Molecular pathotyping platforms for the clubroot pathogen, *Plasmodiophora brassicae*

by

Heather Hillary Tso

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Plant Science

Department of Agricultural, Food and Nutritional Science

University of Alberta

Abstract

Clubroot, caused by the soilborne pathogen *Plasmodiophora brassicae*, is a threat to cruciferous crops worldwide and an important disease of canola (*Brassica napus* L.) in Canada. At present, pathotypes of *P. brassicae* are distinguished phenotypically based on their virulence patterns on host differential sets, including the systems of Williams, Somé et al., the European Clubroot Differential set, and most recently, the Canadian Clubroot Differential and the Sinitic Clubroot Differential sets. While these are frequently used because of their simplicity of application, they are time-consuming, labor-intensive, and can lack sensitivity. Early and preventative pathotype detection is imperative to maximize productivity and promote sustainable crop production. The decreased turnaround time and increased specificity of molecular pathotyping will be valuable for the development of integrated clubroot management strategies, and interest in molecular approaches to complement phenotypic bioassays is increasing. In this study, two rapid and sensitive molecular *P. brassicae* pathotyping assays were developed, the first using RNase H2-dependent PCR (rhPCR) technology, and the second using a modified single base extension technique known as SNaPshot. Both assays clearly distinguished between pathotype clusters. The results correlated fully with whole genome sequencing data *in silico* for all 38 single-spore isolates of *P. brassicae* tested. Additional isolates from pathotyped clubroot galls and from samples in a single-blind test were also identified correctly. The rhPCR assay generated differentiating electrophoretic bands without non-specific amplification. The SNaPshot assay was able to detect down to a 10% relative allelic proportion in template (pathotype) mixtures with both single-spore and field isolates. Collectively, the results demonstrated that the rhPCR-based and single base extension assays developed in this study may be used as fast and reliable diagnostic tools to detect and distinguish between *P. brassicae* pathotype clusters. The ability to

identify pathotypes in a rapid manner will aid in clubroot diagnosis and surveillance activities, complementing traditional bioassays.

Preface

This thesis is an original work of the author, Heather Hillary Tso, under the supervision of Drs. Stephen Strelkov and Leonardo Galindo-González. Ms. Tso conducted all experiments described and prepared first drafts of each chapter. Drs. Strelkov and Galindo-González reviewed and revised all chapters of this thesis. Dr. Galindo-González contributed to the development of the research concept, and provided technical expertise and mentorship. Dr. Strelkov provided guidance throughout the duration of the project and secured research funding. Dr. Gavin Chen provided scientific support as a member of the supervisory committee. Drs. Homa Askarian and Victor Manolii prepared samples used in this study. Mr. Michael Holtz conducted valuable preliminary testing prior to the commencement of this study.

Funding for this research was provided by Results Driven Agriculture Research (RDAR) and Alberta Canola through Project Nos. 2019F023R and 126AR19, respectively, as well as in-kind contributions from the University of Alberta.

A version of *Chapter 2 – "Literature review: pathotyping platforms for* Plasmodiophora brassicae*"* was published as a review article in the journal *Plants*:

- Tso, H.H., Galindo-González, L., & Strelkov, S.E. (2021). Current and Future Pathotyping Platforms for *Plasmodiophora brassicae* in Canada. Plants, 10(7), 1446. https://doi.org/10.3390/plants10071446

A version of *Chapter 3 – "Development of rhPCR and SNaPshot assays to distinguish* Plasmodiophora brassicae *pathotype clusters"* will be submitted for publication as a methods paper:

- Tso, H.H., Galindo-González, L., Askarian, H., Locke, T., & Strelkov, S.E. (2021). Protocol: rhPCR and SNaPshot assays to distinguish *Plasmodiophora brassicae* pathotype clusters. Target journal: Plant Methods

Various portions of this work have been presented at scientific meetings:

- Tri-Society Virtual Conference (Canadian Phytopathological Society, Canadian Society of Agronomy and Canadian Society for Horticultural Science). Jul 2021.
    - Tso H.; Galindo-González L.; Strelkov S.E. A single base extension assay for pathotyping clubroot [*Plasmodiophora brassicae*] of canola.
- 41st Annual Meeting of the Plant Pathology Society of Alberta. Nov 2020.
    - Tso H.; Galindo-González L.; Askarian H.; Holtz M.; Strelkov S.E. Challenges in clubroot pathotype-specific molecular diagnostics.
- 40th Annual Meeting of the Plant Pathology Society of Alberta. Nov 2019.
    - Tso H.; Galindo-González L.; Askarian H.; Holtz M.; Strelkov S.E. Molecular discrimination of two virulent pathotypes of *Plasmodiophora brassicae* using RNase H2-dependent PCR.
- Canadian Phytopathological Society Annual Meeting 2019. Jul 2019.
    - Tso H.; Galindo-González L.; Askarian H.; Holtz M.; Strelkov S.E. A rapid diagnostic tool for pathotype identification of the clubroot pathogen *Plasmodiophora brassicae*.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor, Dr. Stephen Strelkov and mentor, Dr. Leonardo Galindo-González. Thank you for your support, encouragement, patience and guidance, and thank you for pushing me to challenge myself. Both of you have given me the confidence to explore new ideas and trusted me to fully take ownership of this research. I can truly say you have inspired the scientist in me! To Dr. Gavin Chen, thank you for your guidance as member of my graduate committee. To Troy Locke, Cheryl Nargang, Dr. Corey Davis and the entire MBSU team, thank you for your technical expertise and advice! This project would not have been possible without your knowledge. To Dr. Boyd Mori, thank you for serving as the Arm's Length Examiner in my defense. In addition, I would like to thank Results Driven Agriculture Research (RDAR), Alberta Canola, as well the University of Alberta and Alberta Agriculture and Forestry for financial and in-kind support of this research.

Thank you to the entire Plant Pathology Laboratory for welcoming me with open arms and fostering the most amazing work environment. Because of you all I have grown so much personally and professionally over the past three years! To all my lifelong friends, new friends and colleagues, you have made my MSc journey such a memorable and invaluable experience. Every single one of you has made a positive impact in my life, and for that I am beyond grateful. Thank you for the endless support, always cheering me on, and celebrating my accomplishments alongside me!

To my loving family, thank you for believing in and encouraging me to follow my passions! Your constant reassurance and support means the world, and I am forever appreciative for all the contributions you have made. I would not be where I am today without you!

Lastly, I would like to respectfully acknowledge that this work was carried out on Treaty 6 territory, a traditional gathering place for diverse Indigenous peoples including the Cree, Blackfoot, Métis, Nakota Sioux, Iroquois, Dene, Ojibway/Saulteaux/Anishinaabe, Inuit, and many others whose footsteps have marked these lands for centuries, and whose histories, languages, and cultures continue to influence our vibrant community.

Table of Contents

List of Tables

List of Figures

Abbreviations

BLAST	Basic Local Alignment Search Tool
CCD	Canadian Clubroot Differential
CIP	calf intestinal phosphatase
CR	clubroot-resistant
CTAB	cetyltrimethylammonium bromide
ddNTP(s)	dideoxy nucleotide triphosphate(s)
ddPCR	droplet digital polymerase chain reaction
DNA	deoxyribonucleic acid
dNTP(s)	deoxynucleoside triphosphate(s)
ECD	European Clubroot Differential
FI(s)	field isolate(s)
Hi-Di	highly deionized
indel(s)	insertion(s) and/or deletion(s)
ITS	internal transcribed spacer
NCBI	National Center for Biotechnology Information
NGS	next generation sequencing
PCR	polymerase chain reaction
PMA	propidium monoazide
PVY	Potato virus Y
qPCR	quantitative polymerase chain reaction
rhPCR	ribonuclease H2-dependent polymerase chain reaction
RNA	ribonucleic acid

SAP	shrimp alkaline phosphatase
SBE	single base extension
SCD	Sinitic Clubroot Differential
SNP(s)	single nucleotide polymorphism(s)
SSI(s)	single-spore isolate(s)
vcf	variant call format

Chapter 1 – Introduction

1.1 Clubroot disease

Clubroot, caused by the obligate parasite *Plasmodiophora brassicae* Woronin, is one of the most serious soilborne diseases of crucifers worldwide. Crop losses due to clubroot are estimated at 10% to 15% globally (Dixon, 2009). The disease has become a major threat to the $29.9 billion Canadian canola (*Brassica napus* L.) industry (Canola Council of Canada, 2021a), as severely infected canola crops can suffer yield losses of 30% to 100% (Strelkov & Hwang, 2014). The main symptom associated with clubroot is the formation of galls on the host roots, restricting the absorption of the water and nutrients needed to support aboveground growth (Dixon, 2009). Infection by *P. brassicae* reduces canola yield and oil quality as plants become stunted and undergo accelerated flowering and premature senescence (Pageau et al., 2006).

1.1.1 The clubroot outbreak in western Canada

Clubroot was first identified on canola in western Canada in 2003 in 12 fields near Edmonton, Alberta (Tewari et al., 2005). Annual clubroot surveys began in 2004 to monitor the spread of *P. brassicae* and the severity of the disease (Howard et al., 2010). As the number of clubroot-infested fields began to increase, *P. brassicae* was declared a pest under the Agricultural Pests Act of Alberta in April of 2007 (Strelkov et al., 2011); this designation enabled the enforcement of control measures throughout the province aimed at reducing the dissemination and impact of the disease. Nonetheless, *P. brassicae* has continued to spread and is considered one of the most significant problems facing canola growers. Clubroot is now endemic to much of central Alberta (Dixon, 2009), with more than 3000 confirmed field infestations (Strelkov et al., 2020) and the disease having been identified in 44 of the 66 counties or municipal districts where canola is

grown (Strelkov et al., 2021). Epidemiological models had predicted that clubroot could spread beyond Alberta, and since then the disease has been confirmed on canola in Saskatchewan and Manitoba (Strelkov & Hwang, 2014). Clubroot now also occurs in canola crops in Ontario, Canada (Al-Daoud et al., 2018), and in North Dakota, USA (Chapara et al., 2019).

1.1.2 Disease cycle

The clubroot disease cycle begins with resting spores found in infested soils (Wallenhammar, 1996). At the start of the growing season, the resting spores will germinate under favorable weather conditions (good moisture, temperatures > 10 °C) to produce primary zoospores, with the germination rates enhanced in the presence of root exudates. The primary zoospores penetrate and infect young root hairs to begin primary infection (Ayers, 1944). Primary plasmodia are formed within the root hairs and cleave to produce zoosporangia, yielding secondary zoospores (Dixon, 2009). Once released into the soil, the secondary zoospores cause secondary infections by penetrating cortical tissue of the main root. Once in the root cortex, the secondary zoospores develop into intracellular secondary plasmodia, which eventually cleave into the next generation of resting spores (Kageyama & Asano, 2009). These resting spores are released back into the soil as the root tissues deteriorate and galls decompose, serving as inoculum for future infections. Soilborne resting spores of *P. brassicae* are the primary source of inoculum, and the disease is generally regarded as monocyclic (Howard et al., 2010). The resting spores are resilient and can remain viable in the soil for up to 20 years, due to the protection provided by their thick spore walls composed of five layers of chitin and carbohydrates (Moxham & Buczacki, 1983). These walls bestow a high degree of resistance from degradation by environmental factors (Braselton, 1995). In heavily infested fields, it has been estimated that approximately 17 years are required for spore concentrations to drop below detectable levels

(Friberg et al., 2005). More recent studies conducted in Quebec (Peng et al., 2014) and Alberta (Ernst et al., 2019), however, have shown that most of the resting spores disappear from the soil in the two years after cultivation of canola, with a smaller portion persisting for much longer, resulting in a Type III survivorship curve (Rauschert, 2010).

1.2 Emergence of 'new' virulent pathotypes

1.2.1 Clubroot-resistant canola

When clubroot was first identified in western Canada, few management options were available to growers. In 2009, however, the first clubroot-resistant (CR) canola cultivar was released onto the Canadian market, followed quickly by other cultivars from various seed companies. With 40 registered CR cultivars currently available, genetic resistance has become the most important tool for the management of clubroot in Canada (Canola Council of Canada, 2021b). In contrast, other practices recommended as part of an integrated clubroot management plan, including longer rotations out of canola and the sanitization of field machinery (Peng et al., 2014), have not been adopted as widely (Donald & Porter, 2009). Genetic resistance to clubroot relies primarily on single major genes, effective against specific pathotypes or races of *P. brassicae* (Rahman et al., 2014). Hence, the monogenic nature of these cultivars can only provide strong qualitative vertical resistance to a select few pathotypes, while having no effect on other pathotypes (Fredua-Agyeman et al., 2018). The durability and longevity of these CR cultivars are at risk as they are vulnerable to pathotype shifts, which can occur when *P. brassicae* populations are exposed repeatedly to the same resistance source (LeBoldus et al., 2012).

1.2.2 Resistance breakdown

Only four years following the commercial release of the first CR cultivar, 'new' *P. brassicae* pathotypes able to overcome resistance were detected in Alberta (Strelkov et al., 2016). The 2013 Alberta-wide clubroot survey found greater than expected disease severity in six fields sown to CR cultivars (Strelkov et al., 2014). Four *P. brassicae* populations from two of these fields caused significantly increased levels of clubroot in greenhouse trials, and three of these populations were highly virulent across six CR canola cultivars (Strelkov et al., 2016). Sixty-one field populations collected in 2014–2016 could overcome the resistance in at least one CR cultivar (Strelkov et al., 2018). In 2017 and 2018, nine novel pathotypes were discovered throughout western Canada (Hollman et al., 2021), and another four novel pathotypes were discovered in the Peace Country Region of Alberta alone (Strelkov et al., 2021). An additional six novel pathotypes were identified during an investigation of *P. brassicae* single-spore isolates (SSIs) collected from field populations that were virulent on CR cultivars (Askarian et al., 2021b). This suggests an increasing diversity in the virulence of *P. brassicae* strains and a greater prevalence of resistance-breaking pathotypes. While the resistance is often said to have "broken down", the change occurred in the pathotype structure of the *P. brassicae* populations and not in the host cultivars themselves (Rahman et al., 2014). The repeated cultivation of CR cultivars imposed selective pressure that led to virulence shifts in the pathogen, encouraging the proliferation of novel virulent pathotypes (Peng et al., 2014).

1.3 Need for rapid pathotyping

Clubroot is a "disease of cultivation" due to its correlation with the intensive cultivation of susceptible crucifers (Dixon, 2009). The evolving nature of *P. brassicae* populations and the

development of new canola cultivars illustrate the importance of interdisciplinary efforts to synchronize progress in clubroot management systems. With the emergence of new pathotypes overcoming resistance, the need to distinguish rapidly between pathotypes has become a priority in the development of clubroot diagnostics. It is important for growers, the industry, and researchers to understand the distribution and occurrence of pathotypes in order to make informed crop management decisions. Given the diversity in the virulence of *P. brassicae*, the effectiveness of the clubroot resistance found in different canola cultivars can vary greatly. Timely pathotype identification allows canola farmers to select cultivars that carry the appropriate resistance and provide the best protection against the disease. The deployment of resistance effective against pathotypes found in a specific field or region is essential for reducing disease spread and facilitating preventative management practices. In addition, reliable information on the abundance and diversity of existing pathotypes may encourage breeding programs to target development of CR canola cultivars with the appropriate resistance traits.

1.4 Research objectives

Since pathotype identification before and during the cultivation of canola would help to improve clubroot management, rapid molecular detection tools are fundamental. The focus of this thesis is on efforts to pathotype *P. brassicae* for clubroot diagnostics, and includes two main objectives:

(1) To evaluate current and future pathotyping platforms; and

(2) To explore and develop two rapid and sensitive molecular pathotyping assays.

Chapter 2 – Literature review: pathotyping platforms for *Plasmodiophora brassicae*

2.1 Current diagnostic methods

Several diagnostic methods exist for the detection of *Plasmodiophora brassicae* and for the differentiation of *P. brassicae* pathotypes. Although each method has its own limitations, collectively, they all serve a role in the clubroot diagnostic process. This chapter assesses current and future pathotyping platforms for *P. brassicae* in Canada.

2.1.1 Phenotypic approaches

Soil bioassays with bait crops have long been used for the general detection of *P. brassicae*. Bait crops are host plants that stimulate resting spore germination and can become infected. One of the first reported bioassays used cabbage (*B. oleracea*) hosts grown in suspect soil, with the root hairs inspected for the presence of zoosporangia via microscopy after one week (Samuel & Garrett, 1945). This same method was further used to evaluate resting spore survival in soil (Macfarlane, 1952). To ensure that *P. brassicae* detection correlated with gall development, another study grew host plants in suspect soil, with the roots examined for galling after five weeks (Colhoun, 1957; Melville & Hawken, 1967). While bait crops provide reliable viability assessments of *P. brassicae* inoculum in infested soil, these methods are only feasible when inoculum concentrations are greater than 1000 spores per gram of dry soil, as this is the concentration generally required for gall development under greenhouse conditions (Faggian & Strelkov, 2009). Moreover, soil bioassays may not always be practical, since they are time-consuming and labor-intensive, and require large amounts of greenhouse space. The planting of a single, susceptible bait crop also does not provide information on the race or pathotype classification of the inoculum present.

P. brassicae pathotypes are currently distinguished phenotypically based on their virulence patterns in bioassays conducted with host differential sets, where the reactions of the hosts are monitored based on root gall development. These assays aim to capture the occurrence and extent of physiologic specialization in populations of the pathogen (Fredua-Agyeman et al., 2018). Numerous differential sets have been proposed to study pathogenic variability in *P. brassicae* (Buczacki et al., 1975; Pang et al., 2020; Somé et al., 1996; Strelkov et al., 2018; Williams, 1966). The differentials of Williams (1966), one of the earliest and among the most commonly used systems, consists of four differential hosts that can distinguish a theoretical maximum of 16 pathotypes. The European Clubroot Differential (ECD) set was established later and consists of 15 differential hosts, three of which belong to Williams' system (Buczacki et al., 1975). Somé et al. (1996) developed another differential set consisting of three *B. napus* genotypes. The development of the Canadian Clubroot Differential (CCD) set was deemed necessary after the discovery of new virulence phenotypes that overcame resistance in Alberta, Canada (Strelkov et al., 2018). The CCD set consists of 13 host genotypes; these include the differentials of Williams (1966) and Somé et al. (1996), eight hosts from the ECD set, and several *B. napus* cultivars of interest to Canadian canola breeders (Strelkov et al., 2018). Since the differentials of Williams (1966) and Somé et al. (1996) are incorporated entirely into the CCD set, pathotype characterization against these earlier sets is possible with the CCD set (Figure 2.1). Isolates of *P. brassicae* are inoculated onto each host of the CCD set, and disease development is assessed six weeks later (Strelkov et al., 2018). Pathotypes are identified based on the reaction of the inoculated hosts, with distinctive virulence patterns representing individual pathotypes. Each pathotype is assigned a number followed by a letter (for example, 'pathotype 3A'); the numbers correspond to the Williams' (1966) pathotype designation, while the letters

are unique to the CCD set and denote variants of the Williams' pathotypes (Askarian et al., 2021b; Strelkov et al., 2018). Most recently, the Sinitic Clubroot Differential (SCD) set was developed in China using differential hosts with known clubroot-resistance genes to explore the genetic variability of pathotype 4, as defined by the differentials of Williams (Pang et al., 2020). The SCD set was essential when isolates identified as pathotype 4 on the system of Williams were found to exhibit varied virulence patterns on various Chinese cabbage (*B. rapa*) and cabbage (*B. oleracea*) cultivars.

Although the CCD set is effective and has greater differentiating capacity than earlier differential systems, several limitations are associated with the use of differential host sets in general. The methodology is labor-intensive, and sufficient biosecure greenhouse space is required due to the need for replication and pathogen containment. Moreover, the evaluation of pathotypes using differential sets is time-consuming, and the generation of virulence patterns can be affected by the greenhouse conditions and biological growth of the root galls and differential hosts. Typically, the differential hosts are rated about 6 weeks following inoculation, and hence only a comparatively small number of isolates can be tested simultaneously, depending on space availability. Therefore, it is difficult to scale up this methodology for rapid testing or for testing large numbers of samples. There is also a risk of human error and variable results among diagnostic laboratories considering that the identification of virulence patterns requires a high degree of technical expertise (Strelkov et al., 2018). Moreover, the existing clubroot differential systems are based on phenotypic classifications, which may not match genomic variation among pathotypes. Since the genetic basis of resistance in some of the host genotypes is not well defined, it may be difficult to infer genetic relationships between isolates or the specific avirulence/virulence genes found in an isolate based on the pathotype classification. This may be

addressed through the development of near-isogenic lines carrying defined resistance genes for use as differentials, to more appropriately define true 'races' based on the genetics of the host-pathogen interaction. The emergence of new virulence types of *P. brassicae* may also require modification of the differential sets, since ultimately the ability to distinguish pathotypes or races of a pathogen is limited by the effectiveness of the differential hosts.

2.1.2 Microscopy

Histological approaches have long been used to study clubroot. Various staining methods have been employed to visualize *P. brassicae* and host structures. Resting spores can be distinguished from root tissue using a triple staining method (Buczacki & Moxham, 1979); resting spores stain blue, whereas the host tissue stains pink or purple. Another method uses lactophenol cotton blue to stain chitin, a prevalent polymer in the cell wall of *P. brassicae* resting spores (Buczacki & Moxham, 1983). Using methylene blue, resting spores inside the secondary plasmodia are stained a dark blue, while the cell wall of the root tissue stains a contrasting light blue (Sharma et al., 2011). Primary plasmodia and zoospores can be observed with Harris hematoxylin staining followed by a counterstain with eosin Y (Verma et al., 2014). Toluidine blue is used to detect the host resistance response during infection, by staining lignified cell walls blue, since lignification occurs as a defense mechanism (Deora et al., 2013).

Microscopy has also been used to assess inoculum load and viability as well as germination rates. The viability of resting spores can be tested with Evan's blue, in which cell membranes of the dead spores stain blue (Harding et al., 2019; Tanaka et al., 1999), and with acridine orange fluorescent dye, where the viable spores fluoresce green (White & Buczacki, 1979). Distinctive fluorochromes for differential staining are also used; non-viable spores fluoresce red whereas

viable spores fluoresce blue (Takahashi & Yamaguchi, 1988). Germination levels can be estimated with aceto-orcein, which stains ungerminated resting spores (Naiki et al., 1987).

While histological investigations are fundamental for diagnosing the disease as a whole and advancing clubroot research, preventative measures for clubroot remain a challenge, as microscopy-based techniques are not sensitive enough for early pathogen detection. More importantly, they are of no value for pathotyping since cell morphology does not differ between pathotypes.

2.1.3 Molecular approaches

The development of methods for the molecular detection of *P. brassicae* has been an ongoing process over the last three decades (Table 2.1). A polymerase chain reaction (PCR) assay was developed for the general detection of *P. brassicae* in soil (Ito et al., 1999). The primers were designed based on an isopentenyltransferase-like gene specific to *P. brassicae*. To increase sensitivity, another PCR assay was developed based on an internal transcribed spacer (ITS) region of ribosomal deoxyribonucleic acid (DNA) (Faggian et al., 1999). Since there are more copies of this ITS region in the *P. brassicae* genome, detection can occur at lower resting spore concentrations. After initial tests of these primers in artificially infested soils, the primers were later incorporated into an assay for the detection of *P. brassicae* in naturally infested field soils (Wallenhammar & Arwidsson, 2001). The main drawback of these assays (Faggian et al., 1999; Ito et al., 1999; Wallenhammar & Arwidsson, 2001) was their nested PCR design, which involves two amplifications. This requires more time and materials, and there is an increased risk of contamination during sample manipulation in comparison with one-step PCRs. After the start of the clubroot outbreak in western Canada in 2003, a non-nested, one-step PCR assay was

developed for the molecular detection of *P. brassicae* (Cao et al., 2007). The PCR primers were based on a conserved 18S ribosomal ribonucleic acid (RNA) gene that could detect the pathogen in symptomless root tissues as early as three days after inoculation and with concentrations as low as 1×10^3 resting spores per gram of soil. This assay has provided the groundwork for molecular testing of clubroot in Canada, and has since been commercially available from diagnostic laboratories throughout the country (Faggian & Strelkov, 2009). Although these reported PCR assays are effective for clubroot detection, they were initially non-quantitative and therefore could not provide information on levels of soil inoculum or host colonization.

The development of quantitative PCR (qPCR) assays for clubroot detection has been reported over the last decade. In qPCR assays, the initial amount of pathogen DNA is directly correlated with an early or late exponential curve of amplification. Initially, two dye-based assays were reported, one for the quantification of resting spores in plant samples (Sundelin et al., 2010), and another for the quantification of resting spores in seeds harvested from infested fields (Rennie et al., 2011). Both assays were successful in rapid quantification; however, dye-based qPCR is less specific in comparison with probe-based qPCR. Only one target can be investigated at a time with dye-based technologies, as the dye will bind to any DNA fragment amplified in the reaction; therefore, these experiments rely on careful primer design and amplicon selection. In probe-based qPCR, a fluorogenic probe anneals to a specific sequence within the PCR amplicon during the reaction. Two probe-based assays were developed to quantify resting spores in naturally infested soil samples (Deora et al., 2015; Wallenhammar et al., 2012), and another was developed for the quantification of *P. brassicae* in root tissue (Cao et al., 2014). Since probe-based technologies have greater specificity and multiple targets can be detected simultaneously in each sample, they may be more suitable for clubroot quantitative diagnostics. Up to this point,

qPCR assays cannot discriminate viable from non-viable resting spores. To address this issue, propidium monoazide (PMA) was incorporated into a probe-based qPCR assay to prevent the amplification of non-viable resting spores (Al-Daoud et al., 2016). PMA is a photoreactive DNA-binding dye that penetrates the cells of dead membranes and is commonly used for viable microorganism distinction. An assay was designed using droplet digital PCR (ddPCR) to quantify resting spores in soil, and ddPCR was found to be a more versatile tool over existing qPCR assays, since it yielded more accurate results and was less affected by amplification inhibitors (Wen et al., 2020).

While essential for clubroot diagnostics, all of these molecular assays were developed for *P. brassicae* detection and cannot differentiate pathotypes. They consist of primers predominantly associated with the conserved regions of the genome, which are fairly similar among pathotypes.

2.2 Rapid molecular pathotyping

Molecular diagnostic assays may be developed based on genetic variation between *P. brassicae* pathotypes. Several advantages are presented by molecular-based techniques for pathotyping. They can be highly sensitive and rapid, cost-effective in terms of labor, space, and time, biosecurity is not a concern, and the limitation of inter-rater reliability is eliminated. Various potential molecular markers have been evaluated for pathotyping of *P. brassicae*. A presumptive random amplified polymorphic DNA marker specific to pathotype P1, as defined by the differentials of Somé et al. (1996), was identified and converted into a sequence characterized amplified region (Manzanares-Dauleux et al., 2000). The *Cr811* gene was found to be specific to pathotype 5 (Zhang et al., 2015), as defined on the differentials of Williams (1966). Over 1500 single nucleotide polymorphisms (SNPs) differentiating two distinct populations of *P. brassicae*

were identified and characterized based on the pathogenicity of the isolates and their ability to cause disease against CR canola cultivars (Holtz et al., 2018). A region of an 18S internal transcribed spacer sequence was found to be specific to pathotype 5X (Zhou et al., 2018), as defined on the CCD set (Strelkov et al., 2018), and this region was developed into a probe-based qPCR assay to identify pathotype 5X.

2.2.1 Amplicon length distinction

Amplicon length distinction can be used to differentiate pathotypes in PCR-based assays. In this methodology, the insertion/deletion (indel) polymorphism is positioned within the amplicon, and the assay relies on electrophoretic separation of amplified DNA fragments. Pathotype clustering may be performed based on the molecular weight of the band in the gel (Figure 2.2). This technique was used to distinguish defoliating from non-defoliating pathotypes of *Verticillium dahliae* Kleb., the pathogen responsible for the Verticillium wilt disease of cotton (*Gossypium hirsutum* L.) (Erdogan et al., 2013). The assay was found to be effective in identifying genetic relationships among pathotypes. This rapid technique is advantageous in its use of a single primer pair that is conserved among the tested races/pathotypes, and it is a simple conventional PCR method that can be easily adopted by diagnostic laboratories at minimal cost. However, this technique is dependent on the existence of a discriminatory insertion or deletion and an appropriate conserved region for primer design around the amplicon, and these may not be as abundant as necessary to provide sufficient resolution.

2.2.2 SNP-Based distinction

Plant pathologists studying a variety of pathosystems have explored the potential of discriminatory polymorphic regions and SNPs for pathotype detection. Distinctive primers

contain specific polymorphisms corresponding to a subset of pathotypes, resulting in differential PCR amplification (Figure 2.3). The development of SNP-based assays follows a general three-part process: (1) SNP discovery through sequencing and polymorphism detection by bioinformatics approaches and evaluation on a small sample set; (2) validation of SNPs to eliminate sequencing errors; and (3) assay adoption for pathotyping in large populations (Chagné et al., 2007). Several molecular techniques have been tested in plant pathosystems and may be suitable for SNP-based pathotyping in clubroot diagnostics.

Distinguishing polymorphic primers were used to explore the pathotype diversity of *Ascochta rabiei* (Kovatsch.) Arx, the fungus responsible for Ascochyta blight of chickpea (*Cicer arietinum* L.), and this assay was able to differentiate three pathotype clusters (Udupa et al., 1998). Polymorphic primers were also used to distinguish between defoliating and non-defoliating pathotypes of *V. dahliae*, the causal agent of Verticillium wilt of olive (*Olea europaea* L.) (Mercado-Blanco et al., 2002). The expected products were amplified when the primers were used independently although the electrophoretic bands were faint. To increase the sensitivity, the primers were adapted into a nested PCR assay. The nested PCR design increased resolution between defoliating and non-defoliating pathotypes of *V. dahliae*.

In clubroot, six population-specific primer pairs were designed to detect avirulent pathotypes and five were designed to detect virulent pathotypes (Holtz et al., 2021) based on discriminating polymorphic regions identified earlier (Holtz et al., 2018). The avirulent population refers to pathotypes discovered before the 2009 commercialization of CR cultivars that are known to have been present in Alberta for the longest period. The virulent population refers to pathotypes that emerged due to selection pressure imposed by CR cultivars and which can overcome resistance. One primer pair for each population was further developed into a quantitative assay. All eleven

primer pairs were confirmed to be specific to the population they were designed for, and no amplification occurred in non-infested samples. The researchers noted that although the primer pairs are reliable for clubroot diagnostics, they are not as sensitive as earlier reported assays for general detection of *P. brassicae* (Holtz et al., 2021).

Differentiating polymorphic primers provide a rapid and simple conventional PCR-based method that can be easily adopted by a wide range of diagnostic laboratories at minimal cost. Primer pairs for each respective polymorphic cluster need to be designed. Since this is a PCR-based approach, the primers may be optimized into a quantitative assay. There may be a slight increase in running time in carrying out this assay, considering this method involves different primer pairs. The method may be vulnerable to misidentification due to non-specific amplification, false negative results, or technical errors of the PCR. In addition, the longevity of differentiating primers is dependent on the stability of the discriminatory SNPs and the risk for further mutations. SNP-based distinction may also be valuable for a ddPCR assay, a system for absolute quantification used to detect low DNA concentrations (Hindson et al., 2011). With *P. brassicae*, ddPCR could be used to increase detection of low abundance spores.

2.2.3 RNase-H2 dependent PCR

In conventional PCR, differentiation of pathotypes with slight nucleotide variations is challenging as primers may still bind non-specifically depending on PCR conditions, which may lead to false-positive amplification. A novel primer technology known as RNase-H2 dependent PCR (rhPCR), introduced in 2011, provides greater accuracy and sensitivity (Dobosy et al., 2011). rhPCR primers are blocked by a single ribonucleotide residue at the discriminating polymorphic site, preventing amplification by the polymerase (Figure 2.4). The blocked primers

are activated via cleavage of the RNA base by the thermostable RNase H2 enzyme once the ribonucleotide residue has annealed to the template strand. In the case of a mismatch, no cleavage will occur, and the primer remains blocked.

