N., GEOFFROY, M., LUNDBACK, S., DELLAPENNA, D., BUELL, C. R., SHARMA, S. K., MARSHALL, D. F., WAUGH, R., BRYAN, G. J., DESTEFANIS, M., NAGY, I., MILBOURNE, D., THOMSON, S. J., FIERS, M., JACOBS, J. M., NIELSEN, K. L., SONDERKAER, M., IOVENE, M., TORRES, G. A., JIANG, J., VEILLEUX, R. E., BACHEM, C. W., DE BOER, J., BORM, T., KLOOSTERMAN, B., VAN ECK, H., DATEMA, E., HEKKERT, B., GOVERSE, A., VAN HAM, R. C. & VISSER, R. G. 2011. Genome sequence and analysis of the tuber crop potato. *Nature,* 475**,** 189-95.

PRITCHARD, L., HOLDEN, N. J., BIELASZEWSKA, M., KARCH, H. & TOTH, I. K. 2012. Alignment-free design of highly discriminatory diagnostic primer sets for *Escherichia coli* O104:H4 outbreak strains. *PLoS One,* 7**,** e34498.

PRITCHARD, L., HUMPHRIS, S., SADDLER, G. S., PARKINSON, N. M., BERTRAND, V., ELPHINSTONE, J. G., TOTH, I. K. 2013. Detection of phytopathogens of the genus *Dickeya* using a PCR primer prediction pipeline for draft bacterial genome sequences. *Plant Pathology,* 62**,** 587-596.

PUTNAM, M. L. 1995. Evaluation of Selected Methods of Plant-Disease Diagnosis. *Crop Protection,* 14**,** 517-525.

QUENTIN GRAFTON, R., DAUGBJERG, C., EJAZ QURESHI, M. 2015. Towards Food Security by 2050. *Food Security,* 7**,** 179-183.

QUESADA-OCAMPO, L. M., FULBRIGHT, D. W. & HAUSBECK, M. K. 2009. Susceptibility of Fraser Fir to *Phytophthora capsici*. *Plant Dis,* 93**,** 135-141.

QUICK, J., ASHTON, P., CALUS, S., CHATT, C., GOSSAIN, S., HAWKER, J., NAIR, S., NEAL, K., NYE, K., PETERS, T., DE PINNA, E., ROBINSON, E., STRUTHERS, K., WEBBER, M., CATTO, A., DALLMAN, T. J., HAWKEY, P. & LOMAN, N. J. 2015. Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of Salmonella. *Genome Biol,* 16**,** 114.

QUICK, J., LOMAN, N. J., DURAFFOUR, S., SIMPSON, J. T., SEVERI, E., COWLEY, L., BORE, J. A., KOUNDOUNO, R., DUDAS, G., MIKHAIL, A., OUEDRAOGO, N., AFROUGH, B., BAH, A., BAUM, J. H., BECKER-ZIAJA, B., BOETTCHER, J. P., CABEZA-CABRERIZO, M., CAMINO-SANCHEZ, A., CARTER, L. L., DOERRBECKER, J., ENKIRCH, T., DORIVAL, I. G. G., HETZELT, N., HINZMANN, J., HOLM, T., KAFETZOPOULOU, L. E., KOROPOGUI, M., KOSGEY, A., KUISMA, E., LOGUE, C. H., MAZZARELLI, A., MEISEL, S., MERTENS, M., MICHEL, J., NGABO, D., NITZSCHE, K., PALLASH, E., PATRONO, L. V., PORTMANN, J., REPITS, J. G., RICKETT, N. Y., SACHSE, A., SINGETHAN, K., VITORIANO, I., YEMANABERHAN, R. L., ZEKENG, E. G., TRINA, R., BELLO, A., SALL, A. A., FAYE, O., FAYE, O., MAGASSOUBA, N., WILLIAMS, C. V., AMBURGEY, V., WINONA, L., DAVIS, E., GERLACH, J., WASHINGTON, F., MONTEIL, V., JOURDAIN, M., BERERD, M., CAMARA, A., SOMLARE, H., CAMARA, A., GERARD, M., BADO, G., BAILLET, B., DELAUNE, D., NEBIE, K. Y., DIARRA, A., SAVANE, Y., PALLAWO, R. B., GUTIERREZ, G. J., MILHANO, N., ROGER, I., WILLIAMS, C. J., YATTARA, F., LEWANDOWSKI, K., TAYLOR, J., RACHWAL, P., TURNER, D., POLLAKIS, G., HISCOX, J. A., MATTHEWS, D. A., O'SHEA, M. K., JOHNSTON, A. M., WILSON, D., HUTLEY, E., SMIT, E., DI CARO, A., WOELFEL, R., STOECKER, K., FLEISCHMANN, E., GABRIEL, M., WELLER, S. A., KOIVOGUI, L., DIALLO, B., KEITA, S., RAMBAUT, A., FORMENTY, P., et al. 2016. Real-time, portable genome sequencing for Ebola surveillance. *Nature,* 530**,** 228-232.

