

EVOLUTIONARY GENOMIC ANALYSIS OF THE CHARCOAL ROT FUNGUS
MACROPHOMINA PHASEOLINA FOR IMPROVED DISEASE MANAGEMENT UNDER
CLIMATE CHANGE

By

Viviana Ortiz Londoño

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Plant Pathology – Doctor of Philosophy
Ecology, Evolutionary Biology and Behavior – Dual Major

2022

ABSTRACT

Global agricultural production is threatened by several diseases caused by fungal pathogens. Recently increased efforts to characterize genomic diversity in fungal pathogens and the availability of large-scale ecological datasets offer new opportunities for understanding pathogen adaptation. The twin lenses of population genomics and adaptive evolution are powerful frameworks to interpret this data because of characteristics of fungal pathogens in agroecosystems that allow for their rapid evolution. The environment, biotic and abiotic, is a major driver for the evolution of plant pathogens and greatly influences disease outcomes. *Macrophomina phaseolina* causes charcoal rot in many important economic and subsistence crops worldwide. Charcoal rot significantly reduces yield and seed quality of soybean and dry bean and has been recognized as a warm climate-driven disease of increasing concern for crop production under global climate change. Therefore, this dissertation investigated the genetic structure and adaptive potential of *M. phaseolina* to understand how this pathogen responds to hosts, fungicides, and climate and how to best manage and predict charcoal rot disease.

To this end, I first characterized the genetic diversity and genotype-environment associations in *M. phaseolina* filling in fundamental knowledge of population structure and shedding light on climate adaptation. Population genomic analyses of 95 *M. phaseolina* isolates from soybean and dry bean across the continental US, Puerto Rico, and Colombia revealed geographic structure and diversification associated to climate. Phylogenomic and clustering approaches differentiated isolates into two main clades of the US and Colombian-Puerto Rican origins and five divergent genetic clusters within these clades. I identified a predominantly clonal structure in the US and a semi-clonal structure in Colombia and Puerto Rico. Limited genetic differentiation between isolates of soybean and dry bean origins was observed. Estimations of the independent contributions of neutral population structure, space, and climate to genetic variation, revealed that climate significantly contributes to genetic variation between genetic clusters. Genotype-environment associations implicated several genomic regions in *M. phaseolina* adaptation to climate and the loci significantly associated with multivariate climate were found near to genes related to fungal stress responses.

Information on the efficacy of newer fungicides chemistries for charcoal rot management is lack-

ing. Therefore, I characterized the *in-vitro* fungicide sensitivity of *M. phaseolina* to three major chemical classes of single-site fungicides, succinate dehydrogenase inhibitors (SDHI; boscalid) dicarboximides (iprodione) and demethylation inhibitors (DMI; prothioconazole). This study found no isolates in the US, Colombia or Puerto Rico that were insensitive to any of the fungicides tested. Isolates were most sensitive to prothioconazole indicating its potential use for charcoal rot management. Next, mutations in the fungicides target protein genes were investigated. No mutations that associated to levels of sensitivity to boscalid, iprodione and prothioconazole were found among our isolate collection. Finally, a preliminary ecoclimatic suitability model was developed and used to project the climatic suitability of *M. phaseolina* at a global scale. Importantly, this model predicted areas of high climatic suitability which may be at increased risk of disease.

Results from this dissertation work inform and improve charcoal rot management strategies through better understanding of *M. phaseolina* genetic structure and adaptive potential, *in-vitro* efficacy of single-site fungicides and potential disease outcomes under a changing climate. Additionally, this research is expected to contribute to applied issues surrounding plant disease risk prediction, and more broadly predicting short-term evolution of *M. phaseolina* across climates. Ultimately, this research will lead to better understanding of disease outcomes and more efficient management of plant pathogens considering adaptive responses under a changing climate.

Copyright by
VIVIANA ORTIZ LONDOÑO
2022

I dedicate this dissertation to my grandmother Nora Giraldo Castillo and to my mother Fanny Londoño Giraldo. The most loving and strong women in my life.

Grandma, when your children did not have access to education, you became their teacher. You taught them to read, write and do math, all while juggling the hard work of a farm in the middle of the Colombian Andes. I will forever admire your courage and how you always have a smile, even through the most difficult times.

Mom, your love and determination are always with me. You were so determined to keep studying that you went far away from home so you could attend high school. Against all odds, you became the first graduate in the family. You taught me empathy and to stand up for what I believe in.

Abue and mami, you are my greatest source of inspiration, forever.

ACKNOWLEDGEMENTS

First, I would like to thank my advisor Dr. Martin Chilvers for guiding me, encouraging me to keep going and above all for always being a source of confidence. I feel grateful remembering that after each meeting with you I was left reassured that I could make it to the end of this journey. I am thankful to my committee members Dr. Gregory Bonito, Dr. Monique Sakalidis and Dr. Gideon Bradburd for providing expertise and advice throughout my PhD.

I am also very grateful for the support of all my lab members, past and present. Thank you all for your help and advice during my PhD. Especially Dr. Alejandro Rojas, Dr. Zachary Noel and Dr. Austin McCoy whom I'm very grateful for their help while I was getting started with my research and throughout the years. I would also like to thank the scientists at Universidad del Valle and CIAT who mentored me during my undergraduate studies and helped me prepare for graduate school. I am specially grateful to Dr. Soroush Parsa, thank you for taking the time to mentor me and for your support and encouragement to pursue a PhD.

I would like to thank the United Soybean Board, Michigan Soybean Committee, the North Central Soybean Research Program, Project GREEN, and scholarships that enabled me to conduct my PhD research.

Lastly, I would like to thank my friends and family. My parents, my sister and my niece for your unconditional love and support. To my grandparents, aunts and uncles, I am forever grateful for your love. Since I can remember you have supported me in every way possible. Lastly, my husband, thanks for your unconditional love and for always being there for me. I could not have done this without you. You all have been a constant source of inspiration for me and I would not have become the person that I am today without you.

TABLE OF CONTENTS

CHAPTER 1 AN EVOLUTIONARY GENOMICS PERSPECTIVE OF ADAPTATION IN PLANT PATHOGENS	1
BIBLIOGRAPHY	18
CHAPTER 2 POPULATION GENOMIC ANALYSIS REVEALS GEOGRAPHIC STRUCTURE AND CLIMATIC DIVERSIFICATION FOR <i>MACROPHOMINA</i> <i>PHASEOLINA</i> ISOLATED FROM SOYBEAN AND DRY BEAN ACROSS THE US, PUERTO RICO, AND COLOMBIA	23
BIBLIOGRAPHY	60
APPENDIX	68
CHAPTER 3 SENSITIVITY TO SINGLE-SITE FUNGICIDES IN <i>MACROPHOMINA</i> <i>PHASEOLINA</i> POPULATIONS FROM SOYBEAN AND DRY BEAN.....	83
BIBLIOGRAPHY	104
APPENDIX	109
CHAPTER 4 ECOCLIMATIC SUITABILITY AND ADAPTIVE GENOMICS IN <i>MACROPHOMINA PHASEOLINA</i> , THE CHARCOAL ROT PATHOGEN	125
BIBLIOGRAPHY	137
APPENDIX	141
CHAPTER 5 CONCLUDING STATEMENT	145

CHAPTER 1

AN EVOLUTIONARY GENOMICS PERSPECTIVE OF ADAPTATION IN PLANT PATHOGENS

Natural selection is a powerful force for the evolution of living organisms. From humans to microorganisms, adaptations that overcome challenges in the environment are fundamental for the success of populations in diverse environments. Unraveling the genetic basis underlying adaptive phenotypes has for years fascinated scientists and hence has been the focus of many studies. In host-pathogen systems, adaptation is driven mainly by selection pressures imposed by their interaction itself and by the abiotic environment. Most notably, in agroecosystems, such abiotic environmental pressures are related to climate and pesticides use (Stukenbrock et al., 2011). Other evolutionary forces also play an important role in the adaptation of organisms, among them migration is often considered because it can significantly affect adaptation processes. In this chapter, I discuss the implications of each of these factors on pathogen adaptation to plant host and the abiotic environment in agricultural ecosystems and the methodological approaches to study local adaptation. This introductory chapter provides an overview of the population genetics and community ecology concepts that constitute a foundation for the analysis and experiments in this thesis. Additionally, it provides a perspective on the use of genomic data in suitability models of plant pathogens under a changing climate.

1.1 Adaptation in host-pathogen systems

Adaptation is considered a central topic of ecological genetics. Adaptation can be defined as the evolutionary process by which a population becomes better able to live and reproduce in its habitat (Dobzhansky, 1968). Likewise, an adaptive trait has been defined as “a phenotypic trait that has evolved to help an organism deal with something in its environment” (Conner et al., 2004). Adaptation is caused exclusively by natural selection; however, the remaining evolutionary forces, mutation, genetic drift and migration can either accelerate or slow down the development of adaptations (Conner et al., 2004). Ecological factors, biotic and abiotic, are important drivers of natural selection. These selection drivers vary in the space and such ecological heterogeneity results in populations adapted to the local biotic and abiotic conditions (Kawecki and Ebert, 2004). This led me to the concept of local adaptation. In a strict sense, a population is locally adapted when it has a higher relative fitness in their local environment (habitat) relative to any other population introduced to that site (Kawecki and Ebert, 2004). However, local adaptation can be indicated by genetic and phenotypic variation along ecological gradients or contrasting habitats

(Savolainen et al., 2007). For example, there is genetically-based variation in growth rates in response to temperature and latitudinal gradients in fungi, including pathogens (Ellison et al., 2011; Lendenmann et al., 2016).

1.1.1 Coevolution in the host-pathogen system

Plants and pathogens, as biological systems, evolve in response to adaptation in the respective partner. Co-evolutionary interactions, and especially antagonistic ones, impose strong selection on both partners. In this sense, pathogens can act as drivers of natural selection on their hosts, while hosts also impose strong selection on the pathogens by defensive mechanisms (Kawecki and Ebert, 2004; Croll and McDonald, 2016). However, inherent characteristics of pathogens such as large effective population sizes, high mutation rates and short generation times, provide pathogens with strategic advantages to evolve faster than their hosts (Croll and McDonald, 2016). Thus, pathogens are hypothesized to have some advantages in the co-evolutionary race and it is widely accepted that pathogen populations become locally adapted to the local pool of host genotypes (Croll and McDonald, 2016). In agricultural ecosystems, particularly in crop – fungal pathogens interactions, two main considerations will be discussed, the specific characteristics of fungi and those of agricultural ecosystems that makes them a unique system to study local adaptation.

Key fungal characteristics, such as high reproductive potential and extraordinary capacity to disperse and survive, makes fungal pathogens ubiquitous organisms and particularly competent when adapting to new environments (Croll and McDonald, 2016). Global agricultural production is threatened by several diseases caused by fungal pathogens representing the most important cause of crop yield losses, along with diseases caused by oomycetes (Fisher et al., 2012). The crop genetic homogeneity present at the field level contribute to the devastating effects seen in agricultural ecosystems, mainly by favoring local adaptation processes in pathogens. Additionally, agro-ecosystems are managed in quite similar ways even when separated in space and time. For example, because of similarity in practices such as fertilization, irrigation, tillage and pesticide applications, combined with the planting of genetically uniform monocultures, some crop fields show remarkably similar environments on all continents, differing mainly according to the local climate.

1.1.2 Host specialization in plant pathogens

Plant pathogenic fungi in agricultural ecosystems, are considered to be adapted to the local host genotypes, thus they constitute excellent models for identifying the genetic basis of local adaptation (Croll and McDonald, 2016). This adaptation is largely explained by gene for gene interactions (Jones and Dangl, 2006). Several genes have been identified in different fungal pathogens involved in virulence and pathogenicity on their hosts. These pathogens include host-specialized fungi attacking important crops, such as rice, wheat, barley, rye and maize (e.g. *Pyricularia oryzae*, *Zymoseptoria tritici*, *Parastagonospora nodorum*, *Puccinia* spp., *Blumeria* spp., *Ustilago* spp.) (Croll and McDonald, 2016).

The gene-for-gene hypothesis first proposed by Flor (Flor, 1971) designates that avirulence genes in the pathogen are matched by resistance genes in the host. The direct or indirect interaction between the gene products triggers host defense responses that can prevent or reduce the growth of pathogens. Therefore, pathogens on a resistant host are under strong selection and tend to undergo mutation or deletion in the avirulence gene to evolve higher virulence on the host. Due to this highly specific interaction between avirulence and resistance gene products, avirulence genes are expected to play an important role in local adaptation processes in agricultural ecosystems (Croll and McDonald, 2016). Similarly, to the adaptation processes driven by pathogen on their hosts, management practices including the use of resistant host germplasm, affects adaptation processes in pathogens. These practices result in the worldwide distribution of genetically similar or identical crops, thus, selection operating on the local pathogen population can lead to occurrence of the same virulence mutations independently even in the absence of gene flow among the corresponding pathogen populations (Conner et al., 2004; Croll and McDonald, 2016).

1.1.2.1 Host jumps

In plant pathology, a host jump is broadly defined as the process by which a pathogen infects a new previously unaffected host species. In some cases, this process is considered a host jump when the new host is genetically distant (i.e., taxonomically distant, from another class or order) from the original host. In contrast to host shifts, in which the new host is closely related to the old host (Stukenbrock and McDonald, 2008). Common scenarios that favor host jumps in agroecosystems include wild plant species growing nearby field crops, the introduction of new crops into natural ecosystems and the worldwide movement

of infected plant material (Stukenbrock and McDonald, 2008).

A pathogen host jump can be exemplified by *Pyricularia* species on wheat and wild grasses. The wheat blast pathogen *P. graminis-tritici* likely emerged from the *Pyricularia* population infecting the wild grass *Urochloa* or other Brazilian grasses approximately 30 years ago (Grünwald et al., 2016). Multiple host jumps occurred in the Irish potato famine pathogen *Phytophthora infestans* and related species, between plant hosts belonging to four different families (Raffaele et al., 2010). These were favored because these pathogens originated in central Mexico (Goss et al., 2014) which is considered a center of diversity for the genus *Solanum* (Stukenbrock and McDonald, 2008)(Grünwald et al., 2016).

Comparative genomics approaches can detect genomic signatures of a host jump (Grünwald et al., 2016), which are often considered signatures of effector evolution (Dong et al., 2015). After a host jump, the pathogen is expected to adapt to the new host, leading to host specialization (Raffaele et al., 2010) and often to the emergence of a new pathogen species (Dong et al., 2015). Accordingly, a recent host jump may be detected by comparing the genomes of pathogens from host species that represent new and old hosts. The genomes will be very similar except for specific changes in the genomic region that enabled the infection of the new host. These are rapidly evolving genomic regions, repeat-rich and usually containing a lot of effector genes. Thus, some effector genes may be lost because they are not useful anymore in the new host while other effector genes will accumulate mutations that will improve or expand the effector action in the new host (Dong et al., 2015). Such gene loss has been seen after a host jump, in the fungus *Melanopsichium pennsylvanicum* (Sharma et al., 2014). Greater rate of copy number variation of effector genes has been observed among *P. infestans* and related species. Signatures of adaptive evolution identified as having dN/dS ratios >1 (indicative of positive selection) were detected in effector genes of *Phytophthora* clade IC species (Raffaele et al., 2010; Dong et al., 2015).

Importantly, host jumps were proposed as a crucial mechanism for macroevolutionary persistence of host-specialized filamentous pathogens by Raffaele and Kamoun (2012), who described the “jump or die” model in which the survival of a pathogen over long evolutionary timescales depends on the frequency of host jumps. Under this model, host jumps serve as accelerators of effector adaptation and lead to pathogen diversification. Therefore, pathogens with more adaptable genomes, such as those with two-

speed genomes, are more likely to survive as hosts become fully resistant or extinct (Raffaele and Kamoun, 2012).

1.1.3 Effect of haploid vs diploid genome on populations

Ploidy, the number of chromosome sets in an organisms, greatly influences different evolutionary aspects of populations such as the ability of organisms to mask deleterious mutations, the accumulation of deleterious mutations and the rates of adaptation (Gerstein and Otto, 2009). In general, diploid organisms have another layer of genetic variation compared to haploid organisms. Particularly, heterozygosity allows the occurrence of modes of gene actions, which is how genotype affects the phenotype. Additivity and dominance are different modes of gene actions that influence fitness, for example in complete dominance a dominant allele can mask the effect of the recessive allele in a heterozygous organism (Conner et al., 2004). Similarly, overdominance (heterozygote advantage) occurs when the heterozygous genotype has higher fitness than both homozygous genotypes. This may change how an organism responds to its environment and under a given condition may, at least, temporarily increase fitness of diploid heterozygous organisms (Gerstein and Otto, 2009). Thus, overdominance maintains genetic variation in natural populations, and so in this way heterozygosity prevents the accumulation of deleterious mutation in the genome (Conner et al., 2004).

The long-term impact of deleterious mutations on the mean fitness of a population depends almost entirely on the genome-wide deleterious mutation rate and not on the selective disadvantage of the mutations (Gerstein and Otto, 2009; Haldane, 1937). Haploids will have the lowest mutation rate (and lower mutation load). This is because the equilibrium mean fitness of a population is reduced by approximately cU (the “mutation load”), where c is the ploidy level and U is the mutation rate per haploid genome. Therefore, haploids will have higher fitness than diploids, despite that deleterious mutations are masked to some degree in diploids (Gerstein and Otto, 2009).

The effect of ploidy in the rates of adaptation of populations have been investigated using experimental evolution in *Saccharomyces cerevisiae* (Otto and Gerstein, 2008; Gerstein and Otto, 2009; Gerstein et al., 2011; Sharp et al., 2018). Adaptation of an organism to a novel environment depends on the rate in which beneficial mutations are acquired and spread through the population (Todd et al., 2017). The

rate of adaptation is affected by the rate of appearance and fixation of beneficial mutations, the fitness effect of these mutations, the dominance of mutant alleles, and effective population sizes (Gerstein and Otto, 2009; Todd et al., 2017). This is still an area of ongoing investigation, but in general experiments have showed that large asexual haploid populations of *S. cerevisiae* were able to adapt faster than diploids. However, in small populations, haploids and diploids adapted at approximately the same speed, and the advantage of haploidy disappeared (Gerstein and Otto, 2009).

In a recent mutation-accumulation experiment conducted by (Sharp et al., 2018) using *S. cerevisiae*, revealed that haploids were more prone to single-nucleotide mutations (SNMs) and mitochondrial mutations, whereas in diploids larger structural changes were more common (Sharp et al., 2018).

1.1.4 Effect of sexual vs asexual reproduction on populations

Sexual reproduction is known to greatly affect the population structure of organisms and it is a determinant in the evolution of organisms. Plant pathogens, especially fungi with clonal and mixed reproductive systems and highly dynamic genomes constitute remarkable organisms to study the effect of sexual reproduction on different evolutionary aspects. This has led some to consider fungal plant pathogens as model organisms in evolutionary biology, and even as proposed models for investigating cancer cell evolution (Möller and Stukenbrock, 2017). Sexual reproduction is crucial to eukaryotic evolution mainly because it can increase genetic diversity and eliminate deleterious mutations (Ni et al., 2011). Recombination between loci can occur during meiosis, which creates new combinations of alleles at these loci (Conner et al., 2004). These allele combinations may be advantageous under certain ecological conditions, thus allowing rapid adaptability to new environments. Rapid fixation of advantageous mutations, is also enabled by sexual reproduction by increasing the efficacy of natural selection (Möller and Stukenbrock, 2017).

On the contrary, long-term advantages of clonal reproduction include the maintenance of co-adapted allele combinations in the population, and that fit genotypes can be rapidly propagated (Möller and Stukenbrock, 2017). A short-term advantage of clonal reproduction is the ability to rapidly propagate while expending less energy (Ni et al., 2011) which may play an important role in the development of epidemics. However, clonal populations are sometimes considered “evolutionary impaired” because of their inability to recombine advantageous mutations that may occur independently (Möller and Stuken-

brock, 2017). Moreover, deleterious mutations may accumulate in the genome of clonal organisms in an irreversible manner, a process termed “Muller’s ratchet”.

Another factor to consider is that asexual species usually have a lower effective population size (N_e) than sexually reproducing species, as offspring are fundamentally copies of their parents. Thus, the effect of genetic drift is relatively greater compared to populations with large N_e (Möller and Stukenbrock, 2017). Therefore, the smaller the N_e the stronger the selection has to be to counteract the effects of genetic drift (Conner et al., 2004), which may weaken local adaptation processes.

Yet, many clonal species and many fungal pathogens considered to reproduce asexually are common and successful. Approximately one fifth of described fungi are thought to be asexual and clonal (Taylor et al., 1999). A possible explanation for the success of asexual fungi is the “two-speed genome” model proposed for fungi and oomycetes. In this model, genomes have a bipartite architecture with effector genes being associated with compartments enriched in repetitive sequences and transposable elements (Dong et al., 2015), this suggest that high mutation rates in these genome compartments support adaptive evolution by effector innovation (Möller and Stukenbrock, 2017). Other explanations to consider are the occurrence of cryptic sex and recombination as unisexual mating, in which meiotic basidiospores are produced from the fusion of mitotically produced nuclei; and parasexual reproduction, in which there is exchange of genetic material between fused hyphae or cells without meiosis.

Notably, asexual reproduction has evolved independently many times from sexually reproducing ancestors in ascomycete fungi (Taylor et al., 1999). This has led to speculate that clonal population structures in some pathogens, such as in *Verticillium dahliae*, have arisen at least partially because selection imposed by agroecosystems (Milgroom et al., 2014).

1.1.5 The abiotic environment as a driver of natural selection

1.1.5.1 Climate adaptation

Climate fluctuation and particularly temperature are important abiotic factors leading to local adaptation on fungal plant pathogens (Savolainen et al., 2013; Croll and McDonald, 2016). Models of climate change for the coming decades predict increases in global temperature, atmospheric CO₂, ozone and changes in humidity, rainfall and severe weather (Fisher et al., 2012). This is expected to increase the environmental

heterogeneity that already is present across different agricultural systems in different regions of the world. This environmental heterogeneity, acts on genetically different organisms within a population, initially by causing fitness differences among phenotypically different populations and over time mutation and recombination generate populations adapted to the local environment (Fisher et al., 2012; Savolainen et al., 2013).

Thermal adaptation has been researched in several fungal species, including the model fungi *Neurospora crassa* (Ellison et al., 2011) and the powdery mildew pathosystems *Plantago lanceolata*–*Podosphaera plantaginis* (Laine, 2007). Temperature had a profound impact on the trajectory of evolution of *N. crassa* as well as in the co-evolution in the powdery mildew system. In the powdery mildew pathosystem, host and fungal populations were sampled across a natural thermal gradient, and a local vs. foreign experiment was conducted. Host cross-inoculations were conducted using sympatric and allopatric (i.e., local vs. foreign) pathogen populations at three temperatures (i.e., home vs. away environments) using detached leaves in a common garden laboratory environment. Local adaptation patterns differed according to temperature. Pathogen populations from the coolest environment had significantly higher fitness on the sympatric host at the coolest tested temperature, but had lower fitness than allopatric pathogen populations at higher temperatures (Laine, 2007; Croll and McDonald, 2016).

1.1.6 Migration effects on pathogen adaptation

The outcome of whether populations become adapted or not depend on the balance between selection and migration i.e., the levels of gene flow among populations and the strength of selection (Savolainen et al., 2007; Croll and McDonald, 2016). Local adaptation can be hindered for certain conditions, for example by migration rates and recolonization of populations by foreign genotypes. In the context of local adaptation, a high fitness in the local environment also implies a lower fitness in a foreign environment (Savolainen et al., 2013), and thus local adaptation occurs only if the effect of migration does not overwhelm the effect of local selection. Local adaptation may be disfavored by both high and low migration rates. Generally high migration rates overwhelm locally adapted genotypes leading to maladaptation, and low migration rates disfavor local adaptation mainly due to the limited genetic variation that the local population harbor, leaving limited input for selection to act on. Moreover, the unevenness in mi-

gration rates of the pathogen vs. the host also impact local adaptation processes. Pathogens are expected to become more rapidly locally adapted if they have higher migration rates than their hosts (Croll and McDonald, 2016).

1.1.7 Fungicide resistance evolution in plant pathogens

Fungicides play a key role in crop protection. Modern fungicides function primarily by disrupting particular molecular processes and targeting specific proteins, and therefore are often referred to as ‘single-site’ fungicides (Brent and Hollomon, 2007). In contrast, older multi-site fungicides act as general inhibitors affecting many cellular targets (Brent and Hollomon, 2007). The continued use of fungicides may eventually lead to the appearance of resistant pathogen populations. This phenomenon is called ‘acquired resistance’ (Brent and Hollomon, 2007).

Fungal pathogens with rapid reproductive rates and large population sizes are particularly prone to develop fungicide resistance (Lucas et al., 2015). Although several resistance mechanisms are known, the most common one is an alteration of the target site of the fungicide. In single-site fungicides, a single gene mutation can disrupt the target site function and confer resistance or reduced sensitivity (Brent and Hollomon, 2007). In situations in which resistance develops, it can be seen as a qualitative or a quantitative change. In quantitative resistance, the pathogen population shifts gradually towards resistance over time (Brent and Hollomon, 2007). While in qualitative resistance, a bimodal distribution with sensitive and resistant subpopulations is expected. In both cases, there is positive selection for resistant individuals, ultimately leading to resistance in the population if management strategies to limit pathogen exposure are not implemented (Lucas et al., 2015).

1.1.8 Using genomic data to detect population structure and adaptation in plant pathogens

Population structure can be defined as a systematic difference in allele frequencies between subpopulations in a population due to different ancestry (Turchin et al., 2012). Population differentiation occurs when subpopulations are not completely interbreeding and any of the evolutionary forces (mutation, selection, drift, migration) change the allele frequencies within the subpopulation. In other words, when individuals within subpopulations are more closely related than individuals between subpopulations.

Approaches to detect population structure include clustering methods. In clustering methods, in-

dividuals are assigned to populations often by estimating ancestry coefficients or using dimensionality-reduction approaches. Commonly used dimensionality-reduction approaches are principal component analysis (PCA) and discriminant analysis of principal components (DAPC). PCA is a form of multivariate analysis, which involves looking at multiple independent variables simultaneously to understand their contributions to the dependent variable (Abdi and Williams, 2010). PCA is used in identifying population structure to infer the possible number of populations (clusters) without prior knowledge, thus it can be useful to find hidden population structure. PCA is commonly used to convert genetic data into a reduced number of non-correlated variables, called principal components, which summarize the variation between samples closely related individuals can be seen as clusters. DAPC is particularly useful in organisms with clonal reproduction, such as many fungi. DAPC differs from PCA approaches, in which it does require *a priori* defined populations and maximizes the variance between populations, by partitioning the total variance into between-population and within-population components (Thibaut Jombart, Sébastien Devillard).

Model-based clustering approaches use a broad set of algorithms to characterize population structure. Commonly, these algorithms differ in the demographic model adopted, the statistical framework (frequentist or Bayesian), in whether selection is included in the model, among other aspects. Their main advantages are that they may be applied to a wide range of data sets and systems and that most of these methods do not need *a priori* delineation of populations. The main disadvantage is that they often rely on model assumptions. If the assumed model does not reflect the true model, these approaches may lead to false positives or to the incorrect identification of clusters. New approaches have been developed to overcome some of these limitations, such as models that incorporate a spatial component (Bradburd et al., 2018), and PCA-based models (Josephs et al., 2019).

Genetic variation is the input for selection to act and drive adaptation processes. Genomic divergence can be inferred from polymorphisms and fixed differences within and between species. Approaches to infer adaptation processes, rely either on population genetic analyses including reverse ecology approaches, quantitative trait mapping or association studies. Each of these approaches has strengths and limitations and a combination of different strategies would be more informative about adaptive natural selection

than using just one of them. Local adaptation critically depends on selectable genetic variation within local populations. Furthermore, the probability for local adaptation to evolve depends on the genetic architecture of a trait. Phenotypic traits governed by a simple genetic architecture are likely to be more rapidly selected than complex traits (Croll and McDonald, 2016). Similarly, loci with large effects should be favored to contribute to local adaptation as selection acts more rapidly on loci of large effects than small effects (Croll and McDonald, 2016). Strategies to identify loci involved in local adaptation are discussed.

1.1.8.1 Quantitative trait loci mapping

The outcome of host–pathogen interactions is thought to be governed largely by gene-for-gene interactions. However, recent studies showed that virulence can be governed also by quantitative trait loci (QTL) and that many abiotic factors contribute to the outcome of the interaction (Lendenmann et al., 2014; Croll and McDonald, 2016; Lendenmann et al., 2016b; Lendenmann et al., 2016a).

Quantitative trait loci mapping is based on the joint analysis of phenotype and genotype. QTL analysis uses a progeny of crosses between a pair of parental lines (pedigree) segregating for a specific trait, to find association between genotypes and phenotypes. QTL mapping is a powerful approach, however present some limitations. To uncover more variation many crosses and a large sample size are needed. QTL approaches can be time consuming since the progeny needs to be genotyped and phenotyped. Furthermore, extended linkage disequilibrium (LD) is often observed in the progeny, hindering the accurate location of the QTL. A QTL approach was used to investigate thermal adaptation in the fungal pathogen *Zymoseptoria tritici* (Lendenmann et al., 2016b). They identified four QTL associated with temperature sensitivity, containing six candidate genes including a PBS2, encoding a mitogen-activated protein kinase associated with low temperature tolerance in *Saccharomyces cerevisiae*. This study demonstrate a QTL approach can be successfully used in fungi, however, the need of progeny implies that QTL mapping can be applied only to sexual fungi.

1.1.8.2 Association mapping approaches

Association mapping studies are also based on phenotype- genotype associations. However, in contrast to QTL mapping, diverse panels of organisms can be used instead of using progeny populations derived from a parental cross. Advantages of association mapping approaches include that the LD is expected

to be lower than in pedigree-based studies, and multiple different traits can be studied simultaneously. Thus, QTLs can be found in a more accurate way. However, the use of a diversity panel implies the need of correction for population structure. The rates of false positives and false negatives is high and mixed models and correction for multiple hypothesis testing are needed to distinguish real associations from spurious ones. In general, for association mapping studies associations are not necessarily causal and further validation is needed.

1.1.8.3 Genotype–environment associations and redundancy analysis

Genotype–environment association (GEA) methods can be used to identify adaptive loci by correlating genetic data and environmental variables (Lasky et al., 2015; Forester et al., 2018). Multivariate methods in GEA have recently gained attention because their applications to the analysis of large genomic datasets. The multivariate nature of these methods allows the simultaneous analysis of thousands of loci (Forester et al., 2018). One of the most common multivariate approaches used in GEA is redundancy analysis (RDA). RDA is a constrained ordination method that have been used for years in community ecology to examine community composition in relation to environmental variables (Legendre and Legendre, 2012; Forester et al., 2018). In GEA approaches, RDA can be used to disentangle the effects of climatic factors in shaping genetic variation, by modeling sets of molecular markers (e.g. SNPs) as responses to a function of combinations of environmental predictors. RDA has been found to perform better than univariate methods in identifying weak, multilocus selection suggestive of polygenic adaptation (Forester et al., 2018). Partial RDA models, in which the effects of covariables can be removed, have been used to account for underlying population structure in the identification of loci associated with environmental factors in plant and animal systems (Lasky et al., 2012; Forester et al., 2018; Xuereb et al., 2018; Gibson and Moyle, 2020; Capblancq and Forester, 2021)

1.1.8.4 Population genetics and reverse ecology

Local adaptation has been investigated using population genetics with both forward and reverse ecology approaches. Population genetic analyses are based on F_{ST} (Wright fixation index) and linkage disequilibrium (LD) methods to detect candidate loci for local adaptation in the absence of phenotypic traits. These methods can detect outlier loci with an excessive amount of genetic differentiation among pop-

ulations (i.e., F_{ST} outlier analyses; De Mita et al. 2013). The basis is that local selection will exacerbate genetic differentiation at loci under selection compared to the genomic background. Reverse ecology (Li et al., 2008), is coined because the analogy with reverse genetics and implies that prior knowledge about an ecological trait is not necessary, instead first finding the genetic targets of selection and going back to identify the phenotypic differences or the adaptive phenotype. Reverse ecology is especially important to investigate organisms, such as microbes, which are challenging to identify adaptive phenotypes. This will preclude the utilization of associating studies such as Genome Wide Association Studies. Moreover, another challenge is exemplified in fungi, specifically, asexual fungi in which the development of populations to study a specific trait is not possible. Thus, approaches such as QTL analysis are not feasible. Reverse ecology may help overcome these challenges by investigating patterns of genetic diversity within and between populations. Ellison et al., (2011) implemented a reverse ecology approach to investigate temperature adaptation in the model fungus *N. crassa* by using three different population genetics metrics (F_{ST} , Tajima's D, and D_{xy}). They identified regions of genomic divergence, which are those showing low within-population polymorphism and high between-population divergence, and genes associated with response to cold temperature within those regions. However, among the three metrics used, Ellison et al. found that out of a total of 37 regions showing significant signatures of positive selection, only two were identified by all three metrics. This suggests a high proportion of false positives. In fact, it is known that F_{ST} outliers can be seen for reasons other than local adaptation such as deleterious alleles, species-wide selective sweeps and cryptic hybrid zones. Other aspect to consider is that regions identified using reverse ecology constitute just candidate loci of local adaptation, and further functional analysis needs to be done to conclusively identify causal genes.

1.2 *Macrophomina phaseolina* the causal agent of charcoal rot

Macrophomina phaseolina is a seed- and soil-borne fungal pathogen infecting more than 400 host species (Batista, Lopes and Alves, 2021). *M. phaseolina* is haploid, reproduces asexually, and overwinters in soil and crop residue as microsclerotia. Microsclerotia are melanized structures that serve as the primary inoculum to initiate infection in subsequent seasons (Gupta, Sharma and Ramteke, 2012; Islam et al., 2012). Pycnidia have been observed on host plant tissues (Knox-Davies, 1965; Dhingra and Sinclair, 1978; Mihail

and Taylor, 1995; Ma et al., 2010; Gupta et al., 2012). Although conidial suspensions have been used to experimentally inoculate soybean plants, suggesting pycnidia may provide inoculum for secondary infection in the field, their epidemiological significance has yet to be fully defined (Ma et al., 2010; Gupta et al., 2012). Depending on environmental conditions, *M. phaseolina* survives as microsclerotia in soil for up to 15 years (Short et al. 1980; Baird et al. 2003), and for up to 3 years as microsclerotia in symptomatic seeds or as mycelium in asymptomatic seeds (Hartman et al. 1999).

One of the first descriptions of *M. phaseolina* was made in 1890 by Halsted causing disease on sweet potato and the fungus was named *Rhizoctonia bataticola* (Halsted, 1890). Later this fungus was described by Tassi (1901) who named the fungus as *Macrophomina phaseolina* as it is retained today. In 1927, Ashby proposed the name *Macrophomina phaseoli* (Maubl.). Ashby associated the microsclerotia and conidial stage by observing the structures on seedlings of multiple crops. The name *Macrophomina phaseoli* was changed to *Macrophomina phaseolina* (Tassi) G. Goidanich, by Goidanich in 1947. By 1970, there was controversy among researchers over the use of the name, but genera *Macrophomina* and *Macrophoma* were used to refer to the pycnidial stage and *Rhizoctonia* to the sclerotial state.

In 1981, Von Arx introduced the name *Tiarosporella phaseolina* (Tassi) van der Aa and reduced the genus *Macrophomina* to a synonym of *Tiarosporella* Höhn. However, this has largely been ignored by the plant pathological and mycological community (Crous et al., 2006). In 2006, Crous et al., in a comprehensive phylogenetic study of 113 members of the family Botryosphaeriaceae using ribosomal DNA sequences, separated the genera *Macrophomina* and *Tiarosporella*, retaining the genus *Macrophomina* and the name *Macrophomina phaseolina*. The type species of *M. phaseolina* was originally described from *Phaseolus* spp. collected in Italy (Sarr et al., 2014).

Soybean is one of the most economically important crops worldwide, contributing with more than half of the world's total oilseed production (Boerma et al., 2004; Wilson, 2008). Seed oil and protein content makes soybean a valuable source not only for food and feed utilization but also for the industrial production of biofuels (Boerma et al., 2004). Many diseases threaten global soybean production, including charcoal rot, caused by *M. phaseolina*. Charcoal rot severely affects soybean yield under high temperatures and drought conditions (Mengistu et al., 2011). Tropical and subtropical areas, including

the southern US, have been the most affected. However, charcoal rot disease in soybean is now a consistent threat to soybean production in southern and northern US regions (Bradley et al., 2021). Although it is not clear which factors may be driving outbreaks in these regions, climatic changing conditions and resistance overcoming due to pathogen genetic divergence may be involved in the broadening of the geographical range of charcoal rot disease. To date, complete resistance to charcoal rot in soybean is not known and cultural practices and fungicide seed treatments do not provide consistent control to charcoal rot in soybean (Paris et al., 2006; Mengistu et al., 2011; Gillen et al., 2016). The confluence of these factors, makes imperative to investigate the genetic basis for adaptation in *M. phaseolina*.

1.3 Conclusions and dissertation overview

Approaches to study patterns of genetic diversity and adaptation in plant pathogens, as well as their main limitations, were discussed. One of those limitations is the difficulty to distinguish between natural selection and demographic processes. Thus, it is important to carefully consider the experimental design and approaches in light of the biology and epidemiology of the organism under study. Fungal pathogens can reach very high population sizes in a single plant and clonal reproduction and mixed reproduction systems are commonly observed. Furthermore, complex population dynamics and genome architecture are hallmarks of many fungal plant pathogens.

Most computational tools used in population genetics are based on models developed for sexual organisms (Kamvar et al., 2015). Populations that reproduce clonally may violate some of the assumptions underlying the population genetic theory. Moreover, the most widely used model is the Hardy-Weinberg model which assumes diploid, sexual organisms, besides no selection, no mutation, no migration, no drift and random mating between sexes (Hahn, 2019).

An important assumption that is violated in clonal organisms is the random association between alleles at different loci. In several approaches, this assumption allows the prediction of genotype frequencies from the allele frequencies at each locus (Milgroom, 1996). In clonal organisms associations among alleles at several loci are nonrandom and the entire genome may be effectively linked (Anderson and Kohn, 1995). Therefore, with clonal organisms the use of clone-corrected unlinked data is appropriate to avoid bias in diversity estimations due to duplicated genotypes (Kamvar et al., 2015; Milgroom, 2015).

Although, approaches based on genetic diversity metrics are often employed to identify signatures of adaptation in plant pathogens, population genomics and ordination techniques such as redundancy analysis have the potential to accommodate the intrinsic characteristics of fungal pathogens and begin disentangling the effects of selection of those of other evolutionary forces. Such methodological approaches in conjunction with population genomics analyses, constitute powerful tools to identify patterns of genomic diversity and adaptive potential of fungal pathogens.

The focus of this dissertation is to improve our understanding of *M. phaseolina* population structure, adaptation to host and climate and its application to local management practices through using the frameworks and tools of population genomics and community ecology. Additional objectives of this research are to characterize the sensitivity of *M. phaseolina* to fungicides currently used in crop production and provide a preliminary climatic suitability model for the monitoring and prediction of disease risk.