The rhPCR technique has been incorporated into the diagnostic process of several pathosystems (Labbé et al., 2019; McAllister et al., 2018; Zuzak et al., 2017) and for whitefish species identification (Rodgers et al., 2019). An rhPCR assay was developed to identify native from invasive subspecies of common reed (*Phragmites australis* (Cav.) Trin. ex Steud.) (Zuzak et al., 2017), showing correct identification of the subspecies with their respective primer pair, and not with the opposing primer set. No amplification occurred in an rhPCR test without the use of the RNase H2 enzyme, confirming the effectiveness of the block-cleavable technology. In the same study, rhPCR was compared with a previously used restriction fragment length polymorphism protocol. While the level of accuracy was similar, rhPCR was less time-consuming and easier to perform (Zuzak et al., 2017). Another rhPCR assay was developed for the identification and quantification of *Grosmannia clavigera* (Robinson- Jeffrey & R.W.Davidson) Zipfel, Z.W. de Beer & M.J.Wingf., a fungal pathogen of pine trees vectored by mountain pine beetles (*Dendroctonus ponderosae* Hopkins) (McAllister et al., 2018). The quantitative rhPCR assay was able to efficiently and accurately distinguish *G. clavigera* from other species, highlighting the potential of this technique to be used in the diagnostic process of complex phytopathogenic samples. The sensitivity of the rhPCR technique was discussed during the development of a pathotyping assay to discriminate isolates of *Salmonella enterica* serovar Heidelberg, a bacterium responsible for salmonella infection in humans (Labbé et al., 2019). The assay improved resolution and efficiency in isolate discrimination when evaluated against pulsed field gel electrophoresis and phage typing. While the rhPCR pathotyping results were generated as

early as five hours after DNA extraction and isolates were accurately identified, non-specific amplification of isolates containing a non-targeted alternate allele at the discriminatory SNP position occurred in 15% of the reactions. However, these non-specific amplicons were distinguishable from amplicons of targeted isolates via a difference in band intensity. An RNase H2 enzymatic activity error consistent with this result has been previously noted (Dobosy et al., 2011); inaccurate RNase H2 cleavage may occur in a mismatch in the target sequence, but at a much lower frequency relative to the target match. Although not part of a pathosystem diagnostic process, a quantitative rhPCR assay was used for the detection of five closely related whitefish (*Coregonus*) species (Rodgers et al., 2019). Detection had been carried out previously with conventional TaqMan qPCR for other *Coregonus* species; however, sequence variation between these five closely related species was not strong enough for a conventional qPCR assay design. The two techniques were combined to increase specificity by incorporating rhPCR primers into the TaqMan assay. To ensure amplification was specific and to evaluate the rhPCR technique, the TaqMan assay was run in parallel with conventional primers. The addition of the blocked-cleavable technology increased specificity, as no non-targeted amplification occurred.

In clubroot, an rhPCR assay was developed to differentiate a new virulent 'pathotype 3-like' strain of *P. brassicae* from the original pathotype 3 (Yang et al., 2018), as defined on the differentials of Williams (1966). Based on polymorphic regions of two hypothetical protein genes, two primer pairs were designed as a duplex PCR with one primer pair for each gene, and each gene representative of either pathotype 3-like or 3. The primers corresponding to pathotype 3-like produced an amplicon of 135 base pairs, whereas the primers for pathotype 3 produced a larger amplicon of 337 base pairs. The sensitivity of the primers was tested against pre-pathotyped SSIs and field isolates (FIs) of the pathogen. Equal proportions of pathotype 3-like

and pathotype 3 DNA produced bands of comparable intensity. When the pathotype 3-like DNA proportion was greater, its respective band was stronger relative to pathotype 3. Likewise, a greater proportion of pathotype 3 DNA produced stronger bands relative to pathotype 3-like. The researchers tested the assay against four pre-pathotyped field galls representing different counties in Alberta. Each gall produced amplicons with both rhPCR primer pairs, suggesting that field galls are a mixture of virulent and avirulent pathotypes. With the same rhPCR assay, the assay was tested on 79 field galls collected from 22 fields in Alberta (Fu et al., 2019). *P. brassicae* populations from 50 of these galls produced more than one band, confirming the hypothesis that multiple pathotypes co-exist as field populations in a single field gall and that their abundance varies according to their interactions with host plants. The galls were not subjected to phenotypic pathotyping, and therefore the exact pathotypes responsible for the bands were not confirmed.

The above studies demonstrate the potential of the rhPCR technique in SNP discrimination of *P. brassicae* pathotypes as it offers greater resolution and sensitivity in comparison with previously established assays. The combination of rhPCR with other molecular techniques may further increase its SNP-differentiating ability. For instance, rhPCR primers may be designed against a region bearing a higher number of differentiating SNPs to combine the technology with the previously discussed SNP-based distinction. This may increase the reliability and sensitivity of the pathotyping assay. As with most PCR-based assays, the rhPCR primers may be optimized into a quantitative assay. Other than the additional RNase H2 enzyme required, the protocol can be easily adopted by diagnostic laboratories, considering rhPCR is carried out with the same equipment as conventional PCR. The turnaround time is slightly longer in comparison with assays using only one primer pair.

2.2.4 Single base extension

Another SNP-based technology is SNaPshot (CD Genomics, 2021), a single base extension (SBE) reaction that allows the detection of multiple SNPs on multiple DNA templates (Chagné et al., 2007). SNaPshot primers are designed just upstream of the polymorphic base in question. When the primers anneal to the DNA template right before the SNP, Taq polymerase extends the primer by one nucleotide on the 3′ end by selecting the correct complementary base from a pool of fluorescently labelled dideoxy nucleotide triphosphates (ddNTPs), which impairs further extension of the product. The incorporation of the pathotype-specific base at the 3′ end produces a fluorescent signal corresponding to one of four dyes that match each of the four possible bases. The resulting product size is the length of the SNaPshot primer plus the addition of the fluorescent ddNTP base. The SBE product is then separated by capillary electrophoresis inside a genetic analyzer to generate electropherograms, and pathotypes are identified based on peak color and product size.

The SNaPshot protocol is a four-step approach: (1) template preparation; (2) extension reaction; (3) post-extension treatment; and (4) capillary electrophoresis (Figure 2.5). In template preparation, targeted conventional PCR amplification is used to generate the DNA templates containing the SNP. The resulting templates must undergo a purification process to remove PCR primers and unincorporated deoxynucleoside triphosphates (dNTPs) to avoid interference with the extension reaction. The extension reaction takes place, followed by a post-extension treatment of the products. The products are incubated with either shrimp alkaline phosphatase (SAP) or calf intestinal phosphatase (CIP) to remove any unincorporated ddNTPs. To prepare for capillary electrophoresis, purified SNaPshot products and size standards are added into a highly deionized (Hi-Di) formamide solvent. DNA size standards are used to determine the size of the

SNaPshot product. Capillary electrophoresis is conducted inside a genetic analyzer, and the results of this scan are examined using a fragment analysis software.

The SBE technique has been incorporated into the diagnostic process of several different pathosystems. A SNaPshot assay was developed to detect variants of the Potato virus Y (PVY), a pathogen belonging to the *Potyvirus* genus and a threat to crops of the Solanaceae family (Rolland et al., 2008). Similar to *P. brassicae* pathotyping methods, PVY variant identification is usually carried out phenotypically based on disease symptoms on host plants. The SNaPshot technique was advantageous in its ability to characterize variants of mixed samples, and in the amount of starting material required. Reliable detection occurred with as few as 10^2 copies of the PVY genome (Rolland et al., 2008) in comparison with 10^3 copies required in a previously published real time PCR assay (Balme-Sinibaldi et al., 2006). An assay was also designed for the specific detection of race 3 of *Fusarium oxysporum* f.sp. *vasinfectum*, a soilborne fungal pathogen responsible for root rot, vascular wilt, damping-off, and yellowing in a wide range of economically important crops (Egamberdiev et al., 2014). While the SNaPshot technology has not been evaluated for *P. brassicae* pathotyping, it has been used to detect erucic acid, a fatty acid in canola, for marker-assisted selection in canola breeding (Rahman et al., 2008). Canola plants with a two-base deletion in one specific gene produced nominal erucic acid content. This polymorphic locus was used for the development of a SNaPshot assay. It was necessary to develop an assay to identify erucic acid content rapidly at an early plant growth stage, since high levels of erucic acid reduce oil quality and digestibility.

The above studies highlight the reliable and scalable potential of incorporating SNaPshot as a *P. brassicae* pathotyping tool. Several advantages of the SBE technique were consistently noted for the developed assays. The procedure is straightforward and the automation of genetic analyzers

offers convenient data processing with a high degree of accuracy. In comparison with the previously mentioned SNP-based PCRs, where amplification proceeds based on the existence of a specific allele, SNaPshot allows for the detection of up to four allelic variants as bases are distinguished by means of fluorescent ddNTPs. SNaPshot is scalable through a multiplex reaction in which numerous polymorphic regions in the genome can be examined concurrently for efficient and rapid testing. Nonetheless, there are several limitations to this technique. It is a lengthy process due to the number of steps involved, and it requires equipment that may not generally be used by diagnostic laboratories. In addition to basic PCR reagents, a clean-up kit is needed to purify PCR products, SNaPshot reagents are necessary for the extension reaction, SAP or CIP is needed for post-extension treatment, Hi-Di formamide is required as an injection solvent in the genetic analyzer, and a size standard is required to investigate the results of the fluorescent peaks.

2.2.5 Can metabarcoding be used in clubroot?

The advent of next generation sequencing (NGS) and bioinformatics platforms has enabled new lines of research by groups studying a variety of plant pathosystems. NGS services were introduced as a highly efficient and sensitive sequencing platform that overcame the limitations of Sanger sequencing with respect to throughput, making it an appealing strategy for pathogen analyses (Jones et al., 2017). NGS has greater sensitivity to detect low-frequency variants, as it has been able to detect pathogens that were not detected by other molecular tests, especially in early stage infections (Al Rwahnih et al., 2015). By isolating and sequencing small RNAs from co-infected plants using Illumina deep sequencing technology, both the sweet potato feathery mottle virus and the sweet potato chlorotic stunt virus were detected in sweet potato (*Ipomoea batatas* (L.) Lam.) (Kreuze et al., 2009), and multiple viruses infecting a single ornamental plant

were detected (Verdin et al., 2017). Double-stranded RNA from leaves of infected apple plants (*Malus domestica* Borkh) were isolated and sequenced using Illumina technology, resulting in the identification of 12 genotypes of the apple stem pitting virus (Rott et al., 2017). Full genome sequencing was found to be superior in detecting both early stages and low levels of infection by viral pathogens in grapevine (*Vitis vinifera* L.), in comparison with bioassays that are dependent on disease symptoms (Al Rwahnih et al., 2015). Collectively, these studies indicate that NGS technologies allow for the simultaneous detection of multiple pathogens (Kreuze et al., 2009; Verdin et al., 2017) or pathotypes (Rott et al., 2017) in a single sample, and the detection of low-frequency genomes (Al Rwahnih et al., 2015). Full genome sequencing allows for the generation of substantial amounts of data to discover variability among pathotypes, which can be used for diagnostics and metagenomics studies.

Metabarcoding is an NGS approach that offers high quality single nucleotide resolution in a single reaction (Taberlet et al., 2012), and may be adopted into the diagnostic system for *P. brassicae*. The alignment of full pathotype genomes may reveal candidate loci that could be used as pathotype distinctive barcodes for the development of metabarcoding assays. This type of assay may generate masses of short reads of identifying barcodes for *P. brassicae* pathotype characterization, simultaneously detecting multiple pathotypes from a clubroot sample via amplicon-based targeted sequencing. The resulting number of short reads may be proportional to the pathotype composition of a sample. While this may not be absolute quantitative data, it may offer insights on relative pathotype abundance (Lamb et al., 2019). The technique can also provide a representation of the pathotype diversity present in a sample. The identifying barcodes would be designed based on discriminative polymorphic regions. To be functional as a barcode, the region must have sufficient pathotype-level genetic variability, conserved flanking sites for

universal primers, and a short enough sequence length for amplification (Choudhary et al., 2021). It is essential to subject the primers to a comprehensive Basic Local Alignment Search Tool (BLAST) search to ensure specificity to the barcoding region and to *P. brassicae*, and to validate amplification through Sanger sequencing with a select number of samples before mass sequencing.

The protocol is a six-step process: (1) PCR; (2) preparation of sequencing libraries (Head et al., 2014); (3) NGS; (4) filter sequencing reads; (5) sequence assembly; and (6) pathotype identification (Figure 2.6). An initial round of conventional PCR generates the barcoded amplicons and increases the number of DNA copies to be sequenced. Each sample can be labeled with a DNA tag during PCR to identify the origin of sequencing reads (Tedersoo et al., 2019). Multiplexing through DNA tagging allows hundreds of samples to be processed in one sequencing run. Tagged amplicons are then constructed into a sequencing library in preparation for NGS; indexing adapters of known sequences are annealed to the amplicons and the library undergoes a final quantification. Once sequenced, low quality reads are filtered out to reduce sequencing errors and increase accuracy (Baloğlu et al., 2021). The resulting sequencing reads are then assembled and aligned to reference barcodes that are typical of each pathotype, revealing the pathotype of the samples. Incorporating probe-based capture may increase the efficiency of NGS in detecting specific pathotypes. This occurs through the hybridization of a probe designed specifically for a targeted sequence representing a particular pathotype, with the resulting hybridization signal indicating the recognition of the targeted pathotype (Gardiner et al., 2019; Henry et al., 2014; Wit et al., 2015).

Metabarcoding has been proposed and evaluated as a platform for diagnostics in plant pathosystems (Abdelfattah et al., 2018). It was used to analyze the fungal spore composition in

air samples and found to detect a much wider range of pathogens relative to previous methods (Banchi et al., 2018; Nicolaisen et al., 2017). Earlier approaches were insufficient as they could only detect a small fraction of the total fungal diversity. Metabarcoding was also used to investigate *Fusarium* species composition, and it was able to detect 17 species in soil samples and maize (*Zea mays* L.) residue (Cobo-Díaz et al., 2019). The ability to pool hundreds of samples in one sequencing reaction, efficient turnaround time, and increased accuracy were highlighted. Another study tested this approach with *Phytophthora* species in artificially infested soil samples, and found that the sequencing results were comparable with the soil composition (Legeay et al., 2019). Metabarcoding was shown to provide realistic approximations of species abundance when it was used to characterize *Colletotrichum* species on walnut (*Juglans regia* L.) (Da Lio et al., 2018). While metabarcoding has not been established for the clubroot pathogen, an exploration of this method is underway in our research group. We are refining genomic assemblies for 38 SSIs, which will allow us to look for rapidly evolving discriminative polymorphic regions to select for metabarcoding for use in *P. brassicae* pathotyping.

Several limitations are associated with metabarcoding as a diagnostic tool. Bioinformatics expertise and access to bioinformatics facilities are required to analyze the sequencing data competently. Contamination of samples is of greater concern due to the high sensitivity of NGS platforms. Therefore, thorough BLAST testing of the primers against a wide set of microorganisms and preliminary conventional PCR testing of the primers against other species are essential to ensure specificity for the target. Pathotypes in the sample may not be uniformly amplified, as the generation of barcoded amplicons is dependent on conventional PCR. This was evaluated with parasitic soil protist communities and results provided estimations of relative abundance different from those expected (Geisen et al., 2015). Misrepresentation may also occur

during sequencing; three different NGS platforms produced different sequencing outputs when quantitation of artificially assembled fungal communities was evaluated (Castaño et al., 2020). Due to this, it can only provide approximate pathotype proportions. Nonetheless, with the widespread adoption of genome sequencing in plant pathology research and the identification of variant information, and with the flexibility of these methods for a wide range of experimental designs, metabarcoding and NGS technologies may play a substantial role in clubroot diagnostics.

2.2.6 General limitations of molecular approaches

While molecular techniques are the future of clubroot diagnostics, there are some limitations to take into consideration. A possible challenge is finding polymorphisms that provide consistent genomic and phenotypic clustering of pathotypes. The extensively used CCD set and other host differential systems are based on phenotypically distinctive virulence patterns, which may not always be in agreement with DNA sequence variations. In the case of disagreements, pathotype classifications may be modified to incorporate the genomic data or differentiating polymorphic regions must be deliberately chosen to distinguish pathotypes as defined by the CCD or other differential sets. The development of molecular diagnostic assays is dependent on comprehensive sequence databases for the discovery of suitable polymorphic regions, and it would be beneficial to have a reliable genome for each individual pathotype. It may be challenging to assemble variable regions by a reference-based assembly approach, and therefore generating full *de novo* assemblies with long-read technology mixed with short reads may be necessary. Polymorphic regions with the greatest level of diversity will offer the greatest differentiating capacity. Insufficient genetic variability among pathotypes may cause complications in assay design and therefore restrict the use of molecular-based techniques. In addition, defining a unique sequence

for each individual pathotype may not be possible in the same region of the genome. Multiple regions may need to be used to thoroughly differentiate the pathotypes. Primers must be specific enough to avoid amplification of non-target microorganisms and non-target regions of the genome. In the case of amplicon length distinction and metabarcoding, it is imperative for the conserved primers to be generic enough to amplify the DNA of all pathotypes. The cost of adopting new molecular diagnostic tools is also an issue. However, due to the increasing interest in molecular approaches, prices may follow a downward trend as these techniques become more widely used and taken up by diagnostic laboratories.

2.3 Future perspectives

Several PCR-based, SNP-based, and sequencing technologies have been introduced into diagnostic processes of plant pathosystems (Table 2.2). PCR remains the most cost-effective and most widely used molecular technique. Several PCR-based methods may be modified into a quantitative assay, which can allow evaluation of inoculum levels or degree of host colonization. The rhPCR technique has been shown to further increase sensitivity of SNP discrimination over conventional PCR. However, the developmental phase of PCR-based assays may require a lengthy standardization process since numerous factors must be considered to minimize non-specific amplification. The main advantages of SNaPshot are its ability to multiplex samples and to distinguish any of the four alleles at the discriminatory SNP; however, this technique requires higher operational costs and is not quantitative. NGS-based metabarcoding has the greatest sensitivity and scalability; however, it is not absolutely quantitative and routine NGS will require in-depth post-sequencing bioinformatics analysis.

An ideal clubroot diagnostic tool is a mixed strategy of techniques to further increase the sensitivity and accuracy of the assay. An integrated process of metabarcoding, SNP-based distinction, and rhPCR may give the most comprehensive depiction of clubroot samples. Metabarcoding would be used initially for an assessment of pathotype diversity in a clubroot sample because of its conserved primers and high resolution. Once pathotypes are identified, a qPCR assay combining SNP-based distinction and rhPCR would indicate pathotype abundance. In this case, only primers specific to the pathotype(s) identified by metabarcoding are required.

Table 2.1 Molecular assays developed for the general detection of *Plasmodiophora brassicae*.

	Technique	Primer Sequences (5′ to 3′)	
(Ito et al., 1999)	Nested PCR	Outer	PBTZS-2: CCGAATTCGCGTCAGCGTGA [a]
		Inner	PBTZS-3: CCACGTCGATCACGTTGCAAT PBTZS-4: GCTGGCGTTGATGTACTGGAA
(Faggian et al., 1999)	Nested PCR	Outer	PbITS1: ACTTGCATCGATTACGTCCC PbITS2: GGCATTCTCGAGGGTATCAA
		Inner	PbITS6: CAACGAGTCAGCTTGAATGC PbITS7: TGTTTCGGCTAGGATGGTTC
(Wallenhammar & Arwidsson, 2001)	Nested PCR	Outer	PBAW-10: CCCCGGGGATCACGATAAATAACA PBAW-11: GGAAGGCCGCCCAGGACTACC
			PBAW-12: GCCGGCCAGCATCTCCAT PBAW-13: CCCCAGGGTTCACAGCGTTCAA
		Inner	PBTZS-3 (Ito et al., 1999) PBTZS-4 (Ito et al., 1999)
(Cao et al., 2007)	PCR	TC1F: GTGGTCGAACTTCATTAAATTTGGGCTCTT TC1R: TTCACCTACGGAACGTATATGTGCATGTGA	
(Sundelin et al., 2010)	qPCR	Pb4-1: TACCATACCCAGGGCGATT PbITS6 (Faggian et al., 1999)	
(Rennie et al., 2011)	qPCR	DC1F: CCTAGCGCTGCATCCCATAT DC1R: CGGCTAGGATGGTTCGAAAA	

(Wallenhammar et al., 2012)	TaqMan qPCR	PbF: AAACAACGAGTCAGCTTGAATGC PbR: TTCGCGCACAAGCAC TTG (Probe) PbP: CGCGCCATGCGACACTGTTAAATT
(Cao et al., 2014)	TaqMan qPCR	TC1F: GTGGTCGAACTTCATTAAATTTGGGCTCTT RTPbR1a: TCAGCACCGTTTCCGGCTGCTAAGGC (Probe) TCPb1: AAGAAGGAGAAGTCGTAACAAGGTTTC
(Deora et al., 2015)	TaqMan qPCR	PBGFPuv3F: CCTAGCGCTGCATCCCATATCGATGGCCCTGTCCTTTTAC PBGFPuv3R: CGGCTAGGATGGTTCGAAAGTGTAATCCCAGCAGCAGTTA (Probe) GFP1: ACCATTACCTGTCGACACAATCTGCCCT
(Al-Daoud et al., 2016)	PMA-PCR	PBGFPuv3F (Deora et al., 2015) PBGFPuv3R (Deora et al., 2015) (Probe) GFP1 (Deora et al., 2015)
(Wen et al., 2020)	ddPCR	DC1F: CCTAGCGCTGCATCCCATAT DC1mR: CGGCTAGGATGGTTCGAAA (Probe) PB1: /56-FAM/CCA TGTGAA/ZEN/CCGGTGACGTGCG/3IABkFQ/

[a] Primer PBTZS-2 is used as the sole outer primer for this nested PCR for amplifying the fragment from DNA samples (Ito et al., 1999)

Table 2.2 Comparison of proposed pathotyping platforms for *Plasmodiophora brassicae*.

Technique	Efficiency [a]	Specificity	Quantitative Potential	Primers Required	Costs	Main Advantages	Main Disadvantages
Amplicon length distinction	Low	Low	No [b]	1 pair per indel	Low	Conserved primers	Low scalability
SNP-based distinction	Low	Low	Yes	1 or 2 pairs per polymorphic region [c]	Low	Simple procedure	Low scalability and sensitivity
rhPCR	Low	Moderate	Yes	1 or 2 pairs per polymorphic region [c]	Moderate [d]	Simple procedure	Low scalability
SBE	Moderate	Moderate	No	1 pair + 1 SBE primer per SNP [e]	High	Scalable; can detect any allele	Non-quantitative; lengthy procedure
Metabarcoding	High	High	Partially [f]	1 pair per barcoding sequence	Very High	High sensitivity and scalability	High costs and expertise required; lengthy procedure

[a] The scalability and throughput ability of the technique.
[b] Primers are designed against a conserved region with the indel positioned within the amplicon (Figure 2.2). A qPCR assay would not identify the pathotype since amplification occurs regardless of pathotype under investigation.
[c] While only one primer pair is required to identify the SNPs, a second primer pair of alternate alleles would further verify pathotype detection.
[d] RNase H2 enzyme and its dilution buffer is required in addition to basic PCR reagents (Integrated DNA Technologies, 2021b). The price for rhPCR primers is slightly higher than conventional primers (Integrated DNA Technologies, 2021a).
[e] A conserved primer pair is required to generate the template for the SBE reaction in addition to the SBE primer (Figure 2.5).
[f] Metabarcoding cannot provide absolute quantities, only estimates of pathotype proportions.

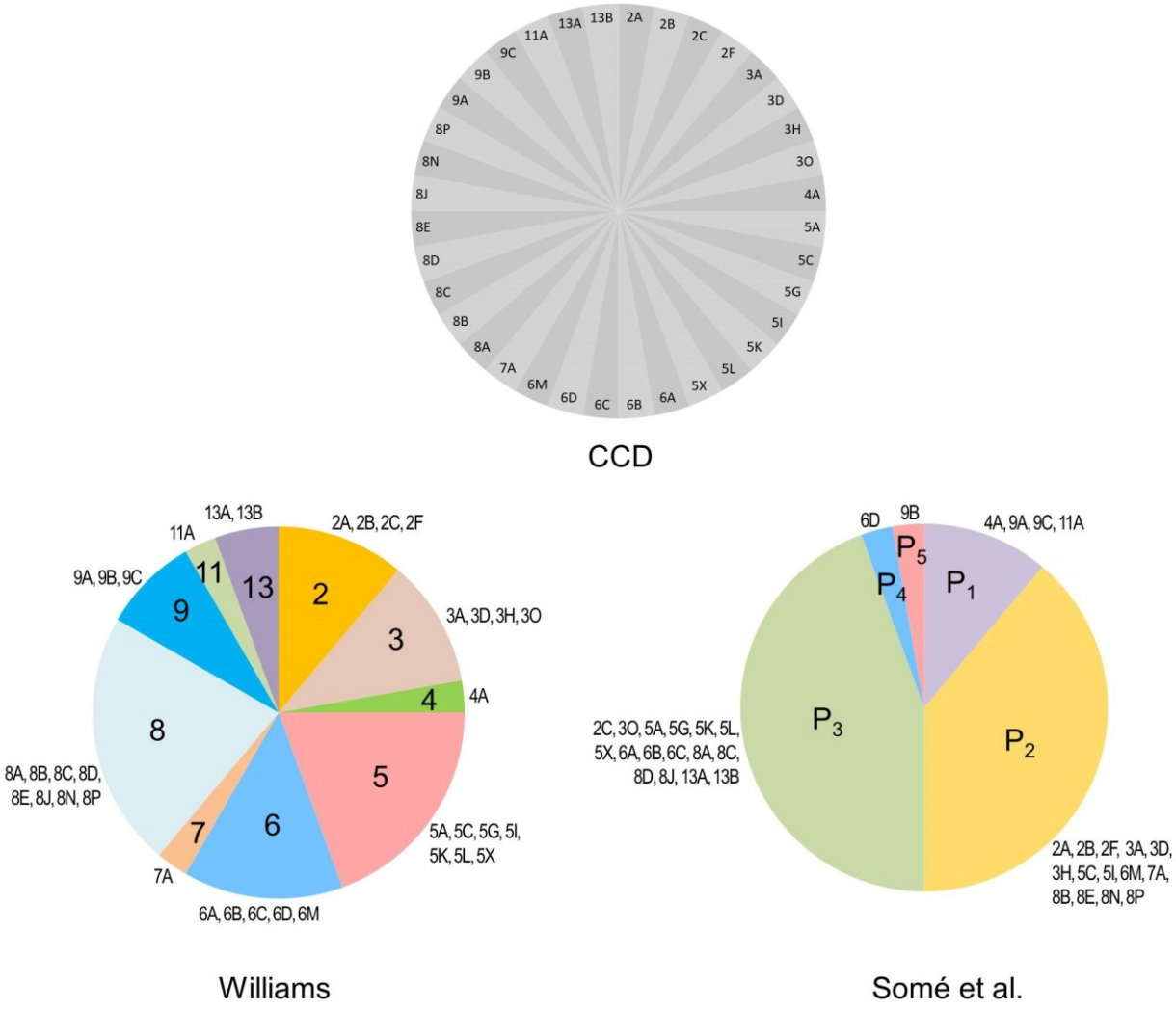

Figure 2.1 Designation of *Plasmodiophora brassicae* pathotypes from Canada as defined on the Canadian Clubroot Differential (CCD) set (Strelkov et al., 2018) in comparison with their classification on the systems of Williams (1966) and Somé et al. (1996). Slices of each pie chart for Williams (1966) and Somé et al. (1996) denote the proportion of each respective pathotype designation by its representation in Canada, with the corresponding CCD pathotypes indicated. Since the CCD set includes the differentials of Williams (1966) and Somé et al. (1996), it is possible to obtain all respective pathotype designations based on the reaction of the CCD hosts. Pathotypes 2B, 2F, 3A, 3D, 3H, 3O, 5C, 5G, 5I, 5K, 5L, 5X, 6M, 8E, 8J, 8N, and 8P were first reported by Strelkov et al. (2018). Pathotypes 2A, 4A, 6A, 6B, 6C, and 7A were first reported by Askarian et al. (2021b). Pathotypes 5A, 8A, 8B, and 8C were first reported by Strelkov et al. (2021). Pathotypes 2C, 6D, 8D, 9A, 9B, 9C, 11A, 13A, and 13B were first reported by Hollman et al. (2021).

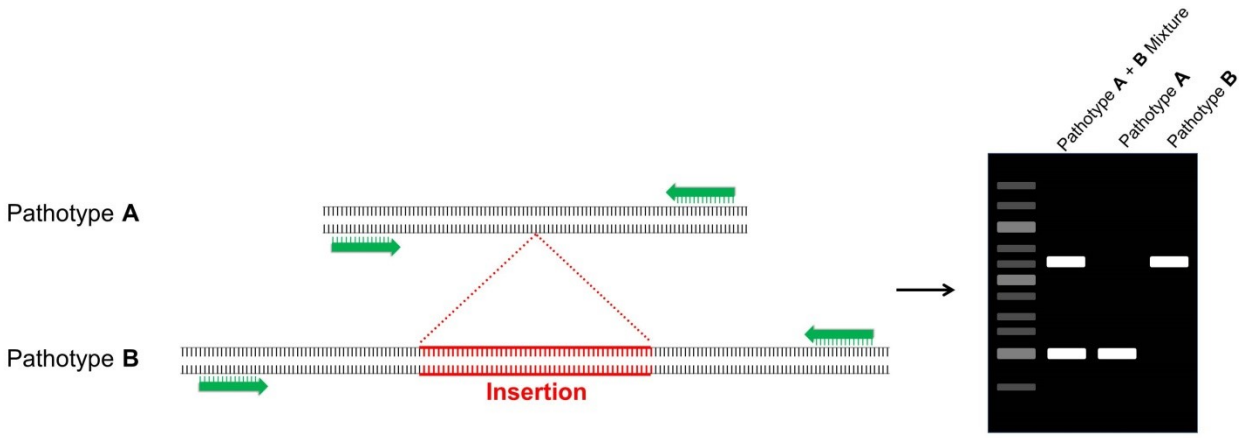

Figure 2.2 This figure summarizes the amplicon length distinction method. One primer pair (shown in green) is used for the detection of both pathotypes, as it is designed against a conserved region. The hypothetical pathotype B has a distinctive insertion (shown in red) within the sequence that will produce a greater amplicon size in comparison with hypothetical pathotype A. The electrophoretic gel presents a noticeable difference in molecular weight of the pathotypes. Both pathotypes are detectable as the polymorphism is located within the amplicon.

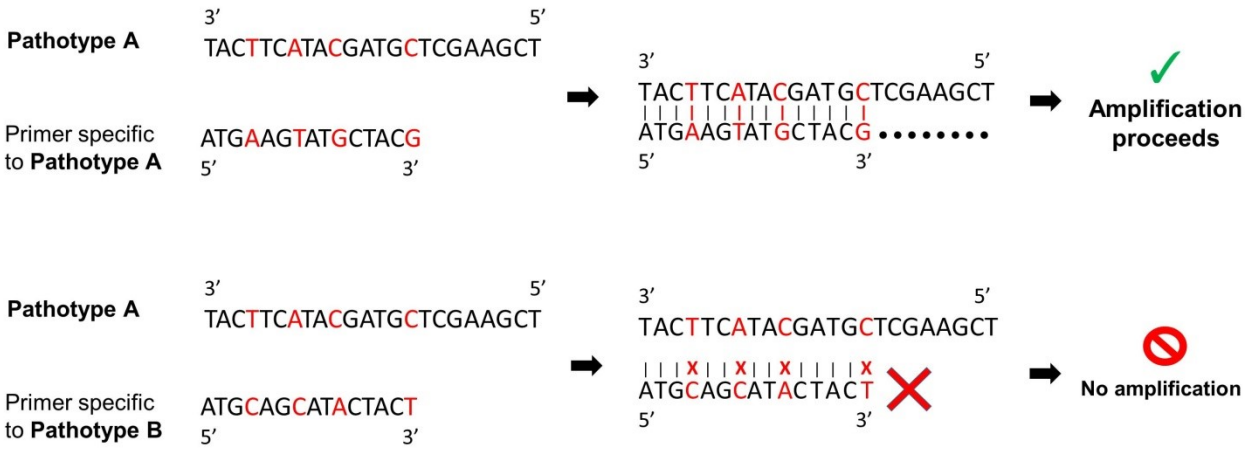

Figure 2.3 This figure summarizes the SNP-based distinction method. Primers are designed to target a distinctive polymorphic sequence specific to a pathotype or pathotype cluster. When a primer specific to the hypothetical pathotype A is used against hypothetical pathotype A, amplification occurs because of the perfect match of primer and DNA template. However, when a primer specific to hypothetical pathotype B is used against hypothetical pathotype A, mismatches prevent amplification.