QUINLAN, A. R. & HALL, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics,* 26**,** 841-2.

QUINN, L., O'NEILL, P. A., HARRISON, J., PASKIEWICZ, K. H., MCCRACKEN, A. R., COOKE, L. R., GRANT, M. R. & STUDHOLME, D. J. 2013. Genome-wide sequencing of *Phytophthora lateralis* reveals genetic variation among isolates from Lawson cypress (*Chamaecyparis lawsoniana*) in Northern Ireland. *FEMS Microbiol Lett,* 344**,** 179-85.

RANDALL, E., YOUNG, V., SIEROTZKI, H., SCALLIET, G., BIRCH, P. R., COOKE, D. E., CSUKAI, M. & WHISSON, S. C. 2014. Sequence diversity in the large subunit of RNA polymerase I contributes to Mefenoxam insensitivity in *Phytophthora infestans*. *Mol Plant Pathol,* 15**,** 664-76.

RESTREPO, S., TABIMA, J. F., MIDEROS, M. F., GRUNWALD, N. J. & MATUTE, D. R. 2014. Speciation in fungal and oomycete plant pathogens. *Annu Rev Phytopathol,* 52**,** 289-316.

RICE, P., LONGDEN, I. & BLEASBY, A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet,* 16**,** 276-7.

RISTAINO, J. B. 2002. Tracking historic migrations of the Irish potato famine pathogen, *Phytophthora infestans*. *Microbes Infect,* 4**,** 1369-77.

RIZZO, D. M., GARBELOTTO, M. & HANSEN, E. M. 2005. *Phytophthora ramorum*: integrative research and management of an emerging pathogen in California and Oregon forests. *Annu Rev Phytopathol,* 43**,** 309-35.

RODRIGUES-LUIZ, G. F., CARDOSO, M. S., VALDIVIA, H. O., AYALA, E. V., GONTIJO, C. M., RODRIGUES, T. S., FUJIWARA, R. T., LOPES, R. S. & BARTHOLOMEU, D. C. 2017. TipMT: Identification of PCR-based taxon-specific markers. *BMC Bioinformatics,* 18**,** 104.

ROSSO, M. L., RUPE, J. C., CHEN, P., MOZZINI, L. A. 2008. Inheritance and Genetic Mapping of Resistance to Damping-off Caused by in "Archer" Soybean. *Crop Science,* 48**,** 2215.

SAMBLES, C., SCHLENZIG, A., O'NEILL, P., GRANT, M. & STUDHOLME, D. J. 2015. Draft genome sequences of *Phytophthora kernoviae* and *Phytophthora ramorum* lineage EU2 from Scotland. *Genom Data,* 6**,** 193-4.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A,* 74**,** 5463-7.

SCHATZ, M. C., DELCHER, A. L. & SALZBERG, S. L. 2010. Assembly of large genomes using second-generation sequencing. *Genome Res,* 20**,** 1165-73.

SCHORNACK, S., VAN DAMME, M., BOZKURT, T. O., CANO, L. M., SMOKER, M., THINES, M., GAULIN, E., KAMOUN, S. & HUITEMA, E. 2010. Ancient class of translocated oomycete effectors targets the host nucleus. *Proc Natl Acad Sci U S A,* 107**,** 17421-6.