BIBLIOGRAPHY

- Abdi H, Williams LJ (2010) Principal component analysis. *Wiley Interdiscip Rev Comput Stat* 2: 433–459
- Anderson JB, Kohn LM (1995) Clonality in soilborne, plant-pathogenic fungi. *Annu Rev Phytopathol* 33: 369–391
- Baird RE, Watson CE, Scruggs M (2003) Relative Longevity of *Macrophomina phaseolina* and Associated Mycobiota on Residual Soybean Roots in Soil. *Plant Dis* 87: 563–566
- Batista E, Lopes A, Alves A (2021) What do we know about botryosphaeriaceae? An overview of a worldwide cured dataset. *Forests*. doi: 10.3390/f12030313
- Boerma HR, Specht JE, Wilcox JR (2004) World distribution and trade of soybean. *Soybeans: Improvement, production, and uses*. American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America, pp 1–14
- Bradburd GS, Coop GM, Ralph PL (2018) Inferring Continuous and Discrete Population Genetic. *Genetics* 210: 33–52
- Bradley CA, Allen TW, Sisson AJ, Bergstrom GC, Bissonnette KM, Bond J, Byamukama E, Chilvers MI, Collins AA, Damicone JP, et al (2021) Soybean Yield Loss Estimates Due to Diseases in the United States and Ontario, Canada, from 2015 to 2019. *Plant Health Prog* 22: 483–495
- Branco S, Bi K, Liao H-L, Gladieux P, Badouin H, Ellison CE, Nguyen NH, Vilgalys R, Peay KG, Taylor JW, et al (2016) Continental-level population differentiation and environmental adaptation in the mushroom *Suillus brevipes*. *Mol Ecol*. doi: 10.1111/mec.13892
- Brent KJ, Hollomon DW (2007) Fungicide resistance in crop pathogens: how can it be managed? 2nd, revised edition.
- Capblancq T, Forester BR (2021) Redundancy analysis: A Swiss Army Knife for landscape genomics. *Methods Ecol Evol* 12: 2298–2309
- Conner JK, Hartl DL, others (2004) A primer of ecological genetics. Sinauer Associates Incorporated
- Croll D, McDonald BA (2016) The genetic basis of local adaptation for pathogenic fungi in agricultural ecosystems. *Mol Ecol*. doi: 10.1111/mec.13870
- Crous PW, Slippers B, Wingfield MJ, Rheeder J, Marasas WFO, Philips AJL, Alves A, Burgess T, Barber P, Groenewald JZ (2006) Phylogenetic lineages in the Botryosphaeriaceae. *Stud Mycol* 55: 235–53
- Dhingra OD, Sinclair JB (1978) Biology and pathology of *Macrophomina phaseolina*. Universidade Federal de Vicosa., Minas Gerais
- Dobzhansky T (1968) On Some Fundamental Concepts of Darwinian Biology. *Evol Biol*. Springer US, Boston, MA, pp 1–34

- Dong S, Raffaele S, Kamoun S (2015) The two-speed genomes of filamentous pathogens: Waltz with plants. *Curr Opin Genet Dev* 35: 57–65
- Ellison CE, Hall C, Kowbel D, Welch J, Brem RB, Glass NL, Taylor JW (2011) Population genomics and local adaptation in wild isolates of a model microbial eukaryote. *Proc Natl Acad Sci U S A* 108: 2831–6
- Fisher MC, Henk Daniel A, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, Gurr SJ (2012) Emerging fungal threats to animal, plant and ecosystem health. *Nature* 484: 186–194
- Flor HH (1971) Current status of the gene-for-gene concept. *Annu Rev Phytopathol* 9: 275–296
- Forester BR, Lasky JR, Wagner HH, Urban DL (2018) Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Mol Ecol* 27: 2215–2233
- Gerstein AC, Cleathero LA, Mandegar MA, Otto SP (2011) Haploids adapt faster than diploids across a range of environments. *J Evol Biol* 24: 531–540
- Gerstein AC, Otto SP (2009) Ploidy and the causes of genomic evolution. *Journal of Heredity* 100: 571–581
- Gibson MJS, Moyle LC (2020) Regional differences in the abiotic environment contribute to genomic divergence within a wild tomato species. *Mol Ecol* 29: 2204–2217
- Gillen AM, Paris RL, Arelli PR, Smith JR, Mengistu A (2016) Registration of soybean germplasm line DBo4-10836 with high yield potential and resistance to soybean cyst nematode. *J Plant Regist* 10: 62
- Goss EM, Tabima JF, Cooke DEL, Restrepo S, Fry WE, Forbes GA, Fieland VJ, Cardenas M, Grünwald NJ (2014) The Irish potato famine pathogen *Phytophthora infestans* originated in central Mexico rather than the Andes. *Proceedings of the National Academy of Sciences* 111: 8791–8796
- Grünwald NJ, McDonald BA, Milgroom MG (2016) Population Genomics of Fungal and Oomycete Pathogens. *Annu Rev Phytopathol* 54: 323–346
- Gupta GK, Sharma SK, Ramteke R (2012) Biology, Epidemiology and Management of the Pathogenic Fungus *Macrophomina phaseolina* (Tassi) Goid with Special Reference to Charcoal Rot of Soybean (*Glycine max* (L.) Merrill). *Journal of Phytopathology* 160: 167–180
- Hartman GL, Rupe JC, Sikora EJ, Domier LL, Davis JA, Steffey KL (2015) Part I. Infectious Diseases. *Compendium of Soybean Diseases and Pests, Fifth Edition*. pp 15–119
- Islam M, Haque M, Islam M, Emdad E, Halim A, Hossen QM, Hossain M, Ahmed B, Rahim S, Rahman M, et al (2012) Tools to kill: Genome of one of the most destructive plant pathogenic fungi *Macrophomina phaseolina*. *BMC Genomics* 13: 493
- Jones JDG, Dangl JL (2006) The plant immune system. *Nature Reviews* 444: 323–329
- Josephs EB, Berg JJ, Ross-Ibarra J, Coop G (2019) Detecting Adaptive Differentiation in Structured Populations with Genomic Data and Common Gardens. *Genetics* 211: 989–1004

Kamvar ZN, Brooks JC, Grünwald NJ (2015) Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front Genet* 6: 208

Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecol Lett* 7: 1225–1241

Knox-Davies PS (1965) Pycnidium Production By *Macrophomina*. *S Afr J Agric Sci* 8: 205–218

Laine AL (2007) Detecting local adaptation in a natural plant-pathogen metapopulation: a laboratory vs. field transplant approach. *J Evol Biol* 20: 1665–1673

Lasky JR, Des Marais DL, McKay JK, Richards JH, Juenger TE, Keitt TH (2012) Characterizing genomic variation of *Arabidopsis thaliana*: The roles of geography and climate. *Mol Ecol* 21: 5512–5529

Lasky JR, Upadhyaya HD, Ramu P, Deshpande S, Hash CT, Bonnette J, Juenger TE, Hyma K, Acharya C, Mitchell SE, et al (2015) Genome-environment associations in sorghum landraces predict adaptive traits. *Sci Adv* 1: 1–14

Legendre P, Legendre L (2012) Canonical analysis. *Developments in Environmental Modelling*. pp 625–710

Lendenmann MH, Croll D, Palma-Guerrero J, Stewart EL, McDonald BA (2016a) QTL mapping of temperature sensitivity reveals candidate genes for thermal adaptation and growth morphology in the plant pathogenic fungus *Zymoseptoria tritici*. *Heredity (Edinb)* 116: 1–11

Lendenmann MH, Croll D, Palma-Guerrero J, Stewart EL, McDonald BA (2016b) QTL mapping of temperature sensitivity reveals candidate genes for thermal adaptation and growth morphology in the plant pathogenic fungus *Zymoseptoria tritici*. *Heredity (Edinb)* 116: 384–394

Lendenmann MH, Croll D, Stewart EL, McDonald BA (2014) Quantitative trait locus mapping of melanization in the plant pathogenic fungus *Zymoseptoria tritici*. *G3 (Bethesda)* 4: 2519–33

Li YF, Costello JC, Holloway AK, Hahn MW (2008) “Reverse Ecology” and the power of population genomics. *Evolution (N Y)* 62: 2984–2994

Lucas JA, Hawkins NJ, Fraaije BA (2015) The Evolution of Fungicide Resistance. *Adv Appl Microbiol* 90: 29–92

Ma J, Hill CB, Hartman GL (2010) Production of *Macrophomina phaseolina* Conidia by Multiple Soybean Isolates in Culture. *Plant Dis* 94: 1088–1092

Mengistu A, Arelli PA, Bond JP, Shannon GJ, Wrather AJ, Rupe JB, Chen P, Little CR, Canaday CH, Newman MA, et al (2011) Evaluation of soybean genotypes for resistance to charcoal rot. *Plant Health Prog*. doi: 10.1094/PHP-2010-0926-01-RS

Mihail JD, Taylor SJ (1995) Interpreting variability among isolates of *Macrophomina phaseolina* in pathogenicity, pycnidium production, and chlorate utilization. *Canadian Journal of Botany* 73: 1596–1603

Milgroom MG (1996) Recombination and the multilocus structure of fungal populations. *Annu Rev Phytopathol* 34: 457–477

Milgroom MG (2015) Population biology of plant pathogens: genetics, ecology, and evolution. APS Press, The American Phytopathological Society St. Paul, MN, USA

Milgroom MG, Jimenez-Gasco M del M, Olivares-Garcia C, Drott MT, Jimenez-Diaz RM (2014) Recombination between clonal lineages of the asexual fungus *Verticillium dahliae* detected by genotyping by sequencing. PLoS One 9: e106740

Möller M, Stukenbrock EH (2017) Evolution and genome architecture in fungal plant pathogens. Nat Rev Microbiol 15: 756–771

Ni M, Feretzaki M, Sun S, Wang X, Heitman J (2011) Sex in Fungi. Annu Rev Genet 45: 405–430

Otto SP, Gerstein AC (2008) The evolution of haploidy and diploidy. Current Biology 18: R1121–R1124

Paris RL, Mengistu A, Tyler JM, Smith JR (2006) Registration of soybean germplasm line DT97–4290 with moderate resistance to charcoal rot. Crop Sci 46: 2324

Raffaele S, Farrer RA, Cano LM, Studholme DJ, MacLean D, Thines M, Jiang RHY, Zody MC, Kunjeti SG, Donofrio NM, et al (2010) Genome evolution following host jumps in the irish potato famine pathogen lineage. Science (1979) 330: 1540–1543

Raffaele S, Kamoun S (2012) Genome evolution in filamentous plant pathogens: why bigger can be better. Nat Rev Microbiol 10: 417–430

Sarr MP, Ndiaye M, Groenewald JZ, Crous PW (2014) Genetic diversity in *Macrophomina phaseolina*, the causal agent of charcoal rot. Phytopathol Mediterr 53: 559564

Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. Nat Rev Genet 14: 807–820

Savolainen O, Pyhäjärvi T, Knürr T, Pyhäjirvi T, Kniirr T (2007) Gene Flow and Local Adaptation in Trees. Source: Annual Review of Ecology, Evolution, and Systematics 38: 595–619

Sharma R, Mishra B, Runge F, Thines M (2014) Gene Loss rather than gene gain is associated with a host jump from monocots to dicots in the smut fungus *Melanopsichium pennsylvanicum*. Genome Biol Evol 6: 2034–2049

Sharp NP, Sandell L, James CG, Otto SP (2018) The genome-wide rate and spectrum of spontaneous mutations differ between haploid and diploid yeast. Proc Natl Acad Sci U S A 115: E5046–E5055

Short G. E., Wyllie T. D., Bristow P. R. (1980) Survival of *Macrophomina phaseolina* in soil and residue of soybean. Phytopathology 70: 13–17

Stukenbrock EH, Bataillon T, Dutheil JY, Hansen TT, Li R, Zala M, McDonald BA, Wang J, Schierup MH (2011) The making of a new pathogen: Insights from comparative population genomics of the domesticated wheat pathogen *Mycosphaerella graminicola* and its wild sister species. Genome Res 21: 2157–2166

Stukenbrock EH, McDonald BA (2008) The Origins of Plant Pathogens in Agro-Ecosystems. Annu Rev Phytopathol 46: 75–100

Taylor JW, Jacobson DJ, Fisher MC (1999) The evolution of asexual fungi: Reproduction, Speciation and Classification. *Annu Rev Phytopathol* 37: 197–246

Thibaut Jombart, Sébastien Devillard FB Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. doi: 10.1371/journal.pcbi.1000455

Todd RT, Forche A, Selmecki A (2017) Ploidy Variation in Fungi: Polyploidy, Aneuploidy, and Genome Evolution. *The Fungal Kingdom*. NIH Public Access, pp 599–61

Turchin MC, Chiang CWK, Palmer CD, Sankararaman S, Reich D, Investigation G, Giant T, Hirschhorn JN (2012) Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nat Genet* 44: 1015–1019

Wilson R (2008) Soybean: market driven research needs. *Genetics and genomics of soybean*

Xuereb A, Kimber CM, Curtis JMR, Bernatchez L, Fortin MJ (2018) Putatively adaptive genetic variation in the giant California sea cucumber (*Parastichopus californicus*) as revealed by environmental association analysis of restriction-site associated DNA sequencing data. *Mol Ecol* 27: 5035–5048

CHAPTER 2

POPULATION GENOMIC ANALYSIS REVEALS GEOGRAPHIC STRUCTURE AND CLIMATIC DIVERSIFICATION FOR *MACROPHOMINA PHASEOLINA* ISOLATED FROM SOYBEAN AND DRY BEAN ACROSS THE US, PUERTO RICO, AND COLOMBIA

2.1 Abstract

Macrophomina phaseolina causes the important disease charcoal rot, which significantly reduces yield and seed quality of soybean and dry bean. Although charcoal rot has been recognized as a warm climate-driven disease of increasing concern under global climate change, knowledge regarding population genetics and climatic variables contributing to the genetic diversity of *M. phaseolina* remains limited. This study conducted genome sequencing for 95 *M. phaseolina* isolates from soybean and dry bean across the continental US, Puerto Rico, and Colombia. Inference on the population structure revealed that the isolates exhibited a discrete genetic clustering at the continental level and a continuous genetic differentiation regionally. Almost all isolates from the US grouped in a clade with a predominantly clonal genetic structure, while most Puerto Rican and Colombian isolates from dry bean were assigned to a separate cluster with higher genetic diversity. Consistently, climate significantly contributed to genomic variation at a continental level with temperature seasonality and precipitation of warmest quarter having the greatest impact. The loci significantly associated with multivariate climate were found closely to the genes related to fungal stress responses, including transmembrane transport, glycoside hydrolase activity and a heat-shock protein, which may mediate climatic adaptation for *M. phaseolina*. On the other hand, limited genome-wide differentiation among populations by hosts was observed. These findings highlight the importance of population genetics and identify candidate genes of *M. phaseolina* that can be used to elucidate the molecular mechanisms that underly climatic adaptation to the changing climate.

2.2 Introduction

Delineating pathogen populations and identifying the factors shaping the patterns of genetic diversity within and among populations allow for inferences about their biology and evolutionary potential. Plant pathogens are often genetically structured in different agricultural landscapes as a result of geographic and environmental differences (Gladieux et al., 2014; McDonald and Stukenbrock, 2016). Among different environments, agroecosystems provide remarkable conditions for rapid adaptation of plant-pathogenic fungi. The abiotic and biotic factors such as genetic crop uniformity of monocultures, the prevalent occurrence of human-mediated migration (Wingfield et al., 2015; Crous et al., 2017), and intrinsic characteristics of fungi such as their mode of reproduction (McDonald and Stukenbrock, 2016) are known

to be strong drivers of genomic divergence and adaptation in plant pathogenic fungi (Stukenbrock et al., 2011; Savolainen, Lascoux and Merilä, 2013; Croll and McDonald, 2016). However, characterizing how selective pressures of abiotic and biotic factors contribute to population genetics of plant-pathogenic fungi remains challenging.

Macrophomina phaseolina is a seed- and soil-borne fungal pathogen that infects more than 400 host species (Batista, Lopes and Alves, 2021), and causes damping off and charcoal rot in many important economic and subsistence crops worldwide, including soybean (*Glycine max*) and dry bean (*Phaseolus vulgaris*) (Dhingra and Sinclair, 1978). During host infection, *M. phaseolina* invades the xylem preventing water uptake, causing wilting and premature plant death with senesced leaves remaining attached to the petioles (Mengistu et al., 2007; Romero Luna et al., 2017). These symptoms can develop rapidly causing extensive yield loss and grain or seed quality reduction (Smith and Carvil, 1997). Charcoal rot of soybean ranked 7th out of 25 pests and pathogens causing global yield losses higher than 1% (Savary et al., 2019), with the potential for yield reductions within individual fields of up to 50% (Wrather et al., 2001). In the US, charcoal rot ranked among the top seven most destructive diseases with economic losses totaling 220 billion dollars from 2010 to 2014 (Allen et al., 2017). Disease is favored by hot and dry conditions (Dhingra and Sinclair, 1974), with colonization in the soybean and dry bean tap root and lower stem being greatest under high temperatures (28°C – 35°C) and low precipitation (Dhingra and Sinclair, 1974; Meyer and Sinclair, 1974; Kendig, Rupe and Scott, 2000; Mengistu, Arelli, et al., 2011; Mengistu, Smith, et al., 2011; Reznikov et al., 2018).

Macrophomina phaseolina is haploid, reproduces asexually, and overwinters in soil and crop residue as abundant, melanized microsclerotia that serve as the primary inoculum to initiate infection in subsequent seasons (Gupta, Sharma and Ramteke, 2012; Islam et al., 2012). Pycnidia are occasionally produced on soybean and other host plants, however, their epidemiological significance has yet to be fully defined (Knox-Davies, 1965; Dhingra and Sinclair, 1978; Mihail and Taylor, 1995; Ma et al., 2010; Gupta et al., 2012). Depending on environmental conditions, *M. phaseolina* may survive as microsclerotia in soil for up to 15 years (Short et al. 1980; Baird et al. 2003), and for up to 3 years as microsclerotia in symptomatic seeds or as mycelium in asymptomatic seeds (Hartman et al. 1999). To date, no clonal lineages or patho-

types have been identified for *M. phaseolina*, despite reports of within-species variation in morphology and pathogenicity (Dhingra and Sinclair, 1973, 1978; Sexton, Hughes and Wise, 2016). Population genetic studies based on microsatellite markers of isolates representing different geographic regions and hosts across the US have found moderate to high genetic diversity and mixed evidence of population structure by host or geography. Although considerable efforts have been focused on ascertaining host specialization, it is generally concluded that there is no strong evidence of this specificity, in which isolates from one plant species can often cause disease in other plant species (G Su et al., 2001; Zveibil et al., 2012; Romero Luna et al., 2017). Nevertheless, genetic similarity of isolates according to host and US regions and some degree of host preference have been noted (G. Su et al., 2001; Jana, Sharma and Singh, 2005; Baird et al., 2010; Saleh et al., 2010; Arias et al., 2011). Notably, a group of *M. phaseolina* isolates obtained from strawberry in California were found to form a species-specific cluster, exhibiting strong host preference for strawberry over other hosts around California (Koike et al., 2016; A. K. Burkhardt et al., 2019).

Studying population genetics using statistical methods that leverage genomic, geographic and environmental data can account for continuous and discrete genetic variation and provide insights into the genetic basis underlying environmental adaptation (Hoban et al., 2016; Bontrager and Angert, 2018; Bradburd, Coop and Ralph, 2018b). These approaches may be used to identify environmental factors driving selection and provide an understanding of how and why pathogen populations vary across space. Population genomics and genotype-environment associations have been applied in numerous studies to resolve the basis of rapid adaptation and identify candidate adaptive loci associated with environmental variation (Lasky et al., 2012; Forester et al., 2018; Xuereb et al., 2018; Gibson and Moyle, 2020; Capblancq and Forester, 2021). However, characterizing population structure and unravelling the effects of continuous or discrete processes on the genetic differentiation remains challenging for many plant-pathogenic fungi.

A major challenge arises because continuous geographic differentiation (e.g. isolation by distance or climatic variation along a gradient) can be confounded with discrete processes such as admixture and long-distance migration (human-mediated migration) which are commonly observed in plant pathogens

(Wingfield et al., 2015; Crous et al., 2017; Tabima et al., 2019; LeBlanc, Cubeta and Crouch, 2021). In addition, collinearity between spatial and environmental variables makes it difficult to elucidate to what extent geographic and environmental differences may be contributing to genetic differentiation. To address these issues, multivariate statistical methods, specifically redundancy analysis (RDA), have been increasingly used to disentangle the effects of environmental factors in shaping genetic variation. RDA is a type of constrained ordination in which a set of SNPs are modeled as responses in a function of combinations of environmental predictors. Because of its ability to evaluate many loci simultaneously, RDA has been found to be superior to traditional mixed-models associations methods in identifying weak, multilocus selection (Forester et al., 2018), suggestive of polygenic adaptation. Furthermore, partial RDA models, in which covariables can be included, has been used to account for underlying population structure in the identification of loci associated with environmental factors for climate adaptation in a variety of systems including plant and animal species (Lasky et al., 2012; Forester et al., 2018; Xuereb et al., 2018; Gibson and Moyle, 2020; Capblancq and Forester, 2021).

Climate fluctuation and temperature in particular, are important abiotic factors leading to local adaptation of plant-associated fungi (Savolainen et al., 2013; Croll and McDonald, 2016), especially in species occupying spatially and climatically heterogeneous environments (Ellison et al., 2011; Branco et al., 2015, 2016; Fitzpatrick and Keller, 2015). *M. phaseolina* is recognized for its different ecological roles as an endophyte, saprotroph, and latent or opportunistic pathogen with broad geographic distribution (Dhingra and Sinclair, 1974; Slippers and Wingfield, 2007; Slippers and Boissin, 2013; Parsa et al., 2016; Crous et al., 2017). Worldwide diseases caused by *M. phaseolina* have re-emerged in recent decades, with outbreaks occurring mostly in tropical and subtropical regions but in temperate regions as well (Leyva-Mir et al., 2015; Casano et al., 2018; Koehler and Shew, 2018; Meena et al., 2018; Nishad et al., 2018; Tančić Živanov et al., 2018; Wang et al., 2020). In the US, charcoal rot of soybean has been primarily an issue in southern states. However, more recently charcoal rot has been reported in northern states such as Wisconsin, New York, Minnesota, and Michigan (Bradley et al., 2003; Brown, 2007; Cummings and Bergstrom, 2013; Elaraby, 2003; Hughes, 2009; Yang and Navi, 2005). Although many factors may influence disease incidence, greater disease and yield losses have been observed in years with high temperature and low soil

moisture (Bradley and Allen, 2014; Allen et al., 2017). When comparing isolates from the northern and southern US states, a recent study concluded that *M. phaseolina* isolates were regionally adapted (Sexton, Hughes and Wise, 2016). Investigations in the context of species within Botryosphaeriaceae suggest that geographical distribution and host affinity dynamics in *M. phaseolina* are strongly influenced by climate due to its broad host range and ecologically diverse roles (Slippers and Wingfield, 2007; Batista, Lopes and Alves, 2021). These factors, together with future extreme rainfall and temperature predicted in the climatic change models (IPCC, 2014), make it critical to better understand the genetic structure and climatic factors as potential selection agents of *M. phaseolina*.

The broad geographic distribution and population dynamics of *M. phaseolina* suggest that populations in the continental US, Puerto Rico and Colombia might have been influenced by a complex environmental and agricultural landscape and may be structured and differentially adapted at a continental or regional level. However, understanding of the population structure of *M. phaseolina* has remained limited. In the present study, the first aim was to better understand the genetic structure in *M. phaseolina* populations isolated from soybean and dry bean across the US, Puerto Rico and Colombia using genome-wide single nucleotide polymorphisms (SNPs). Specifically, the contribution of discrete vs. continuous genetic differentiation was assessed and the hypotheses tested were *M. phaseolina* populations differentiated (i) between geography and (ii) between host within the US, using conventional and spatially explicit population structure analyses. The second aim was to investigate whether climatic variables contribute to patterns of adaptive genetic variation in *M. phaseolina*. Using RDA, the hypotheses tested were (i) specific climatic variables contribute to genetic variation, (ii) climatic variables independently contribute to patterns of genetic variation when accounting for underlying spatial and population structure, and (iii) loci in strong association with multivariate climate can be identified and have roles in driving local adaptation to climate.

2.3 Results

2.3.1 Whole-genome sequencing for 95 *M. phaseolina* isolates

Whole-genome sequences were generated for 95 *M. phaseolina* isolates collected across the US, Puerto Rico, and Colombia, including 52 soybean isolates, 40 dry bean isolates, two strawberry isolates, and one

Ethiopian mustard isolate (Fig 2.1; Supplementary Table A.1). Sequence coverage varied across individual isolates from 5X to 85X, across 93% of the *M. phaseolina* reference genome (JGI Mycocosm, MPI-SDFR-AT-0080 v1.0). A total of 2.8 million SNPs were identified across all isolates, and a mean read depth (DP) of 12X was obtained for all SNPs after filtering. Most SNPs had a mapping quality (MQ) value equal to 60 (94%) and SNPs with MQ values < 60 were removed. The distribution of missing data across the isolates and across the variants was even, with most individuals representing similar missing data (0 – 0.006%), and all variants containing missing data were removed. The final data set contained 76,981 high-quality biallelic SNPs in all isolates, and the data set was retained for all analyses.

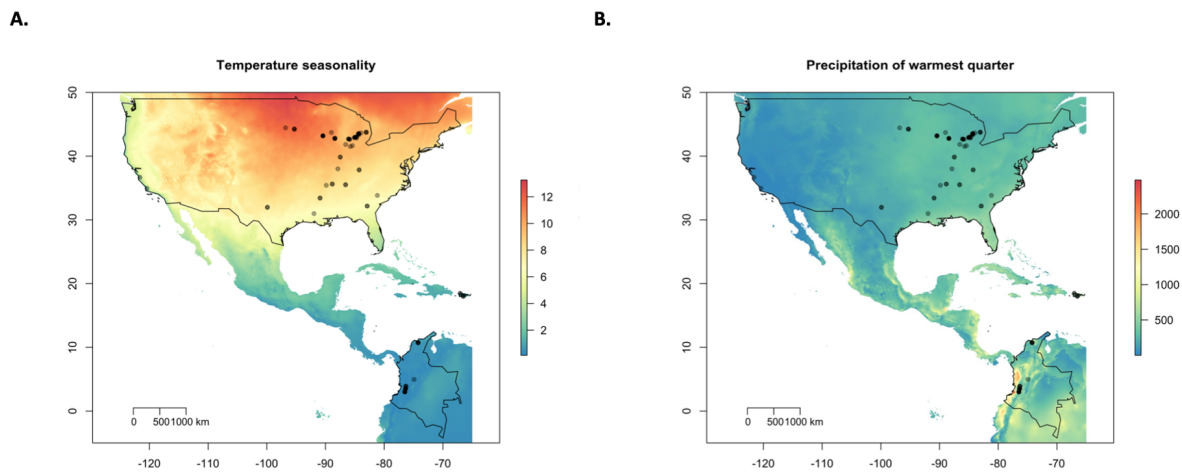


Figure 2.1 Geographic location of the 95 *Macrophomina phaseolina* isolates overlaid on temperature and precipitation variables. (A) Isolate collection sites overlaid on temperature seasonality (standard deviation; °C). Temperature seasonality contributed the most to explaining patterns of spatial genetic variation using redundancy analysis (RDA). (B) Isolates overlain on precipitation warmest quarter (mm). US, Puerto Rico and Colombia are outlined in black.

2.3.2 Phylogenomics differentiated 95 isolates into two main clades of the US and Colombian-Puerto Rican origins

To infer the genetic similarity in *M. phaseolina* isolates across the continental US, Colombia and Puerto Rico, a maximum-likelihood (ML) phylogenetic tree based on the 76,981 SNPs was constructed. Five genetic clusters were identified across the US (n=3), Colombia and Puerto Rico (n=2). Furthermore, a pattern of hierarchical structure differentiating the US and Colombian-Puerto Rican isolates was observed. The ML tree provided strong support (100% bootstrap) for two main clades, hereafter referred to

as US and COLPR, and five well-supported clades within the main clades (Fig. 2.2A). The US isolates M11-12 and M13-26 from California, and TN501 from Louisiana clustered in the COLPR clade, while the Colombian isolates Mph-22, Mph-23, and Mph-49 in the US clade (Fig. 2.2A). Other than these six isolates, all isolates from the US were placed in the US clade, and all isolates from Colombia and Puerto Rico were grouped in the COLPR clade.

There were three subclades (US1A, US1B and US2) within the US clade and two subclades (COLPR1 and COLPR2) within the COLPR clade. The PCA clustered isolates in five distinct groups in agreement with phylogenetic analysis, with little evidence of within group differentiation (Fig. 2.2B). The first PC explains most of the variance (50.6%) and separates out isolates in the US clade from the isolates in the COLPR clade, while the second PC explains 15.5% of the variance dividing isolates into the five groups in the phylogenetic analysis (Fig. 2.2B). An exception was isolate MP258, which in the PCA was grouped in US1B instead of US1A. Since the phylogenetic and PCA clustering revealed essentially the same hierarchical groupings, they were named genetic clusters US1A, US1B, US2, COLPR1 and COLPR2.

US1A isolates represented the predominant group in the US, with most isolates collected in the East North Central and Central regions in the states of Michigan (29), followed by Wisconsin (11), Indiana (5), Tennessee (5) and Kentucky (2). Cluster US1B was represented by isolates from Mississippi (2) and South Carolina (1). US2 isolates represented the second largest group in the US and were mostly collected in the West North Central [Minnesota (3), South Dakota (1)] and South [Texas (2) and Georgia (2)] regions. Also, within this cluster were isolates from Wisconsin (1), Michigan (4), and Kentucky (1). On the other hand, the COLPR1 cluster grouped most isolates from Colombia (11) and Puerto Rico (4) while COLPR2 grouped isolates from Colombia (5), one isolate from Puerto Rico, and three isolates from the US. No evidence of population structure by states was found, which indicated that states do not represent genetic groups and *M. phaseolina* is genetically structured at a broader subcontinental regional extent.

A ML phylogeny rooted with the *M. phaseolina* reference genome was reconstructed using the set of high-quality SNPs. The *M. phaseolina* reference genome was considered as a suitable outgroup based on its European and *Arabidopsis thaliana* origin. The phylogenetic reconstruction with the reference genome as a root revealed the COLPR2 clade as an outgroup to all other clades, while the US clades were

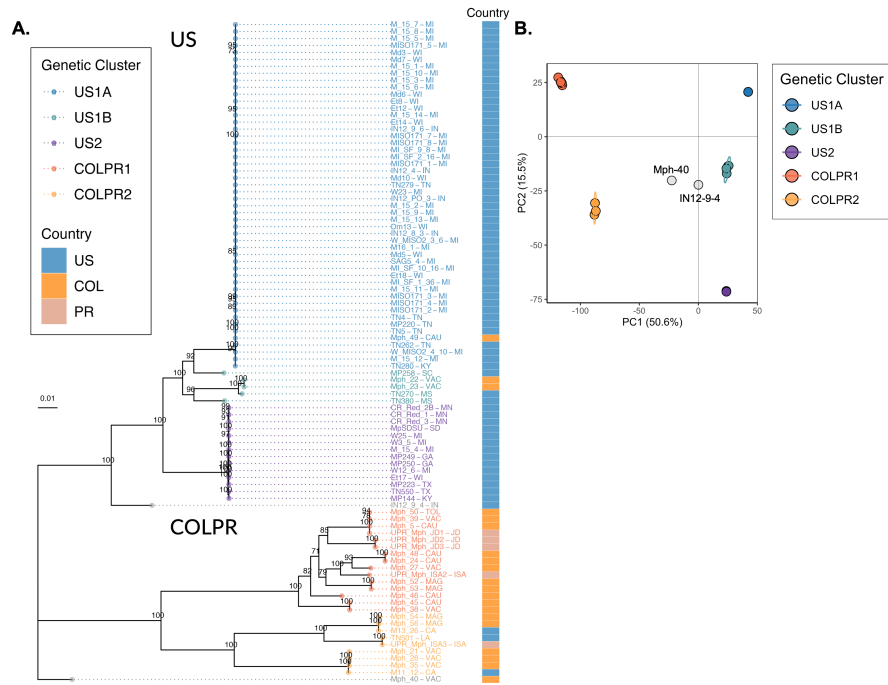


Figure 2.2 Population structure of *Macrophomina phaseolina* in the US, Colombia and Puerto Rico reveals five genetic clusters in a pattern of hierarchical structure. (A) Maximum-likelihood phylogeny reconstructed using 77,465 high-quality SNPs. Bootstrap support values over 70 are shown at nodes. Bootstrapping converged after 400 replicates. Colored tips represent the genetic cluster for each isolate as defined by principal components analysis. The two main clades, US and COLPR, are highlighted by rectangular shading. The country of collection for each isolate is denoted by colored squares at the right bar. (B) Scatterplot from a principal component analysis based on the two first PCs (the eigenvectors of the SNP dataset) for all isolates. Points are colored by membership in the five genetic clusters. Isolate names include states/municipalities codes: CA: California, CAU: Cauca, GA: Georgia, IN: Indiana, ISA: Isabela, JD: Juana Diaz, KY: Kentucky, LA: Louisiana, MAG: Magdalena, MI: Michigan, MN: Minnesota, MS: Mississippi, SC: South Carolina, SD: South Dakota, TN: Tennessee, TOL: Tolima, TX: Texas, VAC: Valle del Cauca, WI: Wisconsin. Country codes: US: United States, COL: Colombia and PR: Puerto Rico.

reconstructed as terminal clades (Supplementary Fig. A.1). The topology of the rooted ML phylogeny indicated the COLPR clades as more diverse than the major US terminal clades (US1A and US2). This higher diversity in COLPR clades was indicated by longer average branch length than in the US clades, representing a higher average number of substitutions per site. Differences in diversity can also be inferred from the PCA clustering. In PC space, 50 isolates in US1A and 14 isolates in US2 genetic clusters clustered effectively on top of each other, while isolates in US1B, COLPR1 and COLPR2, although projected near

each other, clustered distinctively more dispersed (Fig. 2.2B). The placement of COLPR genetic clusters and their higher diversity as compared to US genetic clusters indicates them as potential sources to the US clusters.

To test the relatedness of *M. phaseolina* isolates from soybean and dry bean in US, the host information was mapped to the ML tree (Supplementary Fig. A.1A). Generally, isolates that shared a common host did not cluster within genetic clusters in the US. Isolates collected from soybean and dry bean grouped together in the two larger US genetic clusters (US1A and US2; Supplementary Fig. A.1A). This lack of structure was further supported in a PCA showing overlapping ellipses representing 95% of the isolates from each of the hosts (Supplementary Fig. A.2).

2.3.3 Spatial population structure defines discrete population structure in *M. phaseolina* between the US and Colombia-Puerto Rico and continuous substructure between genetic clusters within US and COLPR clades

To infer the number of distinct genetic groups in *M. phaseolina* while accounting for continuous geographic differentiation, spatial analysis of population structure was conducted using a Bayesian (conStruct) and a model-free matrix factorization (TESS3) framework. Spatial analysis of population structure incorporates geographic distance in the estimation of ancestry coefficients (the proportion of individual isolate's genome originating from the ancestral genetic group, K). The genetic structure of the 95 isolates was explained better by a spatial model of admixture between discrete genetic groups, where isolation by distance was accounted for rather than the non-spatial model. This was indicated by the increase in predictive accuracy in the conStruct spatial models for all tested values of K (referred hereafter as layers in conStruct framework; Supplementary Fig. A.3B). This suggests that isolation by distance or climatic gradients likely play a role in shaping patterns of genetic variation in the sampled isolates.

Spatial population structure description using TESS3 returned the greatest decrease in root mean-squared errors at K=2 (0.087, from 0.318 at K=1 to 0.232 at K=2; Fig. 2.3D) and detected the US and COLPR clades. At K=2, TESS3 spatial estimation strongly assigned 95% of isolates to a single ancestral population (ancestry proportion $Q > 0.8$; Fig. 2.3A). All isolates in the US clade, except for the three isolates collected in Colombia, were identified as being derived from a single ancestral population (represented by blue; Fig. 3A, bottom). Likewise, all COLPR isolates are estimated to have a majority compo-

nent of ancestry from a single source population (represented by orange; Fig. 2.3A, bottom) including the three isolates collected in the US (M_{I11-12} and M_{I13-26} from California, and TN₅₀₁ from Louisiana). The three isolates collected in Colombia grouping in the US clade (M_{ph-22}, M_{ph-23} and M_{ph-49}) were identified as admixed (i.e., to have ancestry from more than one population instead of drawing ancestry mostly [$Q > 0.8$] from a single ancestral population) between the two ancestral groups (Fig. 2.3A, bottom) as well as the two isolates (IN_{I2-9-4} from Indiana and M_{ph-40} from Colombia) placed outside the supported clusters in the ML tree and PCA. At $K=4$, further substructure was detected that generally reflect the genetic clusters within the US and COLPR clades; except that an ancestral population for US_{1B} isolates was not inferred (Fig. 2.3B). The decrease in root mean-squared errors at $K=4$ (0.04; from 0.20 at $K=3$ to 0.16 at $K=4$; Fig. 2.3D) was the second largest value after that at $K=2$, reflecting the hierarchical structure observed in previous analyses. However, although isolates in each genetic cluster (except US_{1B}) were inferred as drawing the most ancestry from their own ancestral population, only 76% of isolates had an ancestry proportion (Q) > 0.80 to a single ancestral population (Fig. 2.3B, bottom), demonstrating weaker assignments than those at $K = 2$.

Consistently, the results from conStruct spatial model with $K = 2$ returned the greatest increase in predictive accuracy and primarily partitioned the isolates in two main groups mostly in line with US and COLPR clades (Supplementary Fig. A.3A). Based on cross-validation results, the predictive accuracy increased with increasing values of K (Supplementary Fig. A.3B), however additional layers beyond $K = 2$ contribute little to total covariance (Supplementary Fig. A.3C). Therefore, supporting two discrete ancestral populations while population substructure can be explained by continuous genetic differentiation. Taken together conStruct and TESS₃ results supported two discrete genetic groups for the US and COLPR main clades and suggested that most isolates within US and COLPR clades can be better described to have ancestry mainly from each single ancestral population. It may therefore be reasonable that the evolutionary processes leading to divergence between genetic clusters within the US (US_{1A}, US_{1B}, US₂) and COLPR (COLPR₁ COLPR₂) clades were associated to isolation by distance or climatic differences rather than different discrete ancestry.

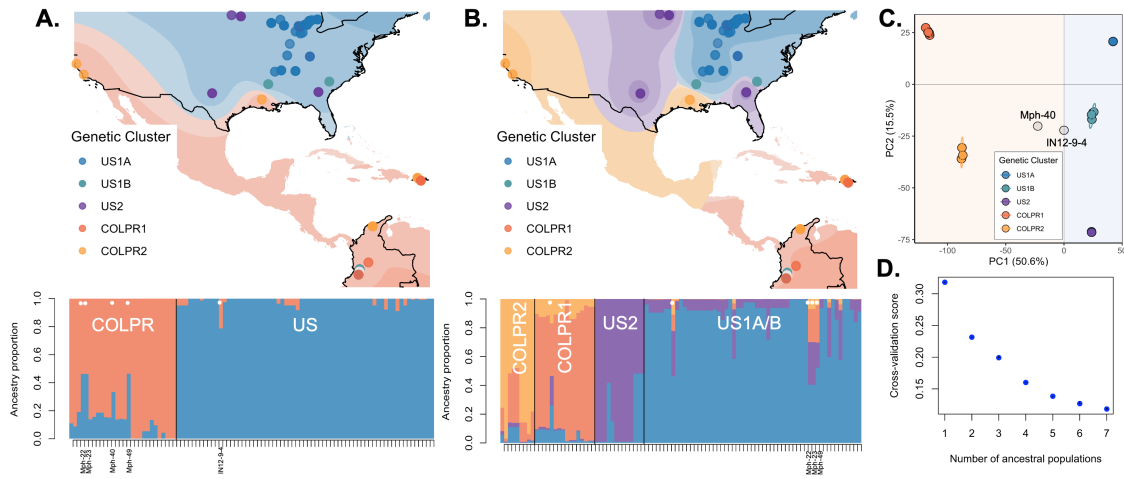


Figure 2.3 Spatial population structure defines discrete population structure in *M. phaseolina* between the US and Colombia-Puerto Rico and continuous substructure between genetic clusters. (A) Isolate membership to ancestral populations identified with TESS₃ using $K = 2$ and (B) $K = 4$. Top: Isolate collection sites overlaid on individual membership, each color representing a population. Each point represents an isolate, points are colored by their assignment to genetic clusters as identified in principal component analysis to show agreement between the methods. Bottom: Ancestry proportions (Q) of all isolates. Isolates identified as admixed (Mph-22, Mph-23, Mph-49, Mph40 and IN129-4) are labeled and indicated with dots. (C) Scatterplot from a principal component analysis for all isolates (from Fig. 2). (D) Values of the TESS₃ cross-validation criterion (root mean-squared errors, RMSE) as a function of the number of ancestral populations ($K = 1$ to $K = 7$).

2.3.4 Genetic diversity and differentiation between the US and COLPR clades and genetic clusters of *M. phaseolina*

To examine genome-wide diversity of *M. phaseolina* within and among clades and genetic clusters, we estimated gene diversity (H_e) and median pairwise genetic distance for each of the clades and genetic clusters. Pairwise genetic distance showed that COLPR isolates had greater genetic distances among isolates than those in the US clade, with a gene diversity (H_e) significantly higher in the COLPR clade (0.236) than the US clade (0.068; Table 2.1) ($H_s.test$, $P = 0.002$). Among clusters, the COLPR₂ cluster has the highest genetic diversity, considering both gene diversity and pairwise genetic distance, followed by COLPR₁, US_{1B}, US₂, and the US_{1A} cluster has the lowest values (Table 2.1). The higher genetic distance among isolates in the US_{1B} cluster as compared to other US clusters, likely reflects that the cluster is only represented by five isolates of which two were collected in Mississippi, two in Colombia and one in South

Carolina.

Table 2.1 Summary statistics for genetic diversity of *Macrophomina phaseolina* clades and genetic clusters. N is number of isolates (sample size); MLG is number of observed multilocus genotypes; eMLG is the number of expected MLG at a sample size of 25 for clades and 5 for genetic clusters based on rarefaction. MLL is number of observed multilocus lineages by population using a bitwise cutoff distance of 0.0001; CF is clonal fraction ($1 - (MLL/N)$). Clone corrected values are shown and indicated by asterisks for indices of genotypic diversity: Shannon-Wiener Index (H^*), Stoddart and Taylor's Index (G^*), Simpson's index (λ^*) and evenness (E_5^*).

Clade, Genetic Cluster	N	Gene diversity (He)	Median pairwise genetic distance	MLG	eMLG	MLL	eMLL	CF	H^*	G^*	λ^*	E_5^*
US	70	0.068	0.089	54	21.75	19	10.7	0.73	1.99	5.53	8	0.702
USiA	50	0.0001	0.00016	34	4.71	6	2.87	0.88	0.935	2.48	0.561	0.883
USiB	5	0.054	0.09	5	5	4	4	0.2	1.332	3.57	0.72	0.922
US2	14	0.001	0.00075	14	5	8	4.26	0.43	1.4	3.99	0.738	0.95
COLPR	25	0.236	0.332	25	25	15	15	0.4	2.64	13.3	0.925	0.942
COLPR1	15	0.084	0.096	15	5	9	4.43	0.4	1.456	4.24	0.756	0.962
COLPR2	9	0.151	0.265	9	5	5	3.73	0.44	1.242	3.35	0.688	0.924

Note: Summary statistics were calculated using the clone-corrected data at 79 MLGs.

To evaluate genotypic diversity both in terms of genotypic richness (the number of observed genotypes) and evenness of distribution of genotypes, the number of multilocus genotypes (MLG) was calculated for each clade and genetic cluster. A MLG was defined as a unique combination of SNPs. Given the large number of 76,981 SNPs and genotyping error rate from NGS data, it is unlikely that a true clone will be represented by an MLG. Thus, to better represent clones, closely related genotypes were collapsed into multilocus lineages (MLLs) based on a Prevosti's genetic distance threshold of 0.0001 (8 SNPs). Of the 95 isolates, 79 had unique genotypes (MLGs) corresponding to 34 MLLs (Table 2.1). eMLG and eMLL are the number of expected MLGs and MLLs based on rarefaction at the lowest common sample size between clades and genetic clusters and were used to allow comparisons across them given their unequal sample sizes. Genotypic richness was highest in the COLPR clade (15 eMLLs) as compared to the US clade (10.7 eMLLs). Among genetic clusters, the COLPR1 cluster had the highest number of eMLLs, followed by US2, USiB, COLPR2 and USiA. This indicates genotypic richness is highest in COLPR1 and lowest in the USiA genetic cluster, in which more than 80% of the isolates were clonal (Table 1, CF). Although, lower genotypic richness is inferred in COLPR2 and USiB as compared to the gene diversity pattern, this may be due to their low sample size. Evenness and the corrected Shannon-Wiener's index,

Stoddart and Taylor's index and Simpson's Index, were all highest in the COLPR clade than in the US clade and followed the same pattern among genetic clusters as with genotypic richness (Table 2.1). Finally, there were no shared MLGs or MLLs among genetic clusters.

Similarly, between countries, significantly higher gene diversity in Colombia (0.263) compared with the US (0.104) (Hs.test, $P = 0.002$). Gene diversity in Puerto Rico (0.163) was intermediate and not significantly different from the US (Hs.test, $P = 0.218$) or Colombia (Hs.test, $P = 0.396$). Pairwise genetic distances, corrected genotypic diversity indices and evenness calculated for each country follow the same pattern of gene diversity (Supplementary Table A.3). To infer migration among countries by tracking genotype flow, MLLs shared among countries were identified. In total three MLLs were shared among countries. The MLL with one isolate from Colombia (Mph-5) and one from Puerto Rico (UPR-Mph-JD1) clustering in COLPR1, the MLL with one isolate from Puerto Rico (UPR-Mph-ISA3) and one from Louisiana (TN501) clustering in COLPR2, and the MLL with one isolate from Colombia (Mph-49) and 19 isolates from US clustering in US1A (Supplementary Fig. A.4). In addition, all populations clustering approaches indicated that Colombian isolates Mph-22 and Mph-23 are the most closely related to the US isolates clustered in US1B, and Californian isolates M13-26 and M11-12 are the most closely related to Colombian isolates clustering in COLPR2. The rooted ML tree indicated isolate Mph-40 (from Colombia) as an outgroup to US clusters and discriminatory analysis of principal components (DAPC) clustered this isolate along with IN12-9-4 (from Indiana) with US1B isolates (Supplementary Fig. A.1). Overall, migration between Colombia, Puerto Rico and US is a likely scenario. To test the hypothesis that genetic clusters of *M. phaseolina* are differentiated, we used hierarchical analysis of molecular variance (AMOVA) and Nei's G_{ST} (an F_{ST} -analogous genetic differentiation measure applicable to haploids). Populations were significantly differentiated among clades, genetic clusters, as well as within genetic clusters ($P < 0.001$; Supplementary Table A.2). AMOVA revealed that most of the total genetic variance was partitioned among US and COLPR clades (47%) and among genetic clusters (42%), and only 11% within genetic clusters. Consistently, very high genetic differentiation was found between US and COLPR clades ($G_{ST} = 0.45$) and among genetic clusters ($G_{ST} = 0.50 - 0.99$; Table 2.2). The COLPR2 ($G_{ST} = 0.50-0.69$) and US1B ($G_{ST} = 0.54-0.69$) clusters had the lowest G_{ST} when compared with any

other cluster. Differentiation was lowest between COLPR₁ – COLPR₂ (GST = 0.50) clusters, and US₁A – US₁B (GST = 0.54) and highest between COLPR₁ – US₁A (GST = 0.80), COLPR₁ – US₂ (GST = 0.81) and US₁A – US₂ (GST = 0.99).

Table 2.2 Population differentiation using Nei's GST pairwise genetic dissimilarity between genetic clusters identified in *Macrophomina phaseolina*.