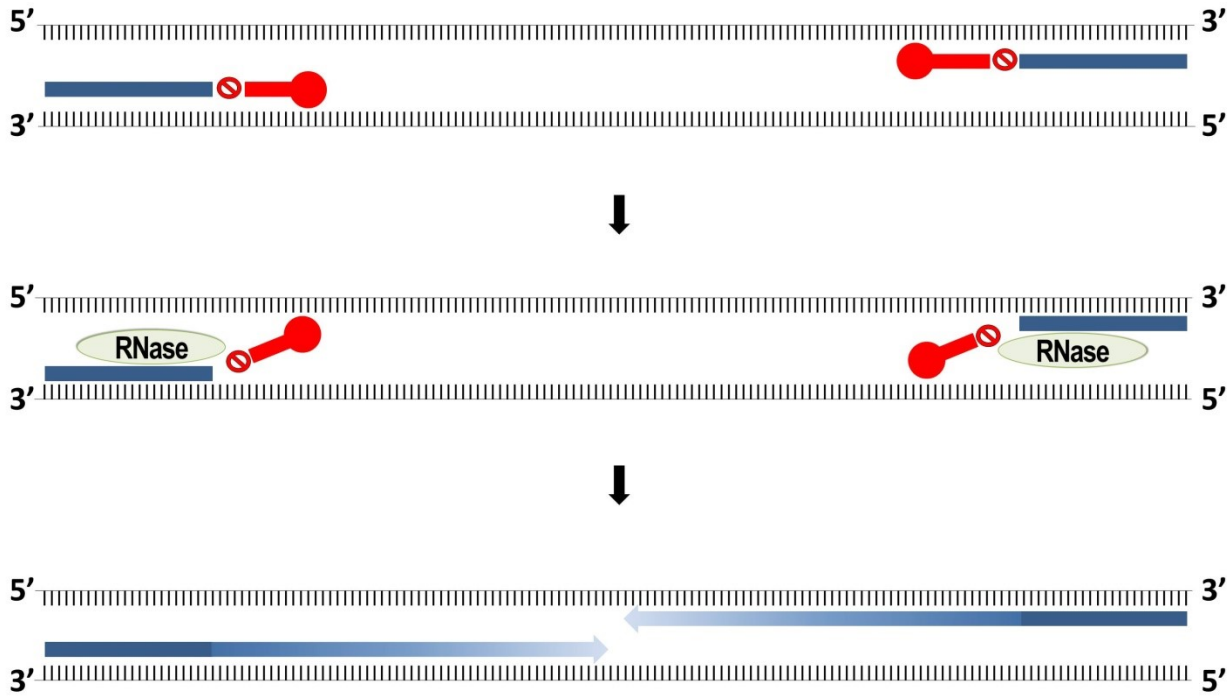

Figure 2.4 This figure summarizes RNase-H2 dependent (rhPCR). The rhPCR primers are blocked by a ribonucleotide residue as represented by the stop symbol, followed downstream by four DNA bases complementary to the template as shown in red. The red circle at the 5′ end of the blocked primer is a propanediol C3 spacer. Once the RNase H2 enzyme comes in and cleaves off the ribonucleotide residue, the primers are activated and extension by DNA polymerase continues.

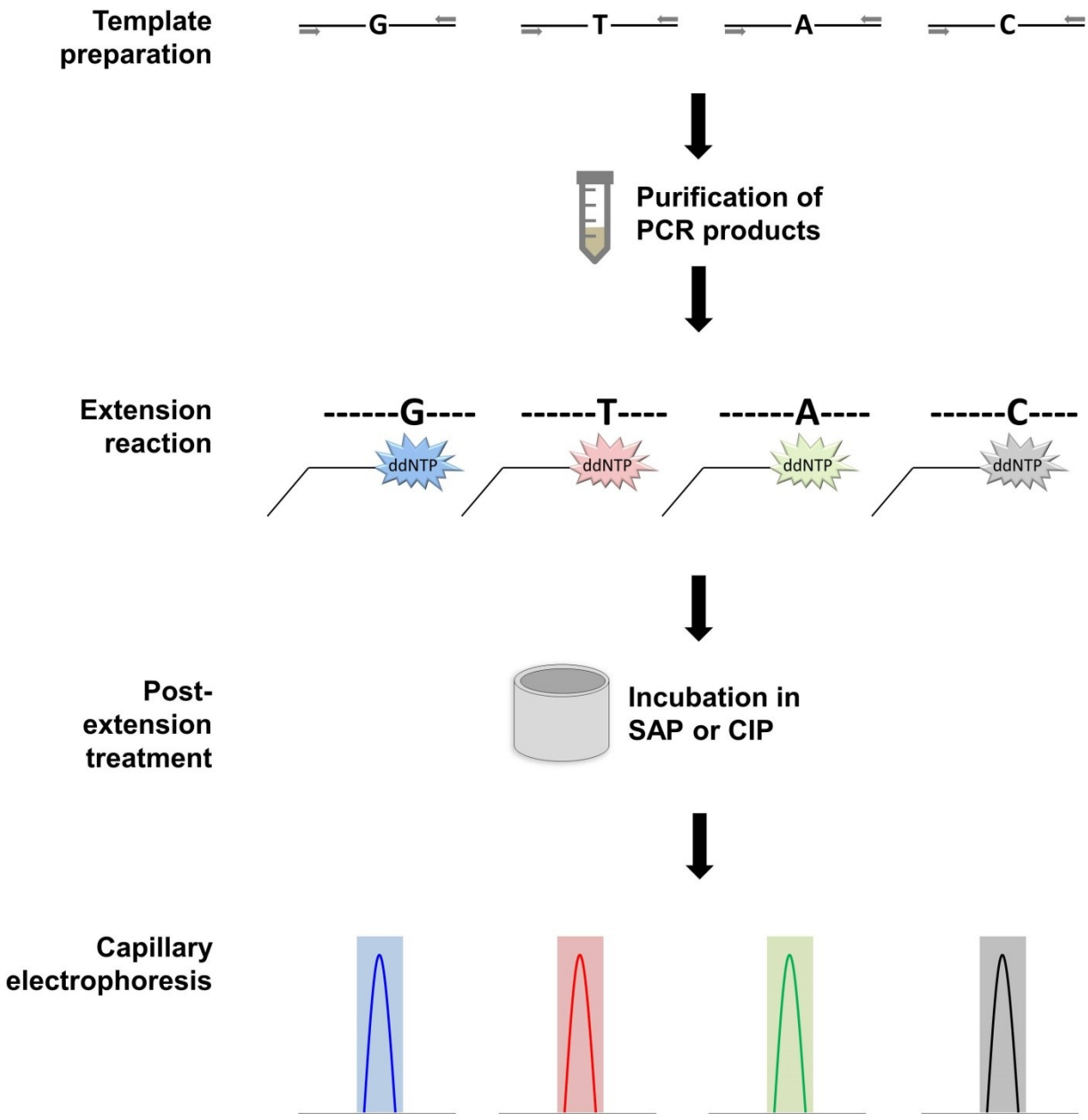

Figure 2.5 This figure summarizes the workflow of the single base extension technique. Samples undergo an initial PCR to generate templates encompassing the single nucleotide polymorphism (SNP). Once templates are purified, the extension reaction occurs. Allele-specific fluorescently labelled ddNTPs are matched to the SNP. Extension products are incubated in either shrimp alkaline phosphatase (SAP) or calf intestinal phosphatase (CIP) to remove unincorporated ddNTPs. Products are scanned via capillary electrophoresis in a genetic analyzer and differentiating SNPs are revealed in the resulting electropherograms. The G fluoresces blue, T fluoresces red, A fluoresces green, and C fluoresces black.

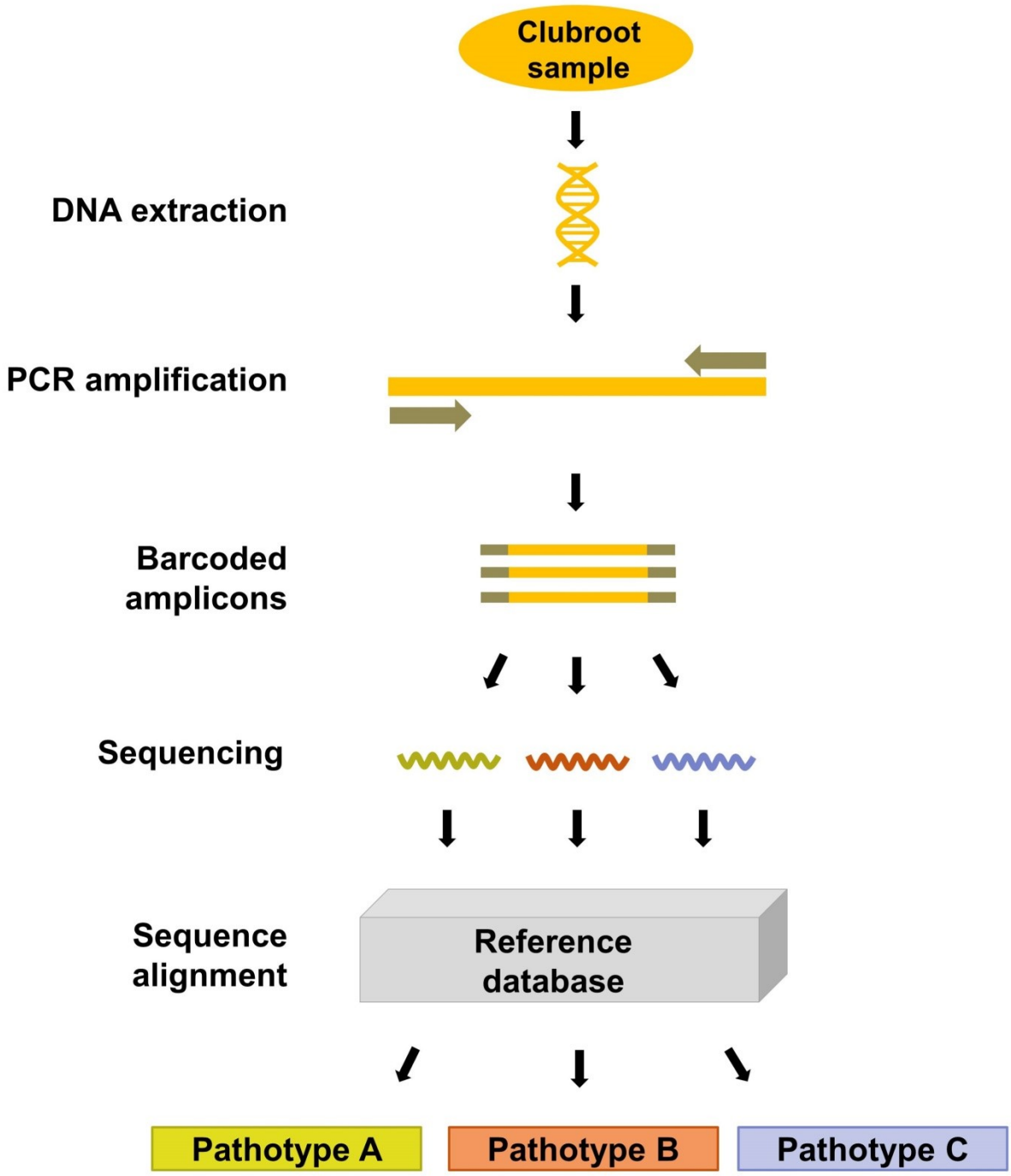

Figure 2.6 This figure summarizes the metabarcoding workflow. Genomic DNA is extracted from *Plasmodiophora brassicae* samples and undergoes an initial PCR to generate barcoded amplicons. Amplicons are prepared for next generation sequencing. The resulting sequencing reads are aligned to the reference barcode database to identify the pathotypes in the sample.

Chapter 3 – Development of rhPCR and SNaPshot assays to distinguish *Plasmodiophora brassicae* pathotype clusters

3.1 Background and introduction

The obligate parasite *P. brassicae* Woronin is the causal agent of clubroot, an important soilborne disease of crucifers worldwide. In Canada, *P. brassicae* is a major constraint in the production of canola, with the disease managed primarily by planting CR cultivars. However, the overuse of CR cultivars has resulted in shifts in the virulence of *P. brassicae* populations and the proliferation of novel pathotypes that can overcome genetic resistance. Resistance-breaking pathotypes of the clubroot pathogen were first identified in 2013 (Strelkov et al., 2016), just four years after the introduction of CR canola to the Canadian market. Since then, novel virulent pathotypes have been identified every year from infected canola plants (Askarian et al., 2021b; Hollman et al., 2021; Strelkov et al., 2018, 2021).

Traditionally, the classification of *P. brassicae* into pathotypes has relied on bioassays conducted with host differential sets. In brief, isolates of *P. brassicae* are inoculated onto a series of differential hosts, and then grouped into pathotypes based on their virulence patterns on these hosts. Various differential systems have been developed, including the hosts of Williams (1966), Somé et al. (1996), the ECD set (Buczacki et al., 1975) and, most recently, the CCD (Strelkov et al., 2018) and SCD (Pang et al., 2020) sets. The CCD set is now the most widely used differential system in Canada, and was developed to improve identification of resistance breaking pathotypes recovered from canola (Askarian et al., 2021b; Strelkov et al., 2018). While effective, the use of any host differential set is time-consuming, labor-intensive, and requires biosecure plant growth facilities. This can limit the number of isolates that can be tested, as well

as the speed at which results can be obtained. Environmental factors and the specific growing conditions may also influence host reactions, thereby potentially reducing the consistency of results. The ability to detect pathotypes quickly and efficiently has become a priority for clubroot management efforts, especially with the rapid emergence of new pathotypes capable of overcoming the resistance in many CR canola cultivars. Molecular assays would facilitate rapid pathotype identification and enable prompt testing of much larger numbers of samples.

Various molecular markers have been explored for *P. brassicae* pathotyping. A random amplified polymorphic DNA marker specific to pathotype P_1, as defined on the differentials of Somé et al. (1996), was identified and converted into a sequence-characterized amplified region (Manzanares-Dauleux et al., 2000). The *Cr811* gene, which was hypothesized to have a role in clubroot pathogenesis, was found to be exclusive to pathotype 5 (Zhang et al., 2015), as defined on the differentials of Williams (1966), and hence could serve a diagnostic purpose. A region of the 18S internal transcribed spacer sequence specific to pathotype 5X, as defined on the CCD set, was used to develop a probe-based qPCR assay for the specific detection of this pathotype (Zhou et al., 2018). Five molecular markers were found to distinguish pathotypes 4, 7, 9, and 11 (Zheng et al., 2018), as classified on the differentials of Williams (1966). Recently, over 1500 SNPs were identified as differentiating two genetically distinct *P. brassicae* populations from Alberta, Canada, enabling the development of population-specific molecular markers (Holtz et al., 2018, 2021). Two rhPCR (Dobosy et al., 2011) primer pairs were also developed to differentiate a new, resistance-breaking "pathotype 3-like strain" of *P. brassicae* from the original pathotype 3 (Yang et al., 2018). However, neither the exact nature of this pathotype 3-like strain, nor its CCD classification, were available. To our knowledge, no rhPCR-based assays capable of distinguishing *P. brassicae* pathotypes as classified on the CCD set have been

reported. Similarly, there are no reports of the use of SNaPshot technology (Chagné et al., 2007) for the identification of *P. brassicae* pathotypes.

The novel allelic discrimination technology, rhPCR, provides greater accuracy and sensitivity relative to conventional PCR (Dobosy et al., 2011). In conventional PCR, the differentiation of pathotypes with slight nucleotide variations is challenging, since primers do not require absolute target specificity. The rhPCR primers are blocked by a single ribonucleotide residue positioned at the discriminatory SNP, and a 3' C3 spacer that prevents polymerase extension activity. Amplification with rhPCR requires perfect binding of primers to the target, allowing differentiation of samples with a single nucleotide difference. The blocked primers are activated via cleavage of the ribonucleotide by the RNase H2 enzyme from *Pyrococcus abyssi*, which removes both the ribonucleotide and the C3 spacer. The enzyme can only unblock the primer if the ribonucleotide is complementary to the template strand as the enzyme binds to RNA-DNA duplexes. In the case of a mismatch, no cleavage will occur, the primer remains blocked, and no amplification occurs during PCR. The rhPCR technology would allow for the detection of pathotypes by differentially permitting the synthesis of amplicons that differ by a SNP.

SNaPshot is a modified sequencing single base extension reaction that enables discrimination based on SNPs (Chagné et al., 2007). Differentiating SNPs are identified based on a fluorescent color corresponding to one of the four possible alleles. SNaPshot primers are positioned one base upstream of the SNP. When the primer anneals to the DNA template, the polymerase extends the primer by one base with a fluorescently labelled ddNTP matching the SNP. The resulting product size is the length of the SNaPshot primer plus the addition of the ddNTP. The SNaPshot product is then analyzed via capillary electrophoresis, and pathotypes are identified based on the

color of the fluorescence and product size. The signal will consist of one fluorescent color if a single allele is present or two or more colors if multiple allelic variants are present.

Here, we report and validate two independent assays based on rhPCR and SNaPshot technologies to differentiate between a pathotype 5X cluster and a pathotype 3H cluster of *P. brassicae,* as defined on the CCD set. Clusters are formed based on the allelic variant in the targeted discriminatory SNP positions used as molecular markers for assay development. Additional pathotypes also were tested and clustered with either pathotype 5X or 3H based on the allelic variant in the discriminatory SNP positions. Pathotype 5X was selected for this study as it is the first and best characterized of the pathotypes able to overcome the resistance in CR canola, while pathotype 3H was included as it is one of the most common pathotypes from canola in western Canada and cannot overcome host resistance (Hollman et al., 2021; Strelkov et al., 2018). Our results suggest that rhPCR and SNaPshot technologies can provide a simple and reliable way to distinguish pathotypes of *P. brassicae* in a rapid manner.

3.2 Materials and methods

3.2.1 SNP selection

Thirty-eight *P. brassicae* single-spore isolates were included in this study. Purified genomic DNA was isolated from resting spores prepared by Askarian et al. (2021a). The DNA was quantified with a Qubit 2.0 DNA HS Assay (Thermo Scientific, Waltham, MA, USA) and DNA quality was assessed with a TapeStation Genomic DNA Assay (Agilent Technologies, CA, USA). The sequencing library was prepared using a KAPA Hyper Prep Kit (Roche, Basel, Switzerland) as per the manufacturer's recommendations. Library quality and quantity were evaluated with a Qubit 2.0 DNA HS Assay (Thermo Scientific) and TapeStation High Sensitivity

D1000 Assay (Agilent Technologies). The prepared sequencing library was then sent to Admera Health (South Plainfield, NJ, USA) for next-generation sequencing on an Illumina® HiSeq X instrument (Illumina, California, USA).

Sequencing reads were aligned to the 2015 e3 reference genome for *P. brassicae* (Schwelm et al., 2015) (European Nucleotide Archive project PRJEB8376). Variants were called from high quality aligned reads using HaploTypeCaller (DePristo et al., 2011; Garrison & Marth, 2012; McKenna et al., 2010; Van der Auwera et al., 2013) with filters of overall read depth equal to or larger than 15 ($DP \geq 15$) and quality equal to or larger than 40 ($GQ \geq 40$) to produce variant call format (vcf) files per each isolate. SOAPdenovo v2.01 (Li et al., 2010; Luo et al., 2012) was used to assemble the reads into draft assemblies. The resulting vcf files were loaded into Integrative Genomics Viewer (Robinson et al., 2011) to visualize polymorphisms and identify candidate SNP loci among the 11 CCD pathotypes represented in our 38 SSIs (Figure 3.1). To confirm the polymorphic region found in contig 1 (Schwelm et al., 2015) that differentiates the 5X pathotype cluster from the 3H cluster (Table 3.1), a conventional PCR primer pair was designed to amplify the region encompassing the SNPs through Sanger sequencing (Sanger et al., 1977). Forward primer SEQ1-43778fw 5'-GCCTGTCGAACGTCTGTT-3' and reverse primer SEQ1-43778rv 5'-ATAAAGTCTGGACACGAGAACG-3' were designed using PrimerQuest (Integrated DNA Technologies, Coralville, IA, USA) with parameters that included primer length ranging between 18-24 bases, GC content ranging between 40-60%, and melting temperature of 60°C. This set produced a 508 base pair amplicon to confirm SNPs used for both the rhPCR and SNaPshot assays. The primers were evaluated for specificity with command line BLAST v. 2.6.0 against the reference e3 *P. brassicae* genome (Schwelm et al., 2015). The argument - task "blastn-short" was used as this task is optimized for short sequences of less than

30 nucleotides. The primers were also subjected to a BLAST search in the National Center for Biotechnology Information (NCBI) online database (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to ensure specificity to *P. brassicae*.

Three SSIs from each cluster were selected for Sanger sequencing to validate the presence of polymorphisms detected using whole genome sequencing. The SSIs ST11 (5X), ST23 (5X) and SR20 (6B) were tested from the reference 5X cluster, and SSIs SL09 (2F), SS48 (3H), and SW30 (3H) were tested from the alternate 3H cluster (Askarian et al., 2021b). PCR analyses were carried out in a 20 µL final volume containing 1X reaction buffer, 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.4 µM of each forward and reverse primers, 1 U Platinum Taq DNA polymerase (Invitrogen, Waltham, Massachusetts, USA), 10 ng of genomic DNA as template, and 13.7 µl nuclease-free water. All reactions were conducted in a Veriti 96-Well Thermal Cycler (Applied Biosystems, Waltham, MA, USA) under the following cycling conditions: 2 min at 94°C, followed by 40 cycles of 30 s at 94°C, 1 min at 63°C, and 1 min at 72°C, with a final 10 min extension at 72°C. Samples were held at 4°C. Four technical replications of each sample were performed. The PCR products from one replicate per each sample were resolved by electrophoresis on a 1% agarose gel to confirm the presence of specific amplification, product size and intensity. The other three replicates were combined and purified using the Wizard SV Gel and PCR Cleanup System (Promega, Madison, WI, USA) following the manufacturer's specifications. The quality and quantity of purified DNA were verified on a NanoDrop spectrophotometer (Thermo Scientific, Waltham, MA, USA), then sent for Sanger sequencing (Sanger et al., 1977) at the Molecular Biology Service Unit, Department of Biological Sciences, University of Alberta (Edmonton, AB, Canada). The resulting sequences were visualized and SNPs were confirmed with Sequencher 5.0 (Gene Codes Corporation, Ann Arbor, MI, USA).

3.2.2 rhPCR

We designed two sets of rhPCR primers to distinguish *P. brassicae* isolates belonging to either one of the pathotype clusters. The reference rhPCR primer pair was designed to amplify isolates of the 5X cluster; it was referred as the reference cluster since the SNPs also belonged to the *P. brassicae* e3 reference genome (Schwelm et al., 2015). The alternate rhPCR primer pair was designed to amplify isolates of the 3H cluster (Table 3.2). In addition to the differentiating SNPs positioned against the ribonucleotide bases, the primers were positioned in a polymorphic region that would allow for multiple SNPs to increase specificity. There were five SNPs between the forward primers and two SNPs between the reverse primers. These sets produced a 230 base pair amplicon. The specificity of the primers was evaluated with command line BLAST v. 2.6.0 against the e3 reference genome (Schwelm et al., 2015). The primers were also subjected to a BLAST search in the NCBI online database (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to ensure specificity to *P. brassicae*.

The specificity of the primers and the rhPCR block-cleavable technology was evaluated against gBlocks gene fragments (Integrated DNA Technologies, Coralville, IA, USA), double-stranded synthetic oligonucleotides. One gBlock was designed to replicate the 5X polymorphic region sequence, and another was designed to replicate the 3H polymorphic sequence. The gBlock gene fragment contained the 230 base pair rhPCR amplicon in its entirety, plus an additional 100 base pairs upstream and downstream from the amplicon. PCR analyses were carried out in a 20 µL final volume containing 1X reaction buffer (Applied Biosystems, Waltham, MA, USA), 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.4 µM of each forward and reverse primer, 1 U Platinum Taq DNA polymerase (Invitrogen, Waltham, MA, USA), 5.2 mU RNase H2 enzyme, and 5 ng gBlock as template. The gBlock testing was run in a Veriti 96-Well Thermal Cycler (Applied Biosystems,

Waltham, MA, USA) under the following cycling conditions: 2 min at 94°C, followed by 12 cycles of 10 s at 94°C and 30 sec at 70°C. Samples were held at 4°C until the PCR products were electrophoresed on a 1% agarose gel. The block-cleavable technology was also tested by repeating the PCR, but with the RNase H2 enzyme excluded from the master mix as a control.

The rhPCR primer pairs were then evaluated and optimized against the SSIs in our collection: 13 isolates belonging to the 5X cluster and 25 isolates belonging to the 3H cluster (Table 3.1). PCR analyses were carried out in a 20 µL final volume containing 1X reaction buffer, 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.4 µM of each forward and reverse primer, 1 U Platinum Taq DNA polymerase (Invitrogen, Waltham, Massachusetts, USA), 5.2 mU RNase H2 (Integrated DNA Technologies, Coralville, IA, USA), and 10 ng genomic DNA as template. The reaction was run in a Veriti 96-Well Thermal Cycler (Applied Biosystems, Waltham, MA, USA) under the following cycling conditions: 2 min at 94°C, followed by 35 cycles of 10 s at 94°C and 30 sec at 70°C. Annealing temperatures and extension times for PCR were determined according to the primer sequence and amplicon size. Samples were held at 4°C until the amplicons were electrophoresed on a 1% agarose gel.

3.2.3 SNaPshot

A conventional PCR primer pair was designed to generate the template for the SNaPshot extension reaction. The primer sites to generate this product were conserved among the 38 SSIs and targeted a region that contained the differentiating SNP. The same forward primer SEQ1-43778fw previously designed for Sanger sequencing was used in conjunction with a newly designed reverse primer SEQ1-43778rv2 5'-CTCGAACTCTTTGTCGTCGTT-3'. This set generated a 305 base pair amplicon in contig 1 of the e3 reference genome (Schwelm et al.,

2015). A SNaPshot primer snpsht1-43778 5'-AAAAAAACGATAACGTCGTGGACGACGGCG-3' was designed upstream of the polymorphic base to distinguish pathotypes. A seven nucleotide non-homologous polyA tail was added to the 5' end to bring the length of the primer to 30 nucleotides long, the minimum length recommended (Applied Biosystems, Waltham, MA, USA). The complementary region between the primer and template was kept at 23 nucleotides, to maintain an annealing temperature of 50°C that matched the annealing temperature (50°C) of the SNaPshot control primer (Applied Biosystems, Waltham, MA, USA). The primer was subjected to reverse phase high performance liquid chromatography purification (Integrated DNA Technologies, Coralville, IA, USA).

All of the SSIs listed in Table 3.1 were also tested in the SNaPshot assay. Template generation was carried out in a 20 µL final volume PCR containing 1X reaction buffer, 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.4 µM of each forward and reverse primer, 1 U Platinum Taq DNA polymerase (Invitrogen, Waltham, Massachusetts, USA), and 10 ng genomic DNA as template. The reaction was run in a Veriti 96-Well Thermal Cycler (Applied Biosystems, Waltham, MA, USA) under the following cycling conditions: 2 min at 94°C, then 40 cycles of 30 s at 94°C, 1 min at 63°C, and 1 min at 72°C, followed by a final 10 min extension at 72°C. Samples were held at 4°C. Four technical replications were included for each sample. The PCR products of a single replicate from each sample were electrophoresed on a 1% agarose gel to confirm the presence of the specific amplicon and product intensity. The other three replicates were combined and purified using the Wizard SV Gel and PCR Cleanup System under manufacturer specifications (Promega, Madison, WI, USA). gBlocks corresponding to each 5X and 3H cluster were also designed and used to run control reactions in parallel. Five nanograms of gBlocks were used as template instead of 10 ng genomic DNA, to reduce the copy number of this region sequence, and

only 12 cycles were conducted in the PCR instead of 40 cycles, as recommended by the manufacturer (Integrated DNA Technologies, Coralville, IA, USA).

The SNaPshot Multiplex Kit (Thermo Scientific, Waltham, MA, USA) was used for the extension reaction in a 10 µL final volume containing 1X master mix (Thermo Scientific, Waltham, MA, USA), 0.2 µM SNaPshot primer, and 0.2 pmol SNaPshot template. The extension reaction was carried out in a Veriti 96-Well Thermal Cycler (Applied Biosystems, Waltham, MA, USA) under the following cycling conditions: 25 cycles of 10 s at 96°C, 5 s at 50°C, and 30 s at 60°C, then held at 4°C. Control reactions with a control template and control primers supplied by the SNaPshot Multiplex Kit (Thermo Scientific, Waltham, MA, USA) were run in parallel under the same cycling conditions. Extension reaction products were then subjected to a post-extension treatment with SAP (New England BioLabs, Ipswich, MA, USA) to remove any unincorporated ddNTPs. One unit of SAP was added to each sample, and then incubated for 60 min at 37°C, followed by 15 min at 75°C, and held at 4°C.

Treated extension products were then prepared in a 96-well plate for capillary electrophoresis. Each injection was performed at a final volume of 10 µL containing 9 µL Hi-Di formamide (Applied Biosystems, Waltham, MA, USA), 0.5 µL GeneScan 120 LIZ size standards (Applied Biosystems, Waltham, MA, USA), and 0.5 µL extension product. The plate was incubated for 5 min at 95°C, and capillary electrophoresis was carried out in an ABI Prism 3730 Genetic Analyzer (Applied Biosystems, Waltham, MA, USA) at the Molecular Biology Service Unit, Department of Biological Sciences, University of Alberta (Edmonton, AB, Canada). The Peak Scanner v1.0 (Applied Biosystems, Waltham, MA, USA) was used to determine the SNP allele based on the resulting fluorescence peak.

3.2.4 Extraction of DNA from root galls for evaluating the rhPCR and SNaPshot assays

The performance of the SNaPshot and rhPCR assays was evaluated with 12 canola root galls representing different FIs and SSIs that had been previously pathotyped using the CCD set (Table 3.3). The *P. brassicae* DNA from the galls was isolated using the cetyltrimethylammonium bromide (CTAB) extraction method (Doyle & Doyle, 1987), followed by phenol-chloroform purification. The CTAB lysis buffer was prepared with 2% CTAB (w/v), 100 mM Tris-HCl (pH 8.0), 20 mM EDTA (pH 8.0), and 1.4 mM NaCl, and the pH of the buffer was adjusted to 8.0 prior to sterilization in an autoclave. The galls were frozen at -80°C for 24 h, and then ground in liquid nitrogen with a mortar and pestle. The resultant ground sample (200 mg from each gall) was transferred into a 2 mL microcentrifuge tube and 600 µL of CTAB extraction buffer was added. The samples were incubated at 60°C for 20 min, during which samples were mixed by inversion every 5 min. After incubation, an equal volume of 600 µL phenol:chloroform:isoamyl alcohol (25:24:1, v/v) was added, vortexed, and centrifuged at 14,000 rpm for 5 min. The top aqueous phase (supernatant) was transferred to a new 2 mL microcentrifuge tube, and was subjected to two more rounds of phenol:chloroform:isoamyl alcohol DNA purification. The purified DNA was then precipitated in 700 µL of 100% ice cold isopropanol; samples were mixed by inversion, placed on ice for 10 min, and then centrifuged at 14,000 rpm for 8 min. The isopropanol was discarded and the precipitated DNA pellet was washed with 500 µL of 80% ice cold ethanol; the sample was vortexed until the pellet detached off the tube, and then centrifuged at 14,000 rpm for 3 min. The ethanol was discarded and the remaining pellet was left at room temperature to air dry. Once dried, the DNA was dissolved and resuspended in 100 µL sterile nuclease-free water. The concentration and purity of each sample were determined with a NanoDrop 1000 spectrophotometer (Thermo Scientific, Waltham, MA,

USA), and DNA integrity was assessed by loading 200 ng per sample onto a 1% agarose gel. Samples were then diluted to a working concentration of 5 ng μL^{-1} with sterile nuclease-free water and stored at -20°C. The samples were tested in the SNaPshot and rhPCR assays under the same conditions as described above.

3.2.5 Testing of relative abundance

Different proportions of mixed isolates were tested to assess the capacity of the SNaPshot assay to determine relative abundances. Three different two-isolate mixtures were evaluated with nine different proportions of 10:90, 20:80, 30:70, 40:60, and 50:50 (Table 3.4). Mixtures were prepared prior to template generation to simulate conditions where a root gall developed as a result of a mixed infection by more than one pathotype. Ten nanograms of total genomic DNA was used for the PCR. The entire SNaPshot assay procedure from template generation to capillary electrophoresis followed the same protocol as described earlier.