SELOSSE, M. A. & ROUSSET, F. 2011. Evolution. The plant-fungal marketplace. *Science,* 333**,** 828-9.

SHABAB, M., SHINDO, T., GU, C., KASCHANI, F., PANSURIYA, T., CHINTHA, R., HARZEN, A., COLBY, T., KAMOUN, S. & VAN DER HOORN, R. A. 2008. Fungal effector protein AVR2 targets diversifying defense-related cys proteases of tomato. *Plant Cell,* 20**,** 1169-83.

SHAH, F. A., ALLEN, N., WRIGHT, C. J. & BUTT, T. M. 2007. Repeated in vitro subculturing alters spore surface properties and virulence of *Metarhizium anisopliae*. *FEMS Microbiol Lett,* 276**,** 60-6.

SHAN, W. & HARDHAM, A. R. 2004. Construction of a bacterial artificial chromosome library, determination of genome size, and characterization of an Hsp70 gene family in *Phytophthora nicotianae*. *Fungal Genet Biol,* 41**,** 369-80.

SHRESTHA, S., HU, J., FRYXELL, R. T., MUDGE, J. & LAMOUR, K. 2014. SNP markers identify widely distributed clonal lineages of *Phytophthora colocasiae* in Vietnam, Hawaii and Hainan Island, China. *Mycologia,* 106**,** 676-85.

SILVAR, C., DUNCAN, J.M., COOKE, D.E.L. ET AL. 2005. Development of specific PCR primers for identification and detection of *Phytophthora capsici* Leon. *Eur J Plant Pathol,* 112.

SILVAR, C., MERINO, F. & DIAZ, J. 2006. Diversity of *Phytophthora capsici* in Northwest Spain: Analysis of Virulence, Metalaxyl Response, and Molecular Characterization. *Plant Dis,* 90**,** 1135-1142.

SKELSEY, P., DANCEY, S. R., PREEDY, K., LEES, A. K., COOKE, D. E. L. 2017. Forecasting the spread of aerially transmitted crop diseases with a binary classifier for inoculum survival. *Plant Pathology,* 67**,** 920-928.

SRIVASTAVA, S. K., ABAD,Z.G., KNIGHT,L.M. AND NAKHLA,M. Unpublished. *Phytophthora citricola* Ex-type Draft genome using ultra long-read MinION 3G Nanopore Sequencing. *Unpublished*.

STAM, R. 2013. *Identification and Characterisation of Phytophthora capsici CRN effector proteins and their targets in planta.* PhD, University of Dundee.

STAM, R., JUPE, J., HOWDEN, A. J., MORRIS, J. A., BOEVINK, P. C., HEDLEY, P. E. & HUITEMA, E. 2013. Identification and Characterisation CRN Effectors in *Phytophthora capsici* Shows Modularity and Functional Diversity. *PLoS One,* 8**,** e59517.

STUDHOLME, D. J., MCDOUGAL, R. L., SAMBLES, C., HANSEN, E., HARDY, G., GRANT, M., GANLEY, R. J. & WILLIAMS, N. M. 2016. Genome sequences of six *Phytophthora* species associated with forests in New Zealand. *Genom Data,* 7**,** 54-6.

STUDHOLME, D. J., PANDA, P., SANFUENTES VON STOWASSER, E., GONZALEZ, M., HILL, R., SAMBLES, C., GRANT, M., WILLIAMS, N. M. & MCDOUGAL, R. L. 2019. Genome sequencing of oomycete

isolates from Chile supports the New Zealand origin of *Phytophthora kernoviae* and makes available the first *Nothophytophthora* sp. genome. *Mol Plant Pathol,* 20**,** 423-431.

STUKENBROCK, E. H. & MCDONALD, B. A. 2008. The origins of plant pathogens in agro-ecosystems. *Annu Rev Phytopathol,* 46**,** 75-100.

SUN, W. X., JIA, Y. J., O'NEILL, N. R., FENG, B. Z. & ZHANG, X. G. 2008. Genetic diversity in *Phytophthora capsici* from eastern China. *Canadian Journal of Plant Pathology-Revue Canadienne De Phytopathologie,* 30**,** 414-424.