Genetic Cluster	US ₁ A	US ₁ B	US ₂	COLPR ₁	COLPR ₂
US ₁ B	0.54				
US ₂	0.99	0.64			
COLPR ₁	0.80	0.69	0.81		
COLPR ₂	0.68	0.59	0.69	0.50	

All other pairwise comparisons had similar intermediate levels of genetic differentiation when compared to any other genetic cluster (GST = 0.63-0.69). The high values of GST in all pairwise comparisons suggest very high differentiation and little migration between genetic clusters. However, US₁A – US₂ GST estimation, which is notably high, was limited in power due to the low levels of gene diversity (H_{exp}) within these genetic clusters. Across the 77,465 loci, there were only 76 and 255 polymorphic loci within US₁A and US₂ clusters, respectively. Thus, low gene diversity (H_{exp}) in US₁A and US₂ subpopulations likely resulted in overestimation of GST in pairwise comparisons of US₁A and US₂ with all other clusters.

2.3.5 *M. phaseolina* is predominantly clonal in the US and semi-clonal to mostly-clonal in Colombia and Puerto Rico

The predominantly star-like topology with little reticulation, in the Neighbor-Net network analysis, is consistent with a clonally reproducing population (Fig. 2.4A). The standardized index of association (I_A) (Brown et al. 1980) was used to estimate the degree of clonality for each of the *M. phaseolina* main populations (US and COLPR clades). The observed I_A distributions for each population were compared to I_A distributions for simulated populations with no linkage, 25%, 50%, 75% and 100% linkage. A predominantly clonal mode of reproduction was inferred in the US and COLPR populations of *M. phaseolina*. The simulated distributions and the different populations were significantly different from each other (analysis of variance ANOVA $df = 6$, $F = 25287$, $P < 0.001$). The distribution of the standardized I_A for

the US population fell within the 75% to 100% range of the linkage simulation (Fig 2.4B). This indicates a mostly clonal mode of reproduction with little potential for recombination. The distribution of the standardized I_A for the COLPR population fell within the 50 to 75% range of the linkage simulation, indicating semi-clonal to mostly clonal reproduction in COLPR clades (Fig. 2.4B). To further investigate the extent to which populations reproduce clonally, the linkage disequilibrium (LD) decay, as measured by the squared correlation coefficient (r^2) was calculated across pairs of loci for each of the clades. LD extends across a much larger distance in the US clade than in the COLPR clade, decaying over the first thousand base pairs, while in the COLPR clade LD decayed over the first hundreds of bases. LD half-decay distance, calculated as the average physical distance over which r^2 decays to half of its initial value was 4000 bp for US clade and 800 bp for COLPR clade (Fig. 2.4C). This indicates a high level of linkage occurs over larger regions of the genome in the US clade versus the COLPR clade. Importantly, although this may provide evidence for less clonal reproduction and higher recombination rates in the COLPR population, interpretation of standardized I_A and LD decay as associated with the frequency of recombination should be done with caution. It is possible that higher LD values did not reflect greater recombination; instead, it may be affected by lower sample size in COLPR and lower diversity in the US clade.

2.3.6 Climate contributes to SNP variation between *M. phaseolina* genetic clusters

To test the hypothesis that climate variation contributes to genetic variation across *M. phaseolina* genetic clusters a redundancy analysis (RDA) was employed. Four climatic variables were identified as significantly predictive of genetic variation using the simple RDA model with forward variable selection. Temperature seasonality (TSsd) was the strongest predictor, explaining 28% of the variation, followed by precipitation of warmest quarter (Pwq), precipitation seasonality (PScv) and mean temperature of warmest quarter (mTwq) (Table 3). Importantly, the climatic variables included in the RDA model were selected by their biological significance and to avoid collinearity with other climatic variables and thus represent a subset of the variables possibly contributing to climate variation. The correlation of these variables with the first two RDA axes suggests their differential contribution to SNP variation among genetic clusters (Fig. 2.5). Spatial structure, represented as distance-based Moran's eigenvectors maps (db-

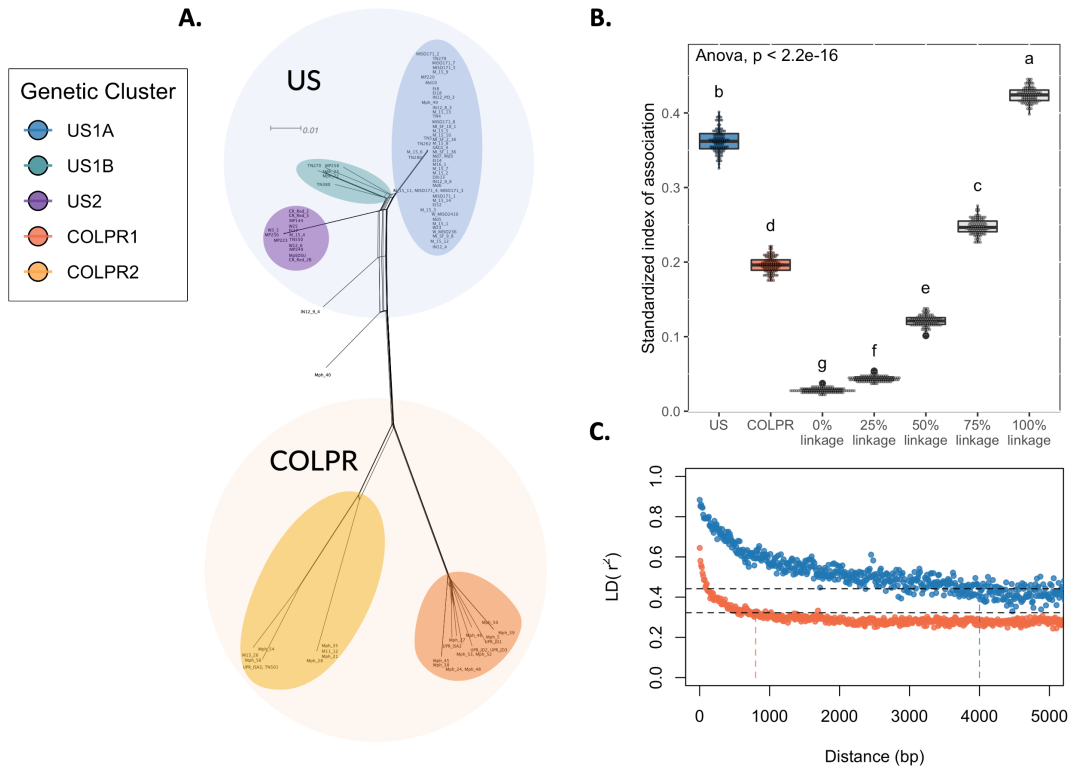


Figure 2.4 *Macrophomina phaseolina* population structure is potentially driven by clonal expansions and rapid divergence. (A) A reticulating phylogenetic network. Neighbor Net method was used to depict conflicting phylogenetic signal. (B) Estimates of linkage disequilibrium for *Macrophomina phaseolina* clades based on observed and simulated distributions of the standardized index of association (I_A). Each boxplot represents the observed distribution of I_A for one of the clades of *M. phaseolina*, compared with the distribution of I_A values for simulated populations with no linkage and 25, 50, 75, and 100% linkage. The letters above each boxplot represent groupings based on Tukey's HSD test. (C) Linkage disequilibrium (LD) decay for predicted populations of *M. phaseolina*, as measured by the squared correlation coefficient (r^2) for all pairs of SNPs calculated over 50 bp windows shown for each population. The dotted black lines give the r^2 decay to half its initial value ($r^2 = 0.44$ and 0.32 in US and COLPR clades, respectively) and the vertical lines indicate the LD half-decay distance for each clade.

MEM), was used to identify climatic variables that are structured in space and to account for the effect of space in variance partitioning of total genomic variation. A total of three spatial variables were identified (dbMEM₁₋₃; Supplementary Fig. A.5). Notably, when accounting for spatial structure (dbMEM₁₋₃ variables), only Pwq, mTwq and precipitation of driest quarter (Pd_q) were significant and accounted for 6% of SNP variation across isolates as determined with forward selection (Table 3), indicating collinearity

between TSsd, PScv and space (i.e., spatially structured TSsd and PScv variation). To identify the spatial variables significantly contributing to genomic variation forward selection was used. Of the three spatial variables, only dbMEM3 was significant explaining 4% of the genomic variation and described broad-scale spatial structure (Supplementary Fig. A.5)

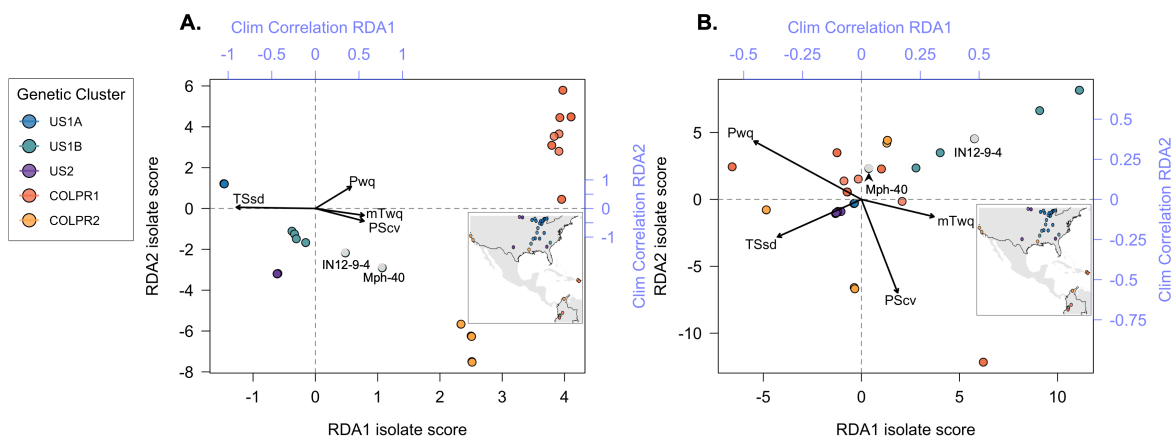


Figure 2.5 Genotype-environment association analyses support the contribution of climate variables to patterns of divergence among *Macrophomina phaseolina* populations across the US, Colombia, and Puerto Rico. Biplot of all isolates scores for the first two RDA axes using (A) Simple RDA (unconditioned) and (B) Partial RDA (conditioned on neutral population structure). Points are colored to show agreement with genetic clusters identified in the PCA (inset). Top and right axes (blue) indicate the correlation of each climate variable with RDA axes 1 and 2, respectively.

Table 2.3 Climatic variables significantly contributing to SNP variation as determined by forward variable selection with simple RDA (redundancy analysis) and partial RDA conditioned on space.

Simple RDA (unconditioned)						Partial RDA (conditioned on space)				
Variable	R ²	Cum R ²	Cum R ² adj	F-value	p-value	Variable	Cum R ² adj	AIC	F-value	p-value
TSsd	0.28	0.28	0.27	36.63	0.001***	Pwq	0.03	638.88	4.96	0.002**
Pwq	0.05	0.33	0.31	6.28	0.001***	mTwq	0.04	637.67	3.06	0.004**
PScv	0.03	0.36	0.34	4.42	0.001***	Pdq	0.06	636.50	2.98	0.010**
mTwq	0.02	0.38	0.36	3.52	0.005**					

***p 0.001, **p 0.01

Partial redundancy analysis (pRDA) was used to estimate the partial contribution of each set of explanatory variables (e.g., climate) while removing the effect of the remaining variable sets (e.g. neutral

population structure and space). Variance partitioning with pRDA revealed that climate (TSsd, Pwq, PScv and mTwq identified by forward selection), neutral population structure (isolate PC scores for the first three axes of a PCA using intergenic SNPs) and space (dbMEM₃ variable identified by forward selection) together significantly explained 72% of the total SNP variance. Nearly half of this variance was uniquely attributable to neutral genetic structure (32%), climate (4%), or space (1%), while the other half of the SNP variation was explained jointly between the three sets of variables (Table 2.4). The effect of climate alone was highly significant and explained 4% of the total genetic variance after removing the effects of neutral population structure and space (Table 2.4). These results support the hypothesis that climate significantly contributes to genetic variation and importantly, suggests that migration, drift, and potentially additional demographic and spatially structured processes (e.g isolation by distance), represented by neutral population structure, play a major role in shaping genomic variation in *M. phaseolina*. Moreover, the large fraction of variation common to climate, population structure and space, emphasizes the importance of accounting for confounded effects in genotype-environment associations, particularly when inferring causal associations.

Table 2.4 Contribution of climate, neutral population structure and space to SNP variation as determined by variance partitioning with partial RDA (redundancy analysis).

Partial RDA model	Inertia (variance)	R ²	p-value	Proportion of explainable variance	Proportion of total variance
Full model: G ~clim. + sp. + struct.	863.8	0.725	0.001***	1.00	0.72
Pure climate: G ~clim. (sp. + struct.)	46.2	0.039	0.001***	0.05	0.04
Pure structure: G ~struct. (clim. + sp.)	387.0	0.325	0.001***	0.45	0.32
Pure space: G ~sp. (clim. + struct.)	6.3	0.005	0.001***	0.01	0.01
Confounded climate/structure/space	424.4			0.49	0.36
Total unexplained	327.9				0.28
Total inertia	1191.7				1.00

***p 0.001

Note: Climate variables are temperature seasonality (TSsd), precipitation of warmest quarter (Pwq), precipitation seasonality (PScv) and mean temperature of warmest quarter (mTwq) as identified with forward selection.

2.3.7 Genotype-environment associations identify candidate SNPs for climatic adaptation

To identify loci that are potentially involved in local adaptation to climatic conditions, SNPs strongly associated with climatic variables were identified using RDA with and without accounting for population

structure. Neutral population structure was used as it uniquely contributed the most to genetic variation. The RDA models, whether accounting for population structure (partial RDA) or not (simple RDA), were globally significant ($p < 0.001$) and the first three RDA axes explained most of the genomic variation associated with climate.

The candidate adaptive loci were identified based on extreme SNPs loadings, ± 3 or ± 4 SD from the mean, on each of the first three axes (Forester et al., 2018). In the partial RDA models, in which the effects of population structure were removed, 49 unlinked SNPs (when using the LD-filtered set and ± 3 SD from the mean; Supplementary Table A.4) and 75 SNPs (using all SNPs and ± 4 SD from the mean; Supplementary Table A.5) strongly associated with climatic variables were identified along the first three RDA axes. Of these SNPs, 15 and 25 (outliers in Fig. 2.6) were identified in the first RDA axis when using the LD-filtered set or all SNPs, respectively, and 20 (19%) in both partial models. The strongest associations include SNPs with predicted effects in the membrane-associated 753275-ankyrin, the 681752-Ksh1 and the 241776-protoporphyrinogen oxidase proteins. Structural modeling of the 753275-ankyrin protein revealed that 598 residues (96% of the sequence) was modelled with 100% homology confidence to the transient receptor potential (TRP) NOMPC (No mechanoreceptor potential C) mechanotransduction channel protein in *Drosophila melanogaster* (chain C, highest scoring template; PDB ID: 5VKQ; data not shown). Other SNPs with top associations are located within or in physical proximity to genes related to transmembrane transport, glycoside hydrolase activity, DNA binding and the gene encoding the 28417-heat shock protein (Table 2.5; Supplementary Table A.6).

Because population structure could not be fully disentangled from climate, as revealed in variance partitioning, the candidate loci obtained with population structure correction represent a conservative set subjected to a reduction in the detection of SNPs truly associated with climate. In the simple RDA model, without correcting for population structure, 91 candidate unlinked SNPs were identified (Supplementary Table A.7). Only two SNPs were identified by both partial RDA and simple RDA models using unlinked SNPs (Supplementary Fig. A.6). This is in line with the high level of collinearity observed between genetic, space and climate (Table 2.4), and highlights the importance of accounting for confounded effects when identifying candidate loci under selection with genotype-environment associations.

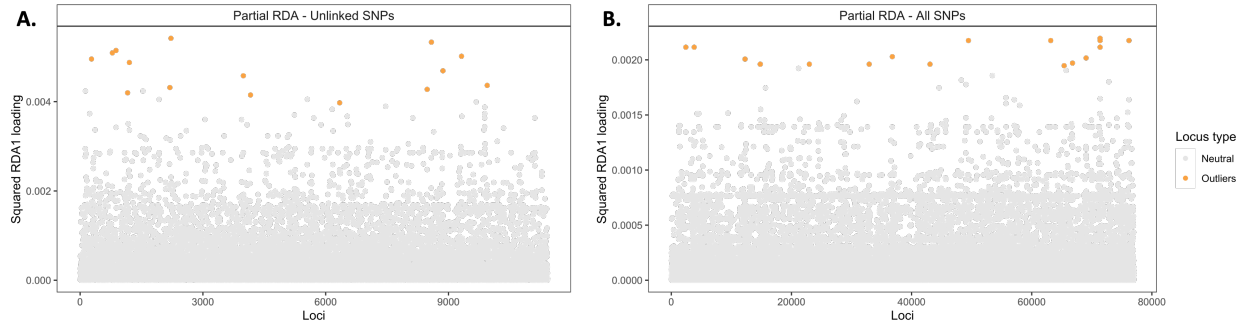


Figure 2.6 Manhattan plot of partial RDA scores. Values of squared SNP loadings for the first RDA axis conditioning on neutral population structure. (A) Fifteen outlier SNPs identified using using 11,421 unlinked SNPs and ± 3 SD from the mean and (B) Twenty-five using all 76,981 SNPs and ± 4 SD from the mean.

Table 2.5 Candidate SNPs and gene models along the first RDA axis, after accounting for neutral population structure using the LD-filtered set of 11,421 SNPs.

SNP position	RDA1 loading	Climate variable	Correlation	SNP category	Distance from locus (bp)	Mycocosm gene location	Mycocosm protein ID	InterPro/KOG Desc	KOG Class/Putative function
scaffold_13:111451	-0.068	TSsd	0.90	Intergenic	3150	scaffold_13:108068-110230	753275	Ankyrin repeat	Cell wall/membrane/envelope biogenesis
scaffold_44:156076	-0.068	TSsd	0.89	Intergenic	3974	scaffold_44:152589-153590	681752	Ksh(Protein kish)	Involved in the early part of the secretory pathway
scaffold_3:1825169	0.070	TSsd	0.62	Missense	0	scaffold_3:1824304-1825956	241776	Protoporphyrinogen oxidase	Coenzyme transport and metabolism/Heme biosynthesis
scaffold_3:1762959	-0.065	TSsd	0.62	Intergenic	1106	scaffold_3:1763965-1765572	763945	None	Unknown
scaffold_6:1672095	-0.074	TSsd	0.61	Intergenic	755	scaffold_6:1672555-1674150	726065	None	Unknown
scaffold_13:484791	-0.064	TSsd	0.59	Intergenic	563	scaffold_13:485238-487511	88264	Transcription factor domain, fungi	DNA binding/Zinc ion binding (Zn(II) ₂ Cys6 transcription factor-related)
scaffold_1:2854576	-0.070	TSsd	0.56	Synonymous	0	scaffold_1:2854139-2855202	36408	GXWVG domain	Unknown/Putative transcription factor Cmr1 homolog
scaffold_59:25310	-0.066	TSsd	0.51	Synonymous	0	scaffold_59:24613-27244	365205	AMP-dependent synthetase/ligase	Lipid transport and metabolism
scaffold_22:420202	-0.063	TSsd	0.50	Intergenic	4892	scaffold_22:414577-415638	787628	DUF1772 family	Unknown
scaffold_48:217394	-0.071	TSsd	0.45	Intergenic	4970	scaffold_48:211321-213431	633816	Glycoside hydrolase, family 1	Carbohydrate transport and metabolism
scaffold_3:931936	0.072	TSsd	0.39	Intergenic	1176	scaffold_3:93112-93948	735628	Allergen V5/TPX-1-related, conserved site	Unknown
scaffold_6:1470614	0.066	TSsd	0.37	Intergenic	14457	scaffold_6:1454967-1456573	382155	Ribonuclease T2-like	RNA processing and modification
scaffold_2:1936912	0.071	TSsd	0.35	Intergenic	443	scaffold_2:1936361-1937105	643334	Thioesterase superfamily	Unknown
scaffold_42:203393	0.073	mTwq	0.29	Synonymous	0	scaffold_42:202488-204519	608404	Cytochrome P450, E-class, group I	Lipid transport and metabolism
scaffold_41:309748	0.065	mTwq	0.29	Intergenic	4841	scaffold_41:303593-305479	582790	Flavin-containing monooxygenase	Secondary metabolites biosynthesis, transport and catabolism

2.4 Discussion

In this study, we describe the population structure of *M. phaseolina* in the continental US, Puerto Rico and Colombia collected from soybean and dry bean fields and the contributions of climatic factors to patterns of genomic diversity among populations. We found that five distinct genetic clusters of *M. phaseolina* evolved across the US, Colombia and Puerto Rico and evidence suggests migration between genetic clusters and countries. To date, population genetic studies in *M. phaseolina* have performed their analyses at the resolution of microsatellites molecular markers and have provided important information on genetic diversity, host and geographic associations in the US (Baird et al., 2010; Arias et al., 2011; Koike et

al., 2016). However, no population-level genomic studies have been conducted to investigate population structure in this widespread pathogen. Here, to our knowledge, we present the first population genomics study to investigate population dynamics and the role of climate in shaping patterns of genomic variation in *M. phaseolina* at a continental and regional scale. This study uses population genomics data to identify multiple strongly differentiated genetic lineages in the US and demonstrated novel population structure in Colombia and Puerto Rico, which previously remained unstudied. Furthermore, our results highlight the importance of within-species genetic variation in understanding pathogens adaptive response to a changing climate and offers new insight with respect to the functional roles of genomic regions potentially underlying adaptation to climate. Notably, this research provides a practical framework for genotype–environment associations studies in *M. phaseolina* and other plant pathogens with complex evolutionary and demographic histories.

The influence of the low number of loci on limiting inferences about *M. phaseolina* population structure is emphasized by recent studies that used microsatellites markers (Baird et al., 2010; Arias et al., 2011). These studies identified genetic groups in the US; however, the genetic groups did not represent lineages (i.e., genetic groups and supported phylogenetic clades). Using population genomics, we provided strong evidence for five distinct genetic clusters of *M. phaseolina* and revealed that genomic variation in this globally distributed pathogen was consistent with a population hierarchically structured at a broad subcontinental regional extent. Two genetically differentiated *M. phaseolina* populations at the US and Colombian-Puerto Rican geographical level (US and COLPR clades) and five distinct genetic clusters representing finer population structure within each of these clades were identified. These genetic clusters, except for USiB, represent strongly supported phylogenetic clades and monophyletic groups, and likely represent different evolutionary lineages of *M. phaseolina*. This distinction is important because the identification of lineages allows the inference of ecological and evolutionary processes in a population-specific manner and underscores the potential for local adaptation in *M. phaseolina* populations.

Our results provide support for regional clustering within the US and a lack of strong grouping at a state level, also observed in previous studies based on microsatellite data (Baird et al., 2010; Arias et al., 2011). The USiA cluster, found in the East North Central and Central region, expands previous studies

confirming that isolates collected from soybean in these regions represent a largely homogeneous population (Arias et al., 2011). This is supported by low gene diversity and pairwise genetic distances found in the US_{1A} genetic cluster in agreement with low diversity detected with microsatellite markers in soybean isolates collected mostly in Tennessee and Missouri (Arias et al., 2011) and midwestern states (group III; Baird et al., 2010). The US₂ genetic cluster found in West North Central and South US regions grouping isolates from Minnesota, South Dakota, Texas, and Georgia is partially consistent with Baird et al. study. Isolates from these states along with isolates from North Dakota represent the majority of a subcluster of group I in Baird et al. Like in the US clusters, grouping at broad geographic regions was observed in COLPR₁ and COLPR₂ clusters. Both COLPR₁ and COLPR₂ clusters grouped isolates from locations across Colombia and Puerto Rico. In COLPR₂, isolates from California and Louisiana grouped closely to isolates from Colombia and Puerto Rico. Although the small sample size from these states (only two isolates collected from strawberry in California and one isolate from soybean in Louisiana) demands that this grouping be reassessed once more isolates are included from these states and hosts in future studies. The clustering of isolates from widespread geographic regions observed in COLPR₂, as well as in US_{1A} and US_{1B} clusters, suggests a role for migration in structuring *M. phaseolina* populations. These results better align our understanding of *M. phaseolina* population structure with a metapopulation model, that predicts regional persistence of populations while local populations are unstable and connected by some level of migration (Hanski, 1998; Milgroom, 2015). The metapopulation dynamics view expands the interpretation of past *M. phaseolina* population structure studies while providing a conceptual basis for the design of future studies.

The presence of multiple distinct genetic clusters in the US and higher genetic diversity in COLPR clusters led us to inquire about whether Colombia and Puerto Rico may serve as potential source populations for US populations. In the rooted ML phylogeny, the reconstruction of COLPR clusters as outgroups to US clusters support this hypothesis. Furthermore, across all analyses we found indications that US_{1B} may serve as a sink population for Colombia and Puerto Rico populations. The US_{1B} genetic cluster grouped isolates from Mississippi and South Carolina along with two Colombian isolates and was the most genetically diverse of the US clusters. Further, US_{1B} was positioned centrally in PCA space,

basal to US_{1A} cluster in the rooted ML phylogeny and was less differentiated, along with COLPR₂, from all other clusters based on GST values. Finally, in DAPC analysis, US_{1B} isolates clustered with IN₁₂₋₉₋₄ and Mph-40 isolates, which are reconstructed intermediate between US and COLPR clades in the rooted ML phylogeny and as admixed in spatial population structure analyses. Although, the high diversity in US_{1B} may be reflective of the grouping of comparatively few isolates from different geographic regions in this cluster. However, when all data are considered, it suggests the US_{1B} cluster geographic region as a potential route of introduction of isolates from Colombia or Puerto Rico to the US. More isolates from the US and other countries would need to be included in future studies to test this hypothesis.

The discrete population structure observed between US and COLPR clades, provides compelling evidence for isolates in each clade drawing ancestry from different ancestral populations. A plausible explanation, supported by our results, for this different ancestry would be a demographic event such as a rare long-distance migration (e.g. introduction event) from the COLPR clusters, leading to a recent bottleneck in the US populations. The high probability assignments observed in US clusters may be consistent with the expected strong recent genetic drift in bottlenecked populations (Lawson, van Dorp and Falush, 2018). In this scenario, we speculate that the diversity in US clusters represent a subset of the diversity of the COLPR genotypes found in Colombia and Puerto Rico. At the finer genetic cluster population structure, isolation by distance provided a potential explanation for the continuous genetic differentiation in spatial population structure analyses. Although, isolation by distance patterns may be observed as part of a variety of underlying biological processes and demographic scenarios (Sexton, Hangartner and Hoffmann, 2014; Milgroom, 2015), it is possible that these patterns reflect a scenario of restricted dispersal in the context of divergence following clonal expansions in the US genetic clusters. For example, both US_{1A} and US₂ genetic clusters are found in Michigan, Wisconsin, and Kentucky, supporting dispersal of isolates among these states. However, high population differentiation indicated by high GST values between genetic clusters, suggest substantial restriction to gene flow. Given the soilborne nature of *M. phaseolina* and limited natural dispersal ability but high potential for anthropogenic mediated dispersal, restricted events of dispersal associated to seed, plant material or farm equipment at limited distances relative to the geographic range of the genetic clusters, seems a likely occurrence (Baird et al., 2010). Similar

isolation by distance patterns has been observed in other soilborne fungal and oomycete pathogens with restricted long-distance dispersal (Grünwald and Hoheisel, 2006; Milgroom et al., 2016).

Diversity was found to be further reduced in US1A genetic cluster as compared to all other clusters. Low diversity and high differentiation are signatures of genetic drift but also selection. If reduced diversity in the US1A genetic cluster was consistent with a clonal expansion following a bottleneck, the divergence and marked low diversity could reflect both genetic drift and selection. Genetic drift is expected to have substantial effects on pathogen populations, because migrations resulting in founder effects and reduced population sizes associated with pathogens survival in soil (Milgroom, 2015). Additionally, we speculate that climatic conditions, particularly strong fluctuations in temperature in the northern US, could impose strong selection on *M. phaseolina* populations in this region. Overall, we believe the genomic signals of discrete and continuous structure that differentiate *M. phaseolina* populations could be reflective of a complex demographic and evolutionary history. Therefore, alternative demographic scenarios, including one of multiple independent introductions, should be considered in future studies ideally applying demographic modelling with a broad geographic and temporal distribution of isolates.

Across all analyses we found support for Colombia and Puerto Rico as potential sources for US *M. phaseolina* populations. Genetic diversity between countries also supported this hypothesis. Whereas Colombian isolates were significantly more diverse than US isolates, diversity in Puerto Rico was intermediate and not significantly different from US or Colombia. These findings may be consistent with the idea of Middle or South America as putative centers of origin for *M. phaseolina* and with its introduction to North America as part of historical crop migrations. For example, common bean Middle American origin, domestication centers in Middle America and South America (Bitocchi et al., 2017) and later movement to the US via the Caribbean, Central and Eastern US (Kelly, 2010), makes likely an explanation for *M. phaseolina* introduction to the US in bean seeds. Pathogen geographic origins have been associated with the centers of diversity of their major crop host. Nonetheless, pathogen origin associated with their hosts' wild relatives, have been also observed in some plant pathogens. For example, a *P. infestans* genetically diverse and sexually reproducing population was found in central Mexico consistent with this pathogen's origin in a secondary center of potato (*Solanum tuberosum*) diversity and potentially

involved in a host jump from native *Solanum* species (Goss et al., 2014). Given *M. phaseolina* host generalist nature, a strict host-pathogen coevolution scenario is not expected (Slippers and Wingfield, 2007), obscuring inferences about its center of origin. In Kansas, isolates collected from wild tallgrass prairie were found more diverse than isolates from maize, soybean and sorghum crops (Saleh et al., 2010). This finding may indicate *M. phaseolina* presence in the US precedes to the introduction of agriculture or it may be explained by connectivity dynamics between natural and agricultural ecosystems contributing to patterns of diversity in *M. phaseolina* populations from these ecosystems (Saleh et al., 2010). Thus, the origin and evolutionary history of *M. phaseolina* is likely more ancient and complex than could be tested with the isolates included in this study, and future studies may benefit from considering the potential involvement of host adaptation from wild hosts.

Genotype tracking provided compelling evidence for migration among the US, Colombia, and Puerto Rico. The MLL consisting of the Colombian isolate Mph-49 and several isolates from the US clustering in US1A, along with the high clonality found in this cluster and the significantly high diversity in Colombia, makes a Colombian source likely. Similarly, the MLL shared between Colombia and Puerto Rico and the MLL between Puerto Rico and Louisiana support migration between countries. Alternatively, the same MLLs could have been introduced independently to US, Puerto Rico, and Colombia, potentially from an ancestral and more diverse population not included in this study. Although this scenario seems less likely, it remains a possibility. Given that besides historical crop migrations, migration as part of international seed exchange is a likely occurrence in *M. phaseolina*, as in other seedborne species and latent pathogens of the Botryosphaeriaceae family (Sakalidis et al., 2013; Crous et al., 2017), we believe that *M. phaseolina* has been spread at least intercontinentally, possibly globally, through seed. However, time, frequency, and directionality of migration between US, Colombia, and Puerto Rico, and the potential for multiple introductions would need to be examined in future studies.

Although various population genetic studies in *M. phaseolina* have found patterns of host associations (Jana, Sharma and Singh, 2005; Baird et al., 2010; Arias et al., 2011; Koike et al., 2016; Reznikov et al., 2018; A. Burkhardt et al., 2019), our results did not find that genetic variation is associated with host in the two major US clusters. Soybean and dry bean isolates grouped together in US1A and US2 clus-

ters. Given that most previous studies support some degree of host preference, and genomic evidence for genes uniquely present in the *M. phaseolina* strawberry genotype further support host preference (A. K. Burkhardt et al., 2019), we suspect that our sampling scheme was not enough to capture clear associations to plant host. A clear limitation in our study was that the host origin was confounded with geographic origin, except for Michigan where isolates were sampled from both soybean and dry bean. The grouping independently of host might also reflect crop rotation and equipment practices implemented in fields. Additionally, it may reflect that the sampled hosts are both legumes. Genetic similarity has been found to be greater among isolates collected from the same host than from hosts in different families (G. Su et al., 2001; Saleh et al., 2010). These results do, nonetheless, have important practical implications for soybean breeding resistance to charcoal rot. In the US1A cluster, the high genetic similarity of isolates collected from soybean and dry bean, may indicate that the use of one or few isolates collected from these crops throughout East North Central and Central US regions may suffice for resistance screening of soybean breeding material. An important limitation to this assumption is that we use a single reference genome approach to characterize genetic diversity and thus accessory genes and other structural variation potentially involved in pathogenesis are not considered (Bertazzoni et al., 2018).

Importantly, the dry bean diversity in research plots from which Colombian and Puerto Rican isolates were collected is a factor likely contributing to their higher genetic diversity as compared to US isolates. In research plots, multiple lines are continually evaluated as part of breeding programs, in contrast to commercial fields in which a single or few varieties are used. This coupled with climatic conditions in Colombia and Puerto Rico that favor year-round inoculum presence in crop residue represent important considerations when interpreting isolate genetic diversity in relation to host origin.

The population structure results suggest that *M. phaseolina* populations lay in-between the clonality-recombination spectrum (Smith et al., 1993). Furthermore, our results suggest that this may occur in a population-specific manner. On one side of the spectrum, we found *M. phaseolina* to have a markedly clonal population structure (Milgroom, 2015). First, most of the intraspecific genetic variation in *M. phaseolina* is explained by differences between clades and genetic clusters, while low genetic variation was observed within genetic clusters. Second, the occurrence of nearly identical genotypes (i.e., MLLs) from

widespread geographic locations found in *M. phaseolina* is in line with a markedly clonal population structure (Milgroom, 2015). On the other end of the spectrum genotypic diversity, network analyses and measures of linkage among loci provided support for recombination within some of the genetic clusters. High levels of genotypic diversity is one of the characteristics reflective of recombination in fungal populations (Milgroom, 1996). The higher genotypic diversity (eMLLs) in USiB, US2, and COLPR clusters, may be consistent with the occurrence of recombination in these clusters. Network analyses account for recombination by allowing to infer homoplasy caused by recombination. The boxes between isolates within genetic clusters in the network and the PHI test supporting recombination within all clusters except for USiA, strengthen this hypothesis. The index of association, I_A , revealed an overall high degree of linkage among SNP markers, in line with a pathogen that reproduces clonally. However, the observed I_A values in the COLPR clade and LD decaying faster in COLPR than in US populations, support the potential occurrence of recombination among isolates within COLPR clusters. Although the problem of smaller sample size in COLPR clusters should be at least partially accounted for by using simulations in I_A analysis and clone-corrected data in LD-decay analysis, particularly half-decay LD values should be interpreted with caution and examined in future studies to determine the extent of recombination in *M. phaseolina* populations.

These results are consistent with the population structure model that lays in between the “strictly clonal” and “epidemic” structure proposed by Maynard Smith et al., in which frequent recombination does not occur between isolates in separate branches of an evolutionary tree but it occurs between isolates within a given branch (Smith et al., 1993). These models have been used to describe the population structure of plant pathogens with mixed modes of reproduction or inferred recombination (Grünwald and Hoheisel, 2006; Milgroom et al., 2014; Milgroom, 2015; Milgroom et al., 2016). While little is known about the occurrence of recombination in *M. phaseolina*, recent studies have started to shed light on potential recombination mechanisms involving parasexuality (Pereira et al., 2018) and horizontal gene transfer mediated by giant mobile genetic elements (Gluck-Thaler et al., 2021). Whether other potential recombination mechanisms occur, and the frequency of recombination in *M. phaseolina* remains an important and exciting area of study.

Partial RDA revealed that nearly half of the SNP variance is confounded between neutral genetic structure, climate, and space. This means that this fraction of the variance cannot be statistically associated to a direct effect of any single set of variables. Importantly, the effects of population structure and space often cannot be independently disentangled from spatially structured process (e.g IBD) or spatially structured environmental variables (Lasky et al., 2015). This study, while highlighting the challenges in assessing genotype-environmental associations, provided an assessment of the fraction of confounded variance and allowed us to start disentangling the effects of climate, spatial, and population structure on genomic variation in *M. phaseolina* populations. The genotype-environment association analyses using partial RDA support our hypothesis that local climatic differences contribute to patterns of adaptive divergence among *M. phaseolina* populations across the US, Colombia, and Puerto Rico. Seasonal variation in temperature and precipitation of warmest quarter, were the primary climatic variables associated with variation of candidate adaptive loci without and after accounting for neutral genetic population structure, respectively. We found SNPs within or in physical proximity to genes with functional annotations related to transmembrane transport, glycoside hydrolase activity and DNA binding. In fungi, genes involved in these activities are known to be important in responses to environmental stressors including temperature, water availability, and oxidative stress (Aguilera, Randez-Gil and Prieto, 2007; Gasch, 2007; Branco et al., 2016). Similarly, among the candidates, we found the 241776-protoporphyrinogen oxidase protein, involved in heme biosynthesis and the putative small heat shock protein 28417-Hsp20. Heme has been shown to regulate several mechanisms during cold-shock in *Saccharomyces cerevisiae* (Abramova et al., 2001) while Hsp20 proteins have been found involved in fungal thermal stress response to both heat and cold (Wu et al., 2016; Wang et al., 2021).

The SNP with the highest correlation with temperature seasonality was located upstream to the 753275-ankyrin repeat protein (Table 5). We found that *M. phaseolina* 753275-ankyrin protein is a predicted homologous to the TRP NOMPC mechanotransduction channel in *Drosophila melanogaster* (Jin et al., 2017). Ankyrin family proteins link membrane proteins, including ion channels, to microtubules of the cytoskeleton by binding of its ankyrin repeat domain. The ankyrin proteins in the NOMPC channel link a displacement of the cytoskeleton to the channel opening, translating external stimuli into intra-

cellular signals (Jin et al., 2017). Moreover, the TRP₁ (transient receptor potential 1) ion channel from the alga *Chlamydomonas reinhardtii*, which shares structural homology to the TRP NOMPC channel, was found to act as thermal sensor, with ankyrin proteins mediating the channel opening in response to increased temperature (McGoldrick et al., 2019). Although there is no structural or functional characterization of the *M. phaseolina* 753275-ankyrin protein, it represents a promising candidate to investigate a potential temperature-related mechanism for environmental stimuli transduction. These findings are consistent with the established roles of proteins in environmental stress responses both specific to fungi and conserved across the tree of life. Although our results cannot confirm whether SNPs are the causal mechanism, the candidate genes could be used in future functional studies. Additionally, common garden experiments could provide support for local adaptation to climate in *M. phaseolina*.

Overall, our observations point to a scenario in which *M. phaseolina*, as other plant pathogens with clonal population structures, is structured in a subcontinental regional stable manner in the face of instability at local scales in line with the metapopulation dynamics perspective. These results are consistent with a scenario of evolution after migration driven by divergence following clonal expansions. The presence of MLLs across countries underscores the potential for a large influence of anthropogenic migration introducing *M. phaseolina* to new environments. The association of genetic divergence with climatic variables and putatively adaptive functions of the genes with SNPs strongly associated that would hypothetically benefit *M. phaseolina* in specific environments, is consistent with potential selection imposed by specific climatic variables. Future studies will be needed to identify the degree to which distinct genetic groups reflect their adaptation to host and climate. Such analyses will benefit from a global sampling collected from diverse hosts in conjunction with multiple reference genomes sequenced with long-read technologies that will allow further characterization of the role of genomic variation, including structural variation, in *M. phaseolina* adaptation to host and the climatic environment.

This knowledge expands the impact that spatial population genomics and genotype-environment associations can have on our ability to characterize adaptive potential in plant pathogens by identifying candidate genes and presents a preliminary and complementary approach to the forward-genetics and phenotypic characterization approaches. The ability to identify candidate genes at a population specific

level in a clonal pathogen presents an opportunity to evaluate candidate genes in a population specific manner, which represents a powerful approach specially in clonal pathogens in which unusually high levels of linkage prevent the application of genome scan methods. Additionally, the RDA approach could be applied using candidate adaptive genetic markers to predict pathogens' "adaptive landscape" representing its adaptive variation for any environment across a geographic range (Capblancq and Forester, 2021). As climate and agricultural challenges become more demanding, the characterization of pathogen adaptation capabilities enabled by population genomics should become increasingly utilized for plant disease risk prediction models specially under adverse future climate scenarios.

2.5 Materials and methods

2.5.1 Isolate collection and DNA preparation

A total of 95 *M. phaseolina* isolates were obtained from culture collections, as well as roots or lower stems of soybean and dry bean plants in production fields (Supplementary Table A.1). There were 52 isolates collected from soybean across a latitudinal range in 13 states, including 38 isolates from a previous study (Sexton, Hughes and Wise, 2016). Forty isolates were collected from dry bean grown in Michigan, Puerto Rico and Colombia. Isolates from Michigan were collected from 2011 to 2017 (Jacobs et al., 2019). Isolates from Puerto Rico and Colombia were collected from research plots at the University of Puerto Rico and at the International Center for Tropical Agriculture (CIAT). Two strawberry isolates collected from California and one isolate from Ethiopian mustard (*Brassica carinata*) were included as host outgroups. Cultures were routinely grown on potato dextrose agar (PDA; Acumedia, Lansing, MI) medium.

For genomic DNA extraction, four 5-mm plugs taken from the edge of the culture were used to inoculate 50 mL of potato dextrose broth amended with chloramphenicol (50 mg/L). The broth was incubated for 7 to 9 d at room temperature. Mycelia were harvested, lyophilized for 24 h and ground using a FastPrep FP120 homogenizer (BIO 101 Savant Instruments, Hobrook, NY). Genomic DNA was extracted from the lyophilized tissue using a modified SDS-based method; briefly, 50 mg of ground mycelia were mixed in lysis buffer (3% SDS (w/v); 100 mM Tris-HCl, pH 8.0; 50 mM EDTA, pH 8.0) followed by phenol/chloroform DNA extraction. The identity of all isolates was confirmed by multigene DNA analysis of the Internal Transcribed Spacer regions for the nuclear rDNA operon (ITS), part of the Trans-

lation Elongation Factor (TEF-1) gene region, and part of the actin (ACT) gene region according to (Sarr et al., 2014). Maximum likelihood analysis of the combined sequence alignment placed all the isolates tested in the *M. phaseolina* cluster. A full heuristic search using the first ten most parsimonious trees and the Neighbor-joining tree as starting trees with 100 random sequence additions was performed in PAUP v4.ob10 (Swofford 2003), to find the maximum likelihood tree (Supplementary Fig. A.7).

2.5.2 Whole genome sequencing and variant calling

Genomic libraries were constructed and each of the isolates were whole-genome sequenced to 23X coverage using a 150 base-pair paired-end strategy on the Illumina HiSeq 4000 platform at the Michigan State University Research Technology Support Facility Genomics Core (East Lansing, MI). The libraries were prepared using the Illumina TruSeq Nano DNA Library Preparation Kit HT. The resulting sequences were quality assessed using FastQC (Andrews et al., 2010) and cleaned using Cutadapt v1.16 (Martin, 2011), with the following parameters: -f fastq, -q 20,20, -trim-n, -m 30, -n 3, -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC, -A AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTA-GATCTCGGTGGTCGCCGTATCATT. After initial quality filtering, the remaining sequences were aligned to the *M. phaseolina* reference genome (JGI Mycocosm, MPI-SDFR-AT-0080 v1.0) using bwa-mem (Heng Li, 2013). The isolate used for the *M. phaseolina* reference genome was collected from natural *Arabidopsis thaliana* populations in France (Mesny et al., 2021). The mapping statistics, genome alignment rate and genome coverage were assessed with SAMtools flagstat (Li et al., 2009). Alignments were sorted and indexed using SAMtools (Li et al., 2009). After mapping, duplicate reads were identified using MarkDuplicates and removed during the variant calling step.