3.2.6 Blind testing

Blind testing was conducted with the rhPCR and SNaPshot assays. While the isolates corresponding to the galls had been previously pathotyped, the experiment was conducted without knowledge of pathotype designations in a single-blind experiment. *P. brassicae* DNA from 16 blinded galls was extracted according to the CTAB method (Doyle & Doyle, 1987) following the same procedure as described earlier. Blinded samples were subjected to both rhPCR and SNaPshot assays under the conditions described above.

3.3 Results

3.3.1 rhPCR

The specificity of the primers and the rhPCR block-cleavable technology was tested with gBlocks gene fragments. The amplification of the gBlocks with rhPCR generated bands of the expected 230 base pair amplicon (Figure 3.2). The primer pair rh1-43812R was specific to the gBlock designed to replicate the reference polymorphic sequence, and yielded no visible PCR products with the alternate gBlock. The primer pair rh1-43812A was specific to the gBlock designed to replicate the alternate polymorphic sequence, and yielded no visible PCR products with the reference gBlock. This confirmed that the rhPCR primer pairs were specific to the targeted polymorphic sequences. No amplification occurred with the no RNase H2 enzyme control (results not shown).

Amplification of *P. brassicae* pathotype SSIs from Table 3.1 generated strong bands of the expected 230 base pair size when using 10 ng of purified genomic DNA template (Figure 3.3). The primer pair rh1-43812R was specific to pathotypes of the reference cluster. Amplification of all 13 SSIs from the reference cluster using the reference primer pair produced single bands and yielded no visible PCR products with the alternate primer pair. In contrast, the primer pair rh1-43812A was specific to pathotypes of the alternate cluster. Amplification of all 25 SSIs from the alternate cluster using the alternate primer pair produced bands, while no visible PCR products were obtained with the reference primer pair. The only exception was with the alternate cluster SSI ST40 classified as pathotype 3A, which produced bands of equal intensity with both primer pairs (Figure 3.3c).

The same 230 base pair amplicons were observed when the two primer pairs were tested against DNA extracted from root galls, and each sample only amplified with one primer pair (Figure 3.4). The sensitivity of the rhPCR assay with DNA extracted from the galls matched that of the DNA from the original SSIs, as the bands were of comparable intensity. Isolates of pathotypes

2F (SACAN-ss3), 3A (F3-14, F185-14, F189-14) and 3H (SACAN-ss1) amplified with the predicted alternate primer pair, and isolates of pathotype 5X (LG-1, LG-2, LG-3) amplified with the predicted reference primer pair. Clustering of pathotypes 5I (ORCA-ss3), 6M (AbotJE-ss1), 8E (F187-14) and 8N (CDCN-ss1) was carried out only after the generation of these results, since these pathotypes were not part of the original SSI collection and we did not have their corresponding sequencing reads. Based on these results, isolates ORCA-ss3, AbotJE-ss1, F187-14 and CDCN-ss1 belong to the alternate cluster.

3.3.2 SNaPshot

The SNaPshot primer snpsht1-43778 correctly produced fluorescent peaks of the expected color for all 38 SSIs, with green corresponding to the reference cluster and blue corresponding to the alternate cluster (Figure 3.5). The SSI ST40, classified as pathotype 3A, yielded both green and blue peaks, showing the existence of both alleles (A and G) in the targeted SNP position. This result is consistent with the results of the rhPCR assay, where amplification of ST40 occurred with both primer pairs, and suggests that this was due to an issue with isolate purity rather than to an error of primer specificity.

The clustering of the DNA of the isolates from canola root galls was consistent with the results of the rhPCR testing (Figure 3.6). Isolates classified as pathotypes 2F (SACAN-ss3), 3A (F3-14, F185-14, F189-14) and 3H (SACAN-ss1) belonging to the alternate cluster produced blue fluorescent peaks, and isolates of pathotype 5X (LG-1, LG-2, LG-3) belonging to the reference cluster produced green fluorescent peaks. Likewise, pathotypes 5I (ORCA-ss3), 6M (AbotJE-ss1), 8E (F187-14) and 8N (CDCN-ss1) were identified as part of the alternate cluster due to their blue fluorescence. This confirmed that the differentiating SNPs selected for assay

development occurred beyond our SSI collection. Furthermore, the sensitivity of the SNaPshot primer with the galled root collections matched that of the DNA from the SSIs as fluorescence peaks were of comparable strengths.

The capacity of the SNaPshot assay to determine the relative abundance of different isolates was assessed with three two-isolate mixtures (Figure 3.7). The first mixture consisted of isolates from our original SSI collection (Figure 3.7a), and the second (Figure 3.7b) and third mixtures (Figure 3.7c) consisted of DNA extracted from root galls. The assay was able to detect a 10% relative allelic proportion in a 10:90 template mixture. However, relative peak strengths were not precisely proportionate to the abundance ratio of the two isolates within each mixture.

3.3.3 Blind testing

The rhPCR and SNaPshot assays were validated in a single-blind study with 16 blinded samples (Table 3.5). Samples were assigned into either the reference or alternate clusters based on the results of the rhPCR amplification and SNaPshot fluorescence peaks. The rhPCR primer pairs produced the expected 230 base pair amplicon, and band intensity was comparable with earlier testing (Figure 3.8). The SNaPshot primer produced either green or blue fluorescence peaks of comparable strength (Figure 3.9). Samples 5, 6, 11, 13, and 16 were revealed to be the same SSIs as in our original collection (Table 3.1), while samples 1, 3, 7, 8, 12, and 14 were revealed to be the same isolates we had previously used for DNA extraction from root galls (Table 3.3). Sample 6, which was revealed as SSI ST40 classified as pathotype 3A, again produced amplicons with both primer pairs in the rhPCR assay and both blue and green peaks with the SNaPshot assay.

3.4 Discussion

The aim of this study was to develop *P. brassicae* pathotyping assays for clubroot diagnostics using discriminating polymorphic regions that differentiate pathotype clusters. Molecular pathotyping of *P. brassicae* has been limited up to this point, as only a few assays and molecular markers have been reported. The rhPCR and the SNaPshot assays developed in this study are much faster than the use of the CCD or any other host differential set, generating same day results from the process of DNA extraction. The technologies behind these two assays show strong potential to be specific and reliable for molecular pathotyping. The SNPs used as molecular markers for the development of the assays were tested and confirmed to be specific to the pathotype clusters from which they were designed. Isolate origin had no effect, since all of the SSIs in our original collection (Table 3.1) and the DNA extracted from the root galls (Table 3.3) resulted in the same level of specificity with both the rhPCR and SNaPshot assays, and yielded the same 230 base pair amplicon with the rhPCR assay. This suggests that the polymorphic region selected here is consistent among all isolates.

Unlike the rhPCR assay developed by Yang et al. (2018), the assay reported here was developed using a collection of *P. brassicae* isolates that had been pathotyped on the CCD set. This allows for a more distinct and potentially relevant clustering of pathotypes from Canada, with the ability to link this clustering to the virulence phenotypes of the pathotypes on the hosts of the CCD. The two rhPCR primer pairs simultaneously and specifically amplified the expected pathotype clusters and produced no amplification of pathotypes of the opposing cluster, demonstrating their high specificity for the SNPs in the selected polymorphic region.

The SNaPshot assay is the first of its kind in clubroot diagnostics as no single base extension assay for the purpose of *P. brassicae* pathotyping has been reported. Template generation with the conventional PCR primer pair produced an amplicon suitable for the extension reaction. The primer sites were conserved among all the SSIs in our original collection and across the pathotyped galls, with the primers consistently producing the expected 305 base pair amplicon. Our selected differentiating SNP and the target site of the SNaPshot primer was adequately situated within the amplicon, as indicated by the successful extension of the SNP. The SNaPshot primer accurately produced green fluorescence peaks for pathotypes of the reference cluster and blue fluorescence peaks for pathotypes of the alternate cluster. The assay was also shown to be sufficiently sensitive to detect both pathotypes in two-isolate mixtures in the relative abundance testing.

The rhPCR and SNaPshot assays were able to differentiate pathotypes of the reference 5X cluster from pathotypes of the alternate 3H cluster; however, the one exception was the SSI ST40 classified as pathotype 3A. With this isolate, amplification of products of comparable intensity was observed with both the reference and alternate rhPCR primer pairs (Figure 3.3c), and extension of the SNaPshot primer produced both blue and green peaks (Figure 3.5). These mixed results from the SSI ST40 were further confirmed in the single-blind study, where ST40 was revealed to be blinded sample 6 (Table 3.5), for which amplification occurred with both rhPCR primer pairs (Figure 3.8), and both blue and green peaks appeared with the SNaPshot primer (Figure 3.9). This indicates that allelic variants of both pathotype clusters are present in the template. In the report where ST40 was first described, it was indicated that while this SSI most closely resembled pathotype 3A, it also displayed characteristics similar to pathotypes 3H, 5X and 6B (Askarian et al., 2021b). As such, the authors of the original study decided to eliminate

SSI ST40 from further testing. Since SSI ST40 was supposedly produced from a single-spore, its heterogeneity could reflect an error in the initial single-spore isolation process (e.g., two resting spores attached together), or perhaps mixing during propagation under greenhouse conditions. It will be important to confirm the purity of isolates prior to subjecting a sample for testing during assay development, to enable the accurate interpretation of results and to prevent false positives or negatives.

The accuracy and sensitivity of the rhPCR and SNaPshot assays should facilitate the analysis of *P. brassicae* field populations for the presence of heterogeneity. For example, the FIs LG-1, LG-2, and LG-3, all of which were classified as pathotype 5X (Strelkov et al., 2016; Strelkov et al., 2018), presented miniscule but notable blue peaks in addition to the expected green peaks with the SNaPshot assay (Figure 3.6). This indicates the presence of another pathotype of the 3H cluster (in much smaller proportions) within the 5X FIs, likely reflecting the coexistence of multiple pathotypes in one field gall (Xue et al., 2008; Fu et al., 2019; Askarian et al. 2021b). The virulence patterns of FIs on differential hosts often reflect the 'predominant' pathotype found in a root gall, and may not capture the extent of heterogeneity in *P. brassicae* populations from the field (Strelkov et al. 2006; Xue et al., 2008). A recent study investigating the purity of pathogen populations collected from field galls confirmed this to be the case, with most representing a mixture of pathotypes and showing some heterogeneity (Fu et al., 2019).

The production of reproducible and accurate data relies upon the DNA extraction and purification methods used. In this study, DNA was collected by means of CTAB extraction (Doyle & Doyle, 1987) followed by phenol-chloroform liquid-liquid purification for all of the SSIs and FIs evaluated, as these methods have been shown to produce high quality DNA for downstream PCR applications. The quality and quantity of the purified DNA were verified on a

NanoDrop spectrophotometer (Thermo Scientific, Waltham, MA, USA). We have not tested the assays against DNA extracted using different methodologies, and therefore, isolates prepared with other extraction methods must be tested to ensure rhPCR amplification and SNaPshot sequencing remain consistent. PCR and sequencing is affected by low yields, low integrity, and impurities in the presence of contaminants and inhibitors (Rådström et al., 2004; Schrader et al., 2012). Since high quality and quantity of template DNA is critical for PCR-based and sequencing studies, appropriate DNA extraction and purification procedures are essential for consistency during both assay optimization and testing. In addition, the primers developed here were designed to identify *P. brassicae* pathotypes present in root galls, and have not been tested for the detection of pathotypes in soil or water samples. The assays should be suitable for evaluation of these types of samples if comparable DNA yield, integrity and quality are obtained during sample preparation.

Initially, the intention of this study was to develop assays to distinguish pathotype 5X from pathotype 3H. However, we found that the discriminatory polymorphic region we selected for our analysis could group many other pathotypes into one of these two main clusters, as listed in Table 3.1. It is possible that the two pathotype clusters observed in this study correspond to the two genetically distinct populations of *P. brassicae* identified by Holtz et al. (2018), with the 5X and 3H clusters correlating with the "virulent" and "avirulent" populations, respectively, of Holtz et al. (2018). Additional testing will be necessary to confirm this hypothesis.

During the initial primer design stage of this study, we were limited to the whole-genome SNP profiles of the 38 SSIs in our collection. Additional pathotypes for which we did not have sequencing reads were only later classified into the clusters, based on the results of the rhPCR and SNaPshot assays. Specifically, the isolates ORCA-ss3, AbotJE-ss1, F187-14 and CDCN-ss1,

corresponding to pathotypes 5I, 6M, 8E and 8N, respectively, were tested without prior knowledge of which cluster they grouped with, as they were not originally considered nor did we have their corresponding whole-genome sequences. This consideration would also apply to any new *P. brassicae* pathotypes identified and tested in the future, as the primers were not initially designed to target their variants. If based on these two assays exclusively, clustering of new pathotypes would depend on the allelic variant in the discriminatory SNP positions, which might or might not be consistent with their CCD designation(s) based on virulence phenotype(s). Moreover, the longevity and stability of the region targeted by these particular rhPCR and SNaPshot primers is not known, since genome rearrangements or shifts in the pathogen population could occur in response to selection pressure. Similarly, mutations in the discriminating SNPs could affect the effectiveness of the assays (Liu et al., 1997; Walsh et al., 1992).

The rhPCR technology has the capacity to be optimized into a multiplex reaction for the simultaneous detection of multiple targets. A different primer pair for each target is required and amplicons must be of different lengths, as multiplexing relies on electrophoretic separation of bands (Henegariu et al., 1997). The target is identified based on the molecular weight of the electrophoretic band, and pathotype clustering is determined based on the presence or absence of that band. Since rhPCR is a PCR-based approach, the capacity to adapt rhPCR primers into a quantitative assay is an advantage of this technique. qPCR provides greater sensitivity for detection of low frequency DNA, since the initial amount of target DNA is directly correlated with an early or late exponential curve of amplification (Cao et al., 2014; Sundelin et al., 2010). A multiplex quantitative rhPCR assay would require the design of additional primer pairs and labeling of probes with distinct fluorophores for each amplicon.

The main advantage of the SNaPshot technology is its capacity to detect up to four alleles by means of fluorescent ddNTPs variants. It would therefore be ideal if a SNaPshot primer is designed against a polymorphic SNP that distinguishes four distinct pathotype clusters. In addition, SNaPshot is scalable through a multiplex reaction, where discriminatory SNPs from several different genomic regions can be examined concurrently. This would facilitate efficient and rapid testing. Differential primer lengths for each targeted SNP are required, however, since the product size of the fluorescence peak is determined by the length of the primer. Product size is measured along the x-axis, and therefore, the size of the product determines the targeted SNP and the peak color determines the allele in that corresponding SNP.

The rhPCR and SNaPshot assays in this study can only distinguish pathotypes of the 5X cluster from the 3H cluster, since the rhPCR primer pairs target only one set of allelic variants and the SNaPshot primer targets one SNP. To be able to further distinguish isolates within the clusters (ideally down to their individual CCD pathotype designations), multiple primers targeting various differential SNPs would need to be designed and multiple reactions would have to be carried out in parallel. In this case, the development of a multiplex reaction would increase efficiency. The sequencing reads of the SSIs in this study were assembled against the 2015 e3 reference genome (Schwelm et al., 2015). We are currently re-aligning the SSI sequencing reads against the 2019 re-assembled e3 reference genome (Stjelja et al., 2019). The 2019 genome is more accurate and reliable than its 2015 counterpart, since it is a more complete database as it is an improved genome assembly with longer continuous sequences (further described in Chapter 4). Moving forward, we will be using the re-aligned whole-genome SNP profiles for assay development.

Table 3.1 Single-spore isolates of *Plasmodiophora brassicae* used during rhPCR assay optimization.

Reference 5X cluster isolates	SC14 (6A); SR20 (6B); SS23 (4A); SS25 (6B); ST11 (5X); ST16 (5X); ST20 (5X); ST23 (5X); ST25 (6B); ST26 (6B); ST29 (6B); ST49 (6B); SW46 (6B)
Alternate 3H cluster isolates	S05 (3D); S16 (2A); S36 (2F); S39 (2F); SA13 (6C); SC07 (2F); SC19 (3H); SC26 (2F); SC50 (3H); SL02 (3H); SL09 (2F); SL36 (3D); SN45 (7A); SR04 (3H); SR42 (3H); SS02 (2A); SS06 (3D); SS11 (2A); SS34 (3D); SS48 (3H); ST27 (3H); ST37 (3H); ST40 (3A); SW09 (3D); SW30 (3H);

* Each single-spore isolate obtained by Askarian et al. (2021b) is given a molecular identification for its isolate name. Canadian Clubroot Differential pathotype designations (Strelkov et al., 2018; Askarian et al., 2021b) are indicated in parentheses after each isolate name.

Table 3.2 The rhPCR primer sequences designed to cluster *Plasmodiophora brassicae* pathotype DNA.

Primer Pair	Primer Name	Primer Sequence (5'-3') [c]
rh1-43812R [a]	rh1-43812Rfw	ACGACGACCCGGACACCATCGCrUAACGC/3SpC3/
	rh1-43812Rrv	TTGGCGATGGGCGCCACCrUGCGAT/3SpC3/
rh1-43812A [b]	rh1-43812Afw	GCGACGTCCCGGACACTATCGTrCAACGC/3SpC3/
	rh1-43812Arv	TTGGCGATGGTCGCCACCrGGCGAT/3SpC3/

[a] Primer pair to amplify pathotypes of the reference cluster (pathotype 5X).
[b] Primer pair to amplify pathotypes of the alternate cluster (pathotype 3H).
[c] The discriminatory SNPs are represented by the bolded alleles. The ribonucleotide residue is shown in bolded blue as represented by the "r", followed by the mismatched nucleotide five bases downstream as shown in black, and followed by the C3 spacer at the 3' end of the primers.

Table 3.3 Field and single-spore isolates of *Plasmodiophora brassicae* used to evaluate the SNaPshot and rhPCR assays.

Isolate Identification	Purity [a]	Pathotype [b]
SACAN-ss3	SSI	2F
F3-14	FI	3A
F185-14	FI	3A
F189-14	FI	3A
SACAN-ss1	SSI	3H
ORCA-ss3	SSI	5I
LG-1	FI	5X
LG-2	FI	5X
LG-3	FI	5X
AbotJE-ss1	SSI	6M
F187-14	FI	8E
CDCN-ss1	SSI	8N

[a] Samples were either field isolates (FI) or had been previously purified as single-spore isolates (SSI) as per Xue et al. (2008) or Askarian et al. (2021b).
[b] Pathotype designations are based on the Canadian Clubroot Differential set (Strelkov et al., 2018; Askarian et al., 2021b).

Table 3.4 *Plasmodiophora brassicae* isolate mixtures generated to assess the capacity of the SNaPshot assay to determine relative abundances.

Isolates [a]	Proportions	Genomic DNA (ng) [b]
SS48(3H) : ST20(5X)	10 : 90	1 : 9
	20 : 80	2 : 8
	30 : 70	3 : 7
	40 : 60	4 : 6
	50 : 50	5 : 5
	60 : 40	6 : 4
	70 : 30	7 : 3
	80 : 20	8 : 2
	90 : 10	9 : 1
F3-14(3A) : LG-1(5X)	10 : 90	1 : 9
	20 : 80	2 : 8
	30 : 70	3 : 7
	40 : 60	4 : 6
	50 : 50	5 : 5
	60 : 40	6 : 4
	70 : 30	7 : 3
	80 : 20	8 : 2
	90 : 10	9 : 1
SACAN-ss1(3H) : LG-2(5X)	10 : 90	1 : 9
	20 : 80	2 : 8
	30 : 70	3 : 7
	40 : 60	4 : 6
	50 : 50	5 : 5
	60 : 40	6 : 4
	70 : 30	7 : 3
	80 : 20	8 : 2
	90 : 10	9 : 1

[a] SS48, ST20, and SACAN-ss1 were single-spore isolates; F3-14, LG-1 and LG-2 were field isolates; pathotype designations based on the Canadian Clubroot Differential set (Strelkov et al., 2018; Askarian et al., 2021b) are indicated in parentheses.
[b] 10 ng total genomic for template generation with primers SEQ1-43778fw and SEQ1-43778rv2.

Table 3.5 Identification of blinded root galls infected by *Plasmodiophora brassicae* used in the single-blind study.

Sample #	Isolate Identification	Purity [a]	Pathotype [b]
1	SACAN-ss3 [c]	SSI	2F
2	ORCA-ss4	SSI	5I
3	AbotJE-ss1 [c]	SSI	6M
4	CDCN-ss3	SSI	8N
5	ST27 [d]	SSI	3H
6	ST40 [d]	SSI	3A
7	LG-2 [c]	FI	5X
8	F3-14 [c]	FI	3A
9	F1-14	FI	3D
10	CDCN # 4-14	FI	3H
11	SW46 [d]	SSI	6B
12	F185-14 [c]	FI	3A
13	SS23 [d]	SSI	4A
14	LG-3 [c]	FI	5X
15	SACAN-03-1	FI	3H
16	SC14 [d]	SSI	6A

[a] Samples were either field isolates (FI) or had been previously purified as single-spore isolates (SSI) as per Xue et al. (2008) or Askarian et al. (2021b).
[b] Pathotype designations are based on the Canadian Clubroot Differential set (Strelkov et al., 2018; Askarian et al., 2021b).
[c] Isolates used in earlier testing of DNA extracted from pathotyped galls (Table 3.3).
[d] SSIs used during the optimization stage of the rhPCR and SNaPshot assays (Table 3.1).

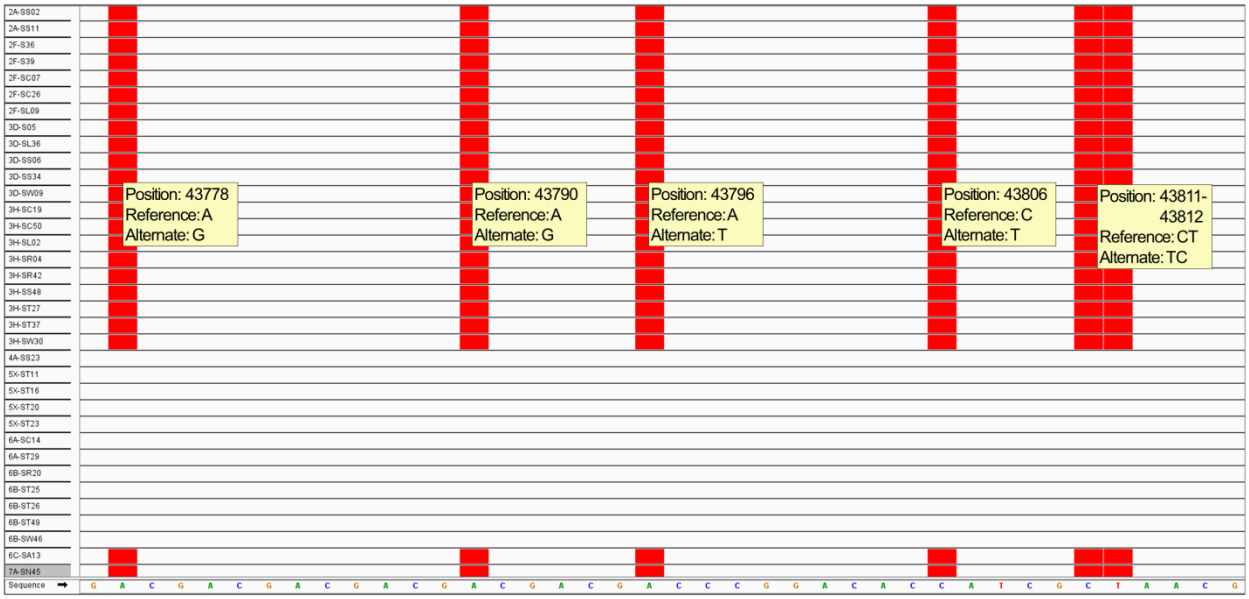

Figure 3.1 Alignment of 38 single-spore isolates of *Plasmodiophora brassicae* used in this study (Table 3.1) with the *P. brassicae* e3 reference genome (Schwelm et al., 2015) using Integrated Genomics Viewer (Robinson et al., 2011). Each row represents an individual single-spore isolate, and the corresponding isolate name is displayed in the left vertical axis. The sequence of the reference genome is shown along the bottom horizontal axis. This presented region belongs to genome coordinates 43777 to 43816 of contig 1, and provides a clear genomic differentiation between the 5X pathotype cluster and the 3H cluster. The red boxes indicate a SNP against the reference genome, and the allelic variant in the SNP position is referred to the "Alternate" allele. Isolates of pathotype 3H belong to the alternate cluster according to this particular polymorphic region. Therefore, the other isolates containing the alternate allelic variants are grouped with pathotype 3H to form the alternate 3H cluster. Likewise, isolates of pathotype 5X belong to the reference cluster. Therefore, the other isolates containing the reference allelic variants are grouped with pathotype 5X to form the reference 5X cluster.

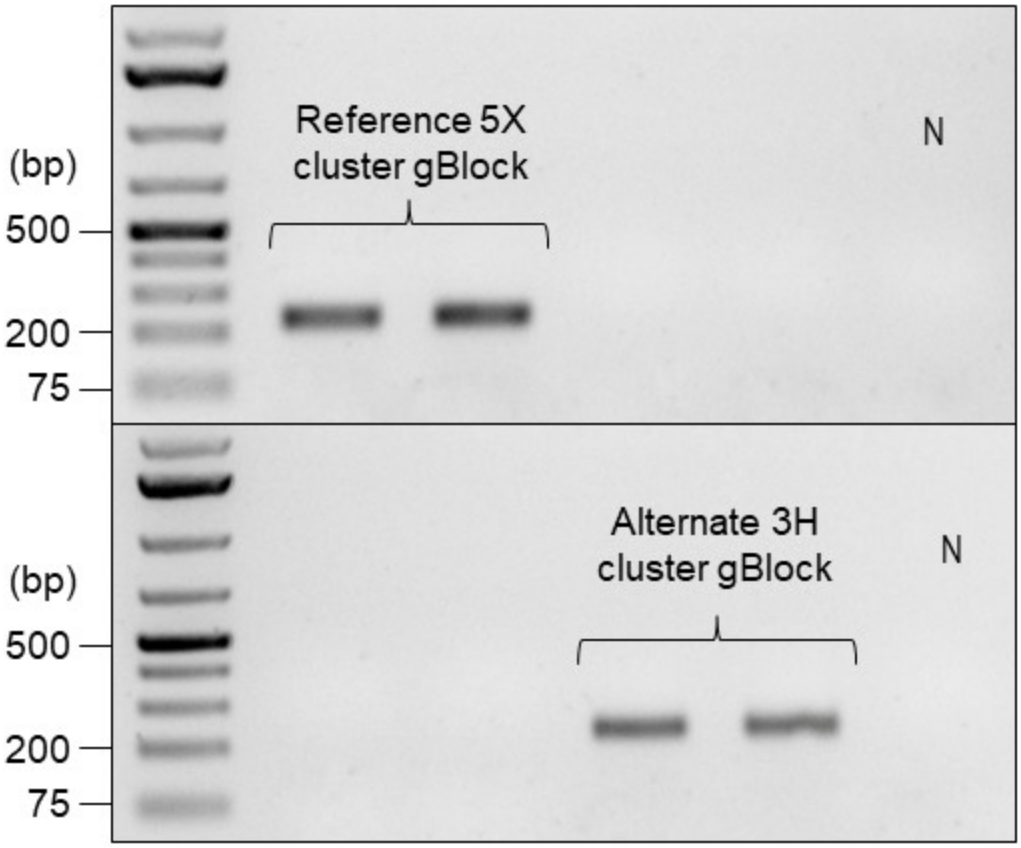

Figure 3.2 Testing of the specificity of the rhPCR assay with gBlocks. Amplicons were resolved by electrophoresis on 1% (w/v) agarose gels prepared with Tris-acetate-EDTA buffer and SYBR Safe gel stain. A GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific, Waltham, MA, USA) was included in the leftmost lane as the marker. The primer pairs rh1-43812R (upper panel) and rh1-43812A (bottom panel) were evaluated against the gBlocks. Two replicates of the reference 5X cluster gBlock and two replicates of the alternate 3H cluster gBlock were ran with each primer pair. The negative control is displayed in the rightmost lane, as represented by the N.

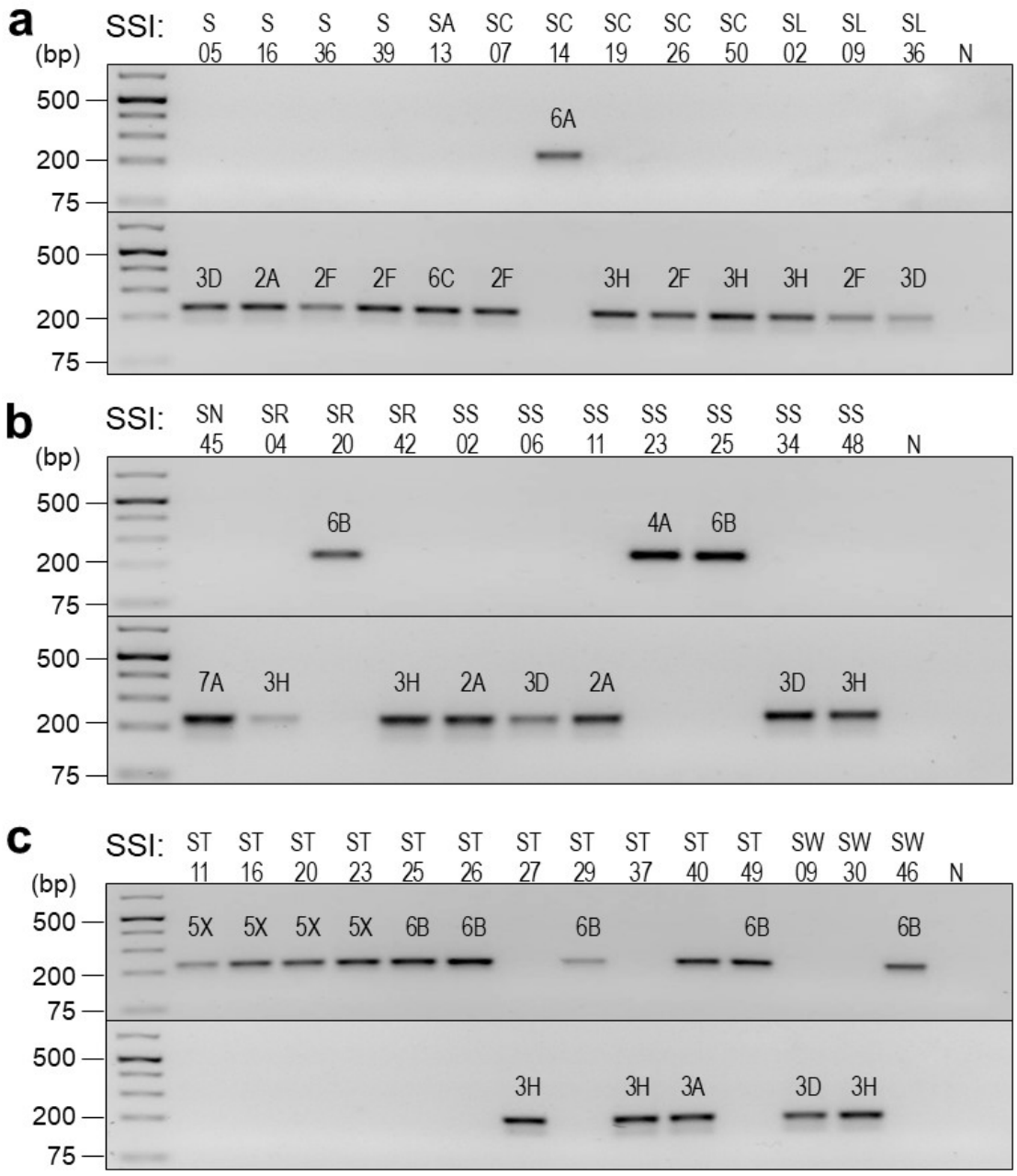

Figure 3.3 Testing of the specificity of the rhPCR assay against single-spore isolates of *Plasmodiophora brassicae*. Amplification products were resolved by electrophoresis on 1% (w/v) agarose gels prepared with Tris-acetate-EDTA buffer and SYBR Safe gel stain. A GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific, Waltham, MA, USA) was included in the leftmost lane as the marker. The primer pairs rh1-43812R (upper panels) and rh1-43812A (bottom panels) were evaluated against single-spore isolates; **a)** S05, S16, S36, S39, SA13, SC07, SC14, SC19, SC26, SC50, SL02, SL09, SL36; **b)** SN45, SR04, SR20, SR42, SS02, SS06, SS11, SS23, SS25, SS34, SS48; **c)** ST11, ST16, ST20, ST23, ST25, ST26, ST27, ST29, ST37, ST40, ST49, SW09, SW30, and SW46. The corresponding pathotype designation of each isolate is indicated above the respective amplicon. The negative control is displayed in the rightmost lane, as represented by the N.