TABIMA, J. F., KRONMILLER, B. A., PRESS, C. M., TYLER, B. M., ZASADA, I. A. & GRUNWALD, N. J. 2017. Whole Genome Sequences of the Raspberry and Strawberry Pathogens *Phytophthora rubi* and *P. fragariae*. *Mol Plant Microbe Interact,* 30**,** 767-769.

TEWARI, S., SHARMA, S. 2018. Molecular Techniques for Diagnosis of Bacterial Plant Pathogens. *Molecular Diversity in the Genomic Era.* Delhi, India.

THILLIEZ, G. 2016. *The Role of Phytophthora Secreted Effectors in Determining Pathogen Host Range.* PhD, University of Dundee.

THILLIEZ, G. J. A., ARMSTRONG, M. R., LIM, T. Y., BAKER, K., JOUET, A., WARD, B., VAN OOSTERHOUT, C., JONES, J. D. G., HUITEMA, E., BIRCH, P. R. J. & HEIN, I. 2019. Pathogen enrichment sequencing (PenSeq) enables population genomic studies in oomycetes. *New Phytol,* 221**,** 1634-1648.

THINES, M. & KAMOUN, S. 2010. Oomycete-plant coevolution: recent advances and future prospects. *Curr Opin Plant Biol,* 13**,** 427-33.

TIAN, M., BENEDETTI, B. & KAMOUN, S. 2005. A Second Kazal-like protease inhibitor from *Phytophthora infestans* inhibits and interacts with the apoplastic pathogenesis-related protease P69B of tomato. *Plant Physiol,* 138**,** 1785-93.

TIAN, M., HUITEMA, E., DA CUNHA, L., TORTO-ALALIBO, T. & KAMOUN, S. 2004. A Kazal-like extracellular serine protease inhibitor from *Phytophthora infestans* targets the tomato pathogenesis-related protease P69B. *J Biol Chem,* 279**,** 26370-7.

TOMATO GENOME, C. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature,* 485**,** 635-41.

TOMPKINS, C. M., TUCKER, C. M. 1937. *Phytophthora* rot of honeydew melon. *J Agr Res,* 54**,** 933-944.

TORNERO, P., GADEA, J., CONEJERO, V. & VERA, P. 1997. Two PR-1 genes from tomato are differentially regulated and reveal a novel mode of expression for a pathogenesis-related gene during the hypersensitive response and development. *Mol Plant Microbe Interact,* 10**,** 624-34.

TORTO, T. A., LI, S., STYER, A., HUITEMA, E., TESTA, A., GOW, N. A., VAN WEST, P. & KAMOUN, S. 2003. EST mining and functional expression assays identify extracellular effector proteins from the plant pathogen *Phytophthora*. *Genome Res,* 13**,** 1675-85.

TURNER, J., O'NEILL, P., GRANT, M., MUMFORD, R. A., THWAITES, R. & STUDHOLME, D. J. 2017. Genome sequences of 12 isolates of the EU1 lineage of Phytophthora ramorum, a fungus-like pathogen that causes extensive damage and mortality to a wide range of trees and other plants. *Genom Data,* 12**,** 17-21.

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., AERTS, A., ARREDONDO, F. D., BAXTER, L., BENSASSON, D., BEYNON, J. L., CHAPMAN, J., DAMASCENO, C. M., 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., NORDBERG, E. K., MACLEAN, D. J., OSPINA-GIRALDO, M. D., MORRIS, P. F., PHUNTUMART, V., PUTNAM, N. H., RASH, S., ROSE, J. K., SAKIHAMA, Y., SALAMOV, A. A., SAVIDOR, A., SCHEURING, C. F., SMITH, B. M., SOBRAL, B. W., TERRY, A., TORTO-ALALIBO, T. A., WIN, J.,

XU, Z., ZHANG, H., GRIGORIEV, I. V., ROKHSAR, D. S. & BOORE, J. L. 2006. *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science,* 313**,** 1261-6.