Single nucleotide polymorphisms (SNPs) of all 95 isolates were predicted using the Genome Analysis Toolkit (GATK) v4.0 (McKenna et al., 2010). Initially, SNPs were called individually with GATK's HaplotypeCaller. GVCF files were combined, and common SNPs jointly identified using CombineGVCFs and GenotypeGVCFs programs. The later using the -new-qual parameter. The combined vcf file was quality filtered using vcfr v1.10.0 package (Knaus and Grünwald, 2017) in R v4.0.0 (R Core Team 2019). To be included in the high-quality set, SNPs were filtered to remove SNPs with a minimum read depth (DP) of <4x and greater than the 95th percentile of each sample DP distribution and exclude SNPs with

minimum threshold mapping quality ($MQ < 60$) and minimum allele frequency ($MAF < 0.02$) which corresponds to the allele presence in at least two isolates. Only variants with no missing data were retained, which corresponds to positions with 0 missing data for all the sequenced isolates. The final high-quality dataset was used in all subsequent analysis. The final vcf was annotated using SnpEff v5.0c (Cingolani, Platts, et al., 2012) and a vcf containing only SNPs in intergenic regions was created using SnpSift v5.0c (Cingolani, Patel, et al., 2012).

2.5.3 Phylogenomics and population genetic structure

The population structure was inferred according to the results from both model-based and model-free clustering methods and phylogenetic inference. The phylogenetic tree was inferred from the full set of high-quality SNPs among the 95 *M. phaseolina* isolates in RAxML-NG v1.01 (Kozlov et al., 2019). The RAxML analysis was performed using the “-all” option which conducted 20 maximum likelihood inferences on the original SNP alignment, standard bootstrapping with automatic determination of the number of replicates (Felsenstein’s bootstrap, FBP; MRE-based bootstopping test) and the subsequent maximum likelihood search. The General-Time-Reversible (GTR) model of nucleotide substitution with GAMMA model of rate heterogeneity and correction for ascertainment bias (GTR+G+ASC_LEWIS) was used. The best-scoring ML tree was used for optimizing all model and branch length parameters and model evaluation. A model-free dimensionality-reduction approach, principal component analysis (PCA), and discriminatory analysis of principal components (DAPC) were also conducted on the full set of SNPs using adegenet package (Jombart, 2008; Jombart and Ahmed, 2011) in R 4.0.0 (R Core Team 2019). To infer population dynamics and reconstruct a rooted *M. phaseolina* phylogeny, the *M. phaseolina* (JGI Mycocosm, MPI-SDFR-AT-0080 v1.0) reference genome was used as outgroup taxon. Maximum likelihood analysis was run in RAxML-NG v1.01 using the “-all” option with automatic bootstrap replicates and the GTR+G+ASC_LEWIS substitution model.

2.5.4 Spatial genetic structure

Bayesian clustering of allele frequencies was implemented in conStruct (Bradburd, Coop and Ralph, 2018). To assess whether population structure was well described by modelling isolates as admixtures between multiple discrete genetic groups or by both discrete and continuous genetic structure, spatial anal-

ysis of population structure was conducted using conStruct (Bradburd, Coop and Ralph, 2018). Spatial analysis in conStruct accounts for isolation by distance by allowing genetic differentiation to increase with geographic distance within discrete genetic groups (layers, K). The data was analyzed treating individual isolates as the unit of analysis, using the spatial models setting K between 1 and 7 with 20000 iterations, and compared these models using cross-validation with 10 replicates. For cross-validation, 90% of loci were used to fit the model and the remaining loci for model evaluation. A geographically constrained least-squares method as implemented in TESS3 (Caye et al., 2016), was used to estimate ancestry coefficients and create interpolation maps based on the coefficients. TESS3 uses a spatially explicit algorithm that can be considered model-free. The algorithm was run using the function “tess3” with K between 1 and 7 and 10 replicates.

2.5.5 Population genetic and genotypic diversity

For each clade and genetic cluster, gene diversity (Nei, 1978) was calculated using the Hs function in the adegenet package (Jombart, 2008; Jombart and Ahmed, 2011). The median estimates of pairwise genetic distance and genotypic diversity indices were calculated within each clade and genetic cluster using the R package poppr v2.9.0 (Kamvar et al. 2014). Genotypic diversity was assessed by calculating the number of multilocus genotypes (MLGs). A MLG was defined as a unique combination of the 76,981 SNPs. MLGs were collapsed into larger groups called multilocus lineages using the average neighbor algorithm and a Prevosti’s distance threshold of 0.0001 (bitwise.dist function; Kamvar et al., 2015). Rarefaction was used to correct for uneven sample sizes using the R package vegan v2.5-6 (Oksanen et al., 2019) and obtain the number of expected MLGs and MLLs (eMLG and eMLL) at the lowest common sample size (i.e., 25 for clades and 5 for genetic clusters). Genotypic diversity indices, Shannon-Wiener Index (H^*), Stoddart and Taylor’s Index (G^*), Simpson’s index (λ^*) and evenness (E_5^*) (Grünwald et al., 2003), were calculated using the R package poppr v2.9.0 (diversity_ci function; Kamvar et al., 2014) based on the number of MLLs in each clade and genetic cluster and correcting for unequal sample sizes based on rarefaction. The function mlg.crosspop in poppr was used to detect the presence of MLGs occurring across populations. Migration was inferred by tracking MLGs across genetic clusters, referred here as genotype flow (McDonald and Linde, 2002).

2.5.6 Population differentiation between genetic clusters and countries

The F_{ST} analog, G_{ST} (Nei 1972, 1973) was calculated from clone-corrected data using vcfR (Knaus and Grunwald 2017) to infer differentiation among genetic clusters. To describe the population dynamics between the US, Puerto Rico and Colombia, the degree of genetic differentiation across *M. phaseolina* samples was measured hierarchically by genetic clusters within clades. Analysis of molecular variance (AMOVA) based on the quasi-Euclidean distance matrix was conducted in poppr v2.9.0 (Kamvar et al. 2014). AMOVA estimates the number of differences summed over loci based on a matrix of distances between individuals and covariance components are used to calculate fixation indices for each hierarchical level, among clades, among genetic clusters and within genetic clusters. Significant differences of fixation indices were determined by 1,000 random permutations (Grunwald and Hoheisel 2006).

2.5.7 Recombination and clonality

To account for potential intraspecific recombination among *M. phaseolina* isolates, a phylogenetic network was built using the Neighbor-Net algorithm as implemented in SplitsTree4 v4.16.1. The extent of clonality was tested by calculating the proportion of significant linkage between pairs of loci, by computing the standardized index of association (I_A , Brown et al. 1980) for each of the main populations (US and COLPR) using poppr v2.9.0 (Kamvar et al. 2014). Linkage disequilibrium is expected in asexual or inbreeding populations and I_A values close to zero are expected for outcrossing populations (Burt et al., 1996). The observed I_A distributions for each population were compared to five simulated recombined distributions (0%, 25%, 50%, 75% and 100% linkage) generated among 76,981 loci and 48 samples (corresponding to the median population size of the two clusters). The observed and simulated I_A values were tested for normality using the Shapiro-Wilk's normality test and an analysis of variance (ANOVA) was conducted to test for significant differences among the distributions. Pairwise comparisons between the I_A simulated distributions and for each population were tested for difference with Tukey's HSD test in R. The extent of clonality was correlated to clonal (100%), mostly clonal (75%), semiclonal (50%, 25%) or sexual (0%) modes of reproduction. Linkage disequilibrium (LD) decay rate was estimated using the physical distance over which LD decays to half its initial value, as measured by the squared correlation coefficient (r^2). The linkage disequilibrium decay was calculated for each clade using the correlation co-

efficient (r^2) in TASSEL v5 (Bradbury et al., 2007) within a window of 50 sites among SNPs using the clone-corrected dataset (79 MLGs). The mean r^2 values, representing the correlation between alleles at two loci within 10 bp of physical distance, were then plotted in R 4.0.0 (R Core Team 2019).

2.5.8 Climatic data

For each isolate, the 19 standard bioclimatic variables available at the WorldClim2 database (Fick Hijmans, 2017) were obtained using 'getData' function from raster R package (Hijmans, 2022). All variables are the average for the years 1970 to 2000 and were obtained at a spatial resolution of 2.5 min (21.5 km²). We used data at a resolution of 2.5 min (21.5 km²), because it corresponds with our sampling design (single isolate samples rather than populations) being at a field or county scale. Coarser resolutions could combine multiple sampling locations into a single spatial grid and finer resolutions (30-s or <30-s), while this may be important for structuring patterns of genetic variation within populations, these data are less suitable for our sampling design and focus on regional to continental-wide patterns. We reduced the number of climatic variables from 19 to five to account for collinearity among them ($|r| > 0.7$) and to represent our hypothesis about the most important factors potentially driving selection. Diseases caused by *M. phaseolina* are more prevalent during hot and dry conditions, therefore temperature and precipitation variables were included. The selected climatic variables were: BIO18 = Precipitation of Warmest Quarter, BIO15 = Precipitation Seasonality (Coefficient of Variation), BIO17 = Precipitation of Driest Quarter, BIO10 = Mean Temperature of Warmest Quarter and BIO4 = Temperature Seasonality (standard deviation *100). Each bioclimatic variable was scaled, centered, and evaluated for inclusion using forward selection with 10,000 permutations using adespatial R package (Dray et al., 2022).

To account for underlying spatial structure (autocorrelation) and reduce spurious GEA, distance-based Moran's eigenvector maps (dbMEM) were generated using sample coordinates in the quickMEM R function (Borcard, Gillet, Legendre, 2018). The dbMEMs are a matrix of axes that capture spatial patterns from multiple angles rather than just a latitudinal or longitudinal vector. Only significant dbMEM axes were selected using forward selection with 1,000 permutations. A simple RDA model and partial RDA model conditioning on space, using only significant dbMEMs, were used to identify the climatic variables significantly contributing to genomic variation and those structured in space.

2.5.9 Variance partitioning and outlier loci identification

To identify potentially adaptive loci, associations between genetic data (loci) and climatic variables hypothesized to drive selection were evaluated using a multivariate method, redundancy analysis (RDA, as implemented by Forester et al., 2016). RDA simultaneously tests multiple loci that covary in response to climatic variables. Partial RDA models were used for variance partitioning and outlier loci identification while correcting for neutral genetic population structure. Variance partitioning analysis was performed with linkage-disequilibrium (LD)-filtered ($r^2 > 0.9$) dataset of 11,421 SNPs. The independent contribution of each set of explanatory variables: climate, neutral population structure or space, was assessed while removing the effect of the remaining variable sets using partial RDA. In outlier loci identification, using a partial RDA is recommended to reduce the number of false-positive detections particularly in scenarios of multilocus adaptation when selective agents are unknown (Forester et al., 2018). On the other hand, partial RDA can lead to high false-negative detections when variance is confounded between climatic variables and neutral population structure (Capblancq and Forester, 2021). Candidate adaptive loci were identified using simple and partial RDA models to examine the extent of this issue. A partial RDA model conditioning on neutral population genetic structure was used for candidate outlier SNPs detection. Outlier loci were identified in the three significant constrained axes as the SNPs having loadings ± 3 or ± 4 SD from the mean score of each constrained axis using both the LD-filtered set of 11,421 SNPs and the full set of 77,465 SNPs, respectively (Forester et al., 2018; Lasky et al., 2012). A simple RDA model, without correcting for population structure, using the LD-filtered set of 11,421 SNPs and outlier loci were identified in the three significant constrained axes as the SNPs having loadings ± 3 SD from the mean score. Gene annotations for the significant candidate SNPs were used to investigate putative adaptive functions, using the annotated vcf.

BIBLIOGRAPHY

Abramova, N. E. et al. (2001) 'Regulatory mechanisms controlling expression of the DAN/TIR mannoprotein genes during anaerobic remodeling of the cell wall in *Saccharomyces cerevisiae*', *Genetics*, 157(3), pp. 1169–1177. doi: 10.1093/genetics/157.3.1169.

Aguilera, J., Randez-Gil, F. and Prieto, J. A. (2007) 'Cold response in *Saccharomyces cerevisiae*: New functions for old mechanisms', *FEMS Microbiology Reviews*, 31(3), pp. 327–341. doi: 10.1111/j.1574-6976.2007.00066.x.

Allen, T. W. et al. (2017) 'Soybean yield loss estimates due to diseases in the United States and Ontario, Canada, from 2010 to 2014', *Plant Health Progress*, 18(1), pp. 19–27. doi: 10.1094/PHP-RS-16-0066.

Andrews, S. (2010) 'FastQC: A quality control tool for high throughput sequence data'. Available at: <http://www.bioinformatics.babraham.ac.uk>.

Arias, R. S. et al. (2011) 'Discriminating microsatellites from *Macrophomina phaseolina* and their potential association to biological functions', *Plant Pathology*, 60(4), pp. 709–718. doi: 10.1111/j.1365-3059.2010.02421.x.

Baird R. E., Watson C. E., Scruggs M (2003) Relative Longevity of *Macrophomina phaseolina* and Associated Mycobiota on Residual Soybean Roots in Soil. *Plant Dis* 87: 563–566

Baird, R. E. et al. (2010) 'Variability of United States isolates of *Macrophomina phaseolina* based on simple sequence repeats and cross genus transferability to related genera within botryosphaeriaceae.', *Mycopathologia*, 170(3), pp. 169–80. doi: 10.1007/s11046-010-9308-3.

Batista, E., Lopes, A. and Alves, A. (2021) 'What do we know about botryosphaeriaceae? An overview of a worldwide cured dataset', *Forests*, 12(3). doi: 10.3390/f12030313.

Bertazzoni, S. et al. (2018) 'Accessories Make the Outfit: Accessory Chromosomes and Other Dispensable DNA Regions in Plant-Pathogenic Fungi', *Molecular Plant-Microbe Interactions*, 31(8), p. MPMI-06-17-0135. doi: 10.1094/MPMI-06-17-0135-FI.

Bitocchi, E. et al. (2017) 'Beans (*Phaseolus* spp.) as a model for understanding crop evolution', *Frontiers in Plant Science*, 8(May), pp. 1–21. doi: 10.3389/fpls.2017.00722.

Bradburd, G. S., Coop, G. M. and Ralph, P. L. (2018) 'Inferring Continuous and Discrete Population Genetic', *Genetics*, 210(September), pp. 33–52. doi: 10.1534/genetics.118.301333.

Bradbury, P. J. et al. (2007) 'TASSEL: software for association mapping of complex traits in diverse samples', *Bioinformatics*. Oxford Univ Press, 23(19), pp. 2633–2635.

Bradley, C. and Allen, T. (2014) 'Estimates of Soybean Yield Reductions Caused by Diseases in the United States', *University of Illinois Outreach and Extension*. University of Illinois Extension Outreach, 2014. Available at: <http://extension.cropsciences.illinois.edu>.

Branco, S. et al. (2015) 'Genetic isolation between two recently diverged populations of a symbiotic fungus', *Molecular Ecology*, 24(11), pp. 2747–2758. doi: 10.1111/mec.13132.

Branco, S. et al. (2016) ‘Continental-level population differentiation and environmental adaptation in the mushroom *Suillus brevipes*’, *Molecular Ecology*. doi: 10.1111/mec.13892.

Burkhardt, A. et al. (2019) ‘Assembly, annotation, and comparison of *Macrophomina phaseolina* isolates from strawberry and other hosts’, Dryad, Dataset. Available at: <https://datadryad.org/stash/dataset/doi:10.5061/dryad.zkhr8935h> (Accessed: 13 August 2021).

Burkhardt, A. K. et al. (2019) ‘Assembly, annotation, and comparison of *Macrophomina phaseolina* isolates from strawberry and other hosts’, *BMC Genomics*. BioMed Central, 20(1), p. 802. doi: 10.1186/s12864-019-6168-1.

Capblancq, T. and Forester, B. R. (2021) ‘Redundancy analysis: A Swiss Army Knife for landscape genomics’, *Methods in Ecology and Evolution*, 12(12), pp. 2298–2309. doi: 10.1111/2041-210X.13722.

Casano, S. et al. (2018) ‘First report of charcoal rot caused by *Macrophomina phaseolina* on hemp (*Cannabis sativa*) varieties cultivated in southern Spain’, *Plant Disease*, p. 1665. doi: 10.1094/PDIS-02-18-0208-PDN.

Caye, K. et al. (2016) ‘TESS3: Fast inference of spatial population structure and genome scans for selection’, *Molecular Ecology Resources*, 16(2), pp. 540–548. doi: 10.1111/1755-0998.12471.

Cingolani, P., Platts, A., et al. (2012) ‘A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff’, *Fly*, 6(2), pp. 80–92. doi: 10.4161/fly.19695.

Cingolani, P., Patel, V. M., et al. (2012) ‘Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, SnpSift’, *Frontiers in Genetics*, 3(MAR), pp. 1–9. doi: 10.3389/fgene.2012.00035.

Croll, D. and McDonald, B. A. (2016) ‘The genetic basis of local adaptation for pathogenic fungi in agricultural ecosystems’, *Molecular Ecology*. doi: 10.1111/mec.13870.

Crous, P. W. et al. (2017) ‘Botryosphaeriaceae: Systematics, pathology, and genetics’, *Fungal Biology*. doi: 10.1016/j.funbio.2017.01.003.

Dhingra, O. D. and Sinclair, J. B. (1973) ‘Location of *Macrophomina phaseoli* on soybean plants related to culture characteristics and virulence’. Edited by F. A. O. of the UN. (Phytopathology), pp. 934–936.

Dhingra, O. D. and Sinclair, J. B. (1974) ‘Survival of *Macrophomina phaseolina* Sclerotia in Soil: Effects of Soil Moisture, Carbon:Nitrogen Ratios, Carbon Sources, and Nitrogen Concentrations’, *Phytopathology*, 65, pp. 236–240.

Dhingra, O. D. and Sinclair, J. B. (1978) *Biology and pathology of Macrophomina phaseolina*. Minas Gerais: Universidade Federal de Vicosa.

Dray, S. et al. (2022) ‘Package “adespatial”’, *Ecological Monographs*, pp. 1–138. doi: 10.1890/11-1183.1.

Ellison, C. E. et al. (2011) ‘Population genomics and local adaptation in wild isolates of a model microbial eukaryote’, *Proceedings of the National Academy of Sciences*, 108(7), pp. 2831–2836. doi: 10.1073/pnas.1014971108.

Fitzpatrick, M. C. and Keller, S. R. (2015) 'Ecological genomics meets community-level modelling of biodiversity: Mapping the genomic landscape of current and future environmental adaptation', *Ecology Letters*, 18(1), pp. 1–16. doi: 10.1111/ele.12376.

Forester, B. R. et al. (2018) 'Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations', *Molecular Ecology*, 27(9), pp. 2215–2233. doi: 10.1111/mec.14584.

Gasch, A. P. (2007) 'Comparative genomics of the environmental stress response in ascomycete fungi', (July), pp. 961–976. doi: 10.1002/yea.

Gibson, M. J. S. and Moyle, L. C. (2020) 'Regional differences in the abiotic environment contribute to genomic divergence within a wild tomato species', *Molecular Ecology*, 29(12), pp. 2204–2217. doi: 10.1111/mec.15477.

Gladieux, P. et al. (2014) 'Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes', *Molecular Ecology*, 23(4), pp. 753–773. doi: 10.1111/mec.12631.

Gluck-Thaler, E. et al. (2021) 'Giant Starship elements mobilize accessory genes in fungal genomes', *bioRxiv*, p. 2021.12.13.472469. Available at: <https://www.biorxiv.org/content/10.1101/2021.12.13.472469v2>

Goss, E. M. et al. (2014) 'The Irish potato famine pathogen *Phytophthora infestans* originated in central Mexico rather than the Andes', *Proceedings of the National Academy of Sciences*, 111(24), pp. 8791–8796. doi: 10.1073/pnas.1401884111.

Grünwald, N. J. et al. (2003) 'Analysis of genotypic diversity data for populations of microorganisms', *Phytopathology*, 93(6), pp. 738–746. doi: 10.1094/PHYTO.2003.93.6.738.

Grünwald, N. J. and Hoheisel, G. A. (2006) 'Hierarchical analysis of diversity, selfing, and genetic differentiation in populations of the oomycete *Aphanomyces euteiches*', *Phytopathology*, 96(10), pp. 1134–1141. doi: 10.1094/PHYTO-96-1134.

Gupta, G. K., Sharma, S. K. and Ramteke, R. (2012) 'Biology, Epidemiology and Management of the Pathogenic Fungus *Macrophomina phaseolina* (Tassi) Goid with Special Reference to Charcoal Rot of Soybean (*Glycine max* (L.) Merrill)', *Journal of Phytopathology*, 160(4), pp. 167–180. doi: 10.1111/j.1439-0434.2012.01884.x.

Hanski, I. (1998) 'Metapopulation dynamics', *Nature*, 396(6706), pp. 41–49. doi: 10.1038/23876.

Heng Li (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, Preprint Oxford University Press. Available at: <https://arxiv.org/pdf/1303.3997.pdf> (Accessed: 21 September 2018).

Hijmans, R. (2022) 'raster: Geographic Data Analysis and Modeling'. Available at: <https://cran.r-project.org/package=raster>.

Islam, Md et al. (2012) 'Tools to kill: Genome of one of the most destructive plant pathogenic fungi *Macrophomina phaseolina*', *BMC Genomics*. *BMC Genomics*, 13(1), p. 493. doi: 10.1186/1471-2164-13-493.

Jacobs, J. L. et al. (2019) 'Determining the Soilborne Pathogens Associated with Root Rot Disease Complex of Dry Bean in Michigan', *Plant Health Progress*, 20(2), pp. 122–127. doi: 10.1094/php-11-18-0076-s.

Jana, T., Sharma, T. R. and Singh, N. K. (2005) 'SSR-based detection of genetic variability in the charcoal root rot pathogen *Macrophomina phaseolina*', *Mycological Research*, 109(1), pp. 81–86. doi: 10.1017/S0953756204001364.

Jin, P. et al. (2017) 'Electron cryo-microscopy structure of the mechanotransduction channel NOMPC', *Nature*. Nature Publishing Group, 547(7661), pp. 118–122. doi: 10.1038/nature22981.

Jombart, T. (2008) 'Adegenet: A R package for the multivariate analysis of genetic markers', *Bioinformatics*, 24(11), pp. 1403–1405. doi: 10.1093/bioinformatics/btn129.

Jombart, T. and Ahmed, I. (2011) 'adegenet 1.3-1: New tools for the analysis of genome-wide SNP data', *Bioinformatics*, 27(21), pp. 3070–3071. doi: 10.1093/bioinformatics/btr521.

Kelly, J. D. (2010) The Story of Bean Breeding, BeanCAP and PBG, (June), pp. 1–30. Available at: <http://www.css.msu.edu>.

Kendig, S. R., Rupe, J. C. and Scott, H. D. (2000) 'Effect of irrigation and soil water stress on densities of *Macrophomina phaseolina* in soil and roots of two soybean cultivars', *Plant Disease*, 84(8), pp. 895–900. doi: 10.1094/PDIS.2000.84.8.895.

Knaus, B. J. and Grünwald, N. J. (2017) 'vcfr: a package to manipulate and visualize variant call format data in R', *Molecular Ecology Resources*, 17(1), pp. 44–53. doi: 10.1111/1755-0998.12549.

Knox-Davies, P. S. (1965) 'Pycnidium production by *Macrophomina*', *S. Afr. J. Agric. Sci*, 8, pp. 205–218.

Koehler, A. M. and Shew, H. D. (2018) 'First Report of Charcoal Rot of Stevia Caused by *Macrophomina phaseolina* in North Carolina', *Plant Disease*. *Plant Disease*, 102(1), pp. 241–241. doi: 10.1094/PDIS-05-17-0693-PDN.

Koike, S. T. et al. (2016) 'Status of *Macrophomina phaseolina* on Strawberry in California and Preliminary Characterization of the Pathogen', *International Journal of Fruit Science*. Taylor and Francis, pp. 1–12. doi: 10.1080/15538362.2016.1195313.

Kozlov, A. M. et al. (2019) 'RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference', *Bioinformatics*, 35(21), pp. 4453–4455. doi: 10.1093/bioinformatics/btz305.

Lasky, J. R. et al. (2012) 'Characterizing genomic variation of *Arabidopsis thaliana*: The roles of geography and climate', *Molecular Ecology*, 21(22), pp. 5512–5529. doi: 10.1111/j.1365-294X.2012.05709.x.

Lasky, J. R. et al. (2015) 'Genome-environment associations in sorghum landraces predict adaptive traits', *Science Advances*, 1(6), pp. 1–14. doi: 10.1126/sciadv.1400218.

Lawson, D. J., van Dorp, L. and Falush, D. (2018) 'A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots', *Nature Communications*. Springer US, 9(1), pp. 1–11. doi: 10.1038/s41467-018-05257-7.

- LeBlanc, N., Cubeta, M. A. and Crouch, J. A. (2021) 'Population genomics trace clonal diversification and intercontinental migration of an emerging fungal pathogen of boxwood', *Phytopathology*, 111(1), pp. 184–193. doi: 10.1094/PHYTO-06-20-0219-FI.
- Leyva-Mir, S. G. et al. (2015) 'First Report of Charcoal Rot of Sugarcane Caused by *Macrophomina phaseolina* in Mexico', *Plant Disease*. *Plant Disease*, 99(4), pp. 553–553. doi: 10.1094/PDIS-06-14-0652-PDN.
- Li, H. et al. (2009) 'The Sequence Alignment/Map format and SAMtools.', *Bioinformatics* (Oxford, England). Oxford University Press, 25(16), pp. 2078–9. doi: 10.1093/bioinformatics/btp352.
- Ma, J., Hill, C. B. and Hartman, G. L. (2010) 'Production of *Macrophomina phaseolina* Conidia by Multiple Soybean Isolates in Culture', *Plant Disease*. The American Phytopathological Society, 94(9), pp. 1088–1092. doi: 10.1094/PDIS-94-9-1088.
- Martin, M. (2011) 'Cutadapt removes adapter sequences from high-throughput sequencing reads', *EMBnet.journal*, 17(1), p. 10. doi: 10.14806/ej.17.1.200.
- McDonald, B. A. and Stukenbrock, E. H. (2016) 'Rapid emergence of pathogens in agro-ecosystems: global threats to agricultural sustainability and food security', *Philosophical Transactions of the Royal Society B: Biological Sciences*. The Royal Society, 371(1709), p. 20160026. doi: 10.1098/rstb.2016.0026.
- McGoldrick, L. L. et al. (2019) 'Structure of the thermo-sensitive TRP channel TRP_{P1} from the alga *Chlamydomonas reinhardtii*', *Nature Communications*. Springer US, 10(1), pp. 1–12. doi: 10.1038/s41467-019-12121-9.
- McKenna, A. et al. (2010) 'The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.', *Genome research*, 20(9), pp. 1297–303. doi: 10.1101/gr.107524.110.
- Meena, B. R. et al. (2018) 'First report of charcoal rot caused by *Macrophomina phaseolina* in *Basella alba* in India', *Plant Disease*, p. 1669. doi: 10.1094/PDIS-01-18-0152-PDN.
- Mengistu, A. et al. (2007) 'Charcoal rot disease assessment of soybean genotypes using a colony-forming unit index', *Crop Science*, 47(December), pp. 2453–2461. doi: 10.2135/cropsci2007.04.0186.
- Mengistu, A., Arelli, P. A., et al. (2011) 'Evaluation of soybean genotypes for resistance to charcoal rot', *Plant Health Progress*. doi: 10.1094/PHP-2010-0926-01-RS.
- Mengistu, A., Smith, J. R., et al. (2011) 'Seasonal progress of charcoal rot and its impact on soybean productivity', *Plant Disease*, 95(9), pp. 1159–1166. doi: 10.1094/PDIS-02-11-0100.
- Mesny, F. et al. (2021) 'Genetic determinants of endophytism in the *Arabidopsis* root mycobiome', *Nature Communications*. Springer US, 12(1), pp. 1–15. doi: 10.1038/s41467-021-27479-y.
- Meyer, W. . and Sinclair, J. B. (1974) 'Factors Affecting Charcoal Rot of Soybean Seedlings', *Phytopathology*, 64, pp. 845–849. doi: DOI: 10.1094/Phyto-64-845.
- Mihail, J. D. and Taylor, S. J. (1995) 'Interpreting variability among isolates of *Macrophomina phaseolina* in pathogenicity, pycnidium production, and chlorate utilization', *Canadian Journal of Botany*,

73(10), pp. 1596–1603. doi: 10.1139/b95-172.

Milgroom, M. G. (1996) 'Recombination and the multilocus structure of fungal populations', Annual Review of Phytopathology. Annual Reviews Inc., 34(1), pp. 457–477. doi: 10.1146/annurev.phyto.34.1.457.

Milgroom, M. G. (2015) Population biology of plant pathogens: genetics, ecology, and evolution. APS Press, The American Phytopathological Society St. Paul, MN, USA.

Milgroom, M. G. et al. (2016) 'Clonal expansion and migration of a highly virulent, defoliating lineage of *Verticillium dahliae*', Phytopathology, 106(9), pp. 1038–1046. doi: 10.1094/PHYTO-11-15-0300-R.

Nishad, I. et al. (2018) 'First Report of Root Rot of *Nepeta cataria* Caused by *Macrophomina phaseolina* in India', Plant Disease. Plant Disease, 102(11), p. 2380. doi: 10.1094/PDIS-04-18-0558-PDN.

Oksanen, J. et al. (2019) 'vegan: Community Ecology Package', pp. 2395–2396. doi: 10.1007/978-94-024-1179-9-301576.

Parsa, S. et al. (2016) 'Fungal endophytes in germinated seeds of the common bean, *Phaseolus vulgaris*', Fungal Biology. doi: 10.1016/j.funbio.2016.01.017.

Reznikov, S. et al. (2018) 'Disease incidence of charcoal rot (*Macrophomina phaseolina*) on soybean in north-western Argentina and genetic characteristics of the pathogen', Canadian Journal of Plant Pathology. Taylor Francis, 40(3), pp. 423–433. doi: 10.1080/07060661.2018.1484390.

Romero Luna, M. P. et al. (2017) 'Advancing Our Understanding of Charcoal Rot in Soybeans', Journal of Integrated Pest Management. Bulletin Texas Agricultural Experimental Station No. 33, 8(1), pp. 526–526. doi: 10.1093/jipm/pmwo20.

Sakalidis, M. L. et al. (2013) 'The challenge of understanding the origin, pathways and extent of fungal invasions: global populations of the *Neofusicoccum parvum*-*N. ribis* species complex Diversity and Distributions', Diversity and Distributions, 19, pp. 873–883. doi: 10.1111/ddi.12030.

Saleh, A. A. et al. (2010) 'Relatedness of *Macrophomina phaseolina* isolates from tallgrass prairie, maize, soybean and sorghum', Molecular Ecology, 19(1), pp. 79–91. doi: 10.1111/j.1365-294X.2009.04433.x.

Sarr, M. P. et al. (2014) 'Genetic diversity in *Macrophomina phaseolina*, the causal agent of charcoal rot', Phytopathologia Mediterranea, 53(3), p. 559564. doi: 10.14601/Phytopathol.

Savary, S. et al. (2019) 'The global burden of pathogens and pests on major food crops', Nature Ecology and Evolution, 3(3), pp. 430–439. doi: 10.1038/s41559-018-0793-y.

Savolainen, O., Lascoux, M. and Merilä, J. (2013) 'Ecological genomics of local adaptation', Nature Reviews Genetics. Nature Publishing Group, 14(11), pp. 807–820. doi: 10.1038/nrg3522.

Sexton, J. P., Hangartner, S. B. and Hoffmann, A. A. (2014) 'Genetic isolation by environment or distance: Which pattern of gene flow is most common?', Evolution, 68(1), pp. 1–15. doi: 10.1111/evo.12258.

Sexton, Z. F., Hughes, T. J. and Wise, K. A. (2016) 'Analyzing isolate variability of *Macrophomina phaseolina* from a regional perspective', Crop Protection. Elsevier Ltd, 81, pp. 9–13. doi: 10.1016/j.cropro.2015.11.012.

Short G. E., Wyllie T. D., Bristow P. R. (1980) Survival of *Macrophomina phaseolina* in soil and residue of soybean. Phytopathology 70: 13–17

Slippers, B. and Boissin, E. (2013) 'Phylogenetic lineages in the Botryosphaeriales: a systematic and evolutionary framework', Studies in Mycology. CBS-KNAW Fungal Biodiversity Centre, 76, pp. 31–49. doi: 10.3114/simoo20.

Slippers, B. and Wingfield, M. J. (2007) 'Botryosphaeriaceae as endophytes and latent pathogens of woody plants: diversity, ecology and impact', Fungal Biology Reviews, 21(2–3), pp. 90–106. doi: 10.1016/j.fbr.2007.06.002.

Smith, G. S. and Carvil, O. N. (1997) 'Field screening of commercial and experimental soybean cultivars for their reaction to *Macrophomina phaseolina*', Plant Disease, 81(4), pp. 363–368. doi: 10.1094/PDIS.1997.81.4.363.

Smith, J. M. et al. (1993) 'How clonal are bacteria?', Proceedings of the National Academy of Sciences of the United States of America, 90(10), pp. 4384–4388. doi: 10.1073/pnas.90.10.4384.

Stukenbrock, E. H. et al. (2011) 'The making of a new pathogen: Insights from comparative population genomics of the domesticated wheat pathogen *Mycosphaerella graminicola* and its wild sister species', Genome Research. Cold Spring Harbor Laboratory Press, 21(12), pp. 2157–2166. doi: 10.1101/gr.118851.110.

Su, G. et al. (2001) 'Host specialization in the charcoal rot fungus, *Macrophomina phaseolina*', Phytopathology. The American Phytopathological Society, 91(2), pp. 120–126. doi: 10.1094/PHTO.2001.91.2.120.

Tabima, J. F. et al. (2019) 'Population genomic analyses reveal connectivity via human-mediated transport across *Populus* plantations in North America and an undescribed sub-population of *Sphaerulina musiva*', Molecular Plant-Microbe Interactions. doi: 10.1094/mpmi-05-19-0131-r.

Tančić Živanov, S. et al. (2018) 'First Report of Charcoal Rot on Zebra Plant (*Aphelandra squarrosa*) Caused by *Macrophomina phaseolina*', Plant Disease. Plant Disease, 102(11), p. 2377. doi: 10.1094/PDIS-03-18-0480-PDN.

Wang, L. et al. (2021) 'Haploid Genome Analysis Reveals a Tandem Cluster of Four HSP20 Genes Involved in the High-Temperature Adaptation of *Coriopsis trogii*', Microbiology Spectrum, 9(1), pp. 1–14. doi: 10.1128/spectrum.00287-21.

Wang, Weiwei et al. (2020) 'First Report of *Macrophomina phaseolina* Causing Stalk Rot of Sacha Inchi (*Plukenetia volubilis*) in China', Plant Disease. Plant Disease, 104(2), pp. 570–570. doi: 10.1094/PDIS-06-19-1312-PDN.

Wingfield, M. J. et al. (2015) 'Planted forest health: The need for a global strategy', Science, 349(6250), pp. 832–836. doi: 10.1126/science.aac6674.

Wrather, J. A. et al. (2001) 'Soybean disease loss estimates for the top ten soybean-producing countries in 1998', Canadian Journal of Plant Pathology, 23(2), pp. 115–121. doi: 10.1080/07060660109506918.

Wu, J. et al. (2016) 'Small heat shock proteins, phylogeny in filamentous fungi and expression analyses in *Aspergillus nidulans*', Gene. Elsevier B.V., 575(2), pp. 675–679. doi: 10.1016/j.gene.2015.09.044.

Xuereb, A. et al. (2018) 'Putatively adaptive genetic variation in the giant California sea cucumber (*Parastichopus californicus*) as revealed by environmental association analysis of restriction-site associated DNA sequencing data', Molecular Ecology, 27(24), pp. 5035–5048. doi: 10.1111/mec.14942.

Zveibil, A. et al. (2012) 'Survival, host-pathogen interaction, and management of *Macrophomina phaseolina* on strawberry in Israel', Plant Disease, 96(2), pp. 265–272. doi: 10.1094/PDIS-04-11-0299.

APPENDIX

SUPPLEMENTARY TABLES AND FIGURES

Table A.1 *Macrophomina phaseolina* isolates used in this study. Isolates were collected across the US, Puerto Rico, and Colombia from soybean, dry bean, strawberry, and Ethiopian mustard.

Isolate ID	Longitude	Latitude	Region	Country	State / Department	Host	Municipality	Genetic Cluster	Collection Year	Source
CR_Red_1	-95.304	44.243	East North Central	US	MN	Soybean	Lamberton, MN	US2		Dean Malvick - UMN
CR_Red_2B	-95.304	44.243	East North Central	US	MN	Soybean	Lamberton, MN	US2		Dean Malvick - UMN
CR_Red_3	-95.304	44.243	East North Central	US	MN	Soybean	Lamberton, MN	US2		Dean Malvick - UMN
Dm13	-88.99	43.707	East North Central	US	WI	Soybean	Markesan, WI	US1A		T. Hughes. Obtained from Kiersten Wise
Etr2	-88.405	42.785	East North Central	US	WI	Soybean	E. Troy, WI	US1A		T. Hughes. Obtained from Kiersten Wise
Etr4	-88.405	42.785	East North Central	US	WI	Soybean	E. Troy, WI	US1A		T. Hughes. Obtained from Kiersten Wise
Etr7	-88.405	42.785	East North Central	US	WI	Soybean	E. Troy, WI	US2		T. Hughes. Obtained from Kiersten Wise
Etr8	-88.405	42.785	East North Central	US	WI	Soybean	E. Troy, WI	US1A		T. Hughes. Obtained from Kiersten Wise
Etr8	-88.405	42.785	East North Central	US	WI	Soybean	E. Troy, WI	US1A		T. Hughes. Obtained from Kiersten Wise
IN12_4	-85.761	41.509	Central	US	IN	Soybean	Benton, IN	US1A		Purdue Plant Diag. Lab. Obtained from Kiersten Wise
IN12_8_3	-85.417	41.642	Central	US	IN	Soybean	Lagrange, IN	US1A		Purdue Plant Diag. Lab. Obtained from Kiersten Wise
IN12_9_4	-87.48	39.851	Central	US	IN	Soybean	Vermillion, IN	NA		Purdue Plant Diag. Lab. Obtained from Kiersten Wise
IN12_9_6	-87.48	39.851	Central	US	IN	Soybean	Vermillion, IN	US1A		Purdue Plant Diag. Lab. Obtained from Kiersten Wise
IN12_PO_3	-87.894	38.002	Central	US	IN	Soybean	Possey, IN	US1A		Purdue Plant Diag. Lab. Obtained from Kiersten Wise
M11_12	-120.434	34.951	West	US	CA	Strawberry	Santa Barbara, CA	COLPR2	2011	Frank Martin
M13_26	-121.612	36.674	West	US	CA	Strawberry	Monterey, CA	COLPR2	2013	Frank Martin
M16_1	-83.963	43.647	East North Central	US	MI	Dry bean	Kawkawlin, MI	US1A	2016	M. Chilvers
MISO171_1	-84.973	42.946	East North Central	US	MI	Soybean	Lyons, MI	US1A	2017	M. Chilvers
MISO171_2	-84.973	42.946	East North Central	US	MI	Soybean	Lyons, MI	US1A	2017	M. Chilvers
MISO171_3	-84.973	42.946	East North Central	US	MI	Soybean	Lyons, MI	US1A	2017	M. Chilvers
MISO171_4	-84.973	42.946	East North Central	US	MI	Soybean	Lyons, MI	US1A	2017	M. Chilvers
MISO171_5	-84.973	42.946	East North Central	US	MI	Soybean	Lyons, MI	US1A	2017	M. Chilvers
MISO171_7	-84.973	42.946	East North Central	US	MI	Soybean	Lyons, MI	US1A	2017	M. Chilvers
MISO171_8	-84.973	42.946	East North Central	US	MI	Soybean	Lyons, MI	US1A	2017	M. Chilvers
MI_SF_10_16	-84.72	42.881	East North Central	US	MI	Soybean	Westphalia, MI	US1A	2011	M. Chilvers
MI_SF_1_36	-86.362	41.828	East North Central	US	MI	Soybean	Galien, MI	US1A	2011	M. Chilvers
MI_SF_2_16	-85.836	42.624	East North Central	US	MI	Soybean	Allegan County, MI	US1A	2011	M. Chilvers
MI_SF_9_8	-84.72	42.881	East North Central	US	MI	Soybean	Westphalia, MI	US1A	2011	M. Chilvers
MP144	-84.27	37.839	Central	US	KY	Soybean	Unknown	US2		R. Baird. Obtained from Kiersten Wise
MP220	-86.58	35.517	Central	US	TN	Soybean	Unknown	US1A		R. Baird. Obtained from Kiersten Wise
MP223	-99.902	31.969	South	US	TX	Soybean	Unknown	US2		R. Baird. Obtained from Kiersten Wise
MP249	-82.9	32.166	Southeast	US	GA	Soybean	Unknown	US2		R. Baird. Obtained from Kiersten Wise
MP250	-82.9	32.166	Southeast	US	GA	Soybean	Unknown	US2		R. Baird. Obtained from Kiersten Wise
MP258	-81.164	33.836	Southeast	US	SC	Soybean	Unknown	US1B		R. Baird. Obtained from Kiersten Wise
M_15_10	-84.339	43.511	East North Central	US	MI	Dry bean	Merrill, MI	US1A	2015	M. Chilvers
M_15_11	-84.339	43.511	East North Central	US	MI	Dry bean	Merrill, MI	US1A	2015	M. Chilvers
M_15_12	-83.061	43.75	East North Central	US	MI	Dry bean	Bad Axe, MI	US1A	2015	M. Chilvers
M_15_13	-83.059	43.75	East North Central	US	MI	Dry bean	Bad Axe, MI	US1A	2015	M. Chilvers
M_15_14	-83.061	43.75	East North Central	US	MI	Dry bean	Bad Axe, MI	US1A	2015	M. Chilvers
M_15_1	-84.419	43.451	East North Central	US	MI	Dry bean	Wheeler, MI	US1A	2015	M. Chilvers
M_15_2	-84.343	43.525	East North Central	US	MI	Dry bean	Merrill, MI	US1A	2015	M. Chilvers
M_15_3	-84.343	43.525	East North Central	US	MI	Dry bean	Merrill, MI	US1A	2015	M. Chilvers
M_15_4	-84.343	43.525	East North Central	US	MI	Dry bean	Merrill, MI	US2	2015	M. Chilvers
M_15_5	-84.343	43.525	East North Central	US	MI	Dry bean	Merrill, MI	US1A	2015	M. Chilvers
M_15_6	-84.342	43.525	East North Central	US	MI	Dry bean	Merrill, MI	US1A	2015	M. Chilvers
M_15_7	-84.342	43.525	East North Central	US	MI	Dry bean	Merrill, MI	US1A	2015	M. Chilvers
M_15_8	-84.342	43.525	East North Central	US	MI	Dry bean	Merrill, MI	US1A	2015	M. Chilvers
M_15_9	-84.339	43.511	East North Central	US	MI	Dry bean	Merrill, MI	US1A	2015	M. Chilvers
Mdio	-90.443	43.185	East North Central	US	WI	Soybean	Muscodia, WI	US1A		T. Hughes. Obtained from Kiersten Wise
Md3	-90.443	43.185	East North Central	US	WI	Soybean	Muscodia, WI	US1A		T. Hughes. Obtained from Kiersten Wise
Md5	-90.443	43.185	East North Central	US	WI	Soybean	Muscodia, WI	US1A		T. Hughes. Obtained from Kiersten Wise
Md6	-90.443	43.185	East North Central	US	WI	Soybean	Muscodia, WI	US1A		T. Hughes. Obtained from Kiersten Wise
Md7	-90.443	43.185	East North Central	US	WI	Soybean	Muscodia, WI	US1A		T. Hughes. Obtained from Kiersten Wise
MPsDSU	-96.801	44.438	West North Central	US	SD	Ethiopian mustard	Brookings County, SD	US2		Febina Mathew - SD state U
Mph_21	-76.312	3.892	Colombia	COL	VAC	Dry bean	Buga, VAC	COLPR2	2010	Gloria Mosquera - CIAT
Mph_22	-76.312	3.892	Colombia	COL	VAC	Dry bean	Buga, VAC	US1B	2010	Gloria Mosquera - CIAT
Mph_23	-76.312	3.892	Colombia	COL	VAC	Dry bean	Buga, VAC	US1B	2010	Gloria Mosquera - CIAT
Mph_24	-76.485	3.012	Colombia	COL	CAU	Dry bean	Santander de Quilichao, CAU	COLPR1	2010	Gloria Mosquera - CIAT
Mph_27	-76.355	3.503	Colombia	COL	VAC	Dry bean	Palмира, VAC	COLPR1	2010	Gloria Mosquera - CIAT
Mph_28	-76.355	3.503	Colombia	COL	VAC	Dry bean	Palмира, VAC	COLPR2	2010	Gloria Mosquera - CIAT
Mph_35	-76.355	3.503	Colombia	COL	VAC	Dry bean	Palмира, VAC	COLPR2	2010	Gloria Mosquera - CIAT
Mph_38	-76.355	3.503	Colombia	COL	VAC	Dry bean	Palмира, VAC	COLPR1	2010	Gloria Mosquera - CIAT
Mph_39	-76.355	3.503	Colombia	COL	VAC	Dry bean	Palмира, VAC	COLPR1	2010	Gloria Mosquera - CIAT
Mph_40	-76.355	3.503	Colombia	COL	VAC	Dry bean	Palмира, VAC	NA	2010	Gloria Mosquera - CIAT
Mph_45	-76.485	3.012	Colombia	COL	CAU	Dry bean	Santander de Quilichao, CAU	COLPR1	2012	Gloria Mosquera - CIAT
Mph_46	-76.485	3.012	Colombia	COL	CAU	Dry bean	Santander de Quilichao, CAU	COLPR1	2012	Gloria Mosquera - CIAT
Mph_48	-76.485	3.012	Colombia	COL	CAU	Dry bean	Santander de Quilichao, CAU	COLPR1	2013	Gloria Mosquera - CIAT
Mph_49	-76.485	3.012	Colombia	COL	CAU	Dry bean	Santander de Quilichao, CAU	US1A	2013	Gloria Mosquera - CIAT
Mph_50	-74.905	4.965	Colombia	COL	TOL	Dry bean	Armero, TOL	COLPR1	2013	Gloria Mosquera - CIAT
Mph_52	-74.224	10.727	Colombia	COL	MAG	Dry bean	Corpoica, MAG	COLPR1	2014	Gloria Mosquera - CIAT
Mph_53	-74.224	10.727	Colombia	COL	MAG	Dry bean	Corpoica, MAG	COLPR1	2014	Gloria Mosquera - CIAT
Mph_54	-74.224	10.727	Colombia	COL	MAG	Dry bean	Corpoica, MAG	COLPR2	2014	Gloria Mosquera - CIAT
Mph_56	-74.224	10.727	Colombia	COL	MAG	Dry bean	Corpoica, MAG	COLPR2	2014	Gloria Mosquera - CIAT
Mph_5	-76.485	3.012	Colombia	COL	CAU	Dry bean	Santander de Quilichao, CAU	COLPR1	2002	Gloria Mosquera - CIAT
SAG5_4	-84.313	43.348	East North Central	US	MI	Soybean	Saginaw County, MI	US1A	2012	M. Chilvers
TN262	-86.58	35.517	Central	US	TN	Soybean	Unknown	US1A		A. Mengistu. Obtained from Kiersten Wise
TN270	-90.932	33.424	South	US	MS	Soybean	Unknown	US1B		A. Mengistu. Obtained from Kiersten Wise
TN279	-88.846	35.622	Central	US	TN	Soybean	Jackson, TN	US1A		A. Mengistu. Obtained from Kiersten Wise
TN280	-84.27	37.839	Central	US	KY	Soybean	Unknown	US1A		A. Mengistu. Obtained from Kiersten Wise
TN380	-90.932	33.424	South	US	MS	Soybean	Stoneville, Mississippi	US1B		A. Mengistu. Obtained from Kiersten Wise
TN4	-88.846	35.622	Central	US	TN	Soybean	Jackson, Tennessee	US1A		A. Mengistu. Obtained from Kiersten Wise
TN501	-91.962	30.984	South	US	LA	Soybean	Unknown	COLPR2		A. Mengistu. Obtained from Kiersten Wise
TN550	-99.902	31.969	South	US	TX	Soybean	Unknown	US2		A. Mengistu. Obtained from Kiersten Wise
TN5	-89.862	35.418	Central	US	TN	Soybean	Ames, TN	US1A		A. Mengistu. Obtained from Kiersten Wise
UPR_Mph_ISA2	-67.056	18.462	Puerto Rico	PR	ISA	Dry bean	Isabela, PR	COLPR1		Consuelo Estevez De Jensen - UPR
UPR_Mph_ISA3	-67.056	18.462	Puerto Rico	PR	ISA	Dry bean	Isabela, PR	COLPR2		Consuelo Estevez De Jensen - UPR
UPR_Mph_JD1	-66.518	18.019	Puerto Rico	PR	JD	Dry bean	Juana Diaz, PR	COLPR1		Consuelo Estevez De Jensen - UPR
UPR_Mph_JD2	-66.518	18.019	Puerto Rico	PR	JD	Dry bean	Juana Diaz, PR	COLPR1		Consuelo Estevez De Jensen - UPR
UPR_Mph_JD3	-66.518	18.019	Puerto Rico	PR	JD	Dry bean	Juana Diaz, PR	COLPR1		Consuelo Estevez De Jensen - UPR
W12_6	-85.999	42.671	East North Central	US	MI	Soybean	Hamilton, MI	US2	2012	M. Chilvers
W23	-84.847	43.001	East North Central	US	MI	Soybean	Pewamo, MI	US1A	2012	M. Chilvers
W25	-85.999	42.671	East North Central	US	MI	Soybean	Hamilton, MI	US2	2012	M. Chilvers
W3_5	-85.999	42.671	East North Central	US	MI	Soybean	Hamilton, MI	US2	2012	M. Chilvers
W_MISO2_3_6	-86.1	42.7	East North Central	US	MI	Soybean	Hamilton, MI	US1A	2012	M. Chilvers
W_MISO2_4_10	-85.1	42.9	East North Central	US	MI	Soybean	Berlin, MI	US1A	2012	M. Chilvers