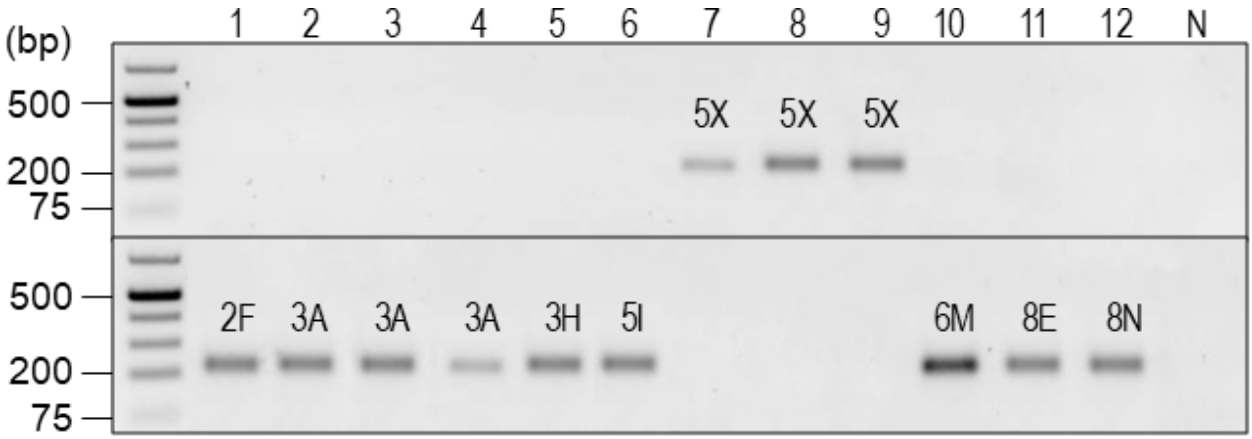

Figure 3.4 Testing of the specificity of the rhPCR assay against DNA extracted from clubroot galls resulting from infection by *Plasmodiophora brassicae* isolates representing different pathotypes. Amplification products were resolved by electrophoresis on 1% (w/v) agarose gels prepared with Tris-acetate-EDTA buffer and SYBR Safe gel stain. A GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific, Waltham, MA, USA) was included in the leftmost lane as the marker. The primer pairs rh1-43812R (upper panel) and rh1-43812A (bottom panel) were evaluated against *P. brassicae* isolates SACAN-ss3, F3-14, F185-14, F189-14, SACAN-ss1, ORCA-ss3, LG-1, LG-2, LG-3, AbotJE-ss1, F187-14, and CDCN-ss1 (lanes 1-12). The corresponding pathotype of each isolate is indicated above the respective amplicon. The negative control is displayed in the rightmost lane, as represented by the N.

Figure 3.5 The differentiating capacity of the SNaPshot assay as displayed by capillary electrophoresis of the single-spore isolates of *Plasmodiophora brassicae* listed in Table 3.1. The SNaPshot primer snpsht1-43778 was designed against a discriminatory SNP containing an allele of A for the 5X pathotype cluster and an allele of G for the 3H pathotype cluster. Allele A fluoresces green, whereas allele G fluoresces blue. Sequencing results were visualized with Peak Scanner v1.0 (Applied Biosystems, Waltham, MA, USA). The strength of the fluorescence peak is measured against the vertical axis.

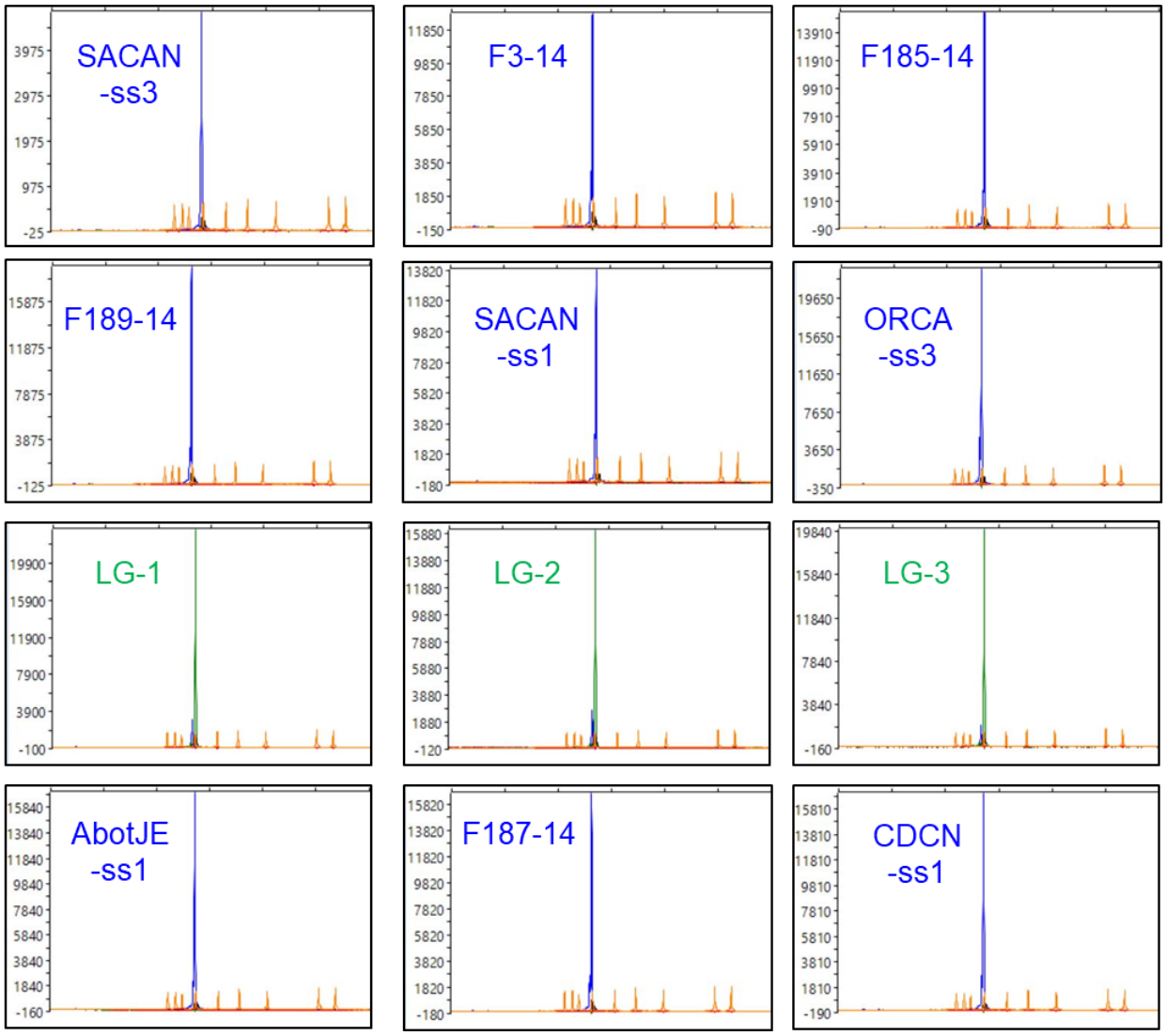

Figure 3.6 The differentiating capacity of the SNaPshot assay displayed by capillary electrophoresis of DNA extracted from isolates of *P. brassicae* from pathotyped galls as listed in Table 3.3. The SNaPshot primer snpsht1-43778 was designed against a discriminatory SNP containing an allele of A for the 5X pathotype cluster and an allele of G for the 3H pathotype cluster. Allele A fluoresces green, whereas allele G fluoresces blue. Sequencing results were visualized with Peak Scanner v1.0 (Applied Biosystems, Waltham, MA, USA). The strength of the fluorescence peak is measured against the vertical axis.

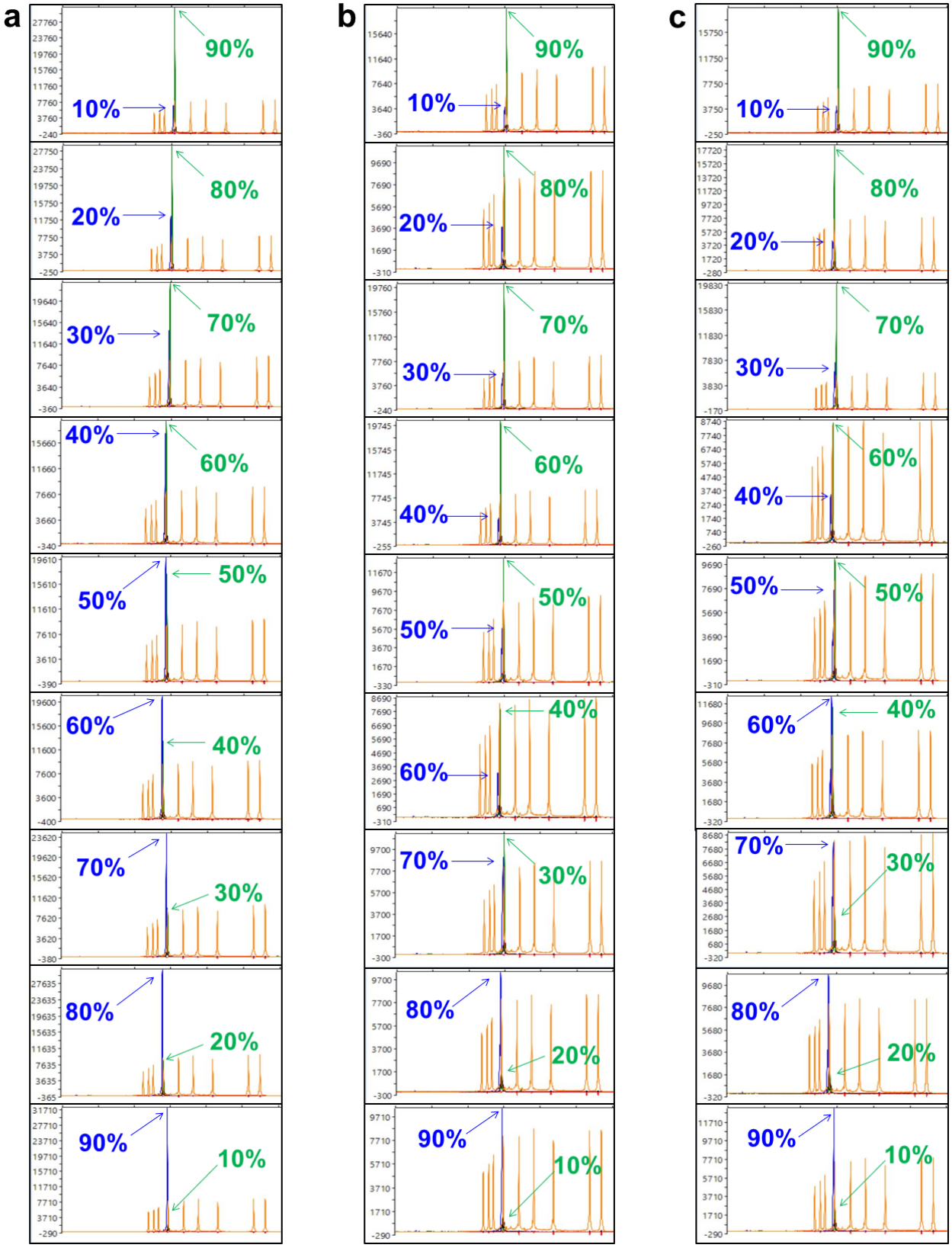
a
b
c
10%
90%
20%
80%
30%
70%
40%
60%
50%
50%
60%
40%
70%
30%
80%
20%
90%
10%

Figure 3.7 The capacity of the SNaPshot assay to determine the relative abundance of mixtures of *Plasmodiophora brassicae* isolates displayed by capillary electrophoresis. The results shown were obtained with the SNaPshot primer snpsht1-43778. Three two-isolate mixtures were evaluated at ratios of 10:90, 20:80, 30:70, 40:60, 50:50, 60:40, 70:30, 80:20, and 10:90. The graphs highlight the relative proportion of each allele, an allele of G for the 3H cluster or an allele of A for the 5X cluster. Allele G fluoresces blue, whereas allele A fluoresces green. **a)** The first mixture included single-spore isolates SS48 and ST20 from our original collection; **b)** the second mixture included the field isolates F3-14 and LG-1 extracted from root galls; and **c)** the third mixture included the single-spore isolate SACAN-ss1 and the field isolate LG-2. Sequencing results were visualized with Peak Scanner v1.0 (Applied Biosystems, Waltham, MA, USA). The strength of the fluorescence peak is measured against the vertical axis.

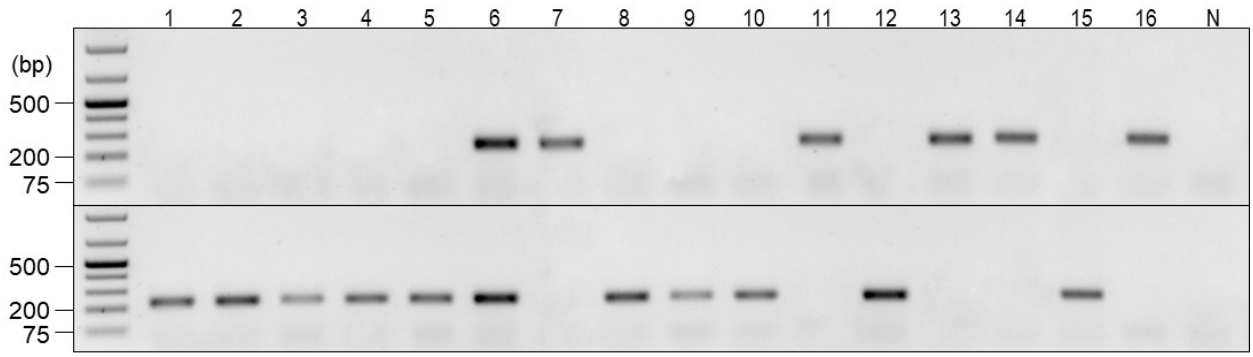

Figure 3.8 Results of a single-blind evaluation of *Plasmodiophora brassicae* field and single-spore isolates with the rhPCR assay. Amplification products were resolved by electrophoresis on a 1% (w/v) agarose gel prepared with Tris-acetate-EDTA buffer and SYBR Safe gel stain. A GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific, Waltham, MA, USA) was included in the leftmost lane as the marker. The primer pairs rh1-43812R (upper panel) and rh1-43812A (bottom panel) were used to cluster the 16 blinded galls. Samples were disclosed to be isolates SACAN-ss3, ORCA-ss4, AbotJE-ss1, CDCN-ss3, ST27, ST40, LG-2, F3-14, F1-14, CDCN #4-14, SW46, F185-14, SS23, LG-3, SACAN-03-1, and SC14 (lanes 1-16). The negative control is displayed in the rightmost lane, as represented by the N.

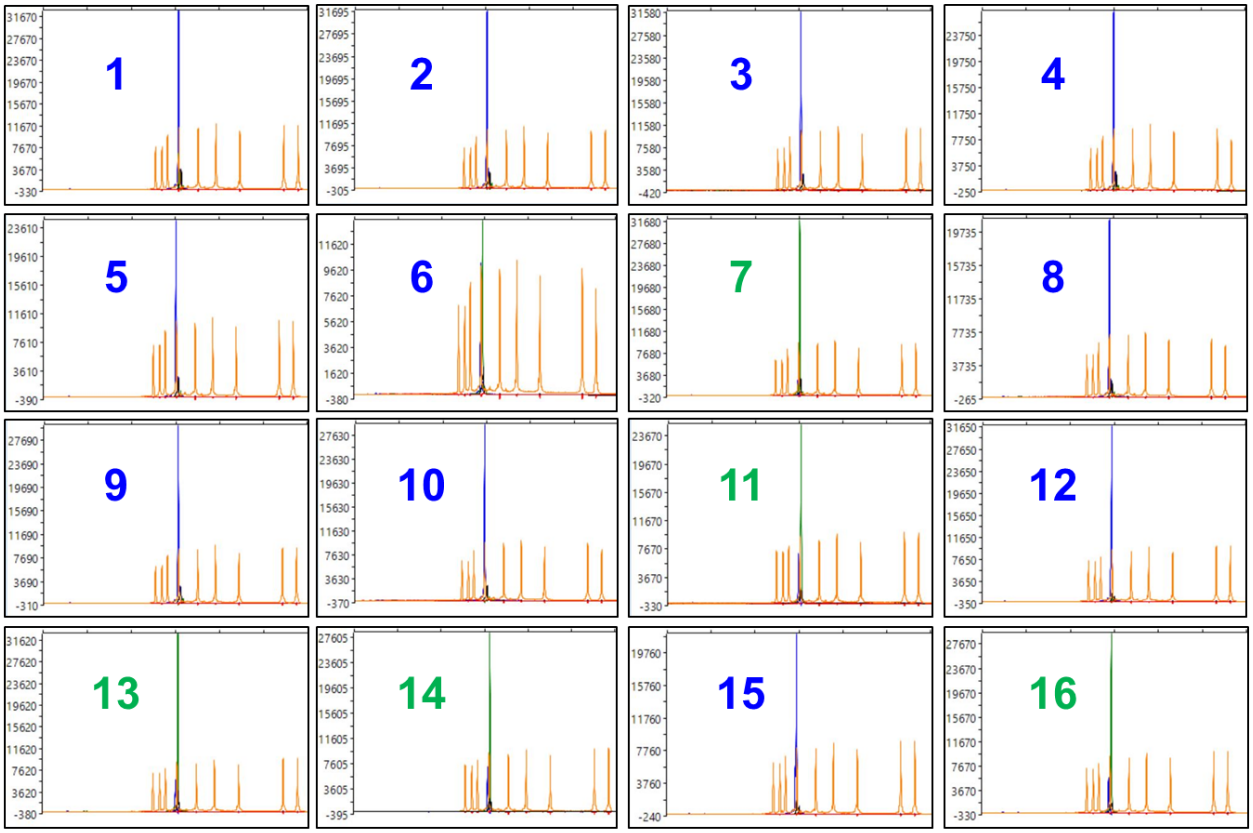

Figure 3.9 Results of the single-blind study with the SNaPshot assay displayed with capillary electrophoresis. The SNaPshot primer snpsht1-43778 was used to cluster the 16 blinded galls. The graphs highlight the discriminatory SNP, either an allele of A for the 5X cluster or an allele of G for the 3H cluster. Allele A fluoresces green, whereas allele G fluoresces blue. Samples were disclosed to be isolates SACAN-ss3, ORCA-ss4, AbotJE-ss1, CDCN-ss3, ST27, ST40, LG-2, F3-14, F1-14, CDCN #4-14, SW46, F185-14, SS23, LG-3, SACAN-03-1, and SC14. The sequencing results were visualized with Peak Scanner v1.0 (Applied Biosystems, Waltham, MA, USA).

Chapter 4 – Evaluating potential heterozygosity in *Plasmodiophora brassicae* genomes

4.1 Background and introduction

The development of molecular assays for pathotyping *P. brassicae* (Chapter 3) began with an *in silico* search of the full genome sequencing reads of our collection of 38 SSIs (Askarian et al., 2021a) to identify discriminative polymorphic regions. Comparative analysis was carried out using the Integrative Genomics Viewer (IGV) program (Robinson et al., 2011) to visualize pathotype variability, as described in Chapter 3. The identification of SNPs among pathotypes was expected; however, numerous putative heterozygous sites were found as well. This was unexpected and somewhat inexplicable, given the haploid nature of *P. brassicae* resting spores (Ingram et al., 1972); in theory, in a haploid organism, the genome should carry only one allele at each position. There were four possible explanations for this apparent heterozygosity: (1) some of the SSIs in our collection were not true SSIs, (2) there exists unexplored diploidy in the isolates, (3) natural mutations or replication errors had occurred during propagation of the isolates, or (4) there were some issues with the original reference genome assembly that we had used to align our reads.

The first hypothesis, that some of the SSIs were not true SSIs, would imply that during the initial preparation of single-spores for inoculation and propagation in a susceptible host (Askarian et al., 2021b), a second undetected resting spore may have been attached or adjacent to the selected resting spore, resulting in a heterogeneous mix (i.e., more than one genome was present in the sample). However, further investigation into the regions bearing putative heterozygosity indicated that the heterozygous positions were not exclusive to just one or a select few isolates, but instead were found among all 38 genomes. If the putative heterozygosity was due to the

presence of more than one genome during the single-spore isolation process, our samples should not have been this equally affected. Given the consistency in heterozygosity observed across all 38 isolates, the hypothesis of a pathotype mixture due to contamination was discarded.

The second hypothesis suggested that we were in fact dealing with unexplored diploidy in *P. brassicae*. While the clubroot pathogen is known to be a haploid organism having one set of chromosomes and homozygous positions (Tommerup & Ingram, 1971), it does pass through a dikaryotic phase during one stage of its life cycle. The life cycle of *P. brassicae* begins with haploid resting spores found in infested soils (Ingram et al., 1972). As described in Chapter 1, favorable conditions and the presence of host root exudates will stimulate the germination of the resting spores, producing primary zoospores. The primary zoospores penetrate and infect the cell walls of young root hairs to begin infection (Ayers, 1944). Primary plasmodia are formed within the root hairs and cleave to produce zoosporangia, yielding secondary zoospores (Dixon, 2009). The secondary zoospores are released back into the soil, from which they re-infect the host by penetrating cortical tissue in the main roots. This is the stage of its life cycle when *P. brassicae* enters a dikaryotic phase. The haploid secondary zoospores inside the root cortex develop into secondary diploid plasmodia. At the end of meiosis II, the diploid plasmodia divide into many daughter protoplasts. Each dikaryotic protoplast contains two haploid nuclei, which are eventually cleaved to produce haploid resting spores (Garber & Aist, 1979). These haploid resting spores are released back into the soil as the infected root tissues decompose (Kageyama & Asano, 2009). Thus, although short-lived, *P. brassicae* does have a diploid phase in cortical tissue as secondary plasmodia. However, the 38 SSIs in our collection were derived from resting spores (Askarian et al., 2021b), a haploid phase that consists of only homozygous positions. Hence, all later stages of the life cycle should have consisted of homozygous genetic material for

each isolate, which would have been derived from a single, original resting spore. It was unlikely, therefore, that the apparent heterozygosity of our samples reflected the occurrence of a diploid phase in *P. brassicae*.

The third hypothesis to explain the putative heterozygosity was that there had occurred natural mutations or replication errors during the propagation of the 38 SSIs of *P. brassicae* in our collection. Although each SSI is originally derived from a single resting spore, each must be increased on a susceptible host to produce sufficient resting spores for additional research and analysis. Through these cycles, the resting spores increase from just one to eventually billions (Hwang et al., 2013). The nature of *P. brassicae* as an obligate parasite means that pathogen biomass cannot be increased *in vitro*. As such, each SSI was subjected to several cycles of infection and meiotic stages, and an undetected replication error or single-point mutation could have occurred during this process. Early errors in replication could lead to different parts of the "clonal" population bearing variance in some positions. However, if the heterozygosity was due to a replication error or mutation during the propagation of the SSIs, we would not expect heterozygosity at the high rate that was observed. Moreover, we would not expect to see the same consistent level of heterozygosity throughout our 38 SSIs. Given the high level and consistency of the heterozygosity, the hypothesis involving mutations or natural replication errors was ruled out.

The final hypothesis was that the heterozygosity observed was an artefact reflecting the collapse of highly analogous reads during assembly of the original (2015) reference genome, and that these reads actually belonged to different regions of the genome. Our sequencing reads were aligned to this original *P. brassicae* e3 reference genome (European Nucleotide Archive project PRJEB8376) (Schwelm et al., 2015). This reference genome was re-characterized in 2019, which

led to a reduction in the number of contigs from 165 to 20 (European Nucleotide Archive; project PRJEB24736) (Stjelja et al., 2019). Contigs are continuous DNA sequences, and a reduction in the number of contigs with longer read lengths indicates a more complete genome and an improved *de novo* assembly. Analysis using the 2019 reference genome could help to identify any issues related to the 2015 reference genome that could have contributed to the detection of putative heterozygosity. Because of this, the objective of this study was to determine if the pre-existing alignments of the 38 SSIs in our collection to the 2015 e3 reference genome (Schwelm et al., 2015) would result into similar patterns when surveying the new 2019 e3 reference genome (Stjelja et al., 2019).

4.2 Materials and methods

4.2.1 Identification of "heterozygous" regions

Similar to the process of discovering SNPs for assay development (Chapter 3.2.1), the resulting vcf files of the 38 SSIs, along with the 2015 e3 reference genome, were loaded to IGV (Robinson et al., 2011) to visualize heterozygosity (Figure 4.1). Twenty regions displaying putative heterozygosity were selected (Table 4.1). The regions contained consistent heterozygous positions at the same coordinates of the genome for most of our 38 SSIs. The regions ranged in length from 74 to 433 nucleotides, and each region contained anywhere from 4 to 17 heterozygous positions.

4.2.2 BLASTn against the 2019 e3 reference genome

The 20 selected regions bearing putative heterozygous positions (Table 4.1) were used to search the 2019 e3 reference genome (Stjelja et al., 2019) using command line BLASTn (Altschul et al.,

1990). BLASTn is used to produce local alignments and identify similar sequences. It involves a nucleotide query sequence and a nucleotide database and, therefore, this search program can directly match oligos to analogous sections of a genome. Each of the 20 putative heterozygous regions was represented by an independent nucleotide query sequence. Query sequences used for this BLAST were configured into a FASTA file (Figure 4.2). The nucleotide database used was the 2019 e3 reference genome.

The BLASTn analysis to the 2019 e3 reference genome was conducted within the command line interface in the Linux operating system. The reference genome assembly file was downloaded from the European Nucleotide Archive (project PRJEB8376). The three required modules were loaded into the command line interface using the "module load" application, and the 2019 reference genome was turned into a BLASTable nucleotide database using the "makeblastdb" application. The BLASTn analysis was executed with our 20 nucleotide query sequences using the "blastn" application (Figure 4.3). The BLASTn output was explored to formulate an explanation of the putative heterozygosity.

4.2.3 BLASTx at NCBI

The nature of each of the 20 regions bearing putative heterozygous positions was further investigated using BLASTx performed through the NCBI online database software tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi) (Figure 4.4). This analysis translates nucleotide query sequences and matches them against protein databases. The non-redundant protein sequence database was selected as it only provides one entry per protein product and identical protein sequences are represented in a single record (Pruitt et al., 2007). We did not restrict the organism search set to *P. brassicae*, in order to broaden our search to all protein sequences recorded in the

database. All of our 20 query sequences (Figure 4.2) were entered individually into the search box and analyzed. The BLASTx output was explored to determine whether any of the identified heterozygous regions belonged to a gene sequence.

4.3 Results

4.3.1 BLASTn

The 20 regions bearing heterozygosity were subjected to a BLASTn search against the 2019 e3 reference genome. All 20 query sequences matched to more than one homozygous region of the genome, ranging from 2 to 20 significant alignments (Table 4.2). The results for query sequences 3_5630-5889, 8_6791-6864 and 16_4488-4574 are displayed (Figures 4.5-4.7).

Query sequence 3_5630-5889 aligned to two different regions (Figure 4.5). The regions belonged to contig 18 in the 2019 reference genome. Both alleles of each of the 12 heterozygous positions were accounted for and found among the two alignments. The 1st allele from all 12 heterozygous positions was found in the first alignment (Figure 4.5b), and the 2nd alleles were all found in the second alignment (Figure 4.5c).

Query sequence 8_6791-6864 aligned to three different regions (Figure 4.6). The regions belonged to contigs 18, 5 and 17 in the 2019 reference genome. Both alleles of each of the 4 heterozygous positions were accounted for and found among the three alignments. The 1st allele from all 4 heterozygous positions was found in the first alignment (Figure 4.6b), and the 2nd alleles were found in the second and third alignments (Figure 4.6c-Figure 4.6d).

Query sequence 16_4488-4574 aligned to two different regions (Figure 4.7). The regions belonged to contigs 20 and 7 in the 2019 reference genome. Both alleles from each of the 6

heterozygous positions were accounted for and were found among the two alignments. The 1st allele from all 6 heterozygous positions was found in the first alignment (Figure 4.7b), and the 2nd alleles were all found in the second alignment (Figure 4.7c).

4.3.2 BLASTx

From the 20 query sequences submitted to NCBI for the BLASTx search, 15 query sequences aligned to gene sequences. The top 10 matched genes for each query sequence are listed in Table 4.3. Query sequence 4_ 5243-5395 matched to 115 genes, the highest amount of matches for any of the 20 query sequences. Query sequences 3_5630-5889, 26_4298-4437, 44_2650-2739, 163_9184-9323, and 146_14984-15102 did not have any alignments. Fourteen query sequences matched to genes of *P. brassicae*. Although query sequence 71-1224 matched to two genes, neither of these genes belonged to the *P. brassicae* genome.

4.4 Discussion

The BLASTn analysis of our 20 query sequences (Figure 4.2) against the improved and reassembled 2019 *P. brassicae* reference genome indicated that the putative heterozygosity observed in the SSIs came from a technical assembly artefact. Since all 20 query sequences aligned to multiple homozygous regions of the 2019 reference genome, and all alleles in the "heterozygous" positions were accounted for (Table 4.2), we hypothesize that the displayed heterozygosity as seen with the 2015 reference genome was artificially generated due to the collapse of highly analogous reads that actually belonged to different regions of the genome. Heterozygosity can result within the misplaced reads in the case of a genome misassembly (Phillippy et al., 2008). In our case, we were able to find this potential error in the assembly by identifying such polymorphisms.

The BLASTx analysis revealed that 15 of our 20 query sequences aligned with protein genes. One particular trend we noticed was that several of our query sequences matched to transposons, or transposable elements (Table 4.3). These genetic elements are repetitive DNA sequences known as "jumping genes" that have the capability to move or transpose from one location to another in the genome (Gao et al., 2015). Since the BLASTn results indicated that our query sequences aligned to multiple regions of the genome, consistent with the description of transposable elements, we hypothesize that regions bearing heterozygosity correspond to transposable elements. Transposons are considered important contributors for gene and genome evolution, since transposon movement can result in mutations, alter gene expression, and induce chromosome rearrangements (Kazazian, 2004). Because of this, it may be useful to carry out an investigation of the nature of these genetic elements in the evolution of the *P. brassicae* genome.

We had originally postulated that the putative heterozygosity was due to unexplored diploidy in *P. brassicae*, rather than a reference genome assembly artefact. Initially, we considered a study to investigate the putative heterozygosity via molecular biology in parallel with our bioinformatics analysis. Our proposed methodology consisted of designing PCR primers to encompass a heterozygous region to amplify and Sanger sequence the targeted heterozygous positions. Based on the discovery of collapsed analogous regions of the genome producing artificial heterozygosity in this chapter, our proposed molecular investigation, if based on the original genome assembly, could have produced misleading results. These primers would have simultaneously annealed to the numerous analogous regions and amplified both the first and second alleles into the same position, which would again falsely indicate heterozygosity in the Sanger sequencing data. This highlights the importance of having a comprehensive reference genome with long reads in determining unrecognized genomic variation. Reliable genomic

information from high quality assemblies greatly supports the identification of accurate structural variants for dependable molecular assay development in *P. brassicae* diagnostics.

Regions bearing heterozygosity were not chosen as candidates for assay development, and therefore, did not interfere with the work described in Chapter 3. All discriminatory SNPs selected for assay development were verified via Sanger sequencing. Although this validation was adequate for the work conducted in this thesis, moving forward, we will need to re-map the raw reads of our SSIs (Table 3.1) against the new 2019 e3 genome, to ensure we are using the most up-to-date and reliable database.

Table 4.1 Regions bearing heterozygosity from the 2015 e3 *Plasmodiophora brassicae* reference genome.