VAN DAMME, M., BOZKURT, T. O., CAKIR, C., SCHORNACK, S., SKLENAR, J., JONES, A. M. & KAMOUN, S. 2012. The Irish potato famine pathogen *Phytophthora infestan*s translocates the CRN8 kinase into host plant cells. *PLoS Pathog,* 8**,** e1002875.

VAN DER HOORN, R. A. & KAMOUN, S. 2008. From Guard to Decoy: a new model for perception of plant pathogen effectors. *Plant Cell,* 20**,** 2009-17.

VAN WEST, P. 2006. *Saprolegina parasitica*, an oomycete pathogen with a fishy appetite: new challenges for an old problem. *Mycologist,* 20**,** 99-104.

VETUKURI, R. R., KUSHWAHA, S. K., SEN, D., WHISSON, S. C., LAMOUR, K. H. & GRENVILLE-BRIGGS, L. J. 2018a. Genome Sequence Resource for the Oomycete Taro Pathogen *Phytophthora colocasiae*. *Mol Plant Microbe Interact,* 31**,** 903-905.

VETUKURI, R. R., TRIPATHY, S., MALAR, C. M., PANDA, A., KUSHWAHA, S. K., CHAWADE, A., ANDREASSON, E., GRENVILLE-BRIGGS, L. J. & WHISSON, S. C. 2018b. Draft Genome Sequence for the Tree Pathogen *Phytophthora plurivora*. *Genome Biol Evol,* 10**,** 2432-2442.

VISSCHERS, I. G. S., PETERS, J. L., TIMMERMANS, L. L. H., EDWARDS, E., FERRATER, J. B., BALATERO, C. H., STRATONGJUN, M., BLEEKER, P. M., VAN HERWIJNEN, Z., GLAWE, G. A., BRUIN, J., VAN DAM, N. M. & MACEL, M. 2019. Resistance to three thrips species in *Capsicum* spp. depends on site conditions and geographic regions. *Journal of Applied Entomology,* 143**,** 929-941.

VOS, P., SIMONS, G., JESSE, T., WIJBRANDI, J., HEINEN, L., HOGERS, R., FRIJTERS, A., GROENENDIJK, J., DIERGAARDE, P., REIJANS, M., FIERENS-ONSTENK, J., DE BOTH, M., PELEMAN, J., LIHARSKA, T., HONTELEZ, J. & ZABEAU, M. 1998. The tomato Mi-1 gene confers resistance to both root-knot nematodes and potato aphids. *Nat Biotechnol,* 16**,** 1365-9.

WANG, S., WELSH, L., THORPE, P., WHISSON, S. C., BOEVINK, P. C. & BIRCH, P. R. J. 2018. The *Phytophthora infestans* Haustorium Is a Site for Secretion of Diverse Classes of Infection-Associated Proteins. *MBio,* 9.

WANG, Z., LANGSTON, D. B., CSINOS, A. S., GITAITIS, R. D., WALCOTT, R. R. & JI, P. 2009. Development of an improved isolation approach and simple sequence repeat markers to characterize *Phytophthora capsici* populations in irrigation ponds in southern Georgia. *Appl Environ Microbiol,* 75**,** 5467-73.

WATERHOUSE, G. 1963. Key to the species of *Phytophthora de Bary. . Mycological Papers,* 92**,** 1-22.

WATERHOUSE, R. M., SEPPEY, M., SIMAO, F. A., MANNI, M., IOANNIDIS, P., KLIOUTCHNIKOV, G., KRIVENTSEVA, E. V. & ZDOBNOV, E. M. 2017. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol*.

YANG, X., TYLER, B. M. & HONG, C. 2017. An expanded phylogeny for the genus *Phytophthora*. *IMA Fungus,* 8**,** 355-384.

YE, W., WANG, Y., SHEN, D., LI, D., PU, T., JIANG, Z., ZHANG, Z., ZHENG, X., TYLER, B. M. & WANG, Y. 2016. Sequencing of the Litchi Downy Blight Pathogen Reveals It Is a *Phytophthora* Species With Downy Mildew-Like Characteristics. *Mol Plant Microbe Interact,* 29**,** 573-83.

ZERBINO, D. R. & BIRNEY, E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res,* 18**,** 821-9.