Table A.2 Hierarchical analysis of molecular variance (AMOVA), partitioning total genetic variance into the following components: between clades, between genetic clusters and within genetic clusters. Clone corrected values are shown. Most of the variance was associated with differences between clades and between genetic clusters.

Source of variation	Variation (%)	p-value	Phi
Between clades (US and COLPR)	46.59	0.001	0.89
Between genetic clusters (US-1A, US-1B, US-2), (COLPR-1, COLPR-2) within clade	42.41	0.001	0.79
Within genetic clusters	10.99	0.001	0.46

Table A.3 Summary statistics for genetic diversity of *Macrophomina phaseolina* by country. N is number of isolates (sample size); MLG is number of observed multilocus genotypes; eMLG is the number of expected MLG at a sample size of 5 based on rarefaction. MLL is number of observed multilocus lineages by population using a bitwise cutoff distance of 0.0001; CF is clonal fraction ($1 - (MLL/N)$). Clone corrected values are shown and indicated by asterisks for indices of genotypic diversity: Shannon-Wiener Index (H^*), Stoddart and Taylor's Index (G^*), Simpson's index (λ^*) and evenness (E_5^*).

Country	N	Gene diversity (H_e)	Median pairwise genetic distance	MLG	eMLG	MLL	eMLL	CF	H^*	G^*	λ^*	E_5^*
US	70	0.104	0.096	54	4.85	21	3.83	0.7	1.27	3.54	0.69	0.921
Colombia	20	0.263	0.355	20	5	12	4.54	0.4	1.47	4.3	0.758	0.966
Puerto Rico	5	0.163	0.101	5	5	4	4	0.2	1.33	3.57	0.72	0.922

Table A.4 Top 49 candidate SNPs along the first three RDA axes, after accounting for neutral population structure using the LD-filtered set of 11,421 SNPs.

RDA axis	SNP position	RDA loading	Climate variable	Correlation	SnpEff SNP category	SnpEff predicted effect	SnpEff annotation locus	SnpEff distance from locus (bp)
1	scaffold_13_111451	-0.068	TSsd	0.90	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_13_108299_108301	3150
1	scaffold_44_156076	-0.068	TSsd	0.89	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_44_152100_152102	3974
1	scaffold_3_1825169	0.070	TSsd	0.62	missense_variant	MODERATE	CDS_scaffold_3_1824442_1825656	0
1	scaffold_3_1762959	-0.065	TSsd	0.62	upstream_gene_variant	MODIFIER	START_CODON_scaffold_3_1764065_1764067	1106
1	scaffold_6_1672095	-0.074	TSsd	0.61	upstream_gene_variant	MODIFIER	START_CODON_scaffold_6_1672850_1672852	755
1	scaffold_13_484791	-0.064	TSsd	0.59	upstream_gene_variant	MODIFIER	START_CODON_scaffold_13_485354_485356	563
1	scaffold_1_2854576	-0.070	TSsd	0.56	synonymous_variant	LOW	CDS_scaffold_1_2854387_2854822	0
1	scaffold_59_25310	-0.066	TSsd	0.51	synonymous_variant	LOW	CDS_scaffold_59_24795_25349	0
1	scaffold_22_420202	-0.063	TSsd	0.50	upstream_gene_variant	MODIFIER	CDS_scaffold_22_414871_415310	4892
1	scaffold_48_217394	-0.071	TSsd	0.45	upstream_gene_variant	MODIFIER	CDS_scaffold_48_212041_212424	4970
1	scaffold_3_91936	0.072	TSsd	0.39	upstream_gene_variant	MODIFIER	START_CODON_scaffold_3_93112_93114	1176
1	scaffold_6_1470614	0.066	TSsd	0.37	intergenic_region	MODIFIER	CDS_scaffold_6_1456157_1456265-START_CODON_scaffold_6_1482386_1482388	0
1	scaffold_2_1936912	0.071	TSsd	0.35	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_2_1936467_1936469	443
1	scaffold_42_203393	0.073	mITwq	0.29	synonymous_variant	LOW	CDS_scaffold_42_203293_204316	0
1	scaffold_41_309748	0.065	mITwq	0.29	downstream_gene_variant	MODIFIER	CDS_scaffold_41_303593_304907	4841
2	scaffold_1_1608893	-0.061	TSsd	0.80	synonymous_variant	LOW	CDS_scaffold_1_1608703_1612256	0
2	scaffold_87_2593	-0.086	TSsd	0.77	upstream_gene_variant	MODIFIER	START_CODON_scaffold_87_3569_3571	976
2	scaffold_58_78470	-0.064	TSsd	0.69	synonymous_variant	LOW	CDS_scaffold_58_78407_78835	0
2	scaffold_12_462369	-0.063	PScv	0.69	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_12_460099_460101	2268
2	scaffold_25_15545	-0.063	PScv	0.69	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_25_121760_121762	3783
2	scaffold_29_158130	-0.063	PScv	0.69	synonymous_variant	LOW	CDS_scaffold_29_157893_158372	0
2	scaffold_71_166465	-0.063	PScv	0.69	upstream_gene_variant	MODIFIER	CDS_scaffold_71_159528_161501	4964
2	scaffold_87_1928	-0.063	PScv	0.69	upstream_gene_variant	MODIFIER	START_CODON_scaffold_87_3569_3571	1641
2	scaffold_2_523738	0.062	PScv	0.60	upstream_gene_variant	MODIFIER	CDS_scaffold_2_518495_519535	4203
2	scaffold_21_52404	-0.062	PScv	0.60	upstream_gene_variant	MODIFIER	START_CODON_scaffold_21_53157_53159	753
2	scaffold_28_275982	-0.062	PScv	0.60	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_28_272569_272571	3411
2	scaffold_9_654713	-0.062	PScv	0.59	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_9_649861_649863	4850
2	scaffold_19_550360	0.061	TSsd	0.58	upstream_gene_variant	MODIFIER	START_CODON_scaffold_19_551212_551214	852
2	scaffold_1_3049711	0.063	PScv	0.54	missense_variant	MODERATE	CDS_scaffold_1_3048856_3052181	0
2	scaffold_54_150712	-0.064	TSsd	0.47	synonymous_variant	LOW	CDS_scaffold_54_150535_151353	0
2	scaffold_27_176390	-0.060	TSsd	0.42	synonymous_variant	LOW	CDS_scaffold_27_175960_176666	0
2	scaffold_23_382823	-0.060	TSsd	0.41	upstream_gene_variant	MODIFIER	CDS_scaffold_23_375724_378157	4666
2	scaffold_6_1522882	-0.074	PScv	0.38	upstream_gene_variant	MODIFIER	CDS_scaffold_6_1517175_1517921	4961
2	scaffold_13_208687	-0.059	mITwq	0.34	synonymous_variant	LOW	CDS_scaffold_13_208376_209227	0
2	scaffold_10_615323	-0.064	PScv	0.33	synonymous_variant	LOW	CDS_scaffold_10_615173_617203	0
2	scaffold_22_158097	-0.067	PScv	0.33	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_22_153686_153688	4409
3	scaffold_1_1378025	0.055	TSsd	0.69	synonymous_variant	LOW	CDS_scaffold_1_1377723_1378766	0
3	scaffold_10_479849	-0.044	TSsd	0.68	upstream_gene_variant	MODIFIER	START_CODON_scaffold_10_483512_483514	3663
3	scaffold_23_382747	0.045	PScv	0.66	upstream_gene_variant	MODIFIER	CDS_scaffold_23_375724_378157	4590
3	scaffold_19_550409	-0.048	TSsd	0.61	upstream_gene_variant	MODIFIER	START_CODON_scaffold_19_551212_551214	803
3	scaffold_22_499377	0.047	TSsd	0.61	downstream_gene_variant	MODIFIER	CDS_scaffold_22_494417_495021	4356
3	scaffold_1_3035852	0.047	PScv	0.48	upstream_gene_variant	MODIFIER	CDS_scaffold_1_3030823_303518	4334
3	scaffold_22_439131	0.045	TSsd	0.46	missense_variant	MODERATE	CDS_scaffold_22_438958_440858	0
3	scaffold_2_1080524	0.045	TSsd	0.43	upstream_gene_variant	MODIFIER	CDS_scaffold_2_1075512_1076051	4473
3	scaffold_23_480413	0.046	Pwq	0.42	upstream_gene_variant	MODIFIER	CDS_scaffold_23_475068_476090	4323
3	scaffold_42_240411	-0.050	TSsd	0.39	synonymous_variant	LOW	CDS_scaffold_42_240295_240429	0
3	scaffold_29_245703	0.049	TSsd	0.39	upstream_gene_variant	MODIFIER	CDS_scaffold_29_240677_241789	3914
3	scaffold_7_1460003	-0.051	TSsd	0.37	synonymous_variant	LOW	CDS_scaffold_7_1459542_1460852	0
3	scaffold_7_1459970	-0.046	TSsd	0.30	synonymous_variant	LOW	CDS_scaffold_7_1459542_1460852	0

Table A.5 Top 75 candidate SNPs along the first three RDA axes, after accounting for neutral population structure using the the full set of 77,465 SNPs.

RDA axis	SNP position	RDA loading	Climate variable	Correlation	SnEff SNP category	SnEff predicted effect	SnEff annotation locus	SnEff distance from locus (bp)
1	scaffold_13_111451	-0.045	TSsd	0.90	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_13_108299_108301	3150
1	scaffold_44_156076	-0.044	TSsd	0.89	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_44_152100_152102	3974
1	scaffold_10_1141677	-0.044	TSsd	0.84	upstream_gene_variant	MODIFIER	CDS_scaffold_10_1136664_1136859	4818
1	scaffold_16_102210	-0.044	TSsd	0.84	synonymous_variant	LOW	CDS_scaffold_16_101948_102440	0
1	scaffold_4_1410133	-0.044	TSsd	0.84	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_4_1407455_1407457	2676
1	scaffold_4_1410339	-0.044	TSsd	0.84	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_4_1407455_1407457	2882
1	scaffold_4_1410344	-0.044	TSsd	0.84	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_4_1407455_1407457	2887
1	scaffold_4_1410417	-0.044	TSsd	0.84	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_4_1407455_1407457	2960
1	scaffold_4_1410558	-0.044	TSsd	0.84	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_4_1407455_1407457	3101
1	scaffold_4_1410743	-0.044	TSsd	0.84	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_4_1407455_1407457	3286
1	scaffold_7_663614	-0.044	TSsd	0.84	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_7_661303_661305	2309
1	scaffold_48_217394	-0.045	TSsd	0.45	upstream_gene_variant	MODIFIER	CDS_scaffold_48_212041_212424	4970
1	scaffold_58_127959	0.047	TSsd	0.43	missense_variant	MODERATE	CDS_scaffold_58_127946_128524	0
1	scaffold_104_79048	0.047	TSsd	0.40	synonymous_variant	LOW	CDS_scaffold_104_78954_79095	0
1	scaffold_104_79165	0.047	TSsd	0.40	upstream_gene_variant	MODIFIER	CDS_scaffold_104_73200_74703	4462
1	scaffold_19_649013	-0.047	TSsd	0.40	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_19_645372_645374	3639
1	scaffold_36_272346	-0.047	TSsd	0.40	synonymous_variant	LOW	CDS_scaffold_36_269772_272918	0
1	scaffold_58_127971	0.047	TSsd	0.40	missense_variant	MODERATE	CDS_scaffold_58_127946_128524	0
1	scaffold_1_2153455	0.046	TSsd	0.38	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_1_2149297_2149299	4156
1	scaffold_1_3063445	-0.046	TSsd	0.38	downstream_gene_variant	MODIFIER	START_CODON_scaffold_1_3058748_3058750	4695
1	scaffold_58_128008	0.046	TSsd	0.38	missense_variant	MODERATE	CDS_scaffold_58_127946_128524	0
1	scaffold_3_1799971	-0.045	TSsd	0.38	synonymous_variant	LOW	CDS_scaffold_3_1799677_1800392	0
1	scaffold_3_1801510	-0.045	TSsd	0.38	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_3_1797900_1797902	3608
1	scaffold_3_1801567	-0.045	TSsd	0.38	synonymous_variant	LOW	CDS_scaffold_3_1801544_1801752	0
1	scaffold_42_203393	0.044	mTwq	0.29	synonymous_variant	LOW	CDS_scaffold_42_203293_204316	0
2	scaffold_87_2593	-0.049	TSsd	0.77	upstream_gene_variant	MODIFIER	START_CODON_scaffold_87_3569_3571	976
2	scaffold_54_28428	-0.045	TSsd	0.73	downstream_gene_variant	MODIFIER	CDS_scaffold_54_23338_23782	4646
2	scaffold_12_462369	-0.046	PScv	0.69	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_12_460099_460101	2268
2	scaffold_12_462379	-0.046	PScv	0.69	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_12_460099_460101	2278
2	scaffold_25_125545	-0.046	PScv	0.69	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_25_121760_121762	3783
2	scaffold_29_158130	-0.046	PScv	0.69	synonymous_variant	LOW	CDS_scaffold_29_157893_158372	0
2	scaffold_71_166465	-0.046	PScv	0.69	upstream_gene_variant	MODIFIER	CDS_scaffold_71_159528_161501	4964
2	scaffold_87_1928	-0.046	PScv	0.69	upstream_gene_variant	MODIFIER	START_CODON_scaffold_87_3569_3571	1641
2	scaffold_27_280895	0.039	PScv	0.69	synonymous_variant	LOW	CDS_scaffold_27_280296_281132	0
2	scaffold_29_188311	-0.041	TSsd	0.67	downstream_gene_variant	MODIFIER	CDS_scaffold_29_182261_184761	3750
2	scaffold_9_654682	-0.041	TSsd	0.67	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_9_649861_649863	4819
2	scaffold_19_360816	-0.041	TSsd	0.67	synonymous_variant	LOW	CDS_scaffold_19_360656_361318	0
2	scaffold_23_382747	-0.042	PScv	0.66	upstream_gene_variant	MODIFIER	CDS_scaffold_23_375724_378157	4590
2	scaffold_29_158052	-0.046	PScv	0.65	synonymous_variant	LOW	CDS_scaffold_29_157893_158372	0
2	scaffold_14_991443	-0.043	PScv	0.64	synonymous_variant	LOW	CDS_scaffold_14_989727_991629	0
2	scaffold_57_80182	-0.042	PScv	0.62	synonymous_variant	LOW	CDS_scaffold_57_80140_80242	0
2	scaffold_2_523738	0.044	PScv	0.60	upstream_gene_variant	MODIFIER	CDS_scaffold_2_518495_519353	4203
2	scaffold_21_52404	-0.044	PScv	0.60	upstream_gene_variant	MODIFIER	START_CODON_scaffold_21_53157_53159	753
2	scaffold_28_275982	-0.044	PScv	0.60	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_28_272569_272571	3411
2	scaffold_62_183467	-0.039	PScv	0.59	missense_variant	MODERATE	CDS_scaffold_62_182889_183615	0
2	scaffold_62_183468	-0.039	PScv	0.59	missense_variant	MODERATE	CDS_scaffold_62_182889_183615	0
2	scaffold_13_172698	-0.040	PScv	0.59	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_13_168216_168218	4480
2	scaffold_9_654713	-0.043	PScv	0.59	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_9_649861_649863	4850
2	scaffold_17_614054	0.039	TSsd	0.55	synonymous_variant	LOW	CDS_scaffold_17_613826_614191	0
2	scaffold_1_3049711	0.040	PScv	0.54	missense_variant	MODERATE	CDS_scaffold_1_304886_3052181	0
2	scaffold_72_169626	-0.039	PScv	0.54	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_72_167171_167173	2453
2	scaffold_2_198659	0.040	TSsd	0.52	missense_variant	MODERATE	CDS_scaffold_2_198491_198675	0
2	scaffold_23_382823	-0.044	TSsd	0.41	upstream_gene_variant	MODIFIER	CDS_scaffold_23_375724_378157	4666
2	scaffold_6_1522882	-0.047	PScv	0.38	upstream_gene_variant	MODIFIER	CDS_scaffold_6_151715_1517921	4961
2	scaffold_7_1460003	0.039	TSsd	0.37	synonymous_variant	LOW	CDS_scaffold_7_1459542_1460852	0
3	scaffold_12_1023368	0.031	TSsd	0.77	synonymous_variant	LOW	CDS_scaffold_12_1023214_1024907	0
3	scaffold_12_1057197	0.031	TSsd	0.77	downstream_gene_variant	MODIFIER	CDS_scaffold_12_1058753_1059094	1556
3	scaffold_16_140381	0.031	TSsd	0.77	synonymous_variant	LOW	CDS_scaffold_16_139893_144227	0
3	scaffold_16_140414	0.031	TSsd	0.77	synonymous_variant	LOW	CDS_scaffold_16_139893_144227	0
3	scaffold_16_421358	-0.031	TSsd	0.77	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_16_417163_417165	4193
3	scaffold_16_421368	-0.031	TSsd	0.77	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_16_417163_417165	4203
3	scaffold_58_220314	-0.031	TSsd	0.77	upstream_gene_variant	MODIFIER	START_CODON_scaffold_58_220995_220997	681
3	scaffold_67_188980	-0.030	TSsd	0.75	synonymous_variant	LOW	CDS_scaffold_67_188284_189448	0
3	scaffold_19_550409	-0.030	TSsd	0.61	upstream_gene_variant	MODIFIER	START_CODON_scaffold_19_551212_551214	803
3	scaffold_15_782911	0.035	TSsd	0.54	synonymous_variant	LOW	CDS_scaffold_15_781874_786184	0
3	scaffold_29_196573	0.031	mTwq	0.44	synonymous_variant	LOW	CDS_scaffold_29_195969_197925	0
3	scaffold_10_1087855	-0.032	TSsd	0.40	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_10_1083679_1083681	4174
3	scaffold_42_225860	0.031	mTwq	0.38	downstream_gene_variant	MODIFIER	START_CODON_scaffold_42_221936_221938	3922
3	scaffold_42_225871	0.031	mTwq	0.38	downstream_gene_variant	MODIFIER	START_CODON_scaffold_42_221936_221938	3933
3	scaffold_70_80051	0.031	mTwq	0.38	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_70_75255_75257	4794
3	scaffold_58_219039	-0.032	mTwq	0.37	upstream_gene_variant	MODIFIER	START_CODON_scaffold_58_220995_220997	1956
3	scaffold_23_297720	-0.031	mTwq	0.36	upstream_gene_variant	MODIFIER	CDS_scaffold_23_29212_293048	4672
3	scaffold_48_217915	0.035	mTwq	0.35	missense_variant	MODERATE	CDS_scaffold_48_217500_218137	0
3	scaffold_42_225897	0.034	mTwq	0.34	downstream_gene_variant	MODIFIER	START_CODON_scaffold_42_221936_221938	3959
3	scaffold_1_1537417	-0.029	mTwq	0.31	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_1_1534130_1534132	3285

Table A.6 Candidate SNPs and gene models along the first RDA axis, after accounting for neutral population structure using the full set of 77,465 SNPs.

SNP position	RDA1 loading	Climate variable	Correlation	SNP category	Distance from locus (bp)	Mycocosm gene location	Mycocosm protein ID	InterPro/KOG Desc
scaffold_13:111451	-0.045	TSsd	0.90	Intergenic	3150	scaffold_13:108068-110230	753275	Ankyrin repeat
scaffold_44:156076	-0.044	TSsd	0.89	Intergenic	3974	scaffold_44:152589-153590	681752	Uncharacterized conserved protein
scaffold_10:1141677	-0.044	TSsd	0.84	Intergenic	40	scaffold_10:1140297-1141637	698606	Glycoside hydrolase, family 5
scaffold_16:102210	-0.044	TSsd	0.84	Synonymous	0	scaffold_16:101708-102507	666336	None
scaffold_4:1410133	-0.044	TSsd	0.84	Intergenic	2676	scaffold_4:1407052-1409077	767171	None
scaffold_4:1410339	-0.044	TSsd	0.84	Intergenic	2882	scaffold_4:1407052-1409077	767171	None
scaffold_4:1410344	-0.044	TSsd	0.84	Intergenic	2887	scaffold_4:1407052-1409077	767171	None
scaffold_4:1410417	-0.044	TSsd	0.84	Intergenic	2960	scaffold_4:1407052-1409077	767171	None
scaffold_4:1410558	-0.044	TSsd	0.84	Intergenic	3101	scaffold_4:1407052-1409077	767171	None
scaffold_4:1410743	-0.044	TSsd	0.84	Intergenic	3286	scaffold_4:1407052-1409077	767171	None
scaffold_7:663614	-0.044	TSsd	0.84	Intergenic	116	scaffold_7:660987-663498	726238	None
scaffold_48:217394	-0.045	TSsd	0.45	Intergenic	11	scaffold_48:215027-217383	714043	Flavin-containing monooxygenase
scaffold_58:127959	0.047	TSsd	0.43	Missense	0	scaffold_58:127946-129509	722169	Glycoside hydrolase, family 5
scaffold_104:79048	0.047	TSsd	0.40	Synonymous	0	scaffold_104:78761-79095	747473	None
scaffold_104:79165	0.047	TSsd	0.40	Intergenic	4462	scaffold_104:71908-75304	59925	Mg2+ transporter protein, CorA-like
scaffold_19:649013	-0.047	TSsd	0.40	Intergenic	3639	scaffold_19:645233-646550	787237	Glycoside hydrolase, family 10
scaffold_36:272346	-0.047	TSsd	0.40	Synonymous	0	scaffold_36:269703-273261	270878	None
scaffold_58:127971	0.047	TSsd	0.40	Missense	0	scaffold_58:127946-129509	722169	Glycoside hydrolase, family 5
scaffold_1:2153455	0.046	TSsd	0.38	Intergenic	79	scaffold_1:2152160-2153376	28417	Alpha crystallin/Hsp20 domain
scaffold_1:3063445	-0.046	TSsd	0.38	Intergenic	377	scaffold_1:3063822-3065544	640605	None
scaffold_58:128008	0.046	TSsd	0.38	Missense	0	scaffold_58:127946-129509	722169	Glycoside hydrolase, family 5
scaffold_3:1799971	-0.045	TSsd	0.38	Synonymous	0	scaffold_3:1799298-1801088	724609	None
scaffold_3:1801510	-0.045	TSsd	0.38	Intergenic	3608	scaffold_3:1797593-1798873	645899	Protein kinase-like domain
scaffold_3:1801567	-0.045	TSsd	0.38	Synonymous	0	scaffold_3:1801258-1802852	645904	None
scaffold_42:203393	0.044	mTwq	0.29	Synonymous	0	scaffold_42:202488-204519	608404	Cytochrome P450, E-class, group I

Table A.7 Top 91 candidate SNPs along the first three RDA axes, without accounting for neutral population structure using the LD-filtered set of 11,421 SNPs.

RDA axis	SNP position	RDA loading	Climate variable	Correlation	Snpeff SNP category	Snpeff predicted effect	Snpeff annotation locus	Snpeff distance from locus (bp)
2	scaffold_10_854700	0.120	PScv	0.43	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_10_851859_851861	2839
2	scaffold_6_234790	-0.123	PScv	0.42	upstream_gene_variant	MODIFIER	CDS_scaffold_6_228361_230947	3843
2	scaffold_71_83900	-0.107	PScv	0.42	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_71_83388_83590	2310
2	scaffold_41_309781	0.118	mTww	0.41	downstream_gene_variant	MODIFIER	CDS_scaffold_41_303593_304907	4874
2	scaffold_6_1418421	0.131	PScv	0.38	downstream_gene_variant	MODIFIER	START_CODON_scaffold_6_1413560_1413562	4859
2	scaffold_24_200868	-0.120	mTww	0.37	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_24_197382_197384	3284
2	scaffold_44_122394	-0.103	PScv	0.36	synonymous_variant	LOW	CDS_scaffold_44_121566_123424	0
2	scaffold_2_1026126	-0.101	PScv	0.31	upstream_gene_variant	MODIFIER	START_CODON_scaffold_2_1029176_1029178	3050
2	scaffold_16_132289	-0.101	mTww	0.30	synonymous_variant	LOW	CDS_scaffold_16_131336_132715	0
2	scaffold_2_218789	-0.103	mTww	0.30	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_2_213830_213832	4957
2	scaffold_5_168493	-0.101	mTww	0.30	upstream_gene_variant	MODIFIER	CDS_scaffold_5_1678605_1679687	4506
2	scaffold_10_672972	-0.102	PScv	0.29	synonymous_variant	LOW	CDS_scaffold_10_672926_673100	0
2	scaffold_21_102468	-0.102	PScv	0.29	intergenic_region	MODIFIER	CDS_scaffold_21_89865_90404-START_CODON_scaffold_21_110672_110674	0
2	scaffold_37_363077	-0.100	mTww	0.27	upstream_gene_variant	MODIFIER	CDS_scaffold_37_358101_358247	4830
2	scaffold_44_307914	-0.101	PScv	0.27	synonymous_variant	LOW	CDS_scaffold_44_307351_308453	0
2	scaffold_6_127167	-0.101	PScv	0.27	missense_variant	MODERATE	CDS_scaffold_6_126424_127606	0
2	scaffold_68_133359	-0.104	PScv	0.27	upstream_gene_variant	MODIFIER	CDS_scaffold_68_130352_130591	4768
3	scaffold_2_211455	0.073	mTww	0.55	upstream_gene_variant	MODIFIER	START_CODON_scaffold_2_212300_212302	845
3	scaffold_21_335353	0.073	mTww	0.55	upstream_gene_variant	MODIFIER	START_CODON_scaffold_21_336487_336489	934
3	scaffold_40_343692	0.073	mTww	0.55	synonymous_variant	LOW	CDS_scaffold_40_342943_344721	0
3	scaffold_15_357578	0.079	mTww	0.54	missense_variant	MODERATE	CDS_scaffold_15_357405_357970	0
3	scaffold_44_201020	0.082	mTww	0.54	upstream_gene_variant	MODIFIER	START_CODON_scaffold_44_204912_204914	3892
3	scaffold_5_221370	0.082	mTww	0.54	synonymous_variant	LOW	CDS_scaffold_5_220717_222345	0
3	scaffold_13_391413	0.074	mTww	0.51	upstream_gene_variant	MODIFIER	CDS_scaffold_13_38752_386447	4966
3	scaffold_17_178704	0.074	mTww	0.51	synonymous_variant	LOW	CDS_scaffold_17_178494_179513	0
3	scaffold_28_120398	0.074	mTww	0.51	synonymous_variant	LOW	CDS_scaffold_28_119753_121192	0
3	scaffold_59_15066	0.074	mTww	0.51	upstream_gene_variant	MODIFIER	START_CODON_scaffold_59_15635_15637	569
3	scaffold_68_115312	0.077	mTww	0.51	missense_variant	MODERATE	CDS_scaffold_68_115232_115678	0
3	scaffold_79_109649	0.074	mTww	0.51	missense_variant	MODERATE	CDS_scaffold_79_109368_109971	0
3	scaffold_11_102345	0.085	mTww	0.50	upstream_gene_variant	MODIFIER	CDS_scaffold_11_101842_1020023	3412
3	scaffold_15_360852	-0.088	mTww	0.50	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_15_358354_358356	2496
3	scaffold_19_393993	0.097	mTww	0.49	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_19_38706_389708	4285
3	scaffold_32_22972	-0.094	mTww	0.49	splice_region_variant&synonymous_variant	LOW	CDS_scaffold_32_22972_23031	0
3	scaffold_44_206095	0.074	mTww	0.49	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_44_201790_201792	4303
3	scaffold_2_273644	0.073	TSsd	0.48	synonymous_variant	LOW	CDS_scaffold_2_173588_173972	0
3	scaffold_24_332185	0.073	TSsd	0.48	upstream_gene_variant	MODIFIER	CDS_scaffold_24_339444_334392	1759
3	scaffold_13_426280	0.076	TSsd	0.47	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_13_424653_424655	1625
3	scaffold_17_627080	0.083	mTww	0.47	upstream_gene_variant	MODIFIER	START_CODON_scaffold_17_627756_627758	676
3	scaffold_3_473402	0.081	mTww	0.47	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_3_468724_468726	4676
3	scaffold_58_76097	0.076	TSsd	0.47	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_58_72657_72659	3458
3	scaffold_6_171488	0.076	TSsd	0.47	missense_variant	MODERATE	CDS_scaffold_6_171215_172006	0
3	scaffold_7_57475	0.081	mTww	0.47	synonymous_variant	LOW	CDS_scaffold_7_57449_58312	0
3	scaffold_3_707374	0.089	mTww	0.46	synonymous_variant	LOW	CDS_scaffold_3_707119_707433	0
3	scaffold_6_151014	0.089	mTww	0.46	upstream_gene_variant	MODIFIER	START_CODON_scaffold_6_1514383_1514385	4269
3	scaffold_62_134355	0.092	mTww	0.46	upstream_gene_variant	MODIFIER	CDS_scaffold_62_1309094_129597	4938
3	scaffold_12_333794	0.079	TSsd	0.45	upstream_gene_variant	MODIFIER	CDS_scaffold_12_328832_329080	4714
3	scaffold_12_460479	0.082	TSsd	0.45	synonymous_variant	LOW	CDS_scaffold_12_460360_461000	0
3	scaffold_21_49304	0.082	TSsd	0.45	upstream_gene_variant	MODIFIER	START_CODON_scaffold_21_51144_51146	1840
3	scaffold_27_375204	0.079	PScv	0.45	upstream_gene_variant	MODIFIER	CDS_scaffold_27_375221_375767	17
3	scaffold_8_83107	0.082	TSsd	0.45	upstream_gene_variant	MODIFIER	CDS_scaffold_8_77182_78164	4943
3	scaffold_13_275316	-0.095	mTww	0.44	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_13_272182_272184	1132
3	scaffold_83_38129	0.088	mTww	0.43	missense_variant	MODERATE	CDS_scaffold_83_37986_38474	0
3	scaffold_10_721881	0.076	mTww	0.42	intergenic_region	MODIFIER	CDS_scaffold_10_691175_692352-START_CODON_scaffold_10_728430_728432	0
3	scaffold_16_424888	0.082	mTww	0.42	upstream_gene_variant	MODIFIER	CDS_scaffold_16_418089_420031	4857
3	scaffold_6_151026	0.085	Pwq	0.42	upstream_gene_variant	MODIFIER	START_CODON_scaffold_6_1514383_1514385	3357
3	scaffold_23_529048	0.077	Pwq	0.41	upstream_gene_variant	MODIFIER	CDS_scaffold_23_523035_524309	4759
3	scaffold_43_121479	0.080	Pwq	0.41	downstream_gene_variant	MODIFIER	CDS_scaffold_43_118170_118541	2938
3	scaffold_19_395821	0.084	Pwq	0.40	missense_variant	MODERATE	CDS_scaffold_19_395633_396036	0
3	scaffold_13_425827	0.086	PScv	0.39	missense_variant	MODERATE	CDS_scaffold_13_424653_426254	0
3	scaffold_44_185863	0.078	mTww	0.39	synonymous_variant	LOW	CDS_scaffold_44_184999_186216	0
3	scaffold_1_1676668	0.075	mTww	0.38	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_1_1674628_1674630	2038
3	scaffold_22_566275	0.084	PScv	0.38	upstream_gene_variant	MODIFIER	CDS_scaffold_22_561257_561331	4944
3	scaffold_27_91501	0.081	PScv	0.38	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_27_86727_86729	4772
3	scaffold_29_316632	0.079	Pwq	0.38	synonymous_variant	LOW	CDS_scaffold_29_316608_316732	0
3	scaffold_36_251653	0.073	Pwq	0.38	upstream_gene_variant	MODIFIER	CDS_scaffold_36_246418_248498	3155
3	scaffold_6_216284	0.076	Pwq	0.38	upstream_gene_variant	MODIFIER	CDS_scaffold_6_210883_211943	4341
3	scaffold_9_644508	0.075	mTww	0.38	upstream_gene_variant	MODIFIER	CDS_scaffold_9_639112_639647	4861
3	scaffold_15_451673	-0.089	Pwq	0.37	missense_variant	MODERATE	CDS_scaffold_15_451651_451894	0
3	scaffold_27_76942	0.090	PScv	0.37	upstream_gene_variant	MODIFIER	CDS_scaffold_27_76991_77398	49
3	scaffold_46_164767	-0.091	Pwq	0.37	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_46_161158_161160	3607
3	scaffold_71_154094	-0.087	mTww	0.37	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_71_151995_151997	2097
3	scaffold_42_135653	0.074	mTww	0.36	synonymous_variant	LOW	CDS_scaffold_42_132413_136111	0
3	scaffold_6_225341	-0.106	mTww	0.36	missense_variant	MODERATE	CDS_scaffold_6_225173_225456	0
3	scaffold_40_256107	-0.087	mTww	0.35	synonymous_variant	LOW	CDS_scaffold_40_255702_256654	0
3	scaffold_13_208687	-0.097	mTww	0.34	synonymous_variant	LOW	CDS_scaffold_13_208376_209227	0
3	scaffold_22_158097	-0.088	PScv	0.33	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_22_153686_153688	4409
3	scaffold_16_1016241	0.077	mTww	0.30	upstream_gene_variant	MODIFIER	CDS_scaffold_16_1030297_1032242	3999
3	scaffold_16_132181	-0.088	mTww	0.28	upstream_gene_variant	MODIFIER	CDS_scaffold_16_1317454_1318444	4837
3	scaffold_45_331999	-0.101	mTww	0.25	synonymous_variant	LOW	CDS_scaffold_45_331450_332788	0
3	scaffold_21_508763	-0.094	PScv	0.24	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_21_5044220_504422	4341
3	scaffold_15_680030	0.080	mTww	0.23	missense_variant	MODERATE	CDS_scaffold_15_679998_680048	0
3	scaffold_3_445922	-0.089	PScv	0.22	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_3_443359_443361	2561
3	scaffold_17_308508	-0.092	mTww	0.21	synonymous_variant	LOW	CDS_scaffold_17_308336_308921	0
3	scaffold_19_209889	0.093	mTww	0.21	missense_variant	MODERATE	CDS_scaffold_19_209760_210050	0
3	scaffold_4_117850	-0.087	mTww	0.21	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_4_113540_113542	4308
3	scaffold_21_39767	0.075	mTww	0.20	missense_variant	MODERATE	CDS_scaffold_21_39612_39944	0
3	scaffold_3_618207	0.079	mTww	0.20	synonymous_variant	LOW	CDS_scaffold_3_617820_619391	0
3	scaffold_46_77132	0.073	mTww	0.20	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_46_73351_73353	3779
3	scaffold_10_898402	-0.100	mTww	0.19	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_10_896048_896050	2352
3	scaffold_15_351654	-0.093	mTww	0.19	upstream_gene_variant	MODIFIER	STOP_CODON_scaffold_15_348114_348116	3538
3	scaffold_15_372338	-0.094	mTww	0.19	missense_variant	MODERATE	CDS_scaffold_15_371961_372617	0

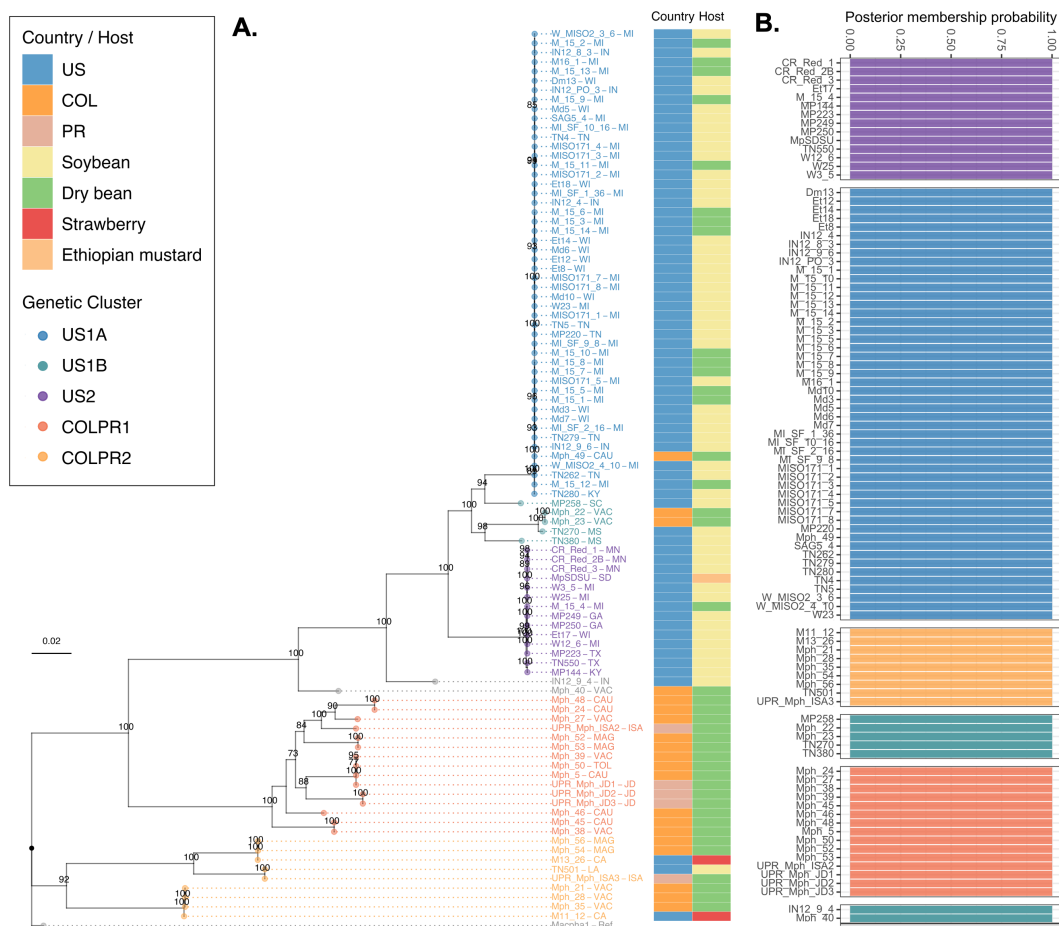


Figure A.1 (A) Rooted phylogeny reconstructed using the *M. phaseolina* reference genome as outgroup. Maximum-likelihood phylogeny reconstructed using 77,465 high-quality SNPs. Bootstrap support values over 70 are shown at nodes. Bootstrapping converged after 600 replicates. Colored tips represent the genetic cluster for each isolate as defined by principal components analysis. Individual isolate names include ANSI/ISO codes for US states, and Colombia and Puerto Rico municipalities: CA: California, CAU: Cauca, GA: Georgia, IN: Indiana, ISA: Isabela, JD: Juana Diaz, KY: Kentucky, LA: Louisiana, MAG: Magdalena, MI: Michigan, MN: Minnesota, MS: Mississippi, SC: South Carolina, SD: South Dakota, TN: Tennessee, TOL: Tolima, TX: Texas, VAC: Valle del Cauca, WI: Wisconsin. ISO country codes: US: United States, COL: Colombia and PR: Puerto Rico. (B) Discriminatory analysis of principal components. Each bar and color indicates the posterior probability membership value per isolate to one of the five genetic clusters.