Contig	Heterozygous region	Sequence length	# heterozygous positions	Heterozygous coordinates	1st allele	2nd allele
3	5630-5889	260	12	5665 5684 5722 5735 5751 5754 5777 5812 5837 5840 5841 5843	C A G C C G T C C A C G	T G A A T A A T A C T A
4	5243-5395	153	11	5298 5302 5308 5309 5330 5338 5340 5345 5349 5353 5366	A A C T T C G C A A A	G G T G C G T T C G G
8	6791-6864	74	4	6798 6826 6839 6853	A C G C	G T A T
16	4488-4574	87	6	4496 4551 4558 4560 4562 4564	T T A C T G	C C C T C T
20	2911-3018	108	8	2938 2940 2957 2976 2985 2988 2999 3006	T T C A T A T C	G C T G A G A T
26	4298-4437	140	9	4304 4330	A A	T G

				4341	A	G
				4362	T	G
				4363	T	G
				4374	A	G
				4376	A	T
				4399	T	C
				4417	G	A
28	280-600	321	4	314	C	G
				394	C	G
				420	T	A
				567	G	A
32	5615-5860	246	14	5653	T	C
				5708	G	A
				5711	A	T
				5714	A	G
				5725	G	C
				5740	G	T
				5753	C	A
				5758	T	C
				5768	C	A
				5803	A	G
				5806	T	G
				5840	T	C
				5841	T	C
				5848	C	T
33	1164-1427	264	4	1224	G	T
				1266	A	G
				1313	T	G
				1387	A	G
39	4640-4858	219	10	4716	C	T
				4719	T	C
				4720	C	T
				4725	T	A
				4753	T	A
				4761	G	C
				4790	A	G
				4800	C	T
				4806	C	T
				4811	A	T
44	2650-2739	90	7	2660	A	G
				2668	C	G
				2674	C	G
				2697	G	C
				2710	G	T
				2713	A	C
				2714	C	T
49	928-1119	192	17	938	A	G
				941	C	T
				950	C	G

				972	T	C
				977	T	C
				994	A	G
				995	T	C
				999	A	C
				1012	C	A
				1022	A	G
				1060	G	A
				1064	T	C
				1068	C	T
				1070	C	T
				1077	T	G
				1088	T	C
				1095	A	T
51	299-472	174	3	342	A	T
				388	T	C
				413	C	G
57	195811-196070	260	7	195849	C	A
				195850	C	T
				195860	A	C
				195878	T	A
				195896	A	C
				195933	A	T
				195974	T	A
63	204-307	104	13	233	G	C
				235	C	T
				236	A	G
				244	G	T
				245	G	C
				246	A	T
				253	G	C
				254	C	G
				260	T	C
				267	A	G
				268	C	G
				275	G	A
				290	G	A
71	1200-1632	433	5	1224	C	T
				1354	A	C
				1530	A	C
				1552	T	G
				1612	T	C
115	2226-2380	155	6	2288	C	T
				2294	G	A
				2311	C	T
				2325	A	T
				2326	T	C
				2345	A	G

146	14984-15102	119	12	15029	A	G
				15030	C	A
				15036	G	A
				15039	C	A
				15043	G	A
				15045	A	T
				15049	A	T
				15055	C	T
				15057	G	C
				15080	T	A
				15085	A	G
				15091	G	A
159	12020-12240	221	5	12052	T	C
				12059	G	A
				12061	C	A
				12127	C	A
				12214	T	A
163	9184-9323	140	13	9234	G	C
				9246	G	A
				9247	T	G
				9250	C	G
				9261	T	C
				9264	C	T
				9272	G	A
				9274	G	A
				9289	A	G
				9296	C	A
				9297	G	T
				9298	G	T
				9304	T	G

Table 4.2 BLASTn results of 20 selected putative heterozygous regions in *Plasmodiophora brassicae* evaluated against the 2019 e3 reference genome.

Query Sequence [a]	**# of significant alignments**	**Contig** [b]	**Coordinates** [b]
3_5630-5889	2	18	437997-438256 504478-504737
4_ 5243-5395	6	3	1281846-1281998
		2	1528844-1528996
		5	752210-752362
		15	719235-719387
		10	1271672-1271824
		4	1267105-1267257
8_6791-6864	3	18	184745-184818
		5	723426-723499
		17	128718-128791
16_4488-4574	2	20	188418-188504
		7	133553-133639
20_2911-3018	2	4	1281201-1281308
		13	884087-884194
26_4298-4437	4	17	27299-27438
		13	9428-9567
		19	15200-15339
		15	3300-3439
28_280-600	5	20	99248-99568 1437614-1437934
		6	986940-987260
		12	448614-448934
		2	50459-50779

32_5615-5860	3	13	17177-17422
		5	1518-1763
		3	5785-6030
33_1164-1427	20	15	707401-707664 687768-688031
		7	461183-461446 444083-444346 469390-469653
		1	589189-589452 602729-602992
		16	170995-171258 149419-149682
		11	316447-316710 325582-325785 334699-334962 346687-346950
		8	1115100-1115363 1103583-1103846
		4	1286366-1286629
		3	1289388-1289651
		2	1525689-1525952
		5	743535-743798 781347-781610
39_ 4640-4858	10	17	3657-3875 9243-9461 14829-15047 20415-20633
		8	1355777-1355995
		6	9460-9678
		12	4636-4854 17213-17431
		13	13847-14065
		14	1028817-1029035
44_2650-2739	2	7	464725-464814

		2	1513060-1513149
49_928-1119	5	16	860228-860419
		15	874806-874997
		18	2032-2223
		7	2117-2308
		20	7164-7355
51_299-472	6	5	775250-775423
		15	693646-693819
		10	1283918-1284091
		20	165879-166052 208428-208601
		2	1502898-1503071
57_195811-196070	8	14	743791-744050
		11	711332-711591 804438-804697
		4	1269804-1270063
		1	664006-664265
		5	208693-208951 768164-768424
		18	556160-556420
63_204-307	19	15	709994-710097 689648-689751 714014-714117
		16	147699-147802 172875-172978
		11	314727-314830 332979-333082 327462-327565 348567-348670
		8	1098880-1098983

			1116980-1117083
		7	442423-442466 471269-442423
		5	779632-779735 745410-745513
		2	1523969-1524072
		3	1291268-1291371
		1	596410-596513 604609-604712
71_1200-1632	2	19	180740-181172
		7	437099-437531
115_2226-2380	4	20	210559-210713
		9	1050766-1050920
		3	1312310-1312464
		6	21430-21584
146_14984-15102	8	3	742531-742649 961898-962016
		20	322110-322228
		19	407376-407494
		13	238126-238244
		9	250362-250480
		8	958735-958853
		7	860606-860724
159_12020-12240	4	15	716171-716391
		10	1268807-1269027
		4	1264240-1264460
		6	1250089-1250309
163_9184-9323	8	19	10815-10954 730383-730522

15	875068-875207
8	1358082-1358161 1358576-1358715
12	1243030-1243169
11	1263914-1264053
1	15262-15401

[a] Corresponding nucleotide sequences are listed in Figure 4.2.
[b] In association with the 2019 e3 reference genome (Stjelja et al., 2019).

Table 4.3 BLASTx matches for 20 selected putative heterozygous regions queried from *Plasmodiophora brassicae*.

Query Sequence [a]	# of significant alignments	Organism	Matched gene	Accession number [b]	E-Value [c]	Identity [d]
3_5630-5889	0	-	-	-	-	-
4_ 5243-5395	115	*Plasmodiophora brassicae*	hypothetical protein PBRA_001014	CEO99109.1	4e-18	35/50(70%)
			hypothetical protein PBRA_009598	CEO95066.1	2e-18	36/50(72%)
		Lasius niger	gag-pol polyprotein	KMQ89051.1	0.004	21/52(40%)
		Tilletia indica	hypothetical protein A4X13_0g5185	KAE8249516.1	0.010	21/47(45%)
			hypothetical protein CF319_g8940	KAE8216286.1	0.011	21/47(45%)
		Chlorella sorokiniana	Retrovirus-related Pol poly from transposon TNT 1-94 isoform B	PRW61567.1	0.016	20/45(44%)
		Oryza sativa	retrotransposon protein	ABA99612.1	0.021	21/49(43%)
			Putative copia-type polyprotein	AAL75752.1	0.021	21/49(43%)
			OSIGBa0134J07.9	CAH66391.1	0.021	21/49(43%)
		Blastomyces percursus	hypothetical protein ACJ73_02992	OJD25639.1	0.024	21/51(41%)
8_6791-6864	2	*Plasmodiophora brassicae*	hypothetical protein PBRA_009600	CEO95068.1	2e-07	24/24(100%)
			unnamed protein product	SPQ96181.1	0.002	20/20(100%)
16_4488-4574	105	*Plasmodiophora*	hypothetical protein PBRA_009599	CEO95067	9e-11	29/29(100%)

		brassicae	hypothetical protein PBRA_009646	CEO95380.1	1e-10	29/29(100%)
			unnamed protein product	SPQ99093.1	2e-05	22/28(79%)
			hypothetical protein PBRA_008863	CEP01920.1	2e-04	20/28(71%)
		Bacillus sp. SRB_331	hypothetical protein B5P42_31315	RAN68804.1	2e-04	19/28(68%)
		Deltaproteobacteria bacterium	hypothetical protein	MAD86319.1	2e-04	21/28(75%)
		Ceraceosorus bombacis	FOG: Transposon-encoded proteins with TYA, reverse transcriptase, integrase domains in various combinations	CEH15898.1	3e-04	19/28(68%)
		Aspergillus oryzae	RNA-directed DNA polymerase	OOO12350.1	8e-04	18/29(62%)
		Trifolium medium	RNA-directed DNA polymerase	MCH82779.1	9e-04	21/29(72%)
		Pseudogymnoascus sp. VKM F-4517	hypothetical protein V498_10608	KFY68550.1	0.001	19/28(68%)
20_2911-3018	94	*Plasmodiophora brassicae*	unnamed protein product	SPR00229.1	7e-14	32/36(89%)
		Trichogramma brassicae	unnamed protein product	CAB0040793.1	2e-06	24/36(67%)
		Vitis vinifera	Retrovirus-related Pol polyprotein from transposon RE1	RVW60793.1	6e-06	19/36(53%)
			hypothetical protein VITISV_019695	CAN70036.1	1e-05	17/33(52%)
		Pochonia chlamydosporia	retrotransposon protein, Ty1-copia subclass	XP_018136168.1	6e-06	21/34(62%)

			reverse transcriptase	XP_018137035.1	7e-06	21/34(62%)
		Trifolium medium	retrovirus-related Pol polyprotein from transposon TNT 1-94	MCH87604.1	1e-05	18/36(50%)
		Trifolium pratense	putative copia-like retrotransposable element	PNY02378.1	2e-05	21/34(62%)
			hypothetical protein L195_g016618	PNX93464.1	3e-05	18/36(50%)
		Beauveria bassiana	retrotransposon protein	XP_008602273.1	2e-05	21/34(62%)
26_4298-4437	0	-	-	-	-	-
28_280-600	4	*Plasmodiophora brassicae*	hypothetical protein PBRA_009676	CEO95668.1	1e-53	85/85(100%)
			hypothetical protein PBRA_006756	CEO98642.1	5e-44	75/86(87%)
			hypothetical protein PBRA_008430	CEP01118.1	2e-10	38/83(46%)
			hypothetical protein PBRA_009521	CEP03636.1	6e-09	39/85(46%)
32_5615-5860	2	*Plasmodiophora brassicae*	hypothetical protein PBRA_009682	CEO95674.1	3e-40	67/67(100%)
			hypothetical protein PBRA_009677	CEO95669.1	9e-34	60/67(90%)
33_1164-1427	114	*Plasmodiophora brassicae*	hypothetical protein PBRA_009683	CEO95675.1	2e-56	87/87(100%)
			hypothetical protein PBRA_007054	CEO98940.1	2e-50	77/87(89%)
			hypothetical protein PBRA_009639	CEO95373.1	8e-49	76/87(87%)
			hypothetical protein PBRA_001015	CEO99110.1	1e-48	82/87(94%)
			hypothetical protein PBRA_009598	CEO95066.1	4e-44	71/87(82%)

			hypothetical protein PBRA_009603	CEO95071.1	6e-39	62/80(78%)
			hypothetical protein PBRA_001014	CEO99109.1	2e-16	37/37(100%)
		Synchytrium endobioticum	hypothetical protein SeLEV6574_g04777	TPX43993.1	2e-15	34/87(39%)
			DNA-directed DNA polymerase	TPX36112.1	2e-11	35/86(41%)
		Bacillus sp. SRB_336	DDE-type integrase/ transposase/ recombinase	WP_142744782.1	2e-12	33/83(40%)
39_ 4640-4858	3	*Plasmodiophora brassicae*	hypothetical protein PBRA_009698	CEO95899.1	1e-40	72/72(100%)
			hypothetical protein PBRA_009702	CEO95903.1	3e-38	72/72(100%)
			hypothetical protein PBRA_009681	CEO95673.1	4e-10	30/69(43%)
44_2650-2739	0	-	-	-	-	-
49_928-1119		*Plasmodiophora brassicae*	hypothetical protein PBRA_009718	CEO96189.1	2e-38	63/63(100%)
			hypothetical protein PBRA_009700	CEO95901.1	5e-18	38/45(84%)
			hypothetical protein PBRA_009719	CEO96190.1	4e-09	27/27(100%)
			hypothetical protein PBRA_006072	CEO97958.1	3e-06	30/66(45%)
			hypothetical protein PBRA_004629	CEO95939.1	3e-04	25/63(40%)
		Didymosphaeria enalia	hypothetical protein CC78DRAFT_281401	KAF2269479.1	9e-05	22/57(39%)
		Caenorhabditis brenneri	hypothetical protein CAEBREN_02540	EGT46873.1	1e-04	24/59(41%)

		Nicrophorus vespilloides	sentrin-specific protease-like	XP_017773058.1	5e-04	21/61(34%)
		Naegleria fowleri	hypothetical protein FDP41_006800	KAF0974190.1	5e-04	22/62(35%)
		Sporormia fimetaria	cysteine proteinase	KAF2742074.1	9e-04	18/57(32%)
51_299-472	112	*Plasmodiophora brassicae*	hypothetical protein PBRA_009599	CEO95067.1	2e-05	21/29(72%)
		Microbotryum silenes-dioicae	BQ5605_C011g06414	SGY12176.1	2e-07	27/57(47%)
		Claviceps purpurea	uncharacterized protein CPUR_08783	CCE34844.1	2e-06	25/56(45%)
		Gaeumannomyces tritici	hypothetical protein GGTG_13370	XP_009229540.1	6e-06	27/55(49%)
		Tilletia caries	hypothetical protein A4X03_0g6391	KAE8251259.1	1e-05	23/45(51%)
		Tilletia laevis	hypothetical protein CF336_g7028	KAE8186358.1	1e-05	23/45(51%)
		Trametes cinnabarina	hypothetical protein BN946_scf184976.g34	CDO69315.1	5e-05	25/56(45%)
		Chaetomium globosum	hypothetical protein CHGG_03939	XP_001223153.1	7e-05	23/47(49%)
		Powellomyces hirtus	hypothetical protein PhCBS80983_g05887	TPX54573.1	2e-04	22/55(40%)
		Armillaria ostoyae	related to transposon-encoded proteins with TYA	SJL12791.1	3e-04	23/57(40%)
57_195811-196070	1	*Plasmodiophora brassicae*	hypothetical protein PBRA_008766	CEP01824.1	2e-08	27/29(93%)
63_204-307	3	*Plasmodiophora*	hypothetical protein PBRA_009602	CEO95070.1	4e-13	34/34(100%)

		brassicae	hypothetical protein PBRA_003564	CEO94751.1	3e-10	29/34(85%)
			hypothetical protein PBRA_001015	CEO99110.1	0.066	17/18(94%)
71_1200-1632	2	*Pisolithus tinctorius*	hypothetical protein M404DRAFT_133309	KIO08830.1	0.001	35/114(31%)
		Paxillus involutus	hypothetical protein PAXINDRAFT_22039	KIJ04667.1	0.013	25/73(34%)
115_2226-2380	45	*Plasmodiophora brassicae*	hypothetical protein PBRA_008592	CEP01650.1	2e-28	51/51(100%)
			hypothetical protein PBRA_002540	CEP02275.1	2e-21	41/42(98%)
			hypothetical protein PBRA_007917	CEP00183.1	6e-21	39/40(98%)
			hypothetical protein PBRA_002393	CEP02128.1	3e-11	30/35(86%)
		Saprolegnia parasitica	hypothetical protein SPRG_16639	XP_012211442.1	2e-07	24/50(48%)
			hypothetical protein SPRG_21975	XP_012212636.1	3e-06	25/50(50%)
		Aphanomyces astaci	hypothetical protein AaE_002323	KAF0772253.1	5e-04	21/52(40%)
		Rhizoclosmatium globosum	hypothetical protein BCR33DRAFT_251222	ORY43852.1	7e-04	22/47(47%)
		Chytriomyces confervae	hypothetical protein CcCBS67573_g09255	TPX57423.1	0.001	21/46(46%)
		Fragilariopsis cylindrus	hypothetical protein FRACYDRAFT_244572	OEU12314.1	0.036	17/46(37%)
146_14984-15102	0	-	-	-	-	-

159_12020-12240	111	*Plasmodiophora brassicae*	hypothetical protein PBRA_009644	CEO95378.1	8e-43	71/73(97%)
			hypothetical protein PBRA_009535	CEP03955.1	3e-41	73/73(100%)
			hypothetical protein PBRA_009602	CEO95070.1	6e-36	65/73(89%)
			hypothetical protein PBRA_009714	CEO96185.1	5e-16	43/73(59%)
		Lasius niger	retroelement pol polyprotein	KMQ87096.1	9e-14	37/71(52%)
		Apolygus lucorum	hypothetical protein GE061_01003	KAE9423402.1	5e-13	35/70(50%)
		Malus baccata	hypothetical protein C1H46_007627	TQE06758.1	9e-12	34/74(46%)
		Nicotiana attenuata	PREDICTED: uncharacterized protein LOC109220472	XP_019240489.1	2e-11	34/73(47%)
		Malus domestica	uncharacterized protein LOC114827448	XP_028965089.1	2e-11	35/74(47%)
		Dichomitus squalens	uncharacterized protein DICSQDRAFT_71809	XP_007370952.1	2e-11	37/69(54%)
163_9184-9323	0	-	-	-	-	-

[a] Corresponding nucleotide sequences are listed in Figure 4.2.
[b] Accession number refers to the unique gene identifier for a sequence record in the NCBI GenBank (https://www.ncbi.nlm.nih.gov/genbank/).
[c] The E-value, or expect value, is the measure of likeliness that sequence similarity is not by random chance. The lower the E-value, or the closer it is to zero, the more "significant" the match is (McGinnis & Madden, 2004).
[d] The identity is a percentage that describes how similar the query is to the aligned sequences (McGinnis & Madden, 2004).

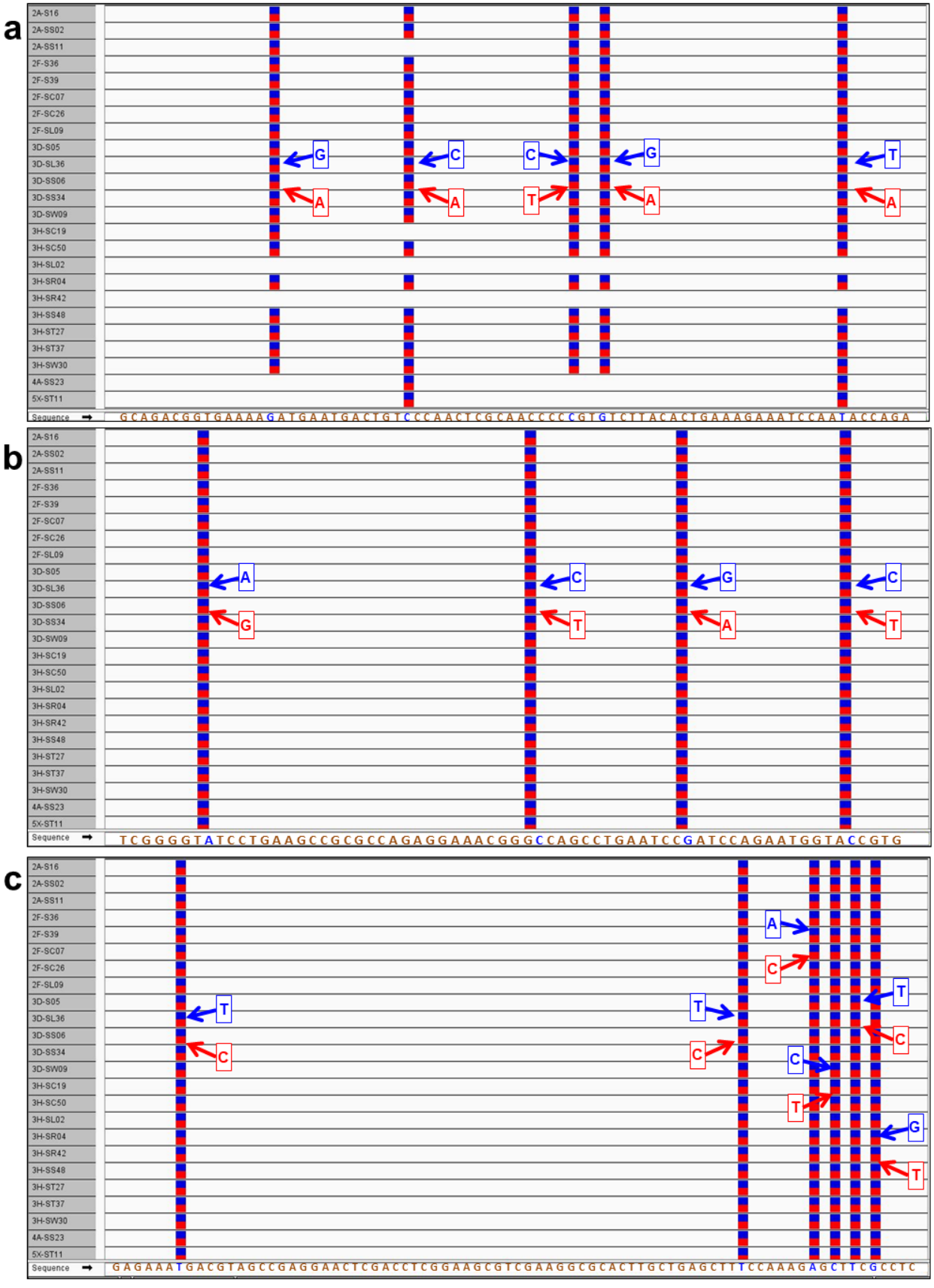
a
2A-S16
2A-SS02
2A-SS11
2F-S36
2F-S39
2F-SC07
2F-SC26
2F-SL09
3D-S05
3D-SL36
3D-SS06
3D-SS34
3D-SW09
3H-SC19
3H-SC50
3H-SL02
3H-SR04
3H-SR42
3H-SS48
3H-ST27
3H-ST37
3H-SW30
4A-SS23
5X-ST11
Sequence
GCAGACGGTGAAAAGATGAATGACTGTCCCAACTCGCAACCCCCGTGTCTTACACTGAAAGAAATCCAATACCAGA
b
2A-S16
2A-SS02
2A-SS11
2F-S36
2F-S39
2F-SC07
2F-SC26
2F-SL09
3D-S05
3D-SL36
3D-SS06
3D-SS34
3D-SW09
3H-SC19
3H-SC50
3H-SL02
3H-SR04
3H-SR42
3H-SS48
3H-ST27
3H-ST37
3H-SW30
4A-SS23
5X-ST11
Sequence
TCGGGGTATCCTGAAGCCGCGCCAGAGGAAACGGGCCAGCCTGAATCCGATCCAGAATGGTACCGTG
c
2A-S16
2A-SS02
2A-SS11
2F-S36
2F-S39
2F-SC07
2F-SC26
2F-SL09
3D-S05
3D-SL36
3D-SS06
3D-SS34
3D-SW09
3H-SC19
3H-SC50
3H-SL02
3H-SR04
3H-SR42
3H-SS48
3H-ST27
3H-ST37
3H-SW30
4A-SS23
5X-ST11
Sequence
GAGAAATGACGTAGCCGAGGAACTCGACCTCGGAAGCGTCGAAGGCGCACTTGCTGAGCTTTCCAAAGAGCTTCGCCTC

Figure 4.1 The visualization of heterozygous positions in the *Plasmodiophora brassicae* genome using Integrative Genomics Viewer (Robinson et al., 2011). The 2015 e3 reference genome (European Nucleotide Archive project PRJEB8376) (Schwelm et al., 2015) is loaded and displayed along the bottom horizontal axis. Each row represents an individual single-spore isolate with the corresponding isolate name displayed in the left vertical axis. The blue and red squares indicate that two alleles exist in that particular position. Three out of 20 of our chosen regions are displayed in this figure. **(a)** This region is found between coordinates 5630-5889 of contig 3, and consists of 12 heterozygous positions in total (only 5 displayed in this figure). **(b)** This region is found between coordinates 6791-6864 of contig 8, and consists of 4 heterozygous positions in total (all displayed). **(c)** This region is found between coordinates 4488-4574 of contig 16, and consists of 6 heterozygous positions in total (all displayed). All contigs and genomic coordinates are associated with the 2015 e3 reference genome.

>3_5630-5889
ACACAAGACCTGAGCCCAGTGCGGAAACAAAACCACAACCGAGAACCACCTGTAAAAAACCGCAAGGACCTATCGGAAGCAGACGGTGAAAAGATGAATGACTGTCCCAACTCGCAA
CCCCCGTGTCTTACACTGAAAGAAATCCAATACCAGACACCGCCCTCTCCTACCAACTACGACCACACCGACACGGACAGTAGCAGTGAGCATACAGTACCCGCCTACGTAGACAAAACC
CACCCCATCCAAATAGAGCCCAG
>4_ 5243-5395
CCTTTGACGAGGAGGCAGGCTTCGCATGGGCACCGGTCCGAACGCTTGGGAATGGAGCCATCGAACTTCATGCCACGACAGCGGTTCTGCAGGATCTGCAGACTGCAATAACCGAGAT
GGCCGATTCGTGCGTGCCACAGATCGGGAGATGCA
>8_6791-6864
TCGGGGTATCCTGAAGCCGCGCCAGAGGAAACGGGCCAGCCTGAATCCGATCCAGAATGGTACCGTGCACGATG
>16_4488-4574
GTGAGAAATGACGTAGCCGAGGAACTCGACCTCGGAAGCGTCGAAGGCGCACTTGCTGAGCTTTCCAAAGAGCTTCGCCTCGCGGAG
>20_2911-3018
CAAGTCGCGATCGATGACGAGCTGCGATCTCTGAACAAGAACCAGACCTGGGAGCTGTGCCCAAGACCAGCTGATCGAAAGGTCATCGTCGGCCGCTGGGTGTTCAAA
>26_4298-4437
GAAAAGATTCGTTTCCATTTTGACCACCCGTGATACTGAAGAAAACATTCAACTACTCAGCGCCTTCGAGGAATGGACACAATTACTATTGCTAAAATACGTAGCAGTTGCTCTAGACTG
CTCGTTTGCTCTATGATTGT
>28_280-600
CGCCAGAAGACAAGAGCGAGGACAATGATGAATCCAATGGCGATGACTCGGATTCAAAAAGCAGTCGGAAGAGGTACAAGCTTGCTCTGCCCAGCGAGTCGGAGACCCGTACCGCCG
TGGATATCGACAATCCTCCCGTGTTTGCCGCCGTCTACCGGGATGCAGCGACAAAGACGGACCGCGTGGTCATCTGTGCGGGGATGTTCGGCATTCCCGACGCCAAGGACGTCCATGTC
GATCTAGATGAGGACGACGCCCAGACCGCCATCATTTCCTATGGCTGGCCGGCTGCGTTATTCTCACCCCAGGTCATGTGCGCGA
>32_5615-5860
TGCGTTGCACGGCGGCGACGGCTTCAGGGACAGCCGTTTTGTTCAGGGCGGCGGCGCGACATTTTGACGCAGGCGGGCAACGTTGTTCTCCAAGCAATTAATGGCCAGATGAGCGGC
CGGGTCCTGAATGAAGCGGAGCATATTGACCATGGACCGCCGGGTGGCGTCTTGAATCTCCAGGCCGAGTAACATGCAGATAGTGCTTCCCAAGCACGTCCAATCGAGTTCCGGCACG
GCATGTGGCGT
>33_1164-1427
GGATGTTGTACAGTATCGCAGAGTAGGACAGCGACCAGGGCCACGTGGAGTCAGGCATGCGAGAGTGGCACCGGATTGAGCTTTCGATACCGCAGAGCGTGTAGTTGACGCTTTCAG
CGCGGCCGTTCTGCTCGGGCGTGTACGGGACATATTTGCGGTGCAGGATGCCGGACTCGTTGTAGAACAGGACGCACGCCTTGGAGTTGTACTCGCCGCCACCGTCAGTCTGGAACGT
CCGGATACCACGTCCTTCCTGGTTGCGAA
>39_ 4640-4858
ACTCCTGGGCGCGCAGGATGTTCGCTGACCCGGACGATCGCCCAGCCCTGTCGGTCCTGCTGGCCCCGGCCTACAACTATCTACGTGCCCTGTGGCTGGATGCCACCCGTGATTTCGAC
GAGATCGCTGACCACCCTACCGCAGCCCTGCAGGACGCCGTCGACAGCCTGTACCGCCACACCACTGACGTCCTTCGCGCCGTTCTCGAGGACATCGCCT
>44_2650-2739
AAAAAACGGCAACGCTCACAAGGTCTCCCACTAAAATAGAGAGAATCGTGAACCCAATCCGCTACTTTCCATCAGGTGCCCGATGAGGCG
>49_928-1119
CGATGCCTGGACGCTACCCGACCTGCTTTGCGCTCGACTCGCAGTTGTATGCGAAGATGATGCAGGATGGCAAGTACGCCTACGCGAATGTGAGAACGTGGGCCAGATGGTTCGCCCA
CGAGCCCGTGTTTGGCTATGACCTCATATTTTTCCCGATCTGTGACGCTAACCATTGGTTCCTGATGGTTGTCC
>51_299-472
TAACTTCGAAAGGCCCGAGCATTTGGACATCCAGCTTACCGCAAGGCCGGCCAGTCGCAATTGCCCCTCGACGGAGCATGACCAACTCCTCGATCTTGAAGTCGGGATGAGGTCCGCGT
TTCCGATCCGCAGCTCGTTTCATGGTGCTGACAGCACGATGCAGTTGGCGGTGTA
>57_195811-196070
CCGTAAACTTCGATGCGCTGTGCCTAGCGCATCCTGTGCCCCCCTGTCCAAAAAAAGGCTTCGAACTTTTTTGCGACCACCGACTACGAGGGCCAGTCAGCTCGACATCATATCAGGCCC
GAATTTCAAGGTCGCAACCACAGCATGGTCCCGATCCGACAGCTGACCGGCGATGTCGTCCGTCCATGTTGAGATTGTAGGGATCGAGATCGGCAACCGTGGACGCTGCTGTCAAGAG
CACACCGTGTGCGGGTCTGTCC
>63_204-307
GGTTCACCTTGATCGGTGATCTCGGCCATGGCAGAGAGCTGGACGAGAAGCGATTCTGCGAACACGACGTCGTTGTACGCCAGCAGGCCGTCGTCGGTGTGCAC
>71_1200-1632
ACGGTAACGACTGTATCGACCCCACTGGAGAGCTTTTCGACGATGAAGAAGACGTTCCCCAGCCCGAATGTATAAACTGTTTTCCGAAGGAGATGCATGGTCCAATGACGTCGAGCCCG
AGGCGGCTGAGTGGCACCCGTATGGATCGAAGGCCATTATCGCTCGGGCAGTTGAAACACATCTGGTTCATTATGGGGACCAAACTCGGCTTCGACTTGCCATCGCTCGAATCTGTCCG
TAGTACCCATCGACGCGTCTGCTTTCGAATCGGCCAGCCCACCAAACAGTTCACCTCCCCACTGGGCAATGGGTACTATGTGAACGACCTATAGCGTTGATTCAACGCGAACTATGCAAG
CCTTTTGTTCGTTCACGGCTGCACATTCATGCAGAGGACTTAGGAACCTCTGTCTGCGAGACGTACCAGGCAGCG
>115_2226-2380
TTCTGTGCCCTGATCGGCGACTACAAGTCGCTGATCCTGCTCGGTGACTGCATTCCCAAGTACGCCCCGTCCATGTTAGCCCGCACGATTGTACGCTACATGGACTGGAAGACCGGCGA
AGAAGGATCTGCACTGCTTGACGTTAATGGCGCCTC
>146_14984-15102
CGTGTTGTGGTTACTTGTTCGTTCTTGAACTAATGACGTTTGCTGACAGGTAGTTCATCGAATGCAACGTCCGGAAAGACTGTGACGTCGCAAGCGTCTCTACATAAGTACTTCGGGCC
>159_12020-12240
GTAACAGCGCAGAGCGATTTGATCTCAGCGTGTGTTGACGACAGACTAACCTTAACGGCGCCTAGGTCCTGGCGTCGCGACTGCCAGTGCACGGGACCGCCGGCCAGCAGAATGATAT
GACCGGAACGGGAACGTCGCGTGTCCGGATCCGAGGCCCAGTCGGCATCGGCGTAAGCGACCAGGCGATCCTCGAGTTGCCGTTCGTATCGAATGCCGTGCGT
>163_9184-9323
CGGCATTTTGCGCACTCCCACGGAGTTCAGTTCTGACTTAGTGAAATCAGGAAATCTGCCCAGTCGCGCTACATTCTTGGCGATACGAGAGGAACGGTGGCACATATGCCAACGGAAAC
ATCCCAGCTGCCTCTGATTTT

Figure 4.2 The FASTA file for the BLASTn search of 20 heterozygous regions from *Plasmodiophora brassicae* (Table 4.1) against the 2019 e3 reference genome (Stjelja et al., 2019). The query names consist of the contig followed by the start and end coordinates of each sequence fragment, according to the 2015 e3 reference genome. The corresponding query sequence to be analyzed is entered directly below the query name. This file was named "Hetero.Blast.fasta".

```
[htso@cedar blast]$ module load nixpkgs/16.09
[htso@cedar blast]$ module load gcc/5.4.0
[htso@cedar blast]$ module load blast+/2.6.0
[htso@cedar blast]$ makeblastdb -in OVEO01.fasta -dbtype nucl
[htso@cedar blast]$ blastn -db OVEO01.fasta -task "blastn-short" -outfmt 0 -out Hetero.NEWe3.Blast.OUT -query Hetero.Blast.fasta
```

Figure 4.3 BLASTn analysis of *Plasmodiophora brassicae* queries conducted in Linux. Each row represents one command line. In the first 3 rows, the three modules required for this analysis were loaded with the "module load" application. In the 4th row, the 2019 e3 reference genome assembly (OVEO01.fasta) was turned into a BLASTable database with the "makeblastdb" application, using the "-dbtype nucl" argument to indicate a nucleotide type database. In the 5th row, the "blastn" application was used to conduct the BLASTn search of our 20 query sequences against the 2019 reference genome. The argument "-db OVEO01.fasta" was used to select our 2019 reference genome as the database. The argument "-task "blastn-short"" was used as this task is optimized for short nucleotide sequences. The argument "-out Hetero.NEWe3.Blast.OUT" named the output file of this BLASTn search as "Hetero.NEWe3.Blast.OUT". The argument "-query Hetero.Blast.fasta" was used to select our FASTA file (Figure 4.2) as the query sequences.