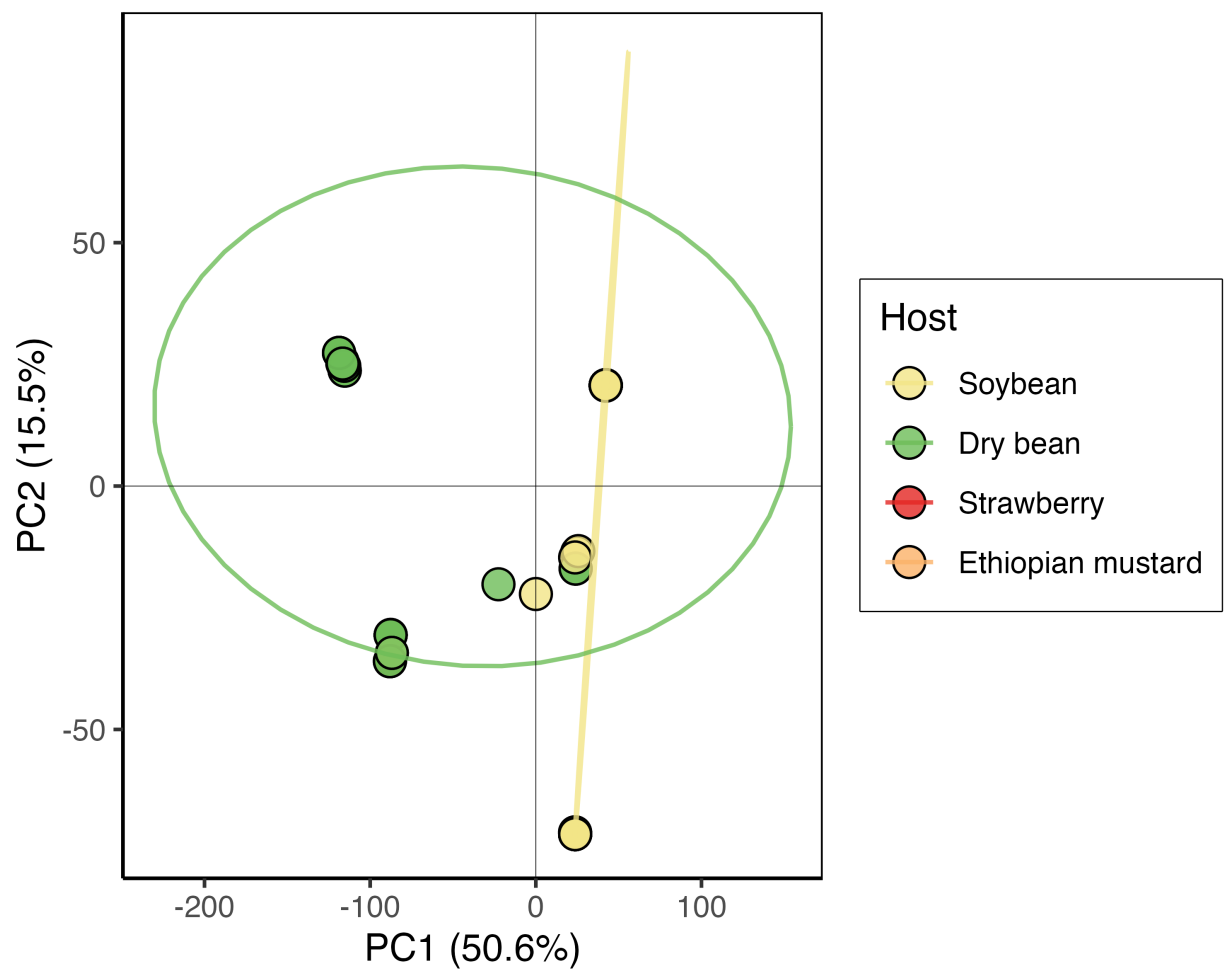


Figure A.2 Principal component analysis (PCA) showing isolate host origin. Scatterplot from a principal component analysis based on the two first PCs (the eigenvectors of the 77,465 SNPs) for all isolates. Points are colored by host from which isolates were collected. Overlapping ellipses representing 95% of the isolates from each of the hosts.

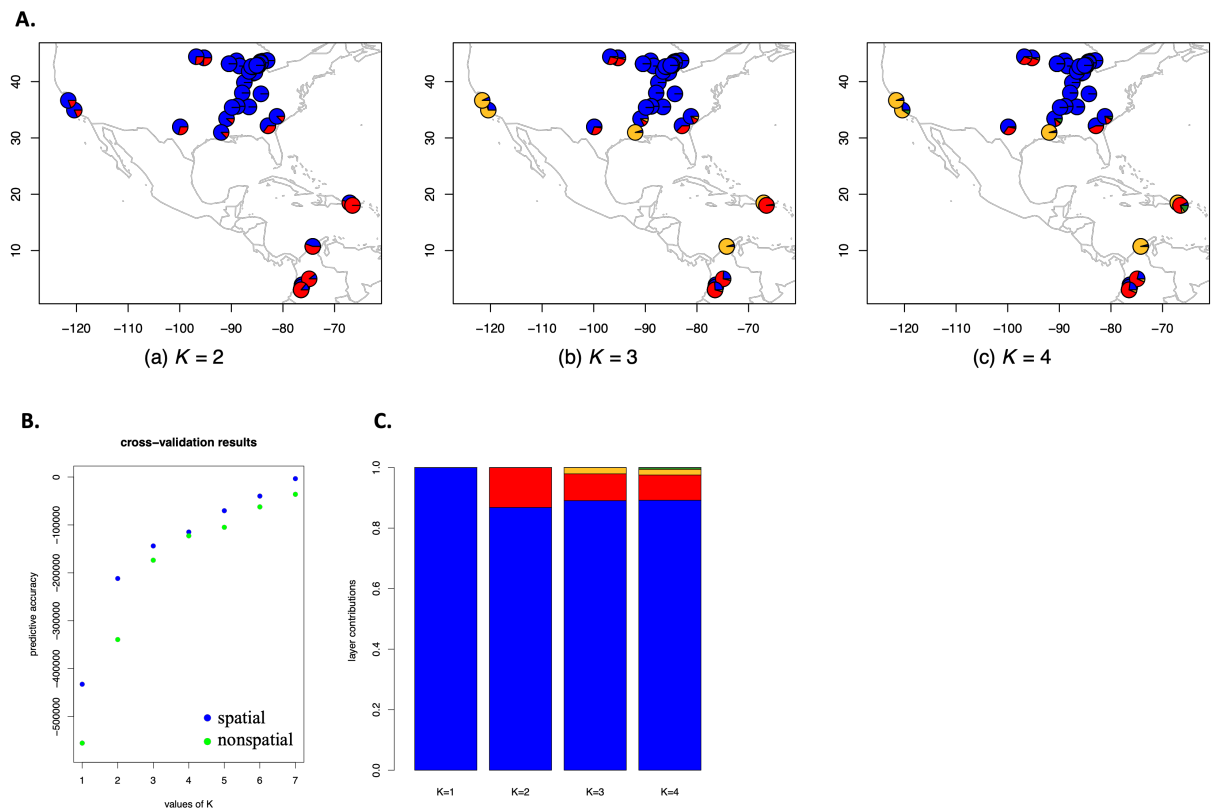


Figure A.3 Spatial population structure using conStruct. (A) Maps of admixture proportions estimated for *M. phaseolina* across the US, Puerto Rico and Colombia using the spatial conStruct model for $K = 2$ to $K = 4$. Pies show mean admixture results for individual isolates within their diameter. (B) Cross-validation predictive accuracy values as a function of the number of layers ($K = 1-7$) for the spatial and nonspatial conStruct models. (C) Layer contributions for $K = 2$ through 4.

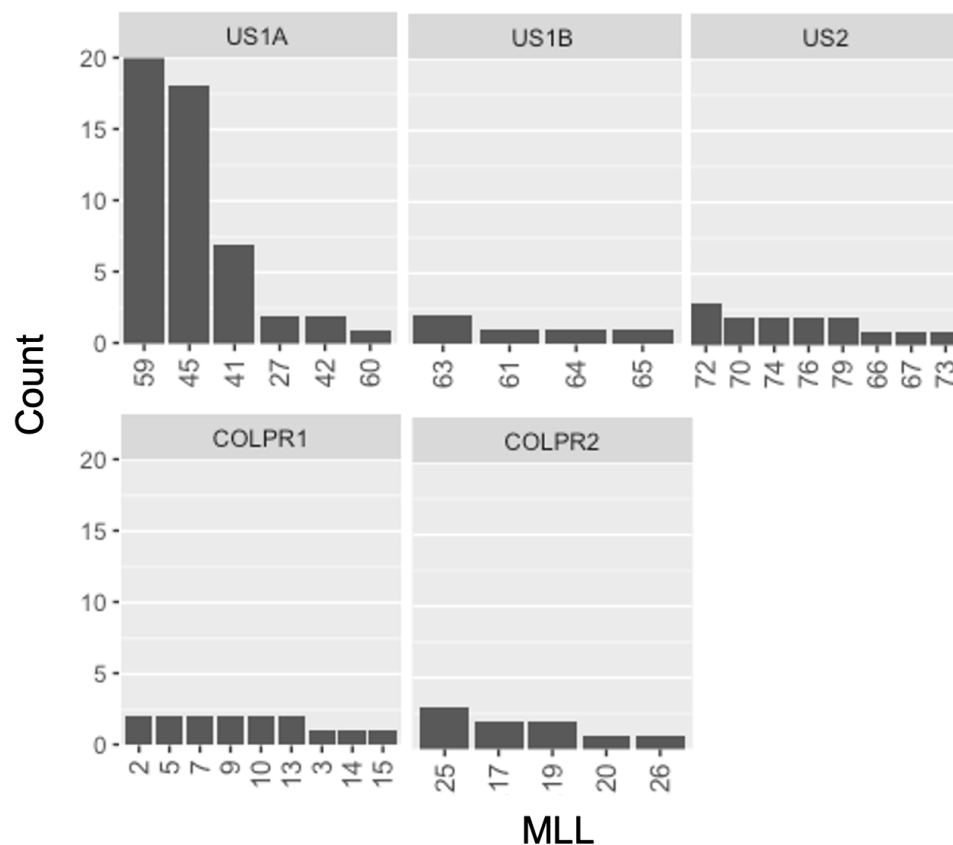


Figure A.4 MLLs shared among countries. MLL 7: one isolate from Colombia (Mph-5) and one from Puerto Rico (UPR-Mph-JDr) clustering in COLPR₁, MLL 17: one isolate from Puerto Rico (UPR-Mph-ISA₃) and one from Louisiana (TN₅₀₁) clustering in COLPR₂, and MLL 59: one isolate from Colombia (Mph-49) and 19 isolates from US clustering in US_{1A}. The two MLLs for isolates IN₁₂₉₋₄ and Mph₄₀ are not shown.

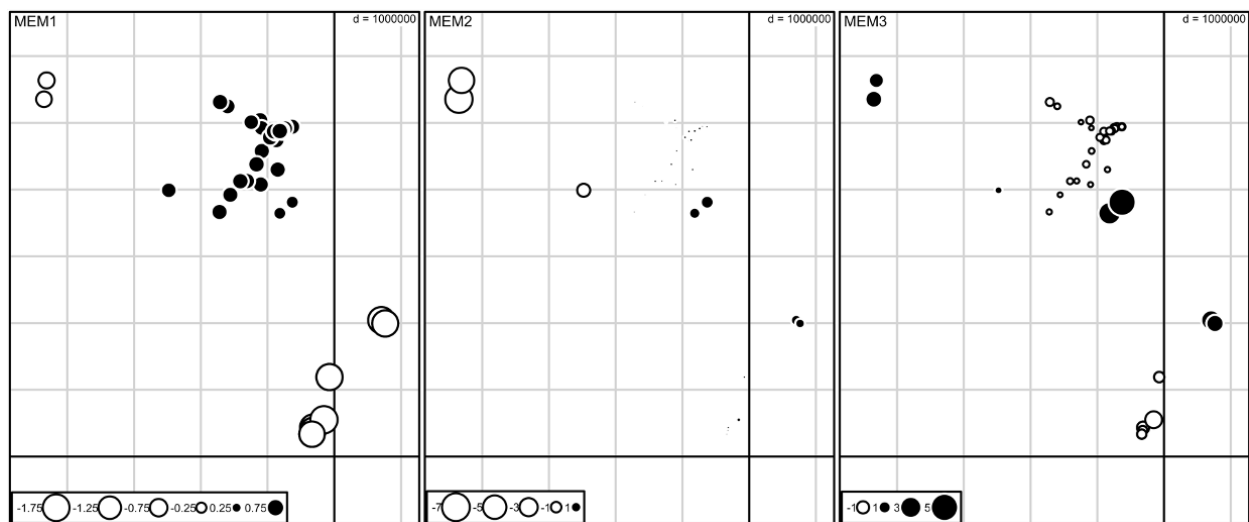


Figure A.5 Spatial structure variables identified using distance-based Moran's eigenvector maps (dbMEMs 1-3). The variable dbMEM 3 identified as significant using forward-variable selection described broad spatial structure. Color and size of the points correspond to the sign (+ or -) and magnitude of the dbMEM variables, respectively.

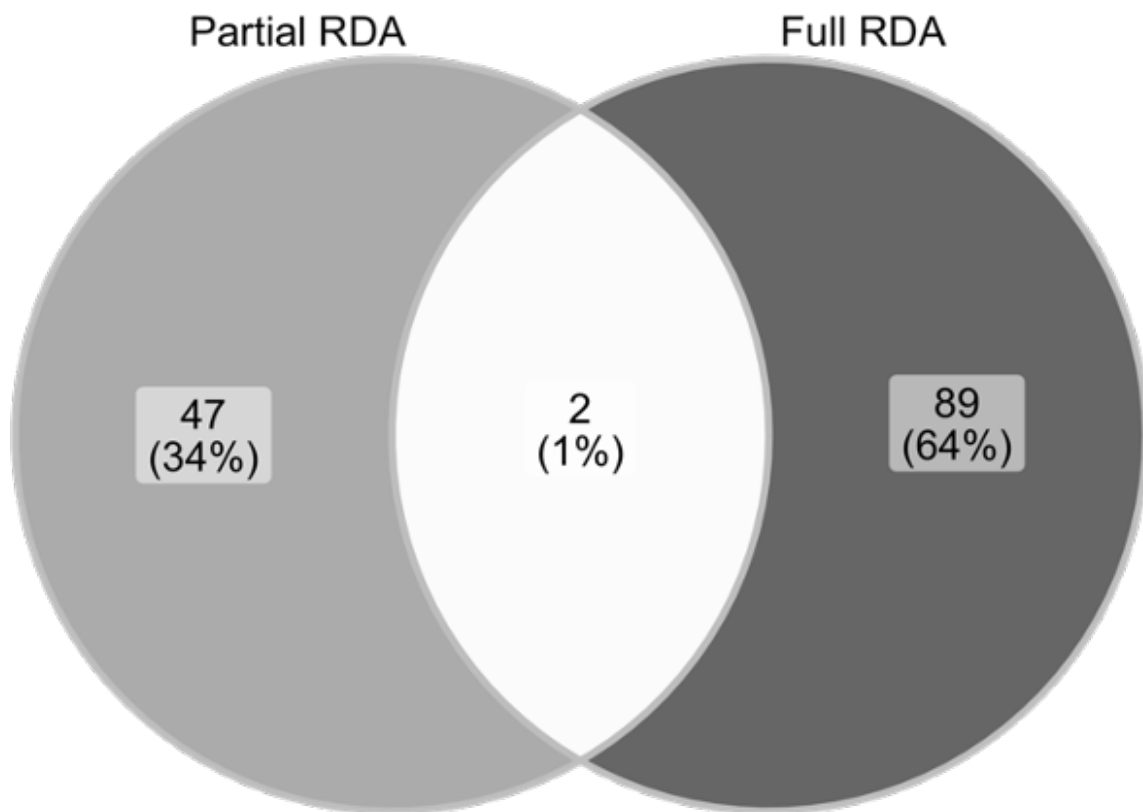


Figure A.6 Venn diagram showing the overlap between outlier loci identified by both partial RDA (constrained on neutral population structure) and full RDA (unconstrained) models using unlinked SNPs (LD-filtered set of 11,421 SNPs).

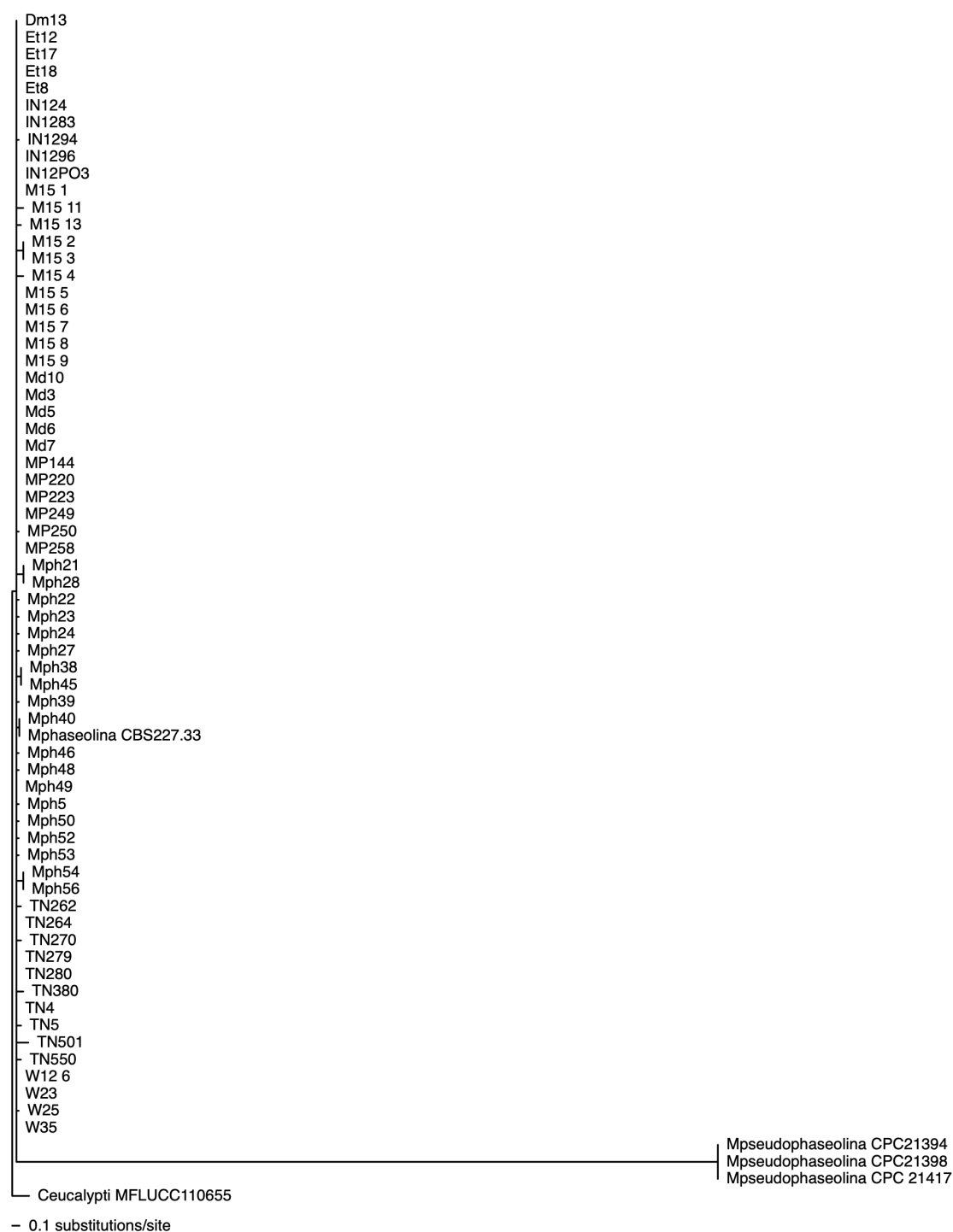


Figure A.7 Maximum-likelihood phylogeny reconstructed using concatenated sequences of the Internal Transcribed Spacer regions for the nuclear rDNA operon (ITS), part of the Translation Elongation Factor (TEF-1) gene region, and part of the actin (ACT) gene region.

CHAPTER 3

SENSITIVITY TO SINGLE-SITE FUNGICIDES IN *MACROPHOMINA PHASEOLINA* POPULATIONS FROM SOYBEAN AND DRY BEAN

3.1 Abstract

Charcoal rot, caused by *Macrophomina phaseolina*, is a soil- and seedborne disease that affects soybean and dry bean production worldwide. Strategies for effectively managing charcoal rot are limited, and management has primarily focused on varietal resistance and cultural practices. Fungicide efficacy studies conducted in past years have focused on older active ingredients and information on the sensitivity of *M. phaseolina* to newer classes of single-site fungicides is lacking. Although not specifically targeting *M. phaseolina*, single-site fungicides are used in soybean and dry bean production as seed treatments, soil applications, and foliar sprays. The *in-vitro* sensitivity of 92 *M. phaseolina* isolates collected from soybean and dry bean in the United States, Puerto Rico and Colombia was assessed for three classes of single-site fungicides widely used in soybean and dry bean production. The relative mycelial growth of *M. phaseolina* isolates challenged against boscalid (SDHI), iprodione (dicarboximide) and prothioconazole (DMI) was used to determine the effective concentration to inhibit mycelial growth by 50% (EC₅₀). All 92 isolates were sensitive to boscalid, iprodione and prothioconazole. Mean EC₅₀ values for boscalid, iprodione, and prothioconazole were 0.51, 0.86 and 0.14 $\mu\text{g ml}^{-1}$ respectively. The full-length nucleotide sequences of fungicide target genes were assembled to investigate mutations in all isolates. Mutations found in target genes did not associate with levels of *M. phaseolina* fungicide sensitivity.

3.2 Introduction

Soil-borne fungal pathogens are a major threat to crops and food security and fungicides are key components of effective disease management to prevent yield loss and ensure high-quality crop production. Since the 1970's fungicide use has increased, partly with the advent of broad-spectrum systemic single-site fungicides such as dicarboximides, sterol biosynthesis inhibitors including demethylation inhibitors (DMIs; azoles), and succinate dehydrogenase inhibitors (SDHIs) (Russell, 2005). Changes in cultural practices such as reduced or no-tillage systems, which add complexity to disease dynamics by favoring pathogen inoculum in crop residue, have further contributed to the increased use of fungicides (Oerke, 2006; Morton and Staub, 2008). Only a few years after the commercial use of fungicides, acquired resistance became a significant threat to their efficacy (Kuck and Russell, 2006; Leadbeater et al., 2019). Therefore, globally, as well as in the US, monitoring for development of resistance is an important com-

ponent for the implementation of effective disease management strategies (Brent and Hollomon, 2007).

Charcoal rot disease, caused by the soil- and seed-borne pathogen *Macrophomina phaseolina*, has been recognized as a threat of increasing importance to soybean (*Glycine max*) and dry bean (*Phaseolus vulgaris*) production in the US and worldwide (Dhingra and Sinclair, 1978; Wrather et al., 2010; Reznikov et al., 2018; Jacobs et al., 2019; Savary et al., 2019; Bradley et al., 2021). In the field, charcoal rot typically develops at reproductive stages of soybean and dry bean. However, infection may occur at emergence and early in the growing season causing up to 100% incidence of seedling infection 2 to 3 weeks after planting causing seedling blight (Hartman et al., 2015a; Hartman et al., 2015b). Seedling disease is most often reported in tropical regions, however in temperate regions damage to soybean seedlings is also observed, particularly under high temperature and low soil moisture conditions (Meyer and Sinclair, 1974; Hartman et al., 2015a). Infection begins, most commonly, with microsclerotia present in soil or plant residue. Microsclerotia germination followed by appressoria development allows host penetration through the root epidermis with subsequent invasion of root and stem tissue. Alternatively, colonization can occur from infected seed. Eventually *M. phaseolina* colonizes the vascular system leading to wilting, necrosis, and plant death (Hartman et al., 2015a). *M. phaseolina* reproduction in infected plants produces abundant microsclerotia which, following plant death and crop harvest, can survive in crop residue and in soil for years (Dhingra and Sinclair, 1974). Although *M. phaseolina* can be a devastating pathogen, it can also colonize plants asymptotically, and it is recognized as an endophyte and latent pathogen in many plant species (Dhingra and Sinclair, 1974; Slippers and Wingfield, 2007; Slippers and Boissin, 2013; Parsa et al., 2016; Crous et al., 2017). Management of charcoal rot in soybean and dry bean relies mostly on host genetic resistance which is limited in both crops and cultural control measures, which may be challenging to implement (Pastor-Corrales et al., 1988; Hartman et al., 2015a; Coser et al., 2017; Romero Luna et al., 2017; Ambachew et al., 2021). Chemical-control strategies are aimed at reducing microsclerotia in soil and limiting host colonization (Romero Luna et al., 2017). Fungicide seed treatments and soil applications can provide protection by delaying colonization and reducing fungal growth within root, stem and vascular tissue (Bradley, 2008). Seed treatments with benomyl (benzimidazole) and carboxin (first generation succinate dehydrogenase inhibitor) showed some effectiveness in reducing incidence of

charcoal rot in dry bean seedlings under greenhouse conditions (Abawi and Pastor-Corrales, 1990). Similarly, soybean seed treated with thiophanate methyl + pyraclostrobin protected plant emergence in field inoculation experiments (Reznikov et al., 2016). Recent studies evaluated the *in-vitro* sensitivity of *M. phaseolina* to different fungicide classes using a single isolate (Tonin et al., 2013; Chaudhary et al., 2017). However, current chemical-control strategies for charcoal rot do not provide consistent effective control and information on the effectiveness of newer fungicides chemistries using a collection of *M. phaseolina* isolates is lacking (Reznikov et al., 2016; Romero Luna et al., 2017; Roth et al., 2021).

In soybean and dry bean, management strategies commonly include the use of single-site fungicides as seed treatments and foliar applications (Hartman et al., 2015b; Lehner et al., 2017; Bandara et al., 2020; Karavidas et al., 2022). Most single-site fungicides target mitochondrial respiration function, the cytoskeleton or ergosterol biosynthesis. The demethylation inhibitors (DMIs) are the most important group of fungicides currently used in crop protection, leading the world fungicide market (Leadbeater et al., 2019). DMIs inhibit the C₁₄-demethylation step of ergosterol biosynthesis interfering with membrane integrity. The succinate dehydrogenase inhibitors (SDHIs) fungicides target the succinate dehydrogenase (mitochondrial complex II in the electron transfer chain), thereby inhibiting fungal respiration. Dicarboximides cause cell death through interference with osmotic signal transduction pathway via inappropriate activation of the osmosensing class III histidine kinase (Motoyama et al., 2005; Yamaguchi and Fujimura, 2005)(Fungicide Resistance Action Committee, FRAC: www.frac.info).

DMIs, SDHIs and dicarboximides are considered either medium or medium to high risk for the development of fungicide resistance (FRAC: www.frac.info) and shifts in fungicide sensitivity have been reported in important crops for these three classes (Brent and Hollomon, 2007; Hartman et al., 2015b; Leadbeater et al., 2019). The most common mechanisms of resistance to DMIs, SDHIs and dicarboximides are changes in the amino acids of the target proteins. Single point mutations in *Sdh* succinate dehydrogenase (Sang and Lee, 2020), *os1* histidine kinase, and *cyp51* C₁₄-demethylase genes are known to confer reduced sensitivity to boscalid, iprodione and prothioconazole, respectively, in several fungal pathogens (FRAC: www.frac.info). In addition, *cyp51* genes overexpression (Schnabel and Jones, 2001; Nikou et al., 2009; Wei et al., 2020) and promoter insertions have been associated with DMI-reduced

sensitivity of phytopathogenic fungi.

Single-site systemic fungicides are highly effective, meaning that most individuals are either killed or inhibited resulting in selection for any resistant individuals (Lucas et al., 2015). Factors such as fungicide distribution in the plant tissues and dilution to non-lethal doses may lead to the development of resistance not only in the target pathogen but in other fungal pathogens or the plant-associated fungal community (Brent and Hollomon, 2007; Chamberlain et al., 2021). Additionally, large-scale homogeneous agricultural systems which often have low crop genetic diversity and can sustain large and rapidly reproducing pathogen populations constitute conducive environments for the evolution of resistance (Brent and Hollomon, 2007).

Selection for resistance can occur in any environment containing fungicides. The risk of fungicide resistance depends mainly on the fungicide mode of action and specificity (e.g. multisite vs. single-site), the biological characteristics of the fungi, such as reproduction mode and rate of reproduction, and agronomic factors related to appropriate fungicide use (Leadbeater et al., 2019). In addition, pathogen demographic history, for example greater inoculum load leading to increases of effective population sizes, and the existence of fitness trade-offs may also play an important role in the development of fungicide resistance (McDonald and Stukenbrock, 2016; Hawkins and Fraaije, 2018). Although evolution of resistance to fungicides has been characterized for many fungal pathogens, there are few studies that assessed the fungicide sensitivity and potential mechanisms of resistance in *M. phaseolina*.

Overall, we consider it likely that *M. phaseolina* is commonly exposed to fungicides used in soybean and dry bean production and that conditions associated with the use of fungicides to protect crops against economically important fungal pathogens could favor the development of fungicide resistance in *M. phaseolina* populations either as direct or off-target effect. We therefore hypothesize that selection for resistance may occur in the internal tissues of plants or seeds treated with fungicides as well as in crop residue and soil containing fungicide residue. Additionally, we hypothesize that populations in tropical and subtropical regions in which environmental conditions could allow for year-round pathogen multiplication and therefore sustain large populations, may be at higher risk of developing resistance. The objectives of this study were *i)* to investigate boscalid, iprodione and prothioconazole *in-vitro* sensitivity of 92 *M.*

phaseolina isolates collected from soybean and dry bean from the US, Colombia, and Puerto Rico, and *ii*) identify mutations in the *Sdh*, *os1* and *cyp51* target genes of *M. phaseolina* isolates and examine their association with levels of *M. phaseolina in-vitro* fungicide sensitivity. For this, we conducted mycelial growth assays and investigated fungicide target genes from *M. phaseolina* whole-genome sequences. This study provides information for effective use of active ingredients in current commercial fungicide formulations and aid in the designing of effective disease management strategies.

3.3 Results

3.3.1 EC₅₀ determination for 35 *M. phaseolina* isolates

The EC₅₀ values of 35 *M. phaseolina* isolates were determined based on mycelial growth on Petri plates amended with different concentrations of boscalid, iprodione and prothioconazole (Table 3.1). Isolate mean EC₅₀ values were not different across experiments, with confidence intervals that overlapped zero, for boscalid (Bos1—Bos2: -0.068 $\mu\text{g ml}^{-1}$; 95% CI -0.478 — 0.104) and iprodione (Ipro1—Ipro2: 0.063 $\mu\text{g ml}^{-1}$; 95% CI -0.045 — 0.213) (Table A.1). For prothioconazole, although the CI did not overlap zero (Pro1—Pro2: -0.065 $\mu\text{g ml}^{-1}$; 95% CI -0.127 — -0.027), mean EC₅₀ difference was less than 0.07 $\mu\text{g ml}^{-1}$, as for boscalid and iprodione (Table A.1). Overall, mean EC₅₀ differences indicate that the EC₅₀ values were consistent across experiments for all fungicides (Figure A.1). The EC₅₀ values for boscalid ranged from 0.12 to 2.77 $\mu\text{g ml}^{-1}$ with mean 0.44 $\mu\text{g ml}^{-1}$, though only two isolates (CR-Red-2B and MP144) had an EC₅₀ more than 1 $\mu\text{g ml}^{-1}$ (Table 3.1). For iprodione, the EC₅₀ of isolates ranged from 0.53 to 1.39 $\mu\text{g ml}^{-1}$ with mean 0.83 $\mu\text{g ml}^{-1}$ (Table 3.1). Isolates were most sensitive to prothioconazole with EC₅₀ values ranging from 0.05 to 0.64 $\mu\text{g ml}^{-1}$, with mean of 0.15 $\mu\text{g ml}^{-1}$ (Table 3.1 and 3.2). These results indicate that *M. phaseolina* isolates evaluated were sensitive to the three fungicides tested. Isolate sensitivity differed across the three fungicides. Isolates were most sensitive to prothioconazole, followed by boscalid and least sensitive to iprodione (Figure 3.2). Mean EC₅₀ differences were 0.39 $\mu\text{g ml}^{-1}$ (95% CI 0.14 — 0.51) and -0.30 $\mu\text{g ml}^{-1}$ (95% CI -0.57 — -0.19) for iprodione and prothioconazole as compared to boscalid, respectively, and -0.69 $\mu\text{g ml}^{-1}$ (95% CI -0.76 — -0.62) for prothioconazole as compared to iprodione (Table A.2).

Table 3.1 Mean EC₅₀ (effective concentration to reduce growth by 50%) estimates for 35 *M. phaseolina* isolates determined from mycelial growth assays in Petri plates amended with different concentrations of fungicides boscalid, iprodione or prothioconazole.

Isolate	Boscalid		Iprodione		Prothioconazole	
	EC ₅₀	Standard Error	EC ₅₀	Standard Error	EC ₅₀	Standard Error
CR_Red_1	0.716	0.001	1.162	0.074	0.165	0.012
CR_Red_2B	2.766	1.267	1.106	0.121	0.089	0.006
CR_Red_3	0.663	0.426	0.871	0.089	0.177	0.064
Dm13	0.527	0.216	0.856	0.057	0.146	0.032
Et12	0.184	0.009	0.780	0.247	0.135	0.037
Et17	0.584	0.133	0.980	0.262	0.154	0.035
Et18	0.480	0.061	0.680	0.005	0.112	0.028
IN12_8_3	0.261	0.027	0.574	0.098	0.094	0.004
IN12_9_4	0.189	0.025	0.588	0.018	0.059	0.028
IN12_9_6	0.256	0.038	0.819	0.187	0.136	0.019
IN12_PO_3	0.253	0.058	0.525	0.078	0.197	0.054
Md3	0.124	0.021	0.650	0.093	0.147	0.057
Md6	0.203	0.022	0.769	0.110	0.124	0.030
MI-SF 1-36	0.216	0.016	0.712	0.056	0.129	0.024
MI-SF 10-16	0.223	0.014	0.786	0.108	0.108	0.009
MI-SF 2-16	0.197	0.007	0.854	0.047	0.110	0.003
MI-SF 9-8	0.190	0.044	0.993	0.079	0.099	0.007
MISO171-1	0.186	0.069	0.810	0.142	0.087	0.013
MISO171-2	0.315	0.079	0.781	0.066	0.091	0.003
MISO171-3	0.170	0.003	0.663	0.064	0.108	0.015
MP144	1.175	0.014	1.394	0.763	0.638	0.142
MP249	0.764	0.233	1.038	0.097	0.156	0.060
MP258	0.290	0.034	1.027	0.051	0.140	0.050
SAG5-4	0.219	0.028	0.984	0.008	0.146	0.012
TN264	0.275	0.109	0.653	0.051	0.237	0.124
TN270	0.496	0.052	0.767	0.101	0.046	0.008
TN4	0.198	0.021	0.657	0.019	0.051	0.013
TN5	0.190	0.009	0.728	0.010	0.200	0.073
TN501	0.674	0.237	0.707	0.143	0.136	0.025
TN550	0.396	0.041	1.000	0.213	0.167	0.005
W-MISO2 3-6	0.409	0.135	0.713	0.087	0.137	0.042
W-MISO2 4-10	0.342	0.053	—	—	0.151	0.032
W23	0.217	0.033	0.645	0.016	0.133	0.025
W25	0.693	0.159	0.900	0.299	0.150	0.022
W3-5	0.475	0.082	1.099	0.204	0.170	0.017

3.3.2 Isolate screening and EC₅₀ prediction using single concentrations

Single concentrations of boscalid 1 µg ml⁻¹, iprodione 1 µg ml⁻¹, or prothioconazole 0.5 µg ml⁻¹ were used to screen the remaining 55 *M. phaseolina* isolates. EC_{50(P)} values for each isolate were predicted after linear regression models based on RMG at the single concentration (Figure A.2). For all fungicides, EC_{50(P)} values were within the range of the EC₅₀ values estimated for the set of 35 *M. phaseolina* isolates initially tested (Table 3.2 and A.3). None of the 55 isolates were found to have EC_{50(P)} values above the threshold to be categorized as less sensitive (1.84, 1.41, and 0.43 µg ml⁻¹ for boscalid, iprodione and prothioconazole, respectively) (Figure 3.3). Consistent with the EC₅₀ estimations for the initial set, isolates were most sensitive to prothioconazole and least sensitive to iprodione (Figure A.3).

Table 3.2 Mean and range EC₅₀ (effective concentration to reduce growth by 50%) estimates and predictions for different sets of *M. phaseolina* isolates determined from mycelial growth assays in Petri plates amended with multiple or single concentrations of boscalid, iprodione or prothioconazole.

Set	EC ₅₀ type	Boscalid			Iprodione			Prothioconazole		
		Mean EC ₅₀	EC ₅₀ range		Mean EC ₅₀	EC ₅₀ range		Mean EC ₅₀	EC ₅₀ range	
			min	max		min	max		min	max
Multiple concentrations 35 isolates ^a	EC ₅₀	0.443	0.124	2.766	0.831	0.525	1.394	0.146	0.046	0.638
Single concentration 55 isolates ^b	EC _{50(P)}	0.404	0.137	1.538	0.758	0.566	1.300	0.108	0.066	0.325
Combined and validation sets 92 isolates ^c	EC ₅₀ and EC _{50(P)}	0.514	0.124	2.766	0.863	0.525	1.902	0.143	0.046	0.638

^aEC₅₀ Estimates for 35 *M. phaseolina* isolates determined from mycelial growth assays in Petri plates amended with different concentrations of boscalid, iprodione or prothioconazole. ^bEC_{50(P)} predictions for 55 *M. phaseolina* isolates determined from mycelial growth assays in Petri plates amended with single concentrations of boscalid 1 µg ml⁻¹, iprodione 1 µg ml⁻¹, or prothioconazole 0.5 µg ml⁻¹. ^cCombined EC₅₀ and EC_{50(P)} values for 92 *M. phaseolina* isolates. This combined data set consists of multiple concentration, single concentration and validation data points, including two additional isolates from the validation set.

Validation of linear regression models for EC_{50(P)} prediction was conducted by estimating EC₅₀ values and correlating them with predicted EC_{50(P)} values in a validation set of 13 isolates (Table A.4). These validation isolates were selected to represent the range of EC₅₀ /EC_{50(P)} values previously determined and included two isolates not tested in any of the previous assays. A significant positive linear relationship was observed between estimated EC₅₀ and predicted EC_{50(P)} values for all fungicides (Pearson's R =

0.97, 0.94 and 0.90 for boscalid, iprodione and prothioconazole, respectively; $P < 0.0001$) (Figure A.4).

3.3.3 Combined EC_{50} and $EC_{50(P)}$ for 92 *M. phaseolina* isolates

EC_{50} and $EC_{50(P)}$ values were combined to analyze the *in-vitro* fungicide sensitivity of the 92 *M. phaseolina* isolates, including data points and isolates from the validation set. The combined EC_{50} and $EC_{50(P)}$ distribution for the 92 isolates ranged from 0.12 to 2.77 $\mu\text{g ml}^{-1}$ with a mean EC_{50} of 0.51 $\mu\text{g ml}^{-1}$ for boscalid. For iprodione, EC_{50} values ranged from 0.53 to 1.90 $\mu\text{g ml}^{-1}$ with a mean value of 0.86 $\mu\text{g ml}^{-1}$. For prothioconazole, EC_{50} ranged from 0.05 to 0.64, with mean of 0.14 $\mu\text{g ml}^{-1}$ (Figure 3.4). Isolates were most sensitive to prothioconazole, followed by boscalid and least sensitive to iprodione, as indicated by mean EC_{50} differences between fungicides for the 92 isolates (Figure 3.5, Table 3.3). While no resistant isolates were identified, isolates with $EC_{50}/EC_{50(P)}$ values above three standard deviations from the mean (1.43, 1.52, and 0.33 $\mu\text{g ml}^{-1}$ for boscalid, iprodione and prothioconazole, respectively) were categorized as less sensitive (Figure 3.4, Table A.5). Less sensitive isolates were CR-Red2B and MpSDSU to boscalid; MP144 to iprodione; and MP144 and MP223 to prothioconazole (Table A.5).

Table 3.3 Mean $EC_{50}/EC_{50(P)}$ differences for 92 *M. phaseolina* isolates across fungicides: boscalid, iprodione and prothioconazole.

Pairwise fungicide comparison	Mean EC_{50} difference	95% CI ^a	
		low	high
Iprodione minus Boscalid	0.392	0.289	0.458
Prothioconazole minus Boscalid	-0.293	-0.397	-0.239
Prothioconazole minus Iprodione	-0.685	-0.732	-0.645

^a95% confidence intervals adjusting for asymmetrical resampling distributions using bias-corrected and accelerated bootstrap (BCa bootstrap).

To examine whether isolates differ in their sensitivity to each fungicide by host or genetic relatedness, $EC_{50}/EC_{50(P)}$ values were examined by isolate soybean or dry bean origin and genetic cluster (as determined in chapter 2). No differences in sensitivity were found between soybean and dry bean isolates for any of the three fungicides (Figure A.5). Isolates collected from strawberry (M11-12 and M13-26) and Ethiopian mustard (MpSDSU) were not included in this analysis because of the low number of isolates from each of these hosts. When analyzed by genetic cluster, isolates in the US1A, US1B and COLPR1 ge-

netic cluster were found to be, on average, more sensitive to the three fungicides, as compared with isolates in US2 and COLPR2 genetic clusters (Figure S6, Table A.6). Although, higher mean EC₅₀ values were obtained for isolates in the US2 and COLPR2 genetic groups, mean EC₅₀ differences between genetic clusters were all < 0.6, 0.4 and 0.07 $\mu\text{g ml}^{-1}$ for boscalid, iprodione and prothioconazole, respectively (Table A.6). All isolates were found to be sensitive to boscalid, iprodione and prothioconazole indicating their potential use for *M. phaseolina* management. Though *in-vitro* assays provide an initial assessment, in-vivo and field efficacy testing is necessary to determine whether they provide protection under field conditions.

3.3.4 Mutations in fungicide target genes and associations to isolate sensitivity

The predicted sequences of SdhB, cyp51, and osi genes of the 92 *M. phaseolina* isolates were obtained to examine whether mutations were associated with fungicide sensitivity. Species sequence alignment of the translated SdhB, cyp51, and osi genes sequences revealed high amino acid identity among all 92 isolates and with the reference sequences. Conservation of the SdhB iron-sulfur subunit was detected across all isolates. The cyp51B sequence was 525 amino acids with the heme-binding domain detected at codons 460-469 including the heme coordinating cysteine at codon 467. The predicted cyp51A sequence was 505 amino acids long. Amino acid sequence alignments of the 92 *M. phaseolina* isolates detected cyp51B mutations in four isolates collected from dry bean in Colombia and two isolates collected from soybean in the US (Table 3.4). Similarly, mutations in cyp51A sequence were detected in 14 isolates collected from dry bean in Colombia or Puerto Rico (Table 3.4). SdhB mutations were identified in three dry bean isolates from Colombia and one isolate from Puerto Rico. None of the mutations were found to be associated with reduced fungicide sensitivity. Structural modeling of the target proteins localized all mutations outside of the binding pocket and with low probability of affecting binding affinity of the fungicide molecules.

3.4 Discussion

This study provides information on the *in-vitro* efficacy against *M. phaseolina* of three single-site fungicides widely used in crop production worldwide. A total of 92 *M. phaseolina* isolates collected mainly from soybean and dry bean across the US, Colombia and Puerto Rico were characterized for their *in-vitro*

Table 3.4 Mutations in fungicide target genes found in *M. phaseolina* isolates. Host and geographic origins of isolates are shown.

Gene	Mutation	Isolate	Origin
cyp51B	V33I	Mph24	Dry bean Colombia
	V86I	Mph21	Dry bean Colombia
	V86I	Mph28	Dry bean Colombia
	I96T	MP258	Dry bean Colombia
	H249Y	IN129-4	Soybean North
	H249Y	MP258	Soybean South
cyp51A	V427F	Mph24, Mph48	Dry bean Colombia
	D479Y	Mph27, Mph38, Mph39, Mph45, Mph50, Mph52, Mph53, Mph5, UPR_ISA2, UPR_JD1, UPR_JD2, UPR_JD3	Dry bean Colombia Dry bean Puerto Rico
sdhB	A20T	Mph24, Mph27, Mph48	Dry bean Colombia
	V150L	UPR_ISA2	Dry bean Puerto Rico

sensitivity to boscalid (SDHI), iprodione (Dicarboximide) and prothioconazole (DMI). This represents the largest assessment of *M. phaseolina* variation in fungicide sensitivity in these countries. Prior studies characterizing the *in-vitro* sensitivity of *M. phaseolina* to different classes of fungicides have focused on older active ingredients or have used a limited number of isolates (Tonin et al., 2013; Chaudhary et al., 2017). We demonstrated that *M. phaseolina* isolates from soybean and dry bean were sensitive to boscalid, iprodione and prothioconazole active ingredients when tested in mycelial growth assays. While isolate variability in EC_{50} values to these fungicides was present, no isolate was insensitive to any of the tested fungicides. Notably, we found significant differences in *M. phaseolina* sensitivity to the three fungicides tested. Prothioconazole was the most efficacious active ingredient in reducing fungal growth (mean $EC_{50}/EC_{50(P)} = 0.14 \mu\text{g ml}^{-1}$), as compared to boscalid (mean $EC_{50}/EC_{50(P)} = 0.51 \mu\text{g ml}^{-1}$) and iprodione (mean $EC_{50}/EC_{50(P)} = 0.86 \mu\text{g ml}^{-1}$). In a study of *M. phaseolina* in Brazil, iprodione inhibited mycelial growth of a soybean isolate ($EC_{50} = 1.13 \mu\text{g ml}^{-1}$) (Tonin et al., 2013). To our knowledge, similar *in-vitro* studies reporting EC_{50} results of boscalid, iprodione or prothioconazole have not been conducted for a collection of *M. phaseolina* isolates.

Our study found no resistant isolates across *M. phaseolina* genetic clusters in the US (US1A, US1B and US2), Colombia and Puerto Rico (COLPR1 and COLPR2). Variation between genetic clusters was

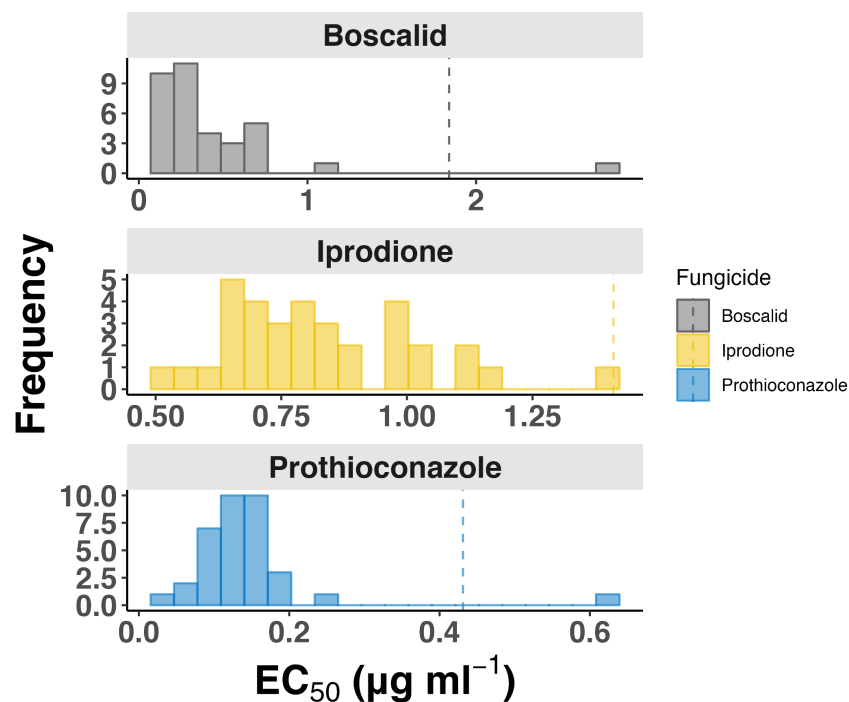


Figure 3.1 EC_{50} distribution of 35 *M. phaseolina* isolates collected from soybean and dry bean. Less sensitive isolates were designated as the ones with an EC_{50} value higher than three standard deviations from the mean.

observed with a general trend of least sensitivity in US2 and COLPR2 clusters as compared to US1A, US1B and COLPR1 cluster across fungicides. We hypothesized that *M. phaseolina* isolates may develop fungicide resistance in the context of soybean and dry bean production. Furthermore, we hypothesized that Colombian and Puerto Rican isolates may have higher risk of developing fungicide resistance as compared to US isolates. The reason for this was that environmental conditions in tropical locations, such as Colombia and Puerto Rico, allow for year-round permanence of *M. phaseolina*, and therefore, the potential maintenance of large pathogen populations in soil and crop residue. Additionally, the higher genetic diversity found in *M. phaseolina* Colombian-Puerto Rican genetic clusters (COLPR1 and COLPR2), as compared to US clusters (US1A, US1B and US2) (Ortiz et al., under review) led us to consider these populations may be at greater risk of developing resistance. However, our results did not support this hypothesis. Instead, we found isolates in all genetic clusters were sensitive to the three fungicides tested

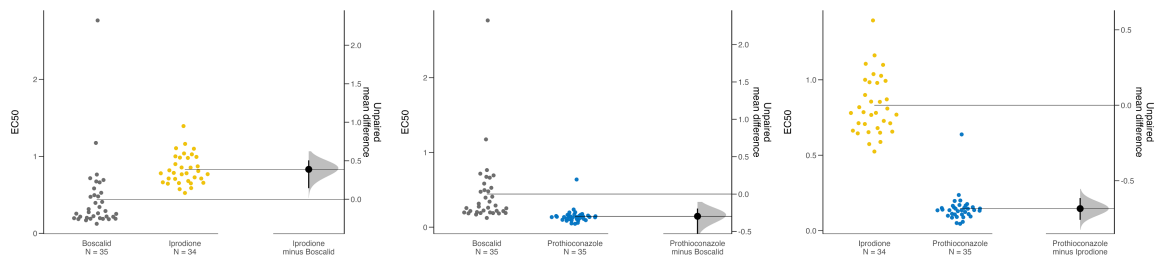


Figure 3.2 Mean EC_{50} differences of 35 *M. phaseolina* isolates across fungicides: (A) Boscalid-Iprodione (B) Boscalid-Prothioconazole (C) Iprodione-Prothioconazole pairwise comparisons. Isolates were most sensitive to prothioconazole, followed by boscalid and least sensitive to iprodione.

and most isolates from Colombia and Puerto Rico (grouped in COLPR₁ genetic cluster) had similar sensitivities to those in other genetic clusters. This suggests that low selection pressure and/or low intrinsic *M. phaseolina* risk for fungicide resistance development. Generally, a single fungicide application is conducted during the growing season in dry bean and soybean, in commercial fields (Hartman et al., 2015b) as well as in experimental plots from where the Colombian and Puerto Rican isolates in this study were collected (Gloria Mosquera and Consuelo Estevez, personal communication). In addition, although population sizes may be high, effective population sizes may remain low due to the mostly clonal nature of *M. phaseolina*.

Resistance to SDHIs, DMIs or Dicarboximide has not been reported in *M. phaseolina* (Sang and Lee, 2020). Overall, this study indicates that *M. phaseolina* has a low risk of developing resistance. A limitation of this study is that isolates from Colombia and Puerto Rico were collected only from dry bean experimental plots. A previous study conducted in a dry bean producing region in Colombia reported that fungicides were applied several times during the growing season (Velasquez et al., 1990). Future studies involving isolates collected from commercial fields in Colombia and Puerto Rico would provide a broader assessment of *M. phaseolina* fungicide sensitivity in these countries.

Information regarding mutations in *M. phaseolina* *Sdh*, *cyp51* and *osi* genes is lacking. In this study we report the predicted *cyp*, *sdhB* and *osi* sequences for 92 *M. phaseolina* isolates. Three paralogs of *cyp51* gene (*cyp51A*, *cyp51B* and *cyp51C*) have been identified in fungi and although their involvement in DMI sensitivity is well known for several fungal phytopathogens (Schnabel and Jones, 2001; Mohd-Assaad et

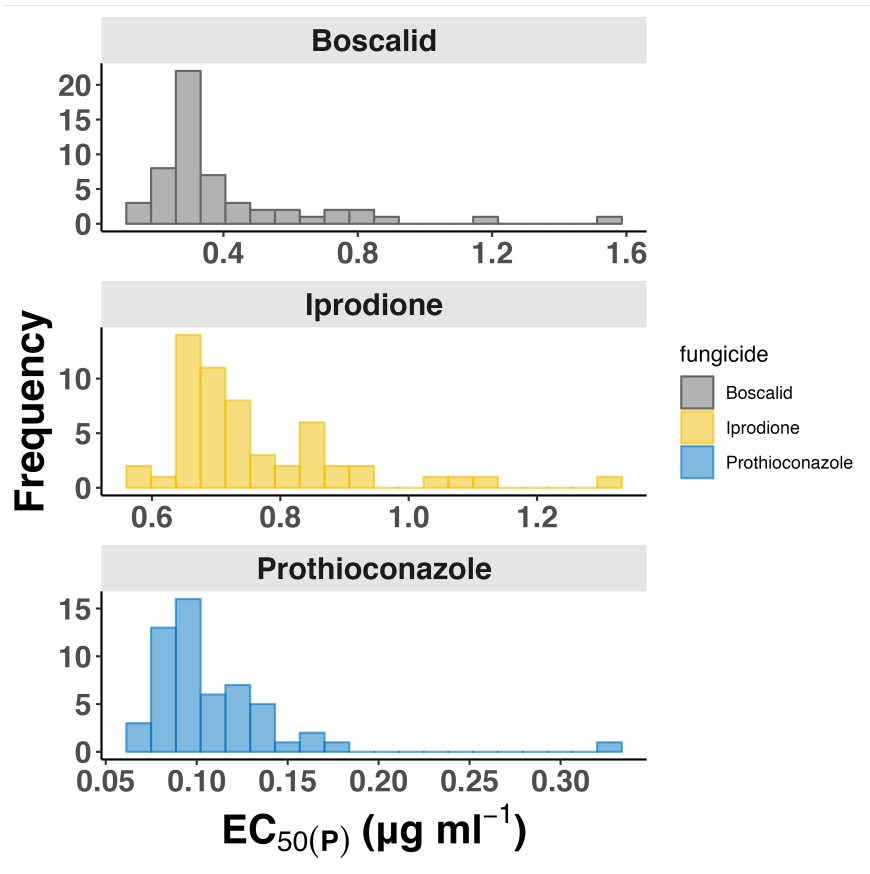


Figure 3.3 Predicted $EC_{50(P)}$ distribution of 55 *M. phaseolina* isolates tested in mycelial growth assays in half-strength PDA plates amended with 1 mg ml⁻¹ of boscalid, 1 mg ml⁻¹ of iprodione, or 0.5 mg ml⁻¹ of prothioconazole.

al., 2016; Zhang et al., 2017; Lestrade et al., 2019; Wei et al., 2020), the occurrence of *cyp51A* and *cyp51C* paralogs is unknown in *M. phaseolina*. The *cyp51A* and *cyp51B* paralogs sequences were present in the 92 *M. phaseolina* isolates tested in this study. *cyp51A* it is thought to play a major role in reduced sensitivity to DMIs, mainly as a functionally redundant mechanism for ergosterol production when fungi are exposed with DMI fungicides (Fan et al., 2013; Liu et al., 2022).

A total of 19 isolates showed mutations occurring in the *cyp51* genes and four isolates in the *SdhB* gene. The *cyp51A* D479Y mutation was the most frequently identified mutation present in twelve isolates of our collection. Interestingly, these isolates were all collected from dry bean in Colombia or Puerto Rico. None of these point mutations were found to be correlated with lower levels of sensitivity. Reduced sensitivity with high resistance factors (strength of resistance) is often observed with mutations located

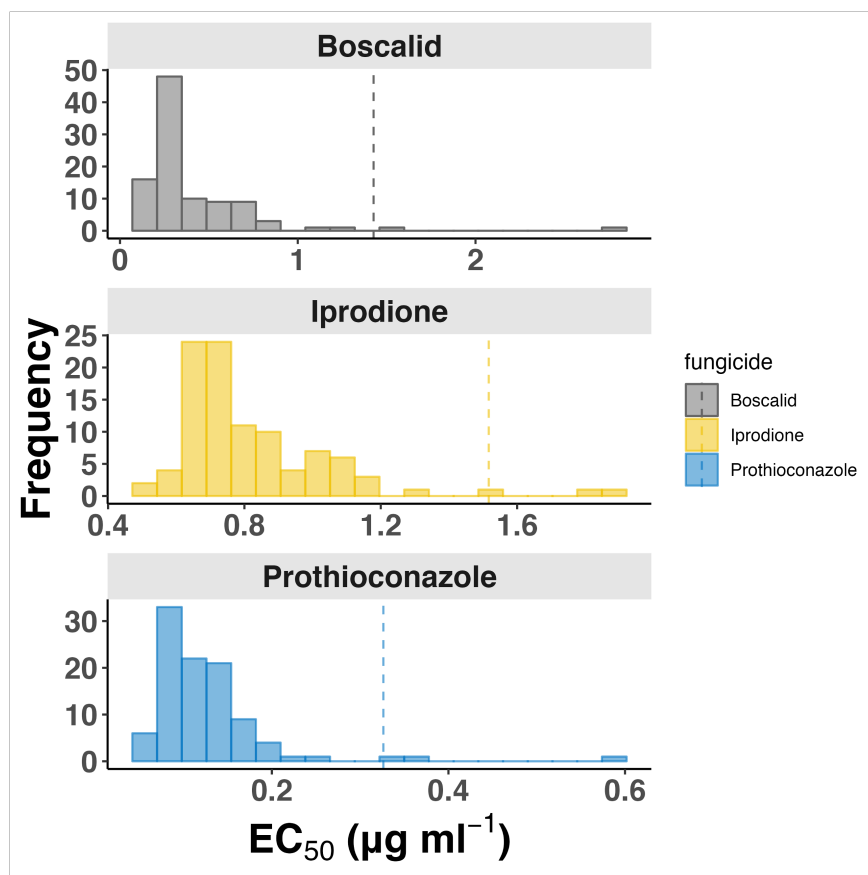


Figure 3.4 Combined EC_{50} and $EC_{50(P)}$ distribution of 92 *M. phaseolina* isolates. This combined data set consists of multiple concentration, single concentration and validation data points, including two additional isolates from the validation set. Less sensitive isolates were designated as the ones with an EC_{50} or $EC_{50(P)}$ value higher than three standard deviations from the mean indicated by the dashed line.

in putative azole molecules recognition sites (e.g V136A, Y131F, Y136F, Y137F, A379G, I381V) whereas mutations in highly conserved regions of the *cyp51* protein close to the heme binding site such as those at codons 459-461 have been correlated with lower resistance factors (Cools et al., 2010; Mullins et al., 2011; Mehl et al., 2019). The *cyp51B* H249Y mutation was identified in two isolates collected from soybean in the US. In *Cryptococcus gattii*, the *cyp51* N249D mutation conferred azole resistance (Gast et al., 2013). Molecular modelling of specific mutations in residues proximal to the binding pocket showed to have differential impact on *cyp51* protein function depending on whether a single mutation was present or in combination with others. The protein function was impacted mainly by alterations in the binding pocket volume. Furthermore, the effect of these mutations on DMI sensitivity was different for certain azole molecules (Cools et al., 2011; Mullins et al., 2011).

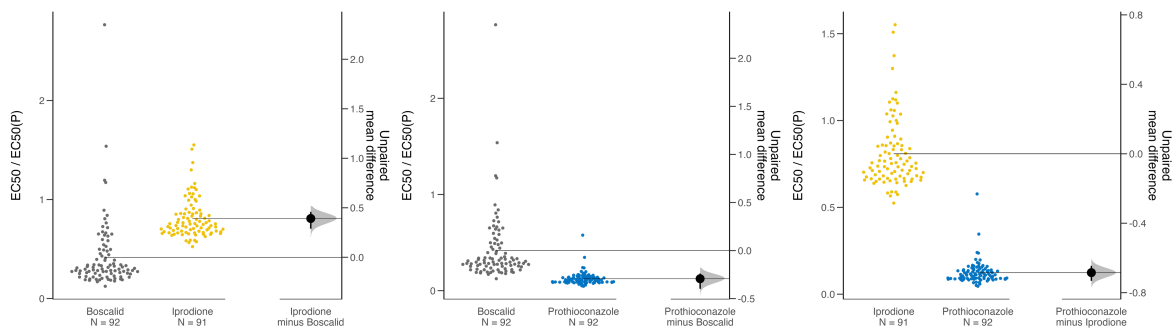


Figure 3.5 Mean $EC_{50} / EC_{50(P)}$ differences of 92 *M. phaseolina* isolates across fungicides: (A) Boscalid-Iprodione (B) Boscalid-Prothioconazole (C) Iprodione-Prothioconazole pairwise comparisons. Isolates were most sensitive to prothioconazole, followed by boscalid and least sensitive to iprodione.

It has been hypothesized that seed treatments may be useful in protecting soybean plants from disease caused by seedborne *M. phaseolina* (Hartman et al., 2015a; Hartman et al., 2015b). However, information regarding the efficacy of active ingredients currently used in commercial fungicide formulations in soybean and dry bean is lacking (Romero Luna et al., 2017). Although in-vivo and field studies would be necessary, our results indicate that formulations with prothioconazole, boscalid or iprodione, may reduce seedling infection originating from infected seeds or inoculum in the soil. In cotton, seed treatments with a commercial formulation of boscalid + pyraclostrobin (Signum) showed efficacy in preventing seedling infection by *M. phaseolina* in field experiments (Cohen et al., 2022). However, this protective effect was observed only for 12 days while roots were exposed to the fungicide in soil (Cohen et al., 2022).

Our data on the *in-vitro* efficacy of prothioconazole suggest that commercial formulations with this active ingredient may be of particular interest for future in-vivo efficacy testing in soybean and dry bean. Currently, fungicides labeled to control charcoal rot in different crops are available, although limited (<http://www.cdms.net>). For instance, a formulation of prothioconazole + fluopyram (Propulse) is labeled for charcoal rot management in soybean. Prothioconazole (Proline) has been shown to suppress plant colonization by *M. phaseolina* and improve yield under field conditions when used in tolerant soybean varieties in inoculated plots as compared to non-inoculated plots (USB report, 2019). Future studies can be aimed at testing prothioconazole efficacy in preventing seedling colonization and charcoal rot disease development as part of integrated management programs incorporating host genetic resistance

and cultural practices. Furthermore, studies investigating novel effective fungicides and monitoring the potential development of resistance to single-site fungicides in *M. phaseolina* populations would be beneficial to charcoal rot management efforts.

3.5 Materials and methods

3.5.1 *Macrophomina phaseolina* isolates and whole-genome sequencing

A total of 92 *M. phaseolina* isolates collected mostly from soybean and dry bean, for which population genomics analysis was conducted, were also used for fungicide sensitivity analysis (Ortiz et al., under review). Species identity of these isolates was confirmed as described previously by sequencing the Internal Transcribed Spacer regions for the nuclear rDNA operon (ITS), part of the Translation Elongation Factor (TEF- α) gene region, and part of the actin (ACT) gene region (Sarr et al., 2014) (Ortiz et al., under review). Briefly, this isolate collection included 52 isolates collected from soybean across the US and 40 isolates collected from dry bean in Michigan, Puerto Rico and Colombia, two isolates from strawberry collected in California and one isolate collected from Ethiopia mustard in the US.

Whole genome sequencing and SNP calling was conducted as described in Ortiz et al., under review. Genomic DNA was extracted from lyophilized mycelia using a modified SDS-based method; as described previously (Ortiz et al., under review). Briefly, hyphal tip cultures grown on potato dextrose agar (PDA) medium were used to produce mycelia on potato dextrose broth. Libraries were prepared using the Illumina TruSeq Nano DNA Library Preparation Kit HT and whole-genome sequencing to 23X coverage using a 150 base-pair paired-end strategy on the Illumina HiSeq 4000 platform at the Michigan State University Research Technology Support Facility Genomics Core (East Lansing, MI) was conducted.

Quality assessment and filtering were conducted using FastQC (Andrews et al., 2010) and Cutadapt v 1.16 (Martin, 2011). Sequences were aligned to the *M. phaseolina* reference genome (JGI Mycosm, MPI-SDFR-AT-0080 v1.0) using bwa-mem (Heng Li, 2013). Single nucleotide polymorphisms (SNPs) of all isolates were predicted using the Genome Analysis Toolkit (GATK) v4.0 (McKenna et al., 2010) pipeline (Ortiz et al., under review). The resulting vcf file was quality filtered using vcfr v1.10.0 package (Knaus and Grünwald, 2017) in R v4.0.0 (R Core Team 2019) (Ortiz et al., under review).

3.5.2 Fungicides

Commercial fungicide formulations of the SDHI boscalid (70% A.I., Endura, BASF corporation, Research Triangle Park, NC), Dicarboximide iprodione (Chipco 26GT 2SC, Bayer, Germany) and DMI prothioconazole (Proline 480 SC, Bayer CropScience, Research Triangle Park, NC) were used. Additional information about these fungicides is presented in Table. Aqueous stock solutions of these fungicides were prepared at 1000 $\mu\text{g ml}^{-1}$ of each respective active ingredient. Serial dilutions from the stock solutions were used to produce final concentrations in half-strength PDA media of boscalid (0.1, 1, 10, 100 and 500 $\mu\text{g ml}^{-1}$), iprodione (0.1, 1, 2.5, 5 and 10 $\mu\text{g ml}^{-1}$) and prothioconazole (0.01, 0.1, 0.5, 1 and 10 $\mu\text{g ml}^{-1}$), except for boscalid highest concentration (500 $\mu\text{g ml}^{-1}$) for which 0.715 g of Endura, per liter of media was used. These concentrations were selected based on preliminary experiments which directed the appropriate fungicide concentrations for fitting a dose-response curve.

3.5.3 Determination of EC_{50} values using mycelial growth inhibition assays

The sensitivity of 35 randomly selected *M. phaseolina* isolates to boscalid, iprodione and prothioconazole was determined based on EC_{50} (effective concentration to reduce growth by 50%) estimates. EC_{50} values for each isolate was determined using mycelial growth inhibition assays on fungicide-amended medium. Before each experiment isolates were recovered from -80°C and grown on potato dextrose agar (PDA; Acumedia, Lansing, MI) in the dark at 35°C for 24 h. Then a mycelial plug from the margin was transferred into a new Petri plate containing PDA and incubated in the dark at 35°C for 43 h. A single 6-mm agar plug taken from the edge of the 43-h old culture was placed mycelial side down on the center of non-amended half-strength PDA plates and plates amended with boscalid, iprodione or prothioconazole at concentrations mentioned above. The plates were incubated in the dark at 35°C for 43 h.

The diameter of each colony was measured in two perpendicular directions with a digital caliper (Absolute Digimatic Caliper, model CD-6" AX, Mitutoyo Corp., Sakado 1-Chome, Japan). Two separate experiments and two replicates (Petri plates) per each experiment were performed for each isolate and fungicide concentration. Isolates with data from at least two replicates were included in all subsequent analyses. Percent relative mycelial growth (RMG) at each concentration was calculated as the percentage of inhibition relative to the control without fungicides ((average colony diameter on fungicide amended

plates / average colony diameter on non-amended plates) X 100).

Absolute EC₅₀ values were calculated using a four parameter log logistic (LL4) dose response model as implemented in R (R Core Team, 2018) in the 'drc' package (Ritz et al., 2016), and following guidelines and workflow provided by (Noel et al., 2018). The LL4 model was used as it was the best fitting model for most isolates as determined by AIC criteria. Less sensitive isolates were designated based on the frequency distribution of the EC₅₀ values as the ones with an EC₅₀ higher than three standard deviations from the mean (EC₅₀ values > 1.84, 1.41, and 0.43 µg ml⁻¹ for boscalid, iprodione and prothioconazole, respectively). To investigate variability across experiments, isolate mean EC₅₀ differences between experiments were estimated using DABEST ('data analysis with bootstrap-coupled estimation') (Ho et al., 2019). Isolate mean EC₅₀ differences between experiments were all less than 0.07 µg ml⁻¹ for all fungicides, therefore experiments were combined in subsequent analyses.

3.5.4 Selection of single screening fungicide concentration and linear regression models

To screen the remaining isolates in a reduced resource-intensive manner, single screening concentrations were determined for boscalid, iprodione and prothioconazole using the EC₅₀ results of 35 *M. phaseolina* isolates. A linear regression analysis between RMG and log-transformed EC₅₀ values of each isolate was performed for the five tested concentrations for each fungicide. The fungicide concentration at which the linear regression model returned the highest correlation coefficient (Pearson's R) and proportion of explained variance (R²) values was selected as the screening concentration for each fungicide. These screening concentrations were found to be 1 µg ml⁻¹ for boscalid and iprodione, and 0.5 µg ml⁻¹ for prothioconazole. While for prothioconazole, 0.1 µg ml⁻¹ and 0.5 µg ml⁻¹ concentrations both had similarly high R and R² values (Pearson's R=0.9, R²=0.8), at 0.5 µg ml⁻¹ most isolates had an RMG below 50% indicating it may differentiate better less sensitive isolates than 0.1 µg ml⁻¹.

3.5.5 Sensitivities and EC_{50(P)} prediction using single screening fungicide concentrations

Sensitivities of each of the remaining 56 isolates were estimated based on RMG on half-strength PDA plates amended with boscalid at 1 µg ml⁻¹, iprodione at 1 µg ml⁻¹, or prothioconazole at 0.5 µg ml⁻¹. Media preparation, inoculation and mycelial growth measurements were conducted using the methods for EC₅₀ estimation described above. Two separate experiments and two replicates (Petri plates) per experiment

were performed for each isolate and fungicide. The linear regression model equations of boscalid ($1 \mu\text{g ml}^{-1}$), iprodione ($1 \mu\text{g ml}^{-1}$), or prothioconazole ($0.5 \mu\text{g ml}^{-1}$) (Figure A.2) were then used to predict an EC_{50} value, hereafter $\text{EC}_{50(P)}$, for each isolate, using the function ‘predict’ in R (R Core Team, 2018).

3.5.6 Validation of linear regression models used to predict $\text{EC}_{50(P)}$ values

A validation set of 13 isolates (Table A.4) was used to assess the performance of linear regression models in predicting $\text{EC}_{50(P)}$ values using a new data set. These validation isolates were selected from those for which an EC_{50} or $\text{EC}_{50(P)}$ was previously estimated in the multiple and single concentration experiments and to represent the range of these values. Additionally, two isolates not tested in any of the previous experiments were included. For these validation isolates, EC_{50} values were determined using the five fungicide concentrations in mycelial growth inhibition assays as described above. Then, the RMG at the screening concentration for each isolate (boscalid [$1 \mu\text{g ml}^{-1}$], iprodione [$1 \mu\text{g ml}^{-1}$], or prothioconazole [$0.5 \mu\text{g ml}^{-1}$]) was used to predict an $\text{EC}_{50(P)}$ value using the linear regression models previously selected. A simple linear regression analysis was used to determine the relationship between estimated EC_{50} and predicted $\text{EC}_{50(P)}$ values for the 13 validation isolates.

3.5.7 Sensitivities of 92 *M. phaseolina* isolates using the combined EC_{50} / $\text{EC}_{50(P)}$ values

All previous data sets, this is the EC_{50} values for the initial 35 isolates, the $\text{EC}_{50(P)}$ values for the 55 isolates in the single concentration experiments and the EC_{50} values for the validation isolates were combined to report the fungicide sensitivity of the 92 *M. phaseolina* isolates. The combined distribution of EC_{50} / $\text{EC}_{50(P)}$ values was used to categorize “less sensitive” isolates. Isolates were designated as less sensitive as the ones with an EC_{50} or and $\text{EC}_{50(P)}$ value higher than three standard deviations from the mean (EC_{50} or $\text{EC}_{50(P)}$ values > 1.43 , 1.52 , and $0.33 \mu\text{g ml}^{-1}$ for boscalid, iprodione, and prothioconazole, respectively).

3.5.8 Sequence and target gene mutation analysis

To explore if the sensitivity of *M. phaseolina* to boscalid (SDHI), iprodione (Dicarboximide) and prothioconazole (DMI) was associated with mutations in their target genes, the complete *SdhB*, *cyp51*, and *os1* genes of the 92 *M. phaseolina* isolates were analyzed. Genome-guided de novo assembly of each isolate was done in Trinity (Grabherr et al., 2011) using whole-genome sequencing Illumina reads with the *M. phaseolina* (JGI Mycocosm, MPI-SDFR-AT-0080 v1.0) as reference genome. Although Trinity was

developed for RNA-seq data, it was used to take advantage of the genome-guided option for de-novo transcript assembly, because it employs a de Bruijn graph approach (used by several whole-genome assembly programs) and can identify transcripts resulting from paralogous genes (Grabherr et al., 2011).

Published sequences of the target genes *SdhB*, *cyp51*, and *osi* (*Shk1*) (Duan et al., 2013) of *Fusarium graminearum*, *Botrytis cinerea*, *Diplodia corticola* and *Sclerotinia sclerotiorum* were used as query sequences to identify the orthologous sequences in the *M. phaseolina* reference genome using the program BLAST. The reference sequence of each target gene was used to retrieve the sequences from the assemblies of the 92 isolates, using a home-made script (<https://github.com/vivianaortiz/>). The amino acid sequences were aligned and analyzed using Geneious software.

3.5.9 Statistical analysis

All data analysis was conducted in R (R Core Team, 2018). Linear regression modeling and model evaluation were done with ‘lm’ and ‘AICctab’ functions (‘bbmle’ package). Isolates with data from at least two replicates were included for analysis. Effect sizes and confidence intervals were estimated using DABEST (Ho et al., 2019). Bootstrap confidence intervals were estimated adjusting for asymmetrical resampling distributions using bias-corrected and accelerated bootstrap (BCa bootstrap) (Efron and Tibshirani, 1993) as implemented in ‘dabestr’ R package (Ho et al., 2019).

BIBLIOGRAPHY

Abawi GS, Pastor-Corrales M (1990) Seed transmission and effect of fungicide seed treatments against *Macrophomina phaseolina* in dry edible beans. *Turrialba* 40: 334–339

Ambachew D, Joshua J, Mmbaga MT, Blair MW (2021) Sources of resistance to common bacterial blight and charcoal rot disease for the production of mesoamerican common beans in the Southern United States. *Plants* 10: 1–21

Bandara AY, Weerasooriya DK, Conley SP, Bradley CA, Allen TW, Esker PD (2020) Modeling the relationship between estimated fungicide use and disease-associated yield losses of soybean in the United States I: Foliar fungicides vs foliar diseases. *PLoS One* 15: 1–21

Bradley CA (2008) Effect of fungicide seed treatments on stand establishment, seedling disease, and yield of soybean in North Dakota. *Plant Dis* 92: 120–125

Bradley CA, Allen TW, Sisson AJ, Bergstrom GC, Bissonnette KM, Bond J, Byamukama E, Chilvers MI, Collins AA, Damicone JP, et al (2021) Soybean Yield Loss Estimates Due to Diseases in the United States and Ontario, Canada, from 2015 to 2019. *Plant Heal Prog* 22: 483–495

Brent KJ, Hollomon DW (2007) Fungicide resistance in crop pathogens: how can it be managed? 2nd, revised edition.

Chamberlain LA, Whitman T, Ané JM, Diallo T, Gaska JM, Lauer JG, Mourtzinis S, Conley SP (2021) Corn-soybean rotation, tillage, and foliar fungicides: Impacts on yield and soil fungi. *F Crop Res*. doi: 10.1016/j.fcr.2020.108030

Chaudhary DH, Pathak DM, Chaudhary MM (2017) In vitro Efficacy of Fungicides against Dry Root Rot (*Macrophomina phaseolina*) of Soybean. *Int J Curr Microbiol App Sci* 6: 1298–1301

Cohen R, Elkabetz M, Paris HS, Gur A, Dai N, Rabinovitz O, Freeman S (2022) Occurrence of *Macrophomina phaseolina* in Israel: Challenges for Disease Management and Crop Germplasm Enhancement. *Plant Dis* 106: 15–25

Cools HJ, Mullins JGL, Fraaije BA, Parker JE, Kelly DE, Lucas JA, Kelly SL (2011) Impact of recently emerged sterol 14 α -demethylase (CYP51) variants of *Mycosphaerella graminicola* on azole fungicide sensitivity. *Appl Environ Microbiol* 77: 3830–3837

Cools HJ, Parker JE, Kelly DE, Lucas JA, Fraaije BA, Kelly SL (2010) Heterologous expression of mutated eburicol 14 α -demethylase (CYP51) proteins of *Mycosphaerella graminicola* to assess effects on azole fungicide sensitivity and intrinsic protein function. *Appl Environ Microbiol* 76: 2866–2872

Coser SM, Reddy RVC, Zhang J, Mueller DS (2017) Genetic Architecture of Charcoal Rot (*Macrophomina phaseolina*) Resistance in Soybean Revealed Using a Diverse Panel. 8: 1–12

Crous PW, Slippers B, Groenewald JZ, Wingfield MJ (2017) Botryosphaeriaceae: Systematics, pathology, and genetics. *Fungal Biol*. doi: 10.1016/j.funbio.2017.01.003

Dhingra OD, Sinclair JB (1974) Survival of *Macrophomina phaseolina* Sclerotia in Soil: Effects of Soil

Moisture, Carbon:Nitrogen Ratios, Carbon Sources, and Nitrogen Concentrations. *Phytopathology* 65: 236–240

Dhingra OD, Sinclair JB (1978) Biology and pathology of *Macrophomina phaseolina*. Universidade Federal de Vicosa., Minas Gerais

Duan Y, Ge C, Liu S, Wang J, Zhou M (2013) A two-component histidine kinase Shk1 controls stress response, sclerotial formation and fungicide resistance in *Sclerotinia sclerotiorum*. *Mol Plant Pathol* 14: 708–718

Efron B, Tibshirani RJ (1993) An Introduction to the Bootstrap. An Introd to Bootstrap. doi: 10.1007/978-1-4899-4541-9

Fan J, Urban M, Parker JE, Brewer HC, Kelly SL, Hammond-Kosack KE, Fraaije BA, Liu X, Cools HJ (2013) Characterization of the sterol 14 α -demethylases of *Fusarium graminearum* identifies a novel genus-specific CYP51 function. *New Phytol* 198: 821–835

Gast CE, Basso LR, Bruzual I, Wong B (2013) Azole resistance in *Cryptococcus gattii* from the Pacific Northwest: Investigation of the role of ERG11. *Antimicrob Agents Chemother* 57: 5478–5485

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29: 644–652

Hartman GL, Rupe JC, Sikora EJ, Domier LL, Davis JA, Steffey KL (2015a) Part I . Infectious Diseases. *Compend. Soybean Dis. Pests, Fifth Ed.* pp 15–119

Hartman GL, Rupe JC, Sikora EJ, Domier LL, Davis JA, Steffey KL (2015b) PART IV: Soybean Disease and Pest Management Strategies. *Compend Soybean Dis Pests, Fifth Ed* 167: 167–173

Hawkins NJ, Fraaije BA (2018) Fitness Penalties in the Evolution of Fungicide Resistance. *Annu Rev Phytopathol* 56: 339–360

He QY, Mason AB, Pakdaman R, Chasteen ND, Dixon BK, Tam BM, Nguyen V, MacGillivray RTA, Woodworth RC (2000) Mutations at the histidine 249 ligand profoundly alter the spectral and iron-binding properties of human serum transferrin N-lobe. *Biochemistry* 39: 1205–1210

Heng Li (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.

Ho J, Tumkaya T, Aryal S, Choi H, Claridge-Chang A (2019) Moving beyond P values: data analysis with estimation graphics. *Nat Methods* 16: 565–566

Jacobs JL, Kelly JD, Wright EM, Varner G, Chilvers MI (2019) Determining the Soilborne Pathogens Associated with Root Rot Disease Complex of Dry Bean in Michigan. *Plant Heal Prog* 20: 122–127

Karavidas I, Ntatsi G, Vougeleka V, Karkanis A, Ntanasi T, Saitanis C, Agathokleous E, Ropokis A, Sabatino L, Tran F, et al (2022) Agronomic Practices to Increase the Yield and Quality of Common Bean (*Phaseolus vulgaris* L.): A Systematic Review. *Agronomy*. doi: 10.3390/agronomy12020271

Knaus BJ, Grünwald NJ (2017) vcfr: a package to manipulate and visualize variant call format data in R. *Mol Ecol Resour* 17: 44–53

Kuck K, Russell PE (2006) FRAC: Combined resistance risk assessment. *Asp Appl Biol* 78: 3

Leadbeater A, Mcgrath MT, Wyenandt CA, Stevenson KL (2019) CHAPTER 1: An Overview of Fungicide Resistance and Resistance Management: History and Future Trends. *Fungic. Resist. North Am.* Second Ed. pp 3–19

Lehner MS, Del Ponte EM, Gugino BK, Kikkert JR, Pethybridge SJ (2017) Sensitivity and efficacy of boscalid, fluazinam, and thiophanate-methyl for white mold control in snap bean in New York. *Plant Dis* 101: 1253–1258

Lestrade PPA, Meis JF, Melchers WJG, Verweij PE (2019) Triazole resistance in *Aspergillus fumigatus*: recent insights and challenges for patient management. *Clin Microbiol Infect* 25: 799–806

Liu J, Jiang J, Guo X, Qian L, Xu J, Che Z, Chen G, Liu S (2022) Sensitivity and resistance risk assessment of *Fusarium graminearum* from wheat to prothioconazole. *Plant Dis* 2097–2104

Lucas JA, Hawkins NJ, Fraaije BA (2015) The Evolution of Fungicide Resistance. *Adv Appl Microbiol* 90: 29–92

McDonald BA, Stukenbrock EH (2016) Rapid emergence of pathogens in agro-ecosystems: global threats to agricultural sustainability and food security. *Philos Trans R Soc B Biol Sci* 371: 20160026

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20: 1297–303

Mehl A, Schmitz H, Stenzel K, Bloomberg J, Lp BC, Carolina N (2019) Chapter 5.

Meyer W., Sinclair JB (1974) Factors Affecting Charcoal Rot of Soybean Seedlings. *Phytopathology* 64: 845–849

Mohd-Assaad N, McDonald BA, Croll D (2016) Multilocus resistance evolution to azole fungicides in fungal plant pathogen populations. *Mol Ecol* 25: 6124–6142

Morton V, Staub T (2008) A Short History of Fungicides. *APSnet Featur Artic* 1–12

Motoyama T, Kadokura K, Ohira T, Ichiishi A, Fujimura M, Yamaguchi I, Kudo T (2005) A two-component histidine kinase of the rice blast fungus is involved in osmotic stress response and fungicide action. *Fungal Genet Biol* 42: 200–212

Mullins JGL, Parker JE, Cools HJ, Togawa RC, Lucas JA, Fraaije BA, Kelly DE, Kelly SL (2011) Molecular modelling of the emergence of azole resistance in *Mycosphaerella graminicola*. *PLoS One*. doi: 10.1371/journal.pone.0020973

Nikou D, Malandrakis A, Konstantakaki M, Vontas J, Markoglou A, Ziogas B (2009) Molecular characterization and detection of overexpressed C-14 alpha-demethylase-based DMI resistance in *Cercospora beticola* field isolates. *Pestic Biochem Physiol* 95: 18–27

Noel ZA, Wang J, Chilvers MI (2018) Significant Influence of EC 50 Estimation by Model Choice and EC 50 Type. *Plant Dis* 102: 708–714

Oerke EC (2006) Crop losses to pests. *J Agric Sci* 144: 31–43

Parsa S, García-Lemos AM, Castillo K, Ortiz V, Becerra López-Lavalle LA, Braun J, Vega FE (2016) Fungal endophytes in germinated seeds of the common bean, *Phaseolus vulgaris*. *Fungal Biol*.