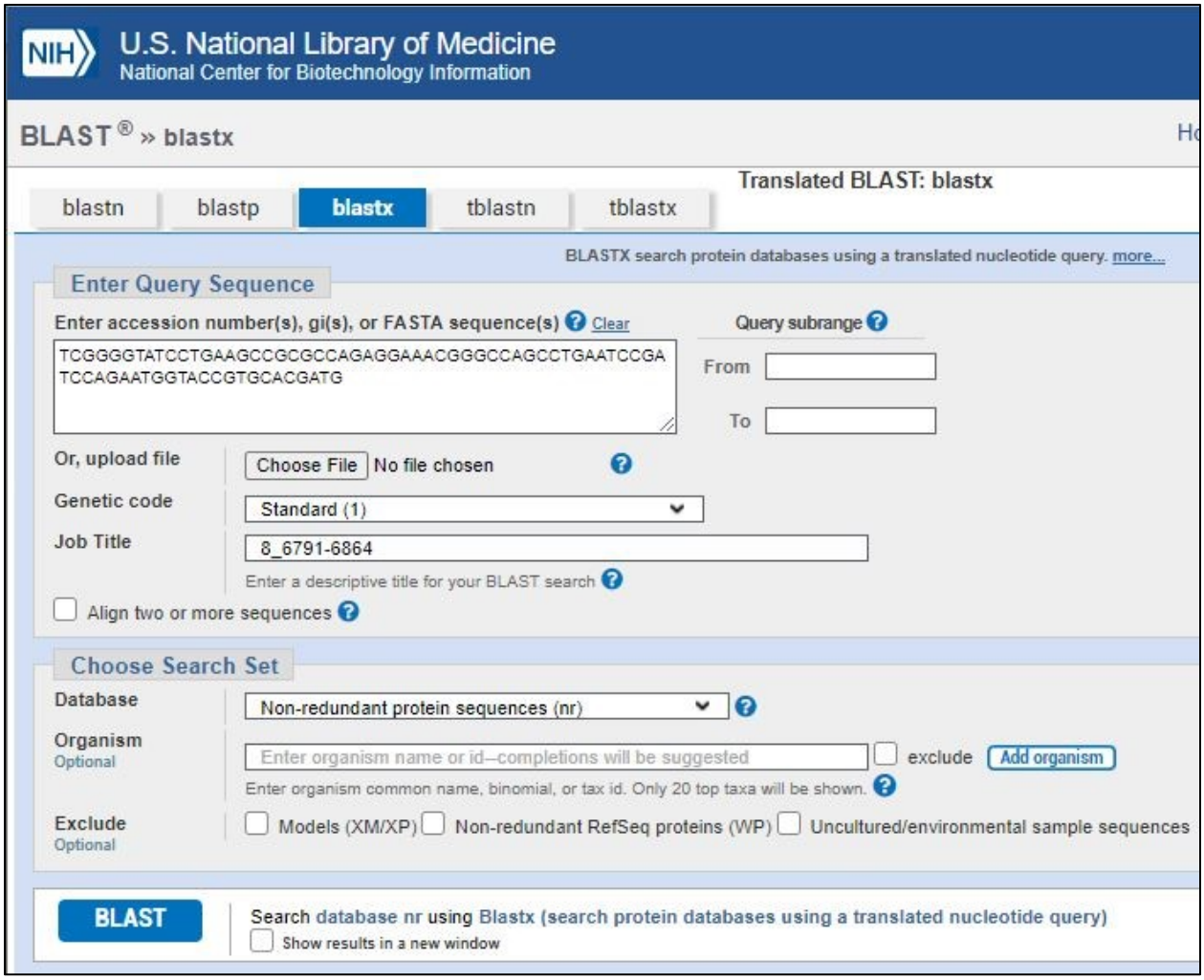

Figure 4.4 BLASTx analysis input in the National Center for Biotechnology Information (NCBI) online database software tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The BLASTx analysis was executed against the "non-redundant protein sequence" database. The entry for only one of the 20 regions is displayed in this figure. Here, the query sequence of coordinates 6791-6864 from contig 8 (job title: "8_6791-6864"), in association with the 2015 *Plasmodiophora brassicae* e3 reference genome, was entered into the search box to be translated and searched.

a

```
ACACAAGACCTGAGCCCAGTGCGGAAACAAAACCA[C/T]AACCGAGAACCACCTGTA[A/G]AAAAC
CGCAAGGACCTATCGGAAGCAGACGGTGAAAA[G/A]ATGAATGACTGT[C/A]CCAACTCGCAACCC
C[C/T]GT[G/A]TCTTACACTGAAAGAAATCCAA[T/A]ACCAGACACCGCCCTCTCCTACCAACTACGAC
CA[C/T]ACCGACACGGACAGTAGCAGTGAG[C/A]AT[A/C][C/T]A[G/A]TACCCGCCTACGTAGACAAAACCCAC
CCCATCCAAATAGAGCCCAG
```

b

```
> ENA|OVEO01000018|OVEO01000018.1 Plasmodiophora brassicae genome
assembly, contig: contig_18
Length=759975

 Score = 515 bits (260),  Expect = 3e-146
 Identities = 260/260 (100%), Gaps = 0/260 (0%)
 Strand=Plus/Minus

Query  1       ACACAAGACCTGAGCCCAGTGCGGAAACAAAACCACAACCGAGAACCACCTGTAAAAAAC  60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  438256  ACACAAGACCTGAGCCCAGTGCGGAAACAAAACCACAACCGAGAACCACCTGTAAAAAAC  438197

Query  61      CGCAAGGACCTATCGGAAGCAGACGGTGAAAAGATGAATGACTGTCCCAACTCGCAACCC  120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  438196  CGCAAGGACCTATCGGAAGCAGACGGTGAAAAGATGAATGACTGTCCCAACTCGCAACCC  438137

Query  121     CCGTGTCTTACACTGAAAGAAATCCAATACCAGACACCGCCCTCTCCTACCAACTACGAC  180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  438136  CCGTGTCTTACACTGAAAGAAATCCAATACCAGACACCGCCCTCTCCTACCAACTACGAC  438077

Query  181     CACACCGACACGGACAGTAGCAGTGAGCATACAGTACCCGCCTACGTAGACAAAACCCAC  240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  438076  CACACCGACACGGACAGTAGCAGTGAGCATACAGTACCCGCCTACGTAGACAAAACCCAC  438017

Query  241     CCCATCCAAATAGAGCCCAG  260
               ||||||||||||||||||||
Sbjct  438016  CCCATCCAAATAGAGCCCAG  437997
```

c

```
> ENA|OVEO01000018|OVEO01000018.1 Plasmodiophora brassicae genome
assembly, contig: contig_18
Length=759975

 Score = 420 bits (212),  Expect = 1e-117
 Identities = 248/260 (95%), Gaps = 0/260 (0%)
 Strand=Plus/Minus

Query  1       ACACAAGACCTGAGCCCAGTGCGGAAACAAAACCACAACCGAGAACCACCTGTAAAAAAC  60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  504737  ACACAAGACCTGAGCCCAGTGCGGAAACAAAACCATAACCGAGAACCACCTGTAGAAAAC  504678

Query  61      CGCAAGGACCTATCGGAAGCAGACGGTGAAAAGATGAATGACTGTCCCAACTCGCAACCC  120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  504677  CGCAAGGACCTATCGGAAGCAGACGGTGAAAAAATGAATGACTGTACCAACTCGCAACCC  504618

Query  121     CCGTGTCTTACACTGAAAGAAATCCAATACCAGACACCGCCCTCTCCTACCAACTACGAC  180
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  504617  CTGTATCTTACACTGAAAGAAATCCAAAACCAGACACCGCCCTCTCCTACCAACTACGAC  504558

Query  181     CACACCGACACGGACAGTAGCAGTGAGCATACAGTACCCGCCTACGTAGACAAAACCCAC  240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  504557  CATACCGACACGGACAGTAGCAGTGAGAATCTAATACCCGCCTACGTAGACAAAACCCAC  504498

Query  241     CCCATCCAAATAGAGCCCAG  260
               ||||||||||||||||||||
Sbjct  504497  CCCATCCAAATAGAGCCCAG  504478
```

Figure 4.5 BLASTn output for query sequence 3_5630-5889 against the 2019 *Plasmodiophora brassicae* e3 reference genome. **(a)** The query sequence is shown with the first alleles highlighted in blue and the second alleles in red. This sequence bearing 12 putative heterozygous positions aligned to two homozygous regions. **(b)** The first alignment belongs to contig 18, and the first alleles of all 12 heterozygous positions were found in this region. **(c)** The second alignment also belongs to contig 18, and the second alleles of all 12 heterozygous positions were found in this region.

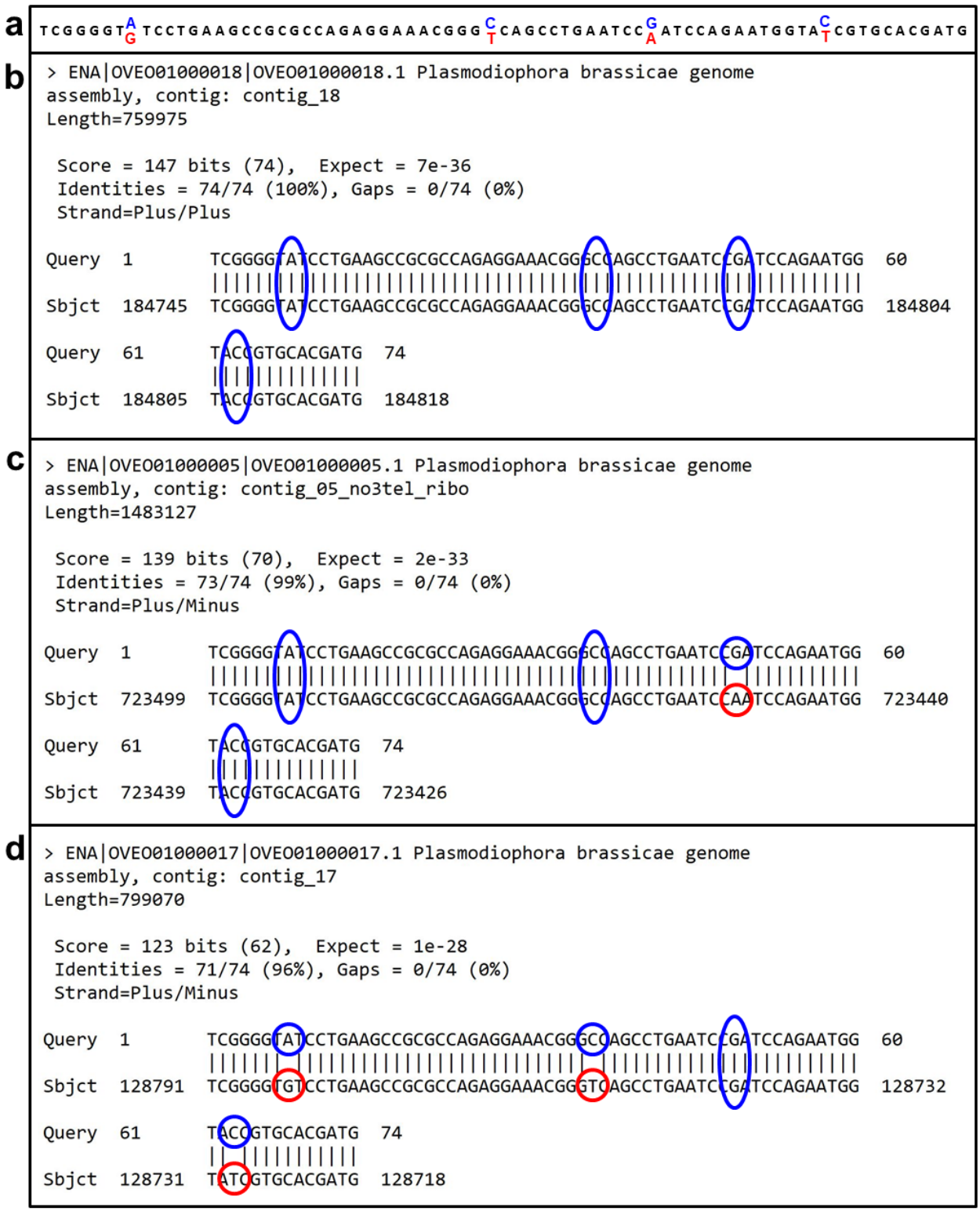
a
TCGGGGTA/G TCCTGAAGCCGCGCCAGAGGAAACGGGC/T AGCCTGAATCCG/A ATCCAGAATGGTAC/T CGTGCACGATG
b
> ENA|OVEO01000018|OVEO01000018.1 Plasmodiophora brassicae genome assembly, contig: contig_18
Length=759975
Score = 147 bits (74), Expect = 7e-36
Identities = 74/74 (100%), Gaps = 0/74 (0%)
Strand=Plus/Plus
Query 1 TCGGGGTATCCTGAAGCCGCGCCAGAGGAAACGGGCCAGCCTGAATCCGATCCAGAATGG 60
Sbjct 184745 TCGGGGTATCCTGAAGCCGCGCCAGAGGAAACGGGCCAGCCTGAATCCGATCCAGAATGG 184804
Query 61 TACCGTGCACGATG 74
Sbjct 184805 TACCGTGCACGATG 184818
c
> ENA|OVEO01000005|OVEO01000005.1 Plasmodiophora brassicae genome assembly, contig: contig_05_no3tel_ribo
Length=1483127
Score = 139 bits (70), Expect = 2e-33
Identities = 73/74 (99%), Gaps = 0/74 (0%)
Strand=Plus/Minus
Query 1 TCGGGGTATCCTGAAGCCGCGCCAGAGGAAACGGGCCAGCCTGAATCCGATCCAGAATGG 60
Sbjct 723499 TCGGGGTATCCTGAAGCCGCGCCAGAGGAAACGGGCCAGCCTGAATCCAATCCAGAATGG 723440
Query 61 TACCGTGCACGATG 74
Sbjct 723439 TACCGTGCACGATG 723426
d
> ENA|OVEO01000017|OVEO01000017.1 Plasmodiophora brassicae genome assembly, contig: contig_17
Length=799070
Score = 123 bits (62), Expect = 1e-28
Identities = 71/74 (96%), Gaps = 0/74 (0%)
Strand=Plus/Minus
Query 1 TCGGGGTATCCTGAAGCCGCGCCAGAGGAAACGGGCCAGCCTGAATCCGATCCAGAATGG 60
Sbjct 128791 TCGGGGTGTCCTGAAGCCGCGCCAGAGGAAACGGGTCAGCCTGAATCCGATCCAGAATGG 128732
Query 61 TACCGTGCACGATG 74
Sbjct 128731 TATCGTGCACGATG 128718

Figure 4.6 BLASTn output for query sequence 8_6791-6864 against the 2019 *Plasmodiophora brassicae* e3 reference genome. **(a)** The query sequence is shown with the first alleles highlighted in blue and the second alleles in red. This sequence bearing 4 putative heterozygous positions aligned to three homozygous regions. **(b)** The first alignment belongs to contig 18, and the first alleles of all 4 heterozygous positions were found in this region. **(c)** The second alignment belongs to contig 5, and 1 out of 4 of the second alleles was found in this region. **(d)** The third alignment belongs to contig 17, and 3 out of the 4 of the second alleles were found in this region.

Figure 4.7 BLASTn output for query sequence 16_4488-4574 against the 2019 *Plasmodiophora brassicae* e3 reference genome. **(a)** The query sequence is shown with the first alleles highlighted in blue and the second alleles in red. This sequence bearing 6 putative heterozygous positions aligned to two homozygous regions. **(b)** The first alignment belongs to contig 20, and the first alleles of all 6 heterozygous positions were found in this region. **(c)** The second alignment belongs to contig 7, and the second alleles of all 6 heterozygous positions were found in this region.

Chapter 5 – Conclusions

Plasmodiophora brassicae, the causal agent of clubroot disease, is one of the major contributors to economic and yield losses in canola production. Continual shifts in the virulence of *P. brassicae* populations due to the selection pressure imposed by the cultivation of CR canola cultivars make clubroot management difficult. Nineteen out of the 36 known *P. brassicae* pathotypes in Canada are virulent on CR cultivars (Fredua-Agyeman et al., 2018); these numbers are expected to increase, both in the number of virulent pathotypes and the total number of pathotypes. This underscores the diversity of *P. brassicae* virulence in existence and the urgent need for integrated and sustainable long-term management strategies. Reliable and early detection of *P. brassicae* pathotypes is one of the most effective actions to allow for informed preventative management decisions. It is important to determine the predominant pathotype in a specific field or region to allow growers to choose the most appropriate cultivar and for breeders to produce the most effective source(s) of resistance.

Current methods to detect *P. brassicae* pathotypes primarily rely on phenotypic bioassays. Although the cornerstone of clubroot diagnostics, these approaches are no longer sufficient due to the extensive time, labor and space required for bioassays and the increasing number of pathotypes. A prerequisite for improved disease surveillance and the development of novel control strategies is to identify virulent pathotypes in a fast, accurate and reliable manner. Chapter 2 provided an update on recent progress on molecular techniques for identification of *P. brassicae* pathotypes, suggested techniques that could be used in future assay development, and evaluated each described method in terms of scalability, sensitivity and accessibility. Chapter 3 described our two developed pathotyping assays, an rhPCR assay and a SNaPshot assay, which could quickly differentiate *P. brassicae* pathotypes. Lastly, Chapter 4 investigated whether our

pre-existing SSI genome sequence alignments with the old 2015 e3 reference genome would result in similar patterns as when surveying the new 2019 e3 reference genome.

5.1 *Plasmodiophora brassicae* pathotyping platforms

An increasing assortment of molecular-based methods is of value for *P. brassicae* pathotyping. The availability of and interest in molecular diagnostics techniques have grown extensively over the last few years, but still there is a long way to go in the development and application of pathotyping assays for clubroot diagnostics. Molecular techniques used in diagnostic laboratories need to be robust, reliable, inexpensive and straightforward, so that they can complement and improve upon traditional techniques. The review in Chapter 2 presented an update on molecular pathotyping assays reported to date, and provided an assessment of techniques such as amplicon length distinction, SNP-based distinction, rhPCR, and single base extension. The powerful new technique of metabarcoding with next-generation sequencing was also proposed. Each described molecular pathotyping platform has its own advantages and limitations, as the methods differ in scalability, sensitivity, accessibility and operational costs. Reductions in the time, space and labor required for pathotyping have remained as main objectives in the development of novel platforms. The rhPCR and SNaPshot technologies discussed in Chapter 2 were explored in Chapter 3 with the development of assays specific for *P. brassicae*.

5.2 Development of rhPCR and SNaPshot assays

Chapter 3 describes the development of two independent rapid and sensitive technologies for *P. brassicae* pathotyping, an rhPCR and a SNaPshot assay. The high-throughput potential and accuracy of both assays makes them promising as SNP-based pathotype identification tools for routine testing of *P. brassicae* pathotypes. As discussed in Chapter 2, rhPCR is a highly sensitive

approach that can be optimized into a quantitative assay, while the main advantages of SNaPshot are its ability to multiplex samples and alleles in a single reaction and the detection of up to four allelic variants. To our knowledge, this is the first report of an rhPCR assay for the detection of *P. brassicae* pathotype clusters as classified by the CCD set, and the first single-base extension assay for the purpose of *P. brassicae* pathotyping. The rhPCR assay described here is currently being used in our research group to identify pathotypes 5X from 3H in controlled greenhouse trials where only characterized isolates of these two pathotypes are used.

5.3 *de novo* assemblies of reference genomes

The sequencing reads from the 38 SSIs used for development of the rhPCR and SNaPshot assays were aligned to the 2015 e3 reference genome (Schwelm et al., 2015) and these reads were used throughout the duration of this thesis. We had noticed that polymorphic patterns identified with the 2015 e3 reference genome were not in agreement when surveying the new 2019 e3 reference genome (Stjelja et al., 2019). Chapter 4 describes the *in silico* bioinformatics analysis of 20 specifically chosen regions that contained putative heterozygosity. The BLAST results indicated that none of the putative heterozygous positions identified with the 2015 reference genome existed in the 2019 reference genome. Each sequence bearing putative heterozygous positions matched various distinct regions of the 2019 reference genome, and these various regions contained only homozygous positions. We hypothesize that regions of nearly identical DNA sequences were mistakenly identified as the same and inaccurately collapsed during the generation of the 2015 reference genome, creating the artificial "heterozygosity" we had originally witnessed. The *de novo* assembly sequencing error presented in the 2015 e3 reference genome was likely corrected during the assembly of the 2019 reference genome. Based on the results of this bioinformatics analysis, we believe that the 2019 *de novo* assembly is a more

complete genome and a more reliable source relative to its 2015 counterpart. Moving forward, we are refining genomic assemblies for our 38 SSIs against the 2019 e3 reference, which will provide us a more accurate pool of discriminative polymorphic regions to select from for future pathotyping assays.

5.4 Future research directions

The research in this thesis provides a good foundation for further analysis of *P. brassicae* pathotyping for clubroot diagnostics. Several follow up studies could immediately stem from this work. The rhPCR primers described in Chapter 3 can be optimized into a quantitative assay with real-time PCR to determine the abundance of each pathotype cluster. Both rhPCR and SNaPshot technologies can be multiplexed to simultaneously target multiple polymorphic regions in a single reaction. This would greatly increase throughput and efficiency. While metabarcoding technologies have not yet been established for clubroot, this approach has potential for *P. brassicae* pathotyping (Chapter 2) and its preliminary evaluation is underway in our research group. Our investigation of metabarcoding will rely on sequencing reads of the 2019 reference genome assembly, to ensure we are using the most accurate and up- to- date database.

The work described in this thesis focused primarily on the differentiation of the pathotype 5X cluster from the pathotype 3H cluster. Pathotype 5X was chosen as it was the first of the resistance-breaking pathotypes to be identified, and is the most characterized virulent pathotype. Pathotype 3H was selected as it is one of the predominant pathotypes in Alberta, and does not overcome the resistance in CR canola. Moving forward, an emphasis should be placed on developing assays that differentiate between other pathotypes and pathotype clusters. Pathotype 3A in particular is of high importance, given its widespread occurrence and high virulence on CR

canola cultivars. The development of pathotyping assays relies heavily on the availability of specific molecular markers for each unique pathotype. The demand for SNP-based pathotyping technologies will continue to grow as additional differentiating SNPs are discovered and validated.

The common goals of increasing production and sustaining economic security are recognized within all sectors of the canola industry. As with every disease, an important criterion that contributes to its economic, ecological, and social consequence is the virulence of the causal agent. Due to the emergence of new virulent pathotypes of *P. brassicae* and the decreased effectiveness of resistant cultivars, clubroot management has increased in complexity. Efforts to standardize a comprehensive diagnostics system will progress in parallel with the generation of improved *P. brassicae* reference databases. Clubroot researchers in search of molecular markers for pathotype detection will make their share of contributions towards the development of additional *P. brassicae* diagnostic tools, as rapid molecular pathotyping assays are dependent on sequence variations and polymorphic regions. Overall, the technique of choice for a rapid molecular pathotyping tool should be accessible by clubroot diagnostic laboratories.

References

Abdelfattah, A., Malacrinò, A., Wisniewski, M., Cacciola, S. O., & Schena, L. (2018). Metabarcoding: A powerful tool to investigate microbial communities and shape future plant protection strategies. *Biological Control*, *120*, 1–10. https://doi.org/10.1016/j.biocontrol.2017.07.009

Al Rwahnih, M., Daubert, S., Golino, D., Islas, C., & Rowhani, A. (2015). Comparison of Next-Generation Sequencing Versus Biological Indexing for the Optimal Detection of Viral Pathogens in Grapevine. *Phytopathology®*, *105*(6), 758–763. https://doi.org/10.1094/PHYTO-06-14-0165-R

Al-Daoud, F., Gossen, B. D., Robson, J., & McDonald, M. R. (2016). Propidium Monoazide Improves Quantification of Resting Spores of *Plasmodiophora brassicae* with qPCR. *Plant Disease*, *101*(3), 442–447. https://doi.org/10.1094/PDIS-05-16-0715-RE

Al-Daoud, F., Moran, M., Gossen, B. D., & McDonald, M. R. (2018). First report of clubroot (*Plasmodiophora brassicae*) on canola in Ontario. *Canadian Journal of Plant Pathology*, *40*(1), 96–99. https://doi.org/10.1080/07060661.2017.1393696

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Askarian, H., Akhavan, A., Galindo-González, L., Hwang, S. F., & Strelkov, S. E. (2021a). Genetic structure of *Plasmodiophora brassicae* populations virulent on clubroot resistant canola (*Brassica napus*). *Plant Disease*. https://doi.org/10.1094/PDIS-09-20-1980-RE

Askarian, H., Akhavan, A., Manolii, V. P., Cao, T., Hwang, S. F., & Strelkov, S. E. (2021b). Virulence Spectrum of Single-Spore and Field Isolates of *Plasmodiophora brassicae*

Able to Overcome Resistance in Canola (*Brassica napus*). *Plant Disease*, *105*(1), 43–52. https://doi.org/10.1094/PDIS-03-20-0471-RE

Ayers, G. W. (1944). Studies on the life history of the club root organism, *Plasmodiophora brassicae*. *Canadian Journal of Research*, *22c*(4), 143–149. https://doi.org/10.1139/cjr44c-012

Balme-Sinibaldi, V., Tribodet, M., Croizat, F., Lefeuvre, P., Kerlan, C., & Jacquot, E. (2006). Improvement of *Potato virus Y* (PVY) detection and quantitation using PVY^{N}- and PVY^{O}-specific real-time RT-PCR assays. *Journal of Virological Methods*, *134*(1), 261–266. https://doi.org/10.1016/j.jviromet.2006.01.019

Baloğlu, B., Chen, Z., Elbrecht, V., Braukmann, T., MacDonald, S., & Steinke, D. (2021). A workflow for accurate metabarcoding using nanopore MinION sequencing. *Methods in Ecology and Evolution*, *n/a*(n/a). https://doi.org/10.1111/2041-210X.13561

Banchi, E., Ametrano, C. G., Stanković, D., Verardo, P., Moretti, O., Gabrielli, F., Lazzarin, S., Borney, M. F., Tassan, F., Tretiach, M., Pallavicini, A., & Muggia, L. (2018). DNA metabarcoding uncovers fungal diversity of mixed airborne samples in Italy. *PLOS ONE*, *13*(3), e0194489. https://doi.org/10.1371/journal.pone.0194489

Braselton, J. P. (1995). Current Status of the Plasmodiophorids. *Critical Reviews in Microbiology*, *21*(4), 263–275. https://doi.org/10.3109/10408419509113543

Buczacki, S. T., & Moxham, S. E. (1979). A triple stain for differentiating resin-embedded sections of *Plasmodiophora brassicae* in host tissues under the light microscope. *Transactions of the British Mycological Society (UK)*. https://agris-fao-org.login.ezproxy.library.ualberta.ca/agris-search/search.do?recordID=XE8000760

Buczacki, S. T., & Moxham, S. E. (1983). Structure of the resting spore wall of *Plasmodiophora brassicae* revealed by electron microscopy and chemical digestion. *Transactions of the British Mycological Society*, *81*(2), 221–231. https://doi.org/10.1016/S0007-1536(83)80073-4

Buczacki, S. T., Toxopeus, H., Mattusch, P., Johnston, T. D., Dixon, G. R., & Hobolth, L. A. (1975). Study of physiologic specialization in *Plasmodiophora brassicae*: Proposals for attempted rationalization through an international approach. *Transactions of the British Mycological Society*, *65*(2), 295–303. https://doi.org/10.1016/S0007-1536(75)80013-1

Canola Council of Canada. (2021a). *Canola industry in Canada, from farm to global markets*. The Canola Council of Canada. https://www.canolacouncil.org/about-canola/industry/

Canola Council of Canada. (2021b). *Clubroot Disease | Canola Encyclopedia*. The Canola Council of Canada. https://www.canolacouncil.org/canola-encyclopedia/diseases/clubroot/

Cao, T., Rennie, D. C., Manolii, V. P., Hwang, S. F., Falak, I., & Strelkov, S. E. (2014). Quantifying resistance to *Plasmodiophora brassicae* in Brassica hosts. *Plant Pathology*, *63*(3), 715–726. https://doi.org/10.1111/ppa.12113

Cao, T., Tewari, J. P., & Strelkov, S. E. (2007). Molecular Detection of *Plasmodiophora brassicae*, Causal Agent of Clubroot of Crucifers, in Plant and Soil. *Plant Disease*, *91*(1), 80–87. https://doi.org/10.1094/PD-91-0080

Castaño, C., Berlin, A., Durling, M. B., Ihrmark, K., Lindahl, B. D., Stenlid, J., Clemmensen, K. E., & Olson, Å. (2020). Optimized metabarcoding with Pacific biosciences enables semi-quantitative analysis of fungal communities. *New Phytologist*, *228*(3). https://doi.org/10.1111/nph.16731

CD Genomics. (2021). *SNaPshot Multiplex System for SNP Genotyping – CD Genomics*. https://www.cd-genomics.com/SNaPshot.html

Chagné, D., Batley, J., Edwards, D., & Forster, J. W. (2007). Single Nucleotide Polymorphism Genotyping in Plants. In N. C. Oraguzie, E. H. A. Rikkerink, S. E. Gardiner, & H. N. De Silva (Eds.), *Association Mapping in Plants* (pp. 77–94). Springer. https://doi.org/10.1007/978-0-387-36011-9_5

Chapara, V., Kalwar, N., Lubenow, L., & Chirumamilla, A. (2019). Prevalence of Clubroot on Canola in North Dakota. *Journal of Agronomy & Agricultural Science*, *2*, 008. https://doi.org/10.24966/AAS-8292/100008

Choudhary, P., Singh, B. N., Chakdar, H., & Saxena, A. K. (2021). DNA barcoding of phytopathogens for disease diagnostics and bio-surveillance. *World Journal of Microbiology and Biotechnology*, *37*(3), 54. https://doi.org/10.1007/s11274-021-03019-0

Cobo-Díaz, J. F., Baroncelli, R., Le Floch, G., & Picot, A. (2019). A novel metabarcoding approach to investigate *Fusarium* species composition in soil and plant samples. *FEMS Microbiology Ecology*, *95*(fiz084). https://doi.org/10.1093/femsec/fiz084