Pastor-Corrales MA, Abawi GS, others (1988) Reactions of selected bean accessions to infection by *Macrophomina phaseolina*. *Plant Dis* 72: 39–41

R Core Team (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Reznikov S, Vellicce GR, González V, de Lisi V, Castagnaro AP, Ploper LD (2016) Evaluation of chemical and biological seed treatments to control charcoal rot of soybean. *J Gen Plant Pathol* 82: 273–280

Reznikov S, Vellicce GR, Mengistu A, Arias RS, Gonzalez V, Lisi V De, Garcia MG, Rocha Carla ML, Pardo EM, Castagnaro AP, Ploper LD (2018) Disease incidence of charcoal rot (*Macrophomina phaseolina*) on soybean in north-western Argentina and genetic characteristics of the pathogen. *Can J Plant Pathol* 40: 423–433

Ritz C, Baty F, Streibig JC, Gerhard D (2016) Dose-Response Analysis Using R. *PLoS One* 10: e0146021

Romero Luna MP, Mueller D, Mengistu A, Singh AK, Hartman GL, Wise KA (2017) Advancing Our Understanding of Charcoal Rot in Soybeans. *J Integr Pest Manag* 8: 526–526

Roth MG, Webster RW, Mueller DS, Chilvers MI, Faske TR, Mathew FM, Bradley CA, Damicone JP, Kabbage M, Smith DL (2021) Integrated Management of Important Soybean Pathogens of the United States in Changing Climate. *J Integr Pest Manag*. doi: 10.1093/jipm/pmaa013

Russell PE (2005) A century of fungicide evolution. *J Agric Sci* 143: 11–25

Sang H, Lee HB (2020) Molecular mechanisms of succinate dehydrogenase inhibitor resistance in phytopathogenic fungi. *Res Plant Dis* 26: 1–7

Savary S, Willocquet L, Pethybridge SJ, Esker P, McRoberts N, Nelson A (2019) The global burden of pathogens and pests on major food crops. *Nat Ecol Evol* 3: 430–439

Schnabel G, Jones AL (2001) The α -Demethylase (CYP51A1) gene is overexpressed in *Venturia inaequalis* strains resistant to myclobutanil. *Phytopathology* 91: 102–110

Slippers B, Boissin E (2013) Phylogenetic lineages in the Botryosphaeriales: a systematic and evolutionary framework. *Stud Mycol* 76: 31–49

Slippers B, Wingfield MJ (2007) Botryosphaeriaceae as endophytes and latent pathogens of woody plants: diversity, ecology and impact. *Fungal Biol Rev* 21: 90–106

Tonin RFB, Avozani A, Danelli ALD, Reis EM, Zoldan SM, Garcés-Fiallos FR (2013) In vitro mycelial sensitivity of *Macrophomina phaseolina* to fungicides. *Pesqui Agropecuária Trop* 43: 460–466

Velasquez JG, Prada P, Henry G. (1990) Snap bean pests and diseases in Sumapaz, Colombia: their present status and implications. *Int. Conf. Sn. Beans Dev. World. Cent. Int. Agric. Trop.*

Wei LL, Chen WC, Zhao WC, Wang J, Wang BR, Li FJ, Wei M Di, Guo J, Chen CJ, Zheng JQ, et al (2020) Mutations and overexpression of CYP51 associated with DMI-resistance in *Colletotrichum gloeosporioides* from Chili. *Plant Dis* 104: 668–676

Wrather A, Shannon G, Balardin R, Carregal L, Escobar R, Gupta GK, Ma Z, Morel W, Ploper D, Tenuta A (2010) Effect of Diseases on Soybean Yield in the Top Eight Producing Countries in 2006. *Plant Heal Prog.* doi: 10.1094/php-2010-0102-01-rs

Yamaguchi I, Fujimura M (2005) Recent topics on action mechanisms of fungicides. *J Pestic Sci* 30: 67–74

Zak O, Aisen P, Crawley JB, Joannou CL, Patel KJ, Rafiq M, Evans RW (1995) Iron Release from Recombinant N-lobe and Mutants of Human Transferrin. *Biochemistry* 34: 14428–14434

Zhang C, Diao Y, Wang W, Hao J, Imran M, Duan H, Liu X (2017) Assessing the risk for resistance and Elucidating the Genetics of *Colletotrichum truncatum* that is only sensitive to some DMI fungicides. *Front Microbiol* 8: 1779

APPENDIX

SUPPLEMENTARY TABLES AND FIGURES

Table A.1 Mean EC₅₀ differences and 95% confidence intervals for 35 *M. phaseolina* isolates across experiments for boscalid, iprodione and prothioconazole. Two separate experiments and two replicates per experiment were performed for each isolate and fungicide.

Experiment 1	Experiment 2	Mean EC ₅₀ difference	95% CI ^a	
			low	high
Bos1	Bos2	-0.068	-0.478	0.104
Ipro1	Ipro2	0.063	-0.045	0.213
Pro1	Pro2	-0.065	-0.127	-0.027

^a95% CI adjusting for asymmetrical resampling distributions using bias-corrected and accelerated bootstrap (BCa bootstrap). Experiments are Bos1 and Bos2 for boscalid, Ipro1 and Ipro2 for iprodione and Pro1 and Pro2 for prothioconazole.

Table A.2 Mean EC₅₀ differences for 35 *M. phaseolina* isolates across fungicides: boscalid, iprodione and prothioconazole.

Pairwise fungicide comparison	Mean EC ₅₀ difference	95% CI ^a	
		low	high
Iprodione minus Boscalid	0.388	0.144	0.505
Prothioconazole minus Boscalid	-0.297	-0.568	-0.194
Prothioconazole minus Iprodione	-0.685	-0.759	-0.615

^a 95% confidence intervals adjusting for asymmetrical resampling distributions using bias-corrected and accelerated bootstrap (BCa bootstrap).

Table A.3 Predicted $EC_{50(P)}$ (effective concentration to reduce growth by 50%) for 55 *M. phaseolina* isolates determined from mycelial growth assays in Petri plates amended with 1 mg ml⁻¹ of boscalid, 1 mg ml⁻¹ of iprodione, or 0.5 mg ml⁻¹ of prothioconazole.

Isolate	Boscalid			Iprodione			Prothioconazole		
	$EC_{50(P)}$	95% CI ^a		$EC_{50(P)}$	95% CI ^a		$EC_{50(P)}$	95% CI ^a	
		low	high		low	high		low	high
Et14	0.297	0.276	0.319	0.663	0.640	0.686	0.079	0.072	0.086
Et8	0.204	0.187	0.222	0.675	0.653	0.698	0.082	0.075	0.089
IN12_4	0.253	0.234	0.273	0.673	0.651	0.697	0.083	0.077	0.0903
M_15_1	0.263	0.244	0.283	0.688	0.666	0.710	0.096	0.090	0.103
M_15_10	0.296	0.276	0.318	0.682	0.660	0.705	0.094	0.087	0.1007
M_15_12	0.331	0.309	0.355	0.736	0.714	0.758	0.108	0.101	0.115
M_15_13	0.273	0.254	0.294	0.748	0.727	0.770	0.109	0.103	0.1164
M_15_14	0.279	0.260	0.301	0.664	0.642	0.688	0.163	0.152	0.1741
M_15_3	0.289	0.269	0.311	0.696	0.674	0.718	0.135	0.127	0.143
M_15_4	0.584	0.536	0.637	0.702	0.681	0.725	0.092	0.086	0.099
M_15_5	0.324	0.302	0.347	0.642	0.618	0.666	0.124	0.117	0.1322
M_15_7	0.396	0.368	0.425	0.700	0.678	0.722	0.129	0.122	0.1375
M_15_8	0.341	0.318	0.366	0.693	0.671	0.716	0.102	0.095	0.1088
M_15_9	0.284	0.264	0.305	0.566	0.540	0.593	0.118	0.111	0.125
M11-12	0.727	0.658	0.803	0.838	0.815	0.863	0.093	0.086	0.1
M13-26	0.310	0.289	0.333	0.722	0.701	0.744	0.161	0.150	0.172
M16_1	0.284	0.264	0.305	0.667	0.645	0.691	0.114	0.107	0.1209
Md10	0.284	0.264	0.305	0.681	0.659	0.704	0.102	0.095	0.109
Md5	0.339	0.316	0.364	0.658	0.635	0.681	0.117	0.110	0.1243
Md7	0.339	0.315	0.363	0.638	0.615	0.663	0.112	0.105	0.1191
MISO171-4	0.274	0.255	0.295	0.626	0.602	0.651	0.131	0.124	0.1395
MISO171-5	0.137	0.122	0.153	0.709	0.687	0.731	0.075	0.069	0.0826
MISO171-7	0.239	0.221	0.258	0.656	0.633	0.680	0.128	0.121	0.1361
MISO171-8	0.291	0.271	0.313	0.761	0.739	0.783	0.137	0.129	0.1456
MP220	0.498	0.461	0.539	0.700	0.678	0.723	0.090	0.084	0.0974
MP223	1.172	1.025	1.341	1.116	1.065	1.169	0.325	0.283	0.3731
MP250	0.651	0.594	0.714	0.799	0.777	0.822	0.086	0.079	0.0929
Mph_21	0.755	0.682	0.835	0.944	0.912	0.976	0.094	0.087	0.1009
Mph_22	0.596	0.546	0.650	1.037	0.996	1.080	0.066	0.060	0.0734
Mph_23	0.429	0.399	0.462	1.100	1.052	1.151	0.067	0.060	0.0743
Mph_27	0.177	0.161	0.194	0.852	0.828	0.877	0.155	0.145	0.1658
Mph_28	0.840	0.753	0.937	0.867	0.841	0.893	0.118	0.111	0.1253
Mph_35	0.806	0.725	0.896	0.904	0.876	0.933	0.118	0.111	0.1251
Mph_38	0.248	0.229	0.268	0.735	0.713	0.757	0.089	0.082	0.0958
Mph_39	0.228	0.210	0.247	0.734	0.712	0.756	0.086	0.079	0.0931
Mph_40	0.272	0.253	0.293	0.754	0.733	0.776	0.087	0.080	0.0936

Table A.3 (cont'd)

Mph_45	0.892	0.796	0.999	1.300	1.223	1.381	0.076	0.069	0.0831
Mph_46	0.319	0.297	0.342	0.833	0.809	0.857	0.115	0.108	0.1226
Mph_48	0.386	0.359	0.414	0.909	0.881	0.938	0.089	0.083	0.0961
Mph_49	0.255	0.236	0.275	0.650	0.626	0.674	0.088	0.081	0.095
Mph_5	0.332	0.310	0.357	0.798	0.776	0.821	0.082	0.075	0.089
Mph_50	0.167	0.152	0.184	0.653	0.630	0.677	0.093	0.087	0.1001
Mph_52	0.454	0.421	0.489	0.737	0.716	0.759	0.093	0.086	0.1
Mph_53	0.299	0.278	0.321	0.757	0.736	0.779	0.086	0.079	0.0928
Mph_54	0.283	0.263	0.304	0.752	0.731	0.774	0.182	0.169	0.1967
MpSDSU	1.538	1.316	1.798	0.893	0.866	0.921	0.081	0.075	0.0886
TN262	0.272	0.253	0.293	0.663	0.640	0.686	0.090	0.084	0.0973
TN279	0.305	0.284	0.328	0.583	0.557	0.610	0.098	0.091	0.1049
TN280	0.281	0.261	0.302	0.649	0.625	0.673	0.126	0.118	0.1336
TN380	0.235	0.217	0.255	0.714	0.692	0.736	0.069	0.062	0.0761
UPR-ISA2	0.327	0.305	0.351	0.723	0.702	0.745	0.084	0.077	0.091
UPR-ISA3	0.543	0.501	0.590	0.832	0.809	0.856	0.132	0.125	0.1407
UPR-JD1	0.217	0.199	0.235	0.667	0.644	0.690	0.092	0.085	0.0987
UPR-JD2	0.443	0.411	0.477	0.859	0.834	0.885	0.105	0.099	0.1124
UPR-JD3	0.388	0.362	0.417	0.706	0.685	0.729	0.092	0.085	0.0988

EC_{50(P)} and 95% confidence intervals were back-transformed. ^a95% confidence intervals adjusting for asymmetrical resampling distributions using bias-corrected and accelerated bootstrap (BCa bootstrap).

Table A.4 Mean EC_{50} (effective concentration to reduce growth by 50%) estimates and $EC_{50(P)}$ predictions for a validation set of 13 *M. phaseolina* isolates determined from mycelial growth assays in Petri plates amended with boscalid, iprodione or prothioconazole.

3*Isolate	Boscalid				Iprodione				Prothioconazole			
	2* $EC_{50} \pm SEa$	$EC_{50(P)}$ (95% CI)b			2* $EC_{50} \pm SEa$	2* $EC_{50(P)}b$	$EC_{50(P)}$ 95% CIb		2* $EC_{50} \pm SEa$	2* $EC_{50(P)}b$	$EC_{50(P)}$ 95% CIb	
		low	high				low	high			low	high
Et17	0.789 $\pm$ 0.199	0.755	0.682	0.836	1.225 $\pm$ 0.225	1.117	1.066	1.171	0.177 $\pm$ 0.034	0.083	0.077	0.090
M_15_7	0.309 $\pm$ 0.056	0.316	0.294	0.339	1.012 $\pm$ 0.060	0.957	0.924	0.991	0.128 $\pm$ 0.014	0.131	0.123	0.139
MISO171-5	0.494 $\pm$ 0.345	0.323	0.301	0.347	0.913 $\pm$ 0.052	0.846	0.822	0.870	0.097 $\pm$ 0.008	0.084	0.078	0.091
MISO171-7	0.196 $\pm$ 0.042	0.215	0.197	0.233	0.876 $\pm$ 0.142	0.863	0.838	0.889	0.105 $\pm$ 0.018	0.101	0.094	0.108
MISO171-8	0.257 $\pm$ 0.066	0.221	0.204	0.240	0.752 $\pm$ 0.220	0.787	0.765	0.809	0.119 $\pm$ 0.045	0.164	0.153	0.175
MP144	1.235 $\pm$ 0.357	1.475	1.266	1.719	1.868 $\pm$ 0.552	1.345	1.261	1.435	0.456 $\pm$ 0.117	0.342	0.296	0.395
MP223	—	2.402	1.978	2.916	1.902 $\pm$ 1.229	1.216	1.151	1.284	0.367 $\pm$ 0.181	0.233	0.210	0.257
Mph_21	0.506 $\pm$ 0.299	0.483	0.447	0.522	1.803 $\pm$ 0.813	1.285	1.210	1.364	0.187 $\pm$ 0.031	0.128	0.121	0.136
Mph_50	0.220 $\pm$ 0.040	0.166	0.151	0.183	0.526 $\pm$ 0.058	0.629	0.605	0.654	0.087 $\pm$ 0.004	0.084	0.077	0.091
Mph_56	0.194 $\pm$ 0.033	0.227	0.209	0.246	1.125 $\pm$ 0.163	0.977	0.942	1.012	0.240 $\pm$ 0.134	0.222	0.202	0.244
W12-6	0.338 $\pm$ 0.070	0.350	0.326	0.376	0.944 $\pm$ 0.127	0.902	0.874	0.930	0.136 $\pm$ 0.012	0.078	0.071	0.085
W25	0.534 $\pm$ 0.159	0.399	0.372	0.429	1.375 $\pm$ 0.130	1.290	1.214	1.370	0.117 $\pm$ 0.024	0.076	0.070	0.084
W3-5	0.507 $\pm$ 0.108	0.408	0.380	0.439	1.139 $\pm$ 0.193	1.037	0.996	1.080	0.150 $\pm$ 0.025	0.075	0.068	0.082

Table A.5 Mean $EC_{50}/EC_{50(P)}$ (effective concentration to reduce growth by 50%) values for 92 *M. phaseolina* isolates determined from mycelial growth assays in Petri plates amended with boscalid, iprodione, or prothioconazole.

Isolate	Boscalid	Iprodione	Prothioconazole
	Mean $EC_{50}/EC_{50(P)}$	Mean $EC_{50}/EC_{50(P)}$	Mean $EC_{50}/EC_{50(P)}$
CR_Red_1	0.716	1.162	0.165
CR_Red_2B	2.766	1.106	0.089
CR_Red_3	0.663	0.871	0.177
Dm13	0.527	0.856	0.146
Et12	0.184	0.780	0.135
Et14	0.297	0.663	0.079
Et17	0.652	1.061	0.162
Et18	0.480	0.680	0.112
Et8	0.204	0.675	0.082
IN12_4	0.253	0.673	0.083
IN12_8_3	0.261	0.574	0.094
IN12_9_4	0.189	0.588	0.059
IN12_9_6	0.256	0.819	0.136
IN12_PO_3	0.253	0.525	0.197
M_15_1	0.263	0.688	0.096
M_15_10	0.296	0.682	0.094
M_15_12	0.331	0.736	0.108
M_15_13	0.273	0.748	0.109
M_15_14	0.279	0.664	0.163
M_15_3	0.289	0.696	0.135
M_15_4	0.584	0.702	0.092
M_15_5	0.324	0.642	0.124
M_15_7	0.352	0.856	0.129
M_15_8	0.341	0.693	0.102
M_15_9	0.284	0.566	0.118
M11-12	0.727	0.838	0.093
M13-26	0.310	0.722	0.161
M16_1	0.284	0.667	0.114
Md10	0.284	0.681	0.102
Md3	0.124	0.650	0.147
Md5	0.339	0.658	0.117
Md6	0.203	0.769	0.124
Md7	0.339	0.638	0.112
MI-SF 1-36	0.216	0.712	0.129
MI-SF 10-16	0.223	0.786	0.108
MI-SF 2-16	0.197	0.854	0.110

Table A.5 (cont'd)

MI-SF 9-8	0.190	0.993	0.099
MISO171-1	0.186	0.810	0.087
MISO171-2	0.315	0.781	0.091
MISO171-3	0.170	0.663	0.108
MISO171-4	0.274	0.626	0.131
MISO171-5	0.315	0.811	0.086
MISO171-7	0.217	0.766	0.117
MISO171-8	0.274	0.756	0.128
MP144	1.195	1.552	0.577
MP220	0.498	0.700	0.090
MP223	1.172	1.509	0.346
MP249	0.764	1.038	0.156
MP250	0.651	0.799	0.086
MP258	0.290	1.027	0.140
Mph_21	0.630	1.373	0.140
Mph_22	0.596	1.037	0.066
Mph_23	0.429	1.100	0.067
Mph_27	0.177	0.852	0.155
Mph_28	0.840	0.867	0.118
Mph_35	0.806	0.904	0.118
Mph_38	0.248	0.735	0.089
Mph_39	0.228	0.734	0.086
Mph_40	0.272	0.754	0.087
Mph_45	0.892	1.300	0.076
Mph_46	0.319	0.833	0.115
Mph_48	0.386	0.909	0.089
Mph_49	0.255	0.650	0.088
Mph_5	0.332	0.798	0.082
Mph_50	0.194	0.589	0.090
Mph_52	0.454	0.737	0.093
Mph_53	0.299	0.757	0.086
Mph_54	0.283	0.752	0.182
Mph_56	0.194	1.125	0.240
MpSDSU	1.538	0.893	0.081
SAG5-4	0.219	0.984	0.146
TN262	0.272	0.663	0.090
TN264	0.275	0.653	0.237
TN270	0.496	0.767	0.046
TN279	0.305	0.583	0.098
TN280	0.281	0.649	0.126
TN380	0.235	0.714	0.069
TN4	0.198	0.657	0.051
TN5	0.190	0.728	0.200

Table A.5 (cont'd)

TN ₅₀₁	0.674	0.707	0.136
TN ₅₅₀	0.396	1.000	0.167
UPR-Mph-ISA ₂	0.327	0.723	0.084
UPR-Mph-ISA ₃	0.543	0.832	0.132
UPR-Mph-JD ₁	0.217	0.667	0.092
UPR-Mph-JD ₂	0.443	0.859	0.105
UPR-Mph-JD ₃	0.388	0.706	0.092
W-MISO _{2 3-6}	0.409	0.713	0.137
W-MISO _{2 4-10}	0.342	—	0.151
W ₁₂₋₆	0.338	0.944	0.136
W ₂₃	0.217	0.645	0.133
W ₂₅	0.640	1.058	0.139
W ₃₋₅	0.491	1.119	0.160

Table A.6 Mean $EC_{50}/EC_{50(P)}$ differences for 92 *M. phaseolina* isolates across genetic clusters. Genetic cluster US1A was used a reference group for comparisons.

Reference	Test	Boscalid			Iprodione			Prothioconazole		
		Mean EC_{50} difference ^a	95% CI ^a		Mean EC_{50} difference	95% CI ^a		Mean EC_{50} difference	95% CI ^a	
			low	high		low	high		low	high
US1A	US1B	0.080	-0.018	0.188	0.142	-0.002	0.275	-0.040	-0.059	-0.015
US1A	US2	0.570	0.357	1.121	0.345	0.238	0.486	0.065	0.018	0.171
US1A	COLPR ₁	0.072	0.001	0.202	0.087	0.014	0.195	-0.021	-0.033	-0.007
US1A	COLPR ₂	0.289	0.109	0.449	0.189	0.078	0.353	0.031	0.005	0.061

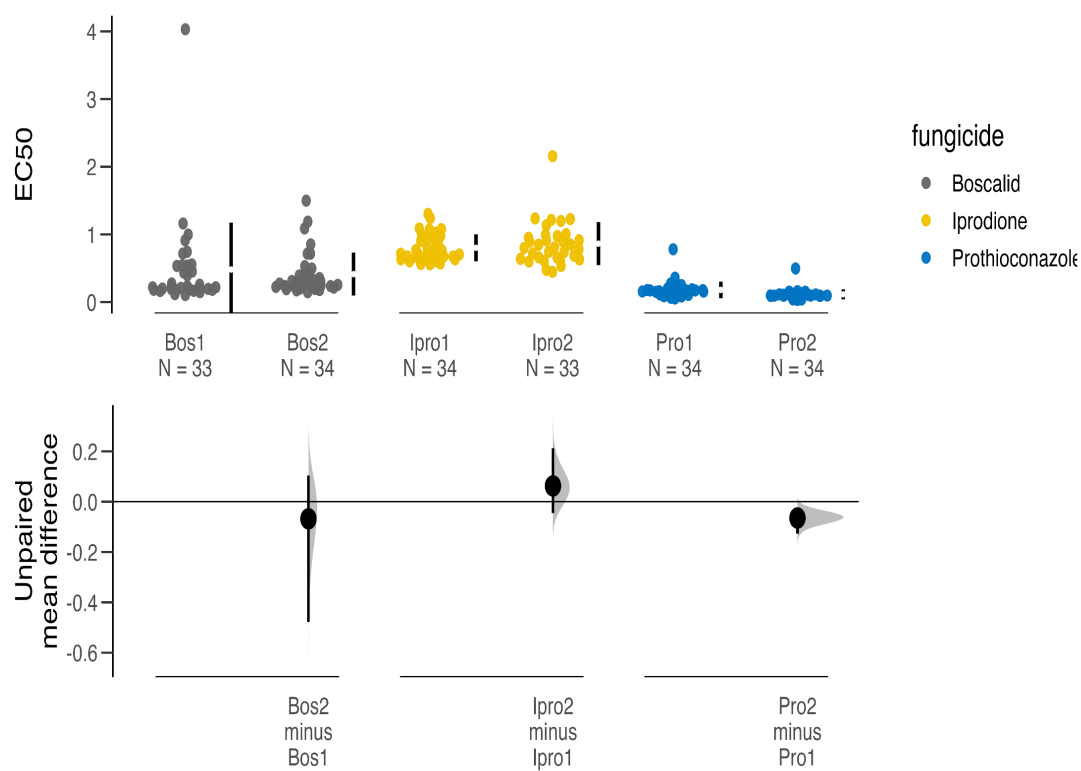


Figure A.1 Mean differences of 35 *M. phaseolina* isolates across experiments. Two separate experiments and two replicates per experiment were performed for each isolate and fungicide. Experiments are Bos1 and Bos2 for boscalid, Ipro1 and Ipro2 for iprodione and Pro1 and Pro2 for prothioconazole. Isolates with data from at least two replicates were included.

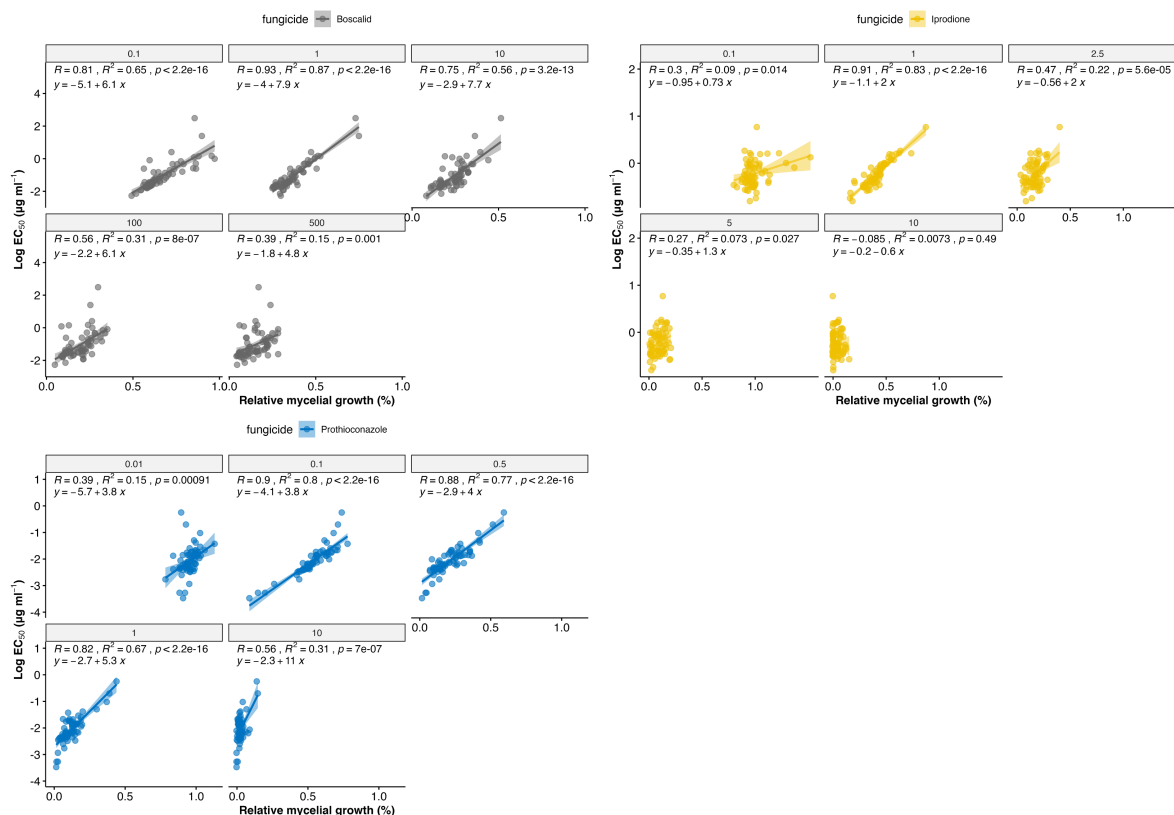


Figure A.2 Linear regression models and correlation analyses of relative mycelial growth and EC_{50} values of 35 *M. phaseolina* isolates. Log Absolute EC_{50} values were used. The line shows the linear regression with 95% confidence interval shaded. Selected concentrations were 1 $\mu\text{g ml}^{-1}$ for boscalid and iprodione and 0.5 $\mu\text{g ml}^{-1}$ for prothioconazole.

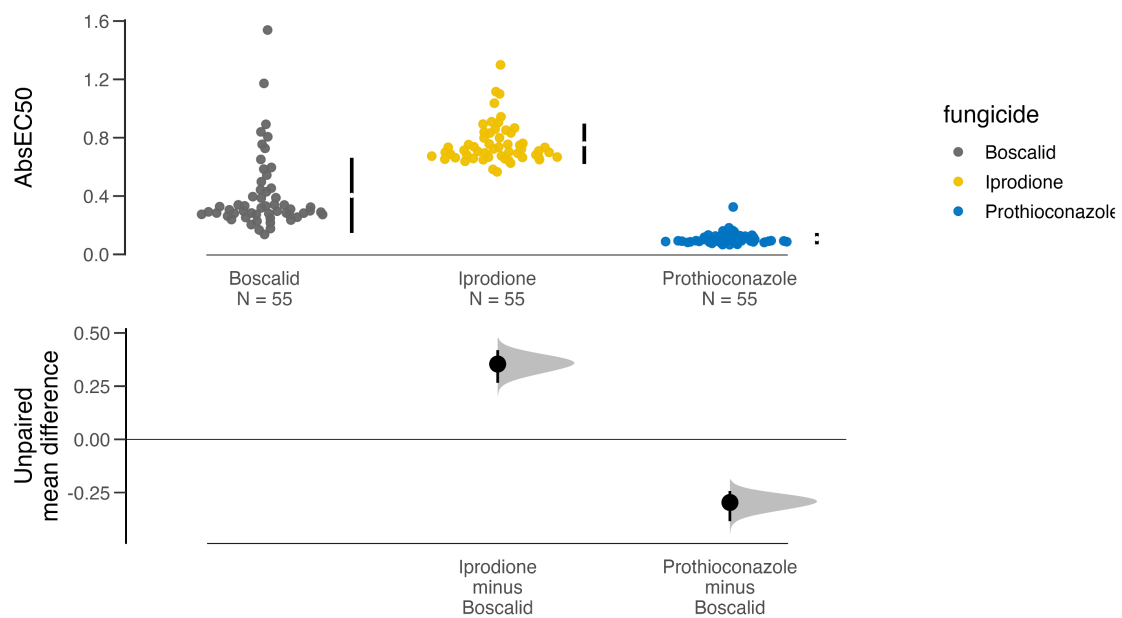


Figure A.3 Mean $EC_{50(P)}$ differences of 55 *M. phaseolina* isolates across fungicides. Iprodione and prothioconazole as compared to boscalid. Isolates were most sensitive to prothioconazole, followed by boscalid and least sensitive to iprodione.

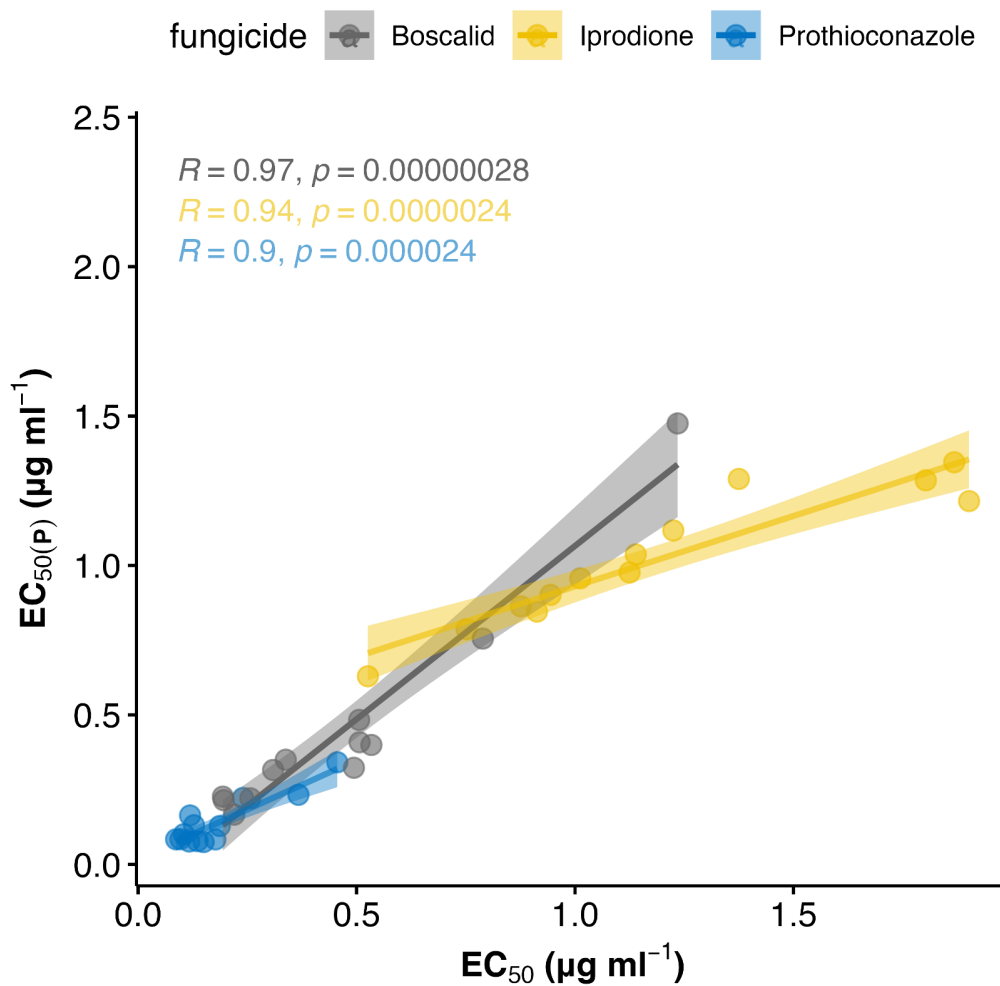


Figure A.4 Correlation of mean EC_{50} (effective concentration to reduce growth by 50%) estimates and $EC_{50(P)}$ predictions for a validation set of 13 *M. phaseolina* isolates determined from mycelial growth assays in Petri plates amended with multiple concentrations boscalid, iprodione or prothioconazole.

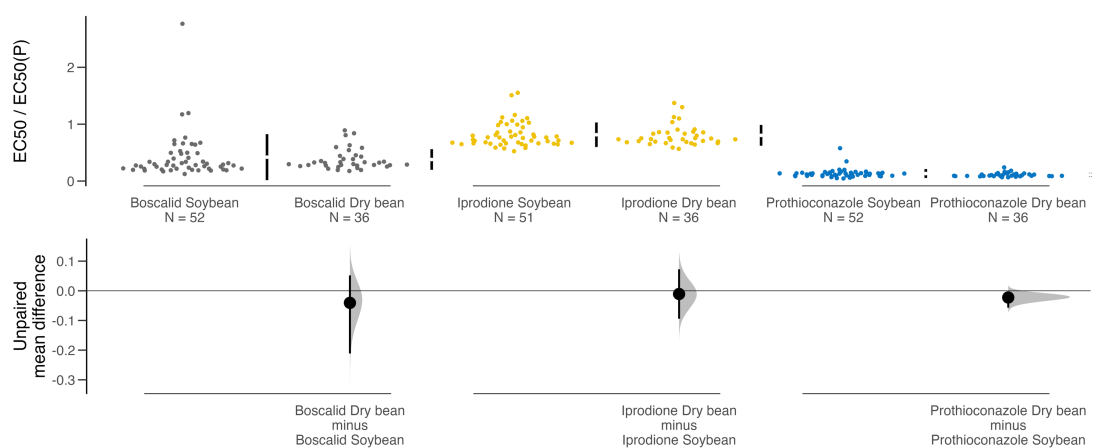


Figure A.5 Mean $EC_{50} / EC_{50(P)}$ differences of *M. phaseolina* isolates by host. Only isolates collected from soybean or dry bean are shown.

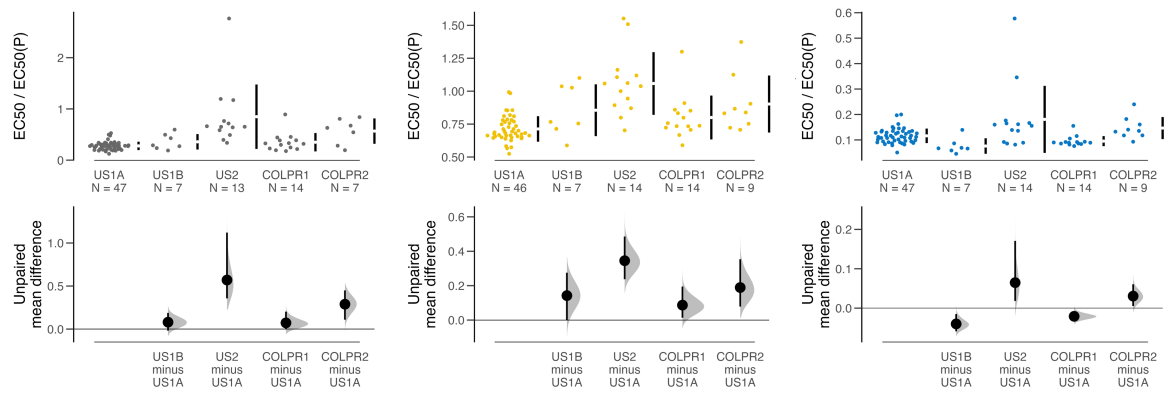


Figure A.6 Mean $EC_{50} / EC_{50(P)}$ differences of 92 *M. phaseolina* isolates by genetic cluster. Isolate TN264 not shown, since there is no information of genetic cluster membership.

CHAPTER 4

ECOCLIMATIC SUITABILITY AND ADAPTIVE GENOMICS IN *MACROPHOMINA PHASEOLINA*, THE CHARCOAL ROT PATHOGEN

4.1 Abstract

Globally, charcoal rot caused by the fungal pathogen *Macrophomina phaseolina* is listed among the top diseases threatening agricultural production. The environment has a profound influence on plant diseases, however the effect of accelerated climate change on disease development is uncertain and host-pathogen system specific. We studied the distribution and genomic adaptive potential of *M. phaseolina*, a major fungal plant pathogen, in relation to climate. We retrieved worldwide occurrences of *M. phaseolina* to develop an explanatory species distribution model using climatically relevant variables. Georeferenced occurrences in the global biodiversity information facility (GBIF) database and information of *M. phaseolina* isolates collected in the US, Colombia and Puerto Rico reported in a previous study were used. Occurrence data and climatic variables were used to identify within species worldwide suitability patterns in *M. phaseolina*. Candidate adaptive loci associated with climatic variation were used to calculate an adaptive index and infer the distribution of adaptive genetic variation in *M. phaseolina*. A global species distribution bioclimatic model for *M. phaseolina* identified areas of high climatic suitability for its occurrence that is consistent with all current records. Notable areas of high suitability were projected in the southern US, north-eastern Argentina, eastern Australia, and southern Europe, where outbreaks were recently reported.

4.2 Introduction

Changes in climate are already affecting disease incidence in agricultural systems (Altizer et al., 2013; Váry et al., 2015; Velásquez et al., 2018). Very often these effects depend on the patterns of climate change and the host-pathogen system. For example, pathogen distribution and crop disease severity are driven to a large extent by particular changing patterns in temperature, rainfall events and humidity (Altizer et al., 2013; Sparks et al., 2014; Velásquez et al., 2018; Yonow et al., 2019; Dudney et al., 2021). Furthermore, responses to changing climate are intricately tied to organisms potential adaptive mechanisms and intra-specific variation in those mechanisms (local adaptation), which in turn are influenced by factors such as gene flow, and phenotypic plasticity (Savolainen et al., 2007; Savolainen et al., 2013; Croll and McDonald, 2016; Waldvogel et al., 2020). Thus, both environmental and evolutionary potential should be investigated and considered when modeling the distribution of species.

Models of climate change for the coming decades predict increases in global temperature, rainfall and severe weather (Fisher et al., 2012). This is expected to increase the climatic variation that already is present across different agricultural systems and regions of the world. To predict how pathogens geographic distribution will be altered under future climate changes it is necessary to understand how the current pathogen distribution depend on climatic factors (Shaw and Osborne, 2011). However, the specific environmental factors that contribute to the current distributions and disease occurrences have not been characterized and species distribution models (SDMs) have not been developed for most plant pathogens (Ireland and Kriticos, 2019).

Charcoal rot, caused by the widespread pathogen *Macrophomina phaseolina*, is listed among the top 10 diseases causing soybean yield losses in the US as well as globally (Allen et al., 2017; Savary et al., 2019; Bradley et al., 2021). Diseases caused by *M. phaseolina* are favored by high temperatures and drought episodes and these conditions are known to play a key role in triggering epidemics (Dhingra and Sinclair, 1974; Meyer and Sinclair, 1974; Kendig et al., 2000; Yang and Navi, 2005; Mengistu et al., 2011a; Mengistu et al., 2011b; Reznikov et al., 2018). In the past few years, a surge in first reports of diseases caused by *M. phaseolina* in a variety of crops and countries have been observed, including hemp in southern Spain (Casano et al., 2018), tomato from Pakistan (Hyder et al., 2018), stevia in North Carolina (Koehler and Shew, 2018), sugarcane in China (Wang et al., 2020), zebra plant in Serbia (Tančić Živanov et al., 2018), catnip in India (Nishad et al., 2018), grapevine in the US (Nouri et al., 2018), strawberry in Italy (Gerin et al., 2018), Malabar spinach in India (Meena et al., 2018) among others. Interest in the interaction of climate-charcoal rot have been rising and the associations between charcoal rot and climate have been examined through review studies (Batista et al., 2021; Cohen et al., 2022). However, predicting the effects of climate change on *M. phaseolina* distribution remains limited and models have not been used to predict the climate suitability of this pathogen.

Species distribution modelling is an important tool in ecology and biogeography to investigate species ranges and factors contributing to their distribution (Sutherland, 2006; Elith and Leathwick, 2009; Juroszek and Von Tiedemann, 2013). SDMs have been used to predict the distributions of plant pathogens as determined by climate (Burgess et al., 2017; Yonow et al., 2019) and to assess the risk of disease (Sparks

et al., 2014) and epidemics (Paini et al., 2016). In addition to environmental conditions, evolutionary processes, including those within-species, are crucial in species response to climate (Jay et al., 2012). Therefore, genomic data is increasingly considered in SDMs being of special interest adaptive genetic variation (Waldvogel et al., 2020). Adaptive genetic variation can provide insights into climate adaptation mechanisms and the potential of rapid local adaptation to occur in the future under climate change. Nevertheless, SDM approaches often encounter challenges incorporating evolutionary information (Waldvogel et al., 2020).

Recent developments in genotype-environment associations using redundancy analysis allow insights into patterns of adaptive variation and can be used for the identification of candidate adaptive genomic loci and adaptive indices in widespread non-model species (Steane et al., 2014; Capblancq and Forester, 2021). These tools have the potential to estimate adaptive indices associated with climatic variation in fungal species in a landscape genomics framework. Indeed, candidate adaptive loci were previously identified in *M. phaseolina* (Ortiz et al., under review) which can be used to calculate adaptive indices in this pathogen. Adaptive indices provide a measure of the adaptive genetic similarity on the landscape as a function of climatic variables values at each location across the landscape (Steane et al., 2014; Capblancq and Forester, 2021). This study investigated the effect of climatic variables in shaping the distribution of *M. phaseolina* on a global scale, incorporating evolutionary projections. The objectives of this study were to describe the climatic suitability and calculate an adaptive genetic-based index of *M. phaseolina* on a global scale. We specifically developed an explanatory global distribution bioclimatic model by associating recorded locations of *M. phaseolina* with climatic variables and projected an adaptive genomic index across the *M. phaseolina* distribution.

4.3 Results

4.3.1 Climatic suitability model

A correlative bioclimatic model based on *M. phaseolina* occurrence data and five climatic variables related to temperature and precipitation was developed using BIOCLIM. The model captured areas of climatic suitability for *M. phaseolina* occurrences in every continent, which is consistent with this pathogen records (Batista et al., 2021). The model mean AUC obtained via cross-validation with presence/pseudo-

absence data was 0.65 (Supplementary Figure A.1). A high AUC indicates that locations with high predicted suitability scores tend to be locations of known presence (i.e., true positive rate). While an AUC score of 0.5 correspond to random predictions. We found useful discriminatory ability of suitable vs. unsuitable areas with our model considering the number of records included in this study and that for presence-background data models the maximum possible AUC is less than 1 (Phillips et al., 2006). The model projected high suitability for localities with low precipitation of driest quarter (BIO17) and precipitation of warmest quarter (BIO18), and high mean temperature of warmest quarter (BIO10) (Figure A.2). These predictions correspond to an expected distribution of higher *M. phaseolina* suitability at warm and dry regions. In the US, predicted suitable regions were concentrated across locations in southern states, including Texas, Oklahoma, Kansas, Arkansas, and Missouri. The highest suitability values in the US were projected in a region of Arizona (southwest US). Although, generally lower than for southern regions, regions of high suitability were projected as well in locations in the East and West North Central regions (Figure 4.1).

In Colombia, regions of intermediate suitability were predicted mainly in the extreme north and eastern plains of Colombia (Caribbean and Orinoquia regions, respectively). Similar, intermediate suitability values were predicted in Puerto Rico and other islands of the Caribbean. A trend of highly suitable values was observed in southern Europe particularly along coastal regions of Spain, France and Italy, and localities of eastern Europe. Notably, the model predicted a large region of high suitability in the north-east of Argentina, referred as the Plata Plain region, with highest values in areas near Buenos Aires and La Pampa provinces. Reports of *M. phaseolina* occurrence and disease outbreaks in soybean, canola and strawberry has been recently observed in northern provinces of Argentina (Gaetán et al., 2006; Bairo et al., 2011; Viejobueno et al., 2017; Reznikov et al., 2018). Likewise, a high suitability is observed in regions of eastern Australia and south-eastern South Africa for which increased charcoal rot incidence has been reported (Hutton et al., 2013; Jordaan et al., 2019a) (Figure 4.1).

4.3.2 Spatial autocorrelation

Precipitation of warmest quarter (BIO18) and precipitation of driest quarter (BIO17) are aggregated with similar precipitation values occurring within approximately 2000 km showing a maximum correlation