Colhoun, J. (1957). A Technique for Examining Soil for the Presence of *Plasmodiophora brassicae* Woron. *Annals of Applied Biology*, *45*(3), 559–565. https://doi.org/10.1111/j.1744-7348.1957.tb05895.x

Da Lio, D., Cobo-Díaz, J. F., Masson, C., Chalopin, M., Kebe, D., Giraud, M., Verhaeghe, A., Nodet, P., Sarrocco, S., Le Floch, G., & Baroncelli, R. (2018). Combined Metabarcoding and Multi-locus approach for Genetic characterization of Colletotrichum species associated with common walnut (*Juglans regia*) anthracnose in France. *Scientific Reports*, *8*(1), 10765. https://doi.org/10.1038/s41598-018-29027-z

Deora, A., Gossen, B. D., Amirsadeghi, S., & McDonald, M. R. (2015). A Multiplex qPCR Assay for Detection and Quantification of *Plasmodiophora brassicae* in Soil. *Plant Disease*, *99*(7), 1002–1009. https://doi.org/10.1094/PDIS-06-14-0608-RE

Deora, A., Gossen, B. D., & McDonald, M. R. (2013). Cytology of infection, development and expression of resistance to *Plasmodiophora brassicae* in canola. *Annals of Applied Biology*, *163*(1), 56–71. https://doi.org/10.1111/aab.12033

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–498. https://doi.org/10.1038/ng.806

Dixon, G. R. (2009). The Occurrence and Economic Impact of *Plasmodiophora brassicae* and Clubroot Disease. *Journal of Plant Growth Regulation*, *28*(3), 194–202. https://doi.org/10.1007/s00344-009-9090-y

Dobosy, J. R., Rose, S. D., Beltz, K. R., Rupp, S. M., Powers, K. M., Behlke, M. A., & Walder, J. A. (2011). RNase H-dependent PCR (rhPCR): Improved specificity and single nucleotide polymorphism detection using blocked cleavable primers. *BMC Biotechnology*, *11*(1), 80. https://doi.org/10.1186/1472-6750-11-80

Donald, C., & Porter, I. (2009). Integrated Control of Clubroot. *Journal of Plant Growth Regulation*, *28*(3), 289. https://doi.org/10.1007/s00344-009-9094-7

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *PHYTOCHEMICAL BULLETIN*, Article RESEARCH. https://worldveg.tind.io/record/33886

Egamberdiev, S. S., Salahutdinov, I. B., Abdullaev, A. A., Ulloa, M., Saha, S., Radjapov, F., Mullaohunov, B., Mansurov, D., Jenkins, J. N., & Abdurakhmonov, I. Y. (2014). Detection of *Fusarium oxysporum* f. Sp. *vasinfectum* race 3 by single-base extension method and allele-specific polymerase chain reaction. *Canadian Journal of Plant Pathology*, *36*(2), 216–223. https://doi.org/10.1080/07060661.2014.905496

Erdogan, O., Nemli, S., Oncu, T., & Tanyolac, B. (2013). Genetic variation among pathotypes of *Verticillium dahliae* Kleb. From cotton in western Turkey revealed by AFLP. *Canadian Journal of Plant Pathology*, *35*(3), 354–362. https://doi.org/10.1080/07060661.2013.809790

Ernst, T. W., Kher, S., Stanton, D., Rennie, D. C., Hwang, S. F., & Strelkov, S. E. (2019). *Plasmodiophora brassicae* resting spore dynamics in clubroot resistant canola (*Brassica napus*) cropping systems. *Plant Pathology*, *68*(2), 399–408. https://doi.org/10.1111/ppa.12949

Faggian, R., Bulman, S. R., Lawrie, A. C., & Porter, I. J. (1999). Specific Polymerase Chain Reaction Primers for the Detection of *Plasmodiophora brassicae* in Soil and Water. *Phytopathology®*, *89*(5), 392–397. https://doi.org/10.1094/PHYTO.1999.89.5.392

Faggian, R., & Strelkov, S. E. (2009). Detection and Measurement of *Plasmodiophora brassicae*. *Journal of Plant Growth Regulation*, *28*(3), 282–288. https://doi.org/10.1007/s00344-009-9092-9

Fredua-Agyeman, R., Hwang, S. F., Strelkov, S. E., Zhou, Q., & Feindel, D. (2018). Potential loss of clubroot resistance genes from donor parent *Brassica rapa* subsp. *rapifera* (ECD 04) during doubled haploid production. *Plant Pathology*, *67*(4), 892–901. https://doi.org/10.1111/ppa.12816

Friberg, H., Lagerlöf, J., & Rämert, B. (2005). Germination of *Plasmodiophora brassicae* resting spores stimulated by a non-host plant. *European Journal of Plant Pathology*, *113*(3), 275. https://doi.org/10.1007/s10658-005-2797-0

Fu, H., Yang, Y., Mishra, V., Zhou, Q., Zuzak, K., Feindel, D., Harding, M. W., & Feng, J. (2019). Most *Plasmodiophora brassicae* Populations in Single Canola Root Galls from Alberta Fields are Mixtures of Multiple Strains. *Plant Disease*, *104*(1), 116–120. https://doi.org/10.1094/PDIS-06-19-1235-RE

Gao, D., Jiang, N., Wing, R. A., Jiang, J., & Jackson, S. A. (2015). Transposons play an important role in the evolution and diversification of centromeres among closely related species. *Frontiers in Plant Science*, *6*, 216. https://doi.org/10.3389/fpls.2015.00216

Garber, R. C., & Aist, J. R. (1979). The ultrastructure of meiosis in *Plasmodiophora brassicae* (Plasmodiophorales). *Canadian Journal of Botany*, *57*(22), 2509–2518. https://doi.org/10.1139/b79-297

Gardiner, L.-J., Brabbs, T., Akhunov, A., Jordan, K., Budak, H., Richmond, T., Singh, S., Catchpole, L., Akhunov, E., & Hall, A. (2019). Integrating genomic resources to present full gene and putative promoter capture probe sets for bread wheat. *GigaScience*, *8*(giz018). https://doi.org/10.1093/gigascience/giz018

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *ArXiv:1207.3907 [q-Bio]*. http://arxiv.org/abs/1207.3907

Geisen, S., Laros, I., Vizcaíno, A., Bonkowski, M., & Groot, G. A. de. (2015). Not all are free-living: High-throughput DNA metabarcoding reveals a diverse community of protists parasitizing soil metazoa. *Molecular Ecology*, *24*(17), 4556–4569. https://doi.org/10.1111/mec.13238

Harding, M. W., Hill, T. B., Yang, Y., Daniels, G. C., Hwang, S. F., Strelkov, S. E., Howard, R. J., & Feng, J. (2019). An Improved Evans Blue Staining Method for Consistent, Accurate Assessment of *Plasmodiophora brassicae* Resting Spore Viability. *Plant Disease*, *103*(9), 2330–2336. https://doi.org/10.1094/PDIS-05-18-0855-RE

Head, S. R., Komori, H. K., LaMere, S. A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D. R., & Ordoukhanian, P. (2014). Library construction for next-generation sequencing: Overviews and challenges. *BioTechniques*, *56*(2), 61–77. https://doi.org/10.2144/000114133

Henegariu, O., Heerema, N. a., Dlouhy, S. r., Vance, G. h., & Vogt, P. h. (1997). Multiplex PCR: Critical Parameters and Step-by-Step Protocol. *BioTechniques*, *23*(3), 504–511. https://doi.org/10.2144/97233rr01

Henry, I. M., Nagalakshmi, U., Lieberman, M. C., Ngo, K. J., Krasileva, K. V., Vasquez-Gross, H., Akhunova, A., Akhunov, E., Dubcovsky, J., Tai, T. H., & Comai, L. (2014). Efficient Genome-Wide Detection and Cataloging of EMS-Induced Mutations Using Exome Capture and Next-Generation Sequencing. *The Plant Cell*, *26*(4), 1382–1397. https://doi.org/10.1105/tpc.113.121590

Hindson, B. J., Ness, K. D., Masquelier, D. A., Belgrader, P., Heredia, N. J., Makarewicz, A. J., Bright, I. J., Lucero, M. Y., Hiddessen, A. L., Legler, T. C., Kitano, T. K., Hodel, M. R., Petersen, J. F., Wyatt, P. W., Steenblock, E. R., Shah, P. H., Bousse, L. J., Troup, C. B., Mellen, J. C., … Colston, B. W. (2011). High-Throughput Droplet Digital PCR System for Absolute Quantitation of DNA Copy Number. *Analytical Chemistry*, *83*(22), 8604–8610. https://doi.org/10.1021/ac202028g

Hollman, K. B., Hwang, S. F., Manolii, V. P., & Strelkov, S. E. (2021). Pathotypes of *Plasmodiophora brassicae* collected from clubroot resistant canola (*Brassica napus* L.) cultivars in western Canada in 2017-2018. *Canadian Journal of Plant Pathology*. https://doi.org/10.1080/07060661.2020.1851893

Holtz, M. D., Hwang, S. F., Manolii, V. P., Strelkov, I. S., & Strelkov, S. E. (2021). Development of molecular markers to identify distinct populations of *Plasmodiophora brassicae*. *European Journal of Plant Pathology*, *159*(3), 637–654. https://doi.org/10.1007/s10658-020-02194-4

Holtz, M. D., Hwang, S. F., & Strelkov, S. E. (2018). Genotyping of *Plasmodiophora brassicae* reveals the presence of distinct populations. *BMC Genomics*, *19*(1), 254. https://doi.org/10.1186/s12864-018-4658-1

Howard, R. J., Strelkov, S. E., & Harding, M. W. (2010). Clubroot of cruciferous crops – new perspectives on an old disease. *Canadian Journal of Plant Pathology*, *32*(1), 43–57. https://doi.org/10.1080/07060661003621761

Hwang, S. F., Ahmed, H. U., Zhou, Q., Rashid, A., Strelkov, S. E., Gossen, B. D., Peng, G., & Turnbull, G. D. (2013). Effect of susceptible and resistant canola plants on *Plasmodiophora brassicae* resting spore populations in the soil. *Plant Pathology*, *62*(2), 404–412. https://doi.org/10.1111/j.1365-3059.2012.02636.x

Ingram, D. S., Tommerup, I. C., & Brian, P. W. (1972). The life history of *Plasmodiophora brassicae* Woron. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *180*(1058), 103–112. https://doi.org/10.1098/rspb.1972.0008

Integrated DNA Technologies. (2021a). *RhPCR primers*. Integrated DNA Technologies. https://www.idtdna.com/pages/products/qpcr-and-pcr/custom-primers/rhpcr-primers

Integrated DNA Technologies. (2021b). *RNase H2 Enzyme*. Integrated DNA Technologies. https://www.idtdna.com/pages/products/qpcr-and-pcr/master-mixes-reagents/rnase-h-enzyme

Ito, S., Maehara, T., Maruno, E., Tanaka, S., Kameya-Iwaki, M., & Kishi, F. (1999). Development of a PCR-based Assay for the Detection of *Plasmodiophora brassicae* in Soil. *Journal of Phytopathology*, *147*(2), 83–88. https://doi.org/10.1046/j.1439-0434.1999.147002083.x

Jones, S., Baizan-Edge, A., MacFarlane, S., & Torrance, L. (2017). Viral Diagnostics in Plants Using Next Generation Sequencing: Computational Analysis in Practice. *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.01770

Kageyama, K., & Asano, T. (2009). Life Cycle of *Plasmodiophora brassicae*. *Journal of Plant Growth Regulation*, *28*(3), 203. https://doi.org/10.1007/s00344-009-9101-z

Kazazian, H. H. (2004). Mobile Elements: Drivers of Genome Evolution. *Science*, *303*(5664), 1626–1632. https://doi.org/10.1126/science.1089670

Kreuze, J. F., Perez, A., Untiveros, M., Quispe, D., Fuentes, S., Barker, I., & Simon, R. (2009). Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: A generic method for diagnosis, discovery and sequencing of viruses. *Virology*, *388*(1), 1–7. https://doi.org/10.1016/j.virol.2009.03.024

Labbé, G., Rankin, M. A., Robertson, J., Moffat, J., Giang, E., Lee, L. K., Ziebell, K., MacKinnon, J., Laing, C. R., Parmley, E. J., Agunos, A., Daignault, D., Bekal, S., Chui, L., MacDonald, K. A., Hoang, L., Slavic, D., Ramsay, D., Pollari, F., … Johnson, R. P. (2019). Targeting discriminatory SNPs in *Salmonella enterica* serovar Heidelberg

genomes using RNase H2-dependent PCR. *Journal of Microbiological Methods*, *157*, 81–87. https://doi.org/10.1016/j.mimet.2018.12.021

Lamb, P. D., Hunter, E., Pinnegar, J. K., Creer, S., Davies, R. G., & Taylor, M. I. (2019). How quantitative is metabarcoding: A meta-analytical approach. *Molecular Ecology*, *28*(2), 420–430. https://doi.org/10.1111/mec.14920

LeBoldus, J. M., Manolii, V. P., Turkington, T. K., & Strelkov, S. E. (2012). Adaptation to Brassica Host Genotypes by a Single-Spore Isolate and Population of *Plasmodiophora brassicae* (Clubroot). *Plant Disease*, *96*(6), 833–838. https://doi.org/10.1094/PDIS-09-11-0807

Legeay, J., Husson, C., Cordier, T., Vacher, C., Marcais, B., & Buée, M. (2019). Comparison and validation of Oomycetes metabarcoding primers for Phytophthora high throughput sequencing. *Journal of Plant Pathology*, *101*(3), 743–748. https://doi.org/10.1007/s42161-019-00276-9

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 Research*, *20*(2), 265–272. https://doi.org/10.1101/gr.097261.109

Liu, Q., Thorland, E. C., & Sommer, S. S. (1997). Inhibition of PCR Amplification by a Point Mutation Downstream of a Primer. *BioTechniques*, *22*(2), 292–300. https://doi.org/10.2144/97222st01

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., … Wang, J.

(2012). SOAPdenovo2: An empirically improved memory-efficient short-read *de novo* assembler. *GigaScience*, *1*(1). https://doi.org/10.1186/2047-217X-1-18

Macfarlane, I. (1952). Factors Affecting the Survival of *Plasmodiophora brassicae* Wor. In the Soil and Its Assessment by a Host Test. *Annals of Applied Biology*, *39*(2), 239–256. https://doi.org/10.1111/j.1744-7348.1952.tb00903.x

Manzanares-Dauleux, M. J., Barret, P., & Thomas, G. (2000). Development of a Pathotype Specific SCAR Marker in *Plasmodiophora brassicae*. *European Journal of Plant Pathology*, *106*(8), 781–787. https://doi.org/10.1023/A:1026586803761

McAllister, C. H., Fortier, C. E., St Onge, K. R., Sacchi, B. M., Nawrot, M. J., Locke, T., & Cooke, J. E. K. (2018). A novel application of RNase H2-dependent quantitative PCR for detection and quantification of *Grosmannia clavigera*, a mountain pine beetle fungal symbiont, in environmental samples. *Tree Physiology*, *38*(3), 485–501. https://doi.org/10.1093/treephys/tpx147

McGinnis, S., & Madden, T. L. (2004). BLAST: At the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Research*, *32*(suppl_2), W20–W25. https://doi.org/10.1093/nar/gkh435

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Melville, S. C., & Hawken, R. H. (1967). Soil Testing for Club Root in Devon and Cornwall. *Plant Pathology*, *16*(4), 145–147. https://doi.org/10.1111/j.1365-3059.1967.tb00393.x

Mercado-Blanco, J., Rodríguez-Jurado, D., Pérez-Artés, E., & Jiménez-Díaz, R. M. (2002). Detection of the Defoliating Pathotype of *Verticillium dahliae* in Infected Olive Plants by Nested PCR. *European Journal of Plant Pathology*, *108*(1), 1–13. https://doi.org/10.1023/A:1013994827836

Moxham, S. E., & Buczacki, S. T. (1983). Chemical composition of the resting spore wall of *Plasmodiophora brassicae*. *Transactions of the British Mycological Society*, *80*(2), 297–304. https://doi.org/10.1016/S0007-1536(83)80013-8

Naiki, T., Dixon, G. R., & Ikegami, H. (1987). Quantitative estimation of spore germination of *Plasmodiophora brassicae*. *Transactions of the British Mycological Society*, *89*(4), 569–572. https://doi.org/10.1016/S0007-1536(87)80091-8

Nicolaisen, M., West, J. S., Sapkota, R., Canning, G. G. M., Schoen, C., & Justesen, A. F. (2017). Fungal Communities Including Plant Pathogens in Near Surface Air Are Similar across Northwestern Europe. *Frontiers in Microbiology*, *8*. https://doi.org/10.3389/fmicb.2017.01729

Pageau, D., Lajeunesse, J., & Lafond, J. (2006). Impact of clubroot [*Plasmodiophora brassicae*] on the yield and quality of canola. *Canadian Journal of Plant Pathology*, *28*(1), 137–143. https://www.cabdirect.org/cabdirect/abstract/20063152269

Pang, W., Liang, Y., Zhan, Z., Li, X., & Piao, Z. (2020). Development of a Sinitic Clubroot Differential Set for the Pathotype Classification of *Plasmodiophora brassicae*. *Frontiers in Plant Science*, *11*. https://doi.org/10.3389/fpls.2020.568771

Peng, G., Lahlali, R., Hwang, S. F., Pageau, D., Hynes, R. K., McDonald, M. R., Gossen, B. D., & Strelkov, S. E. (2014). Crop rotation, cultivar resistance, and fungicides/biofungicides

for managing clubroot (*Plasmodiophora brassicae*) on canola. *Canadian Journal of Plant Pathology*, *36*(sup1), 99–112. https://doi.org/10.1080/07060661.2013.860398

Phillippy, A. M., Schatz, M. C., & Pop, M. (2008). Genome assembly forensics: Finding the elusive mis-assembly. *Genome Biology*, *9*(3), R55. https://doi.org/10.1186/gb-2008-9-3-r55

Pruitt, K. D., Tatusova, T., & Maglott, D. R. (2007). NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, *35*(suppl_1), D61–D65. https://doi.org/10.1093/nar/gkl842

Rådström, P., Knutsson, R., Wolffs, P., Lövenklev, M., & Löfström, C. (2004). Pre-PCR processing. *Molecular Biotechnology*, *26*(2), 133–146. https://doi.org/10.1385/MB:26:2:133

Rahman, H., Peng, G., Yu, F., Falk, K. C., Kulkarni, M., & Selvaraj, G. (2014). Genetics and breeding for clubroot resistance in Canadian spring canola (*Brassica napus* L.). *Canadian Journal of Plant Pathology*, *36*(sup1), 122–134. https://doi.org/10.1080/07060661.2013.862571

Rahman, M., Sun, Z., McVetty, P. B. E., & Li, G. (2008). High throughput genome-specific and gene-specific molecular markers for erucic acid genes in *Brassica napus* (L.) for marker-assisted selection in plant breeding. *Theoretical and Applied Genetics*, *117*(6), 895–904. https://doi.org/10.1007/s00122-008-0829-9

Rauschert, E. (2010). Survivorship Curves. *Nature Education Knowledge*, *3*(10), 18. https://www.nature.com/scitable/knowledge/library/survivorship-curves-16349555/

Rennie, D. C., Manolii, V. P., Cao, T., Hwang, S. F., Howard, R. J., & Strelkov, S. E. (2011). Direct evidence of surface infestation of seeds and tubers by *Plasmodiophora brassicae*

and quantification of spore loads. *Plant Pathology*, *60*(5), 811–819. https://doi.org/10.1111/j.1365-3059.2011.02449.x

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative Genomics Viewer. *Nature Biotechnology*, *29*(1), 24–26. https://doi.org/10.1038/nbt.1754

Rodgers, T. W., Olson, J. R., & Mock, K. E. (2019). Use of RNase H-dependent PCR for discrimination and detection of closely related species from environmental DNA. *Methods in Ecology and Evolution*, *10*(7), 1091–1096. https://doi.org/10.1111/2041-210X.13187

Rolland, M., Glais, L., Kerlan, C., & Jacquot, E. (2008). A multiple single nucleotide polymorphisms interrogation assay for reliable Potato virus Y group and variant characterization. *Journal of Virological Methods*, *147*(1), 108–117. https://doi.org/10.1016/j.jviromet.2007.08.022

Rott, M., Xiang, Y., Boyes, I., Belton, M., Saeed, H., Kesanakurti, P., Hayes, S., Lawrence, T., Birch, C., Bhagwat, B., & Rast, H. (2017). Application of Next Generation Sequencing for Diagnostic Testing of Tree Fruit Viruses and Viroids. *Plant Disease*, *101*(8), 1489–1499. https://doi.org/10.1094/PDIS-03-17-0306-RE

Samuel, G., & Garrett, S. D. (1945). The infected root-hair count for estimating the activity of *Plasmodiophora brassicae* Woron. In the soil. *Annals of Applied Biology*, *32*(2), 96–101. https://doi.org/10.1111/j.1744-7348.1945.tb06767.x

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, *74*(12), 5463–5467. https://doi.org/10.1073/pnas.74.12.5463

Schrader, C., Schielke, A., Ellerbroek, L., & Johne, R. (2012). PCR inhibitors – occurrence, properties and removal. *Journal of Applied Microbiology*, *113*(5), 1014–1026. https://doi.org/10.1111/j.1365-2672.2012.05384.x

Schwelm, A., Fogelqvist, J., Knaust, A., Jülke, S., Lilja, T., Bonilla-Rosso, G., Karlsson, M., Shevchenko, A., Dhandapani, V., Choi, S. R., Kim, H. G., Park, J. Y., Lim, Y. P., Ludwig-Müller, J., & Dixelius, C. (2015). The *Plasmodiophora brassicae* genome reveals insights in its life cycle and ancestry of chitin synthases. *Scientific Reports*, *5*(1), 11153. https://doi.org/10.1038/srep11153

Sharma, K., Gossen, B. D., & McDonald, M. R. (2011). Effect of Temperature on Cortical Infection by *Plasmodiophora brassicae* and Clubroot Severity. *Phytopathology®*, *101*(12), 1424–1432. https://doi.org/10.1094/PHYTO-04-11-0124

Somé, A., Manzanares, M. J., Laurens, F., Baron, F., Thomas, G., & Rouxel, F. (1996). Variation for virulence on *Brassica napus* L. amongst *Plasmodiophora brassicae* collections from France and derived single-spore isolates. *Plant Pathology*, *45*(3), 432–439. https://doi.org/10.1046/j.1365-3059.1996.d01-155.x

Stjelja, S., Fogelqvist, J., Tellgren-Roth, C., & Dixelius, C. (2019). The architecture of the *Plasmodiophora brassicae* nuclear and mitochondrial genomes. *Scientific Reports*, *9*(1), 15753. https://doi.org/10.1038/s41598-019-52274-7

Strelkov, S. E., & Hwang, S. F. (2014). Clubroot in the Canadian canola crop: 10 years into the outbreak. *Canadian Journal of Plant Pathology*, *36*(sup1), 27–36. https://doi.org/10.1080/07060661.2013.863807

Strelkov, S. E., Hwang, S. F., Howard, R. J., Hartman, M., & Turkington, T. K. (2011). Progress towards the Sustainable Management of Clubroot (*Plasmodiophora brassicae*) of Canola

on the Canadian Prairies. *Prairie Soils & Crops Journal*, *4*, 114–121. https://prairiesoilsandcrops.ca/articles/volume-4-13-screen.pdf

Strelkov, S. E., Hwang, S. F., Manolii, V. P., Cao, T., & Feindel, D. (2016). Emergence of new virulence phenotypes of *Plasmodiophora brassicae* on canola (*Brassica napus*) in Alberta, Canada. *European Journal of Plant Pathology*, *145*(3), 517–529. https://doi.org/10.1007/s10658-016-0888-8

Strelkov, S. E., Hwang, S. F., Manolii, V. P., Cao, T., Fredua-Agyeman, R., Harding, M. W., Peng, G., Gossen, B. D., McDonald, M. R., & Feindel, D. (2018). Virulence and pathotype classification of *Plasmodiophora brassicae* populations collected from clubroot resistant canola (*Brassica napus*) in Canada. *Canadian Journal of Plant Pathology*, *40*(2), 284–298. https://doi.org/10.1080/07060661.2018.1459851

Strelkov, S. E., Hwang, S. F., Manolii, V. P., Turnbull, G. D., Fredua-Agyeman, R., Hollman, K. B., & Kaus, S. (2021). Characterization of clubroot (*Plasmodiophora brassicae*) from canola (*Brassica napus*) in the Peace Country of Alberta, Canada. *Canadian Journal of Plant Pathology*, *43*(1), 155–161. https://doi.org/10.1080/07060661.2020.1776931

Strelkov, S. E., Manolii, V. P., Harding, M. W., Daniels, G. C., Nuffer, P., & Hwang, S. F. (2020). The occurrence and spread of clubroot on canola in Alberta in 2019. Canadian Plant Disease Survey 2020 Vol. 100: Disease Highlights 2019. *Canadian Journal of Plant Pathology*, *42*(sup1), 117–120. https://doi.org/10.1080/07060661.2020.1752524

Strelkov, S. E., Manolii, V. P., Harding, M. W., Hwang, S. F., Poscente, N., Lisowski, S. L. I., Pugh, C. A., & Burke, D. A. (2014). The occurrence of clubroot on canola in Alberta in 2013. *Canadian Plant Disease Survey*, *94*, 158–161.

Sundelin, T., Christensen, C. B., Larsen, J., Møller, K., Lübeck, M., Bødker, L., & Jensen, B. (2010). In Planta Quantification of *Plasmodiophora brassicae* Using Signature Fatty Acids and Real-Time PCR. *Plant Disease*, *94*(4), 432–438. https://doi.org/10.1094/PDIS-94-4-0432

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, *21*(8), 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Takahashi, K., & Yamaguchi, T. (1988). A Method for Assessing the Pathogenic Activity of Resting Spores of *Plasmodiophora brassicae* by Fluorescence Microscopy. *Japanese Journal of Phytopathology*, *54*(4), 466–475. https://doi.org/10.3186/jjphytopath.54.466

Tanaka, S., Kochi, S., Kunita, H., Ito, S., & Kameya-Iwaki, M. (1999). Biological Mode of Action of the Fungicide, Flusulfamide, Against *Plasmodiophora brassicae* (clubroot). *European Journal of Plant Pathology*, *105*(6), 577–584. https://doi.org/10.1023/A:1008722005570

Tedersoo, L., Drenkhan, R., Anslan, S., Morales-Rodriguez, C., & Cleary, M. (2019). High-throughput identification and diagnostics of pathogens and pests: Overview and practical recommendations. *Molecular Ecology Resources*, *19*(1), 47–76. https://doi.org/10.1111/1755-0998.12959

Tewari, J. P., Strelkov, S. E., Orchard, D., Hartman, M., Lange, R. M., & Turkington, T. K. (2005). Identification of clubroot of crucifers on canola (*Brassica napus*) in Alberta. *Canadian Journal of Plant Pathology*, *27*(1), 143–144. https://doi.org/10.1080/07060660509507206

Tommerup, I. C., & Ingram, D. S. (1971). The Life-Cycle of *Plasmodiophora brassicae* Woron. In Brassica Tissue Cultures and in Intact Roots. *New Phytologist*, *70*(2), 327–332. https://doi.org/10.1111/j.1469-8137.1971.tb02531.x

Udupa, S. M., Weigand, F., Saxena, M. C., & Kahl, G. (1998). Genotyping with RAPD and microsatellite markers resolves pathotype diversity in the ascochyta blight pathogen of chickpea. *Theoretical and Applied Genetics*, *97*(1), 299–307. https://doi.org/10.1007/s001220050899

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in Bioinformatics*, *43*(1), 11.10.1-11.10.33. https://doi.org/10.1002/0471250953.bi1110s43

Verdin, E., Wipf-Scheibel, C., Gognalons, P., Aller, F., Jacquemond, M., & Tepfer, M. (2017). Sequencing viral siRNAs to identify previously undescribed viruses and viroids in a panel of ornamental plant samples structured as a matrix of pools. *Virus Research*, *241*, 19–28. https://doi.org/10.1016/j.virusres.2017.05.019

Verma, S. S., Rahman, M. H., Deyholos, M. K., Basu, U., & Kav, N. N. V. (2014). Differential Expression of miRNAs in *Brassica napus* Root following Infection with *Plasmodiophora brassicae*. *PLOS ONE*, *9*(1), e86648. https://doi.org/10.1371/journal.pone.0086648

Wallenhammar, A.-C. (1996). Prevalence of *Plasmodiophora brassicae* in a spring oilseed rape growing area in central Sweden and factors influencing soil infestation levels. *Plant Pathology*, *45*(4), 710–719. https://doi.org/10.1046/j.1365-3059.1996.d01-173.x

Wallenhammar, A.-C., Almquist, C., Söderström, M., & Jonsson, A. (2012). In-field distribution of *Plasmodiophora brassicae* measured using quantitative real-time PCR. *Plant Pathology*, *61*(1), 16–28. https://doi.org/10.1111/j.1365-3059.2011.02477.x

Wallenhammar, A.-C., & Arwidsson, O. (2001). Detection of *Plasmodiophora brassicae* By PCR in Naturally Infested Soils. *European Journal of Plant Pathology*, *107*(3), 313–321. https://doi.org/10.1023/A:1011224503200

Walsh, P. S., Erlich, H. A., & Higuchi, R. (1992). Preferential PCR amplification of alleles: Mechanisms and solutions. *Genome Research*, *1*(4), 241–250. https://doi.org/10.1101/gr.1.4.241

Wen, R., Lee, J., Chu, M., Tonu, N., Dumonceaux, T., Gossen, B. D., Yu, F., & Peng, G. (2020). Quantification of *Plasmodiophora brassicae* Resting Spores in Soils Using Droplet Digital PCR (ddPCR). *Plant Disease*, *104*(4), 1188–1194. https://doi.org/10.1094/PDIS-03-19-0584-RE

White, J. G., & Buczacki, S. T. (1979). Observations on suppression of clubroot by artificial or natural heating of soil. *Transactions of the British Mycological Society*, *73*(2), 271–275. https://doi.org/10.1016/S0007-1536(79)80111-4

Williams, P. H. (1966). A system for the determination of races of *Plasmodiophora brassicae* that infect Cabbage and Rutabaga. *Phytopathology*, *56*(6), 624–626. https://www.cabdirect.org/cabdirect/abstract/19661105237

Wit, P. D., Pespeni, M. H., & Palumbi, S. R. (2015). SNP genotyping and population genomics from expressed sequences – current advances and future possibilities. *Molecular Ecology*, *24*(10), 2310–2323. https://doi.org/10.1111/mec.13165

Xue, S., Cao, T., Howard, R. J., Hwang, S. F., & Strelkov, S. E. (2008). Isolation and Variation in Virulence of Single-Spore Isolates of *Plasmodiophora brassicae* from Canada. *Plant Disease*, *92*(3), 456–462. https://doi.org/10.1094/PDIS-92-3-0456

Yang, Y., Zuzak, K., Harding, M. W., Strelkov, S. E., Hwang, S. F., Feindel, D., & Feng, J. (2018). DNA Sequence Dimorphisms in Populations of the Clubroot Pathogen *Plasmodiophora brassicae*. *Plant Disease*, *102*(9), 1703–1707. https://doi.org/10.1094/PDIS-02-18-0225-RE

Zhang, H., Feng, J., Manolii, V. P., Strelkov, S. E., & Hwang, S. F. (2015). Characterization of a Gene Identified in Pathotype 5 of the Clubroot Pathogen *Plasmodiophora brassicae*. *Phytopathology*®, *105*(6), 764–770. https://doi.org/10.1094/PHYTO-10-14-0270-R

Zheng, J., Wang, X., Li, Q., Yuan, S., Wei, S., Tian, X., Huang, Y., Wang, W., & Yang, H. (2018). Characterization of Five Molecular Markers for Pathotype Identification of the Clubroot Pathogen *Plasmodiophora brassicae*. *Phytopathology*®, *108*(12), 1486–1492. https://doi.org/10.1094/PHYTO-11-17-0362-R

Zhou, Q., Hwang, S. F., Strelkov, S. E., Fredua-Agyeman, R., & Manolii, V. P. (2018). A molecular marker for the specific detection of new pathotype 5-like strains of *Plasmodiophora brassicae* in canola. *Plant Pathology*, *67*(7), 1582–1588. https://doi.org/10.1111/ppa.12868

Zuzak, K., Yang, Y., Kimmel, N., Harding, M. W., Feindel, D., & Feng, J. (2017). Identification of native and invasive subspecies of common reed (*Phragmites australis*) in Alberta, Canada, by RNase-H-dependent PCR. *Botany*. https://doi.org/10.1139/cjb-2017-0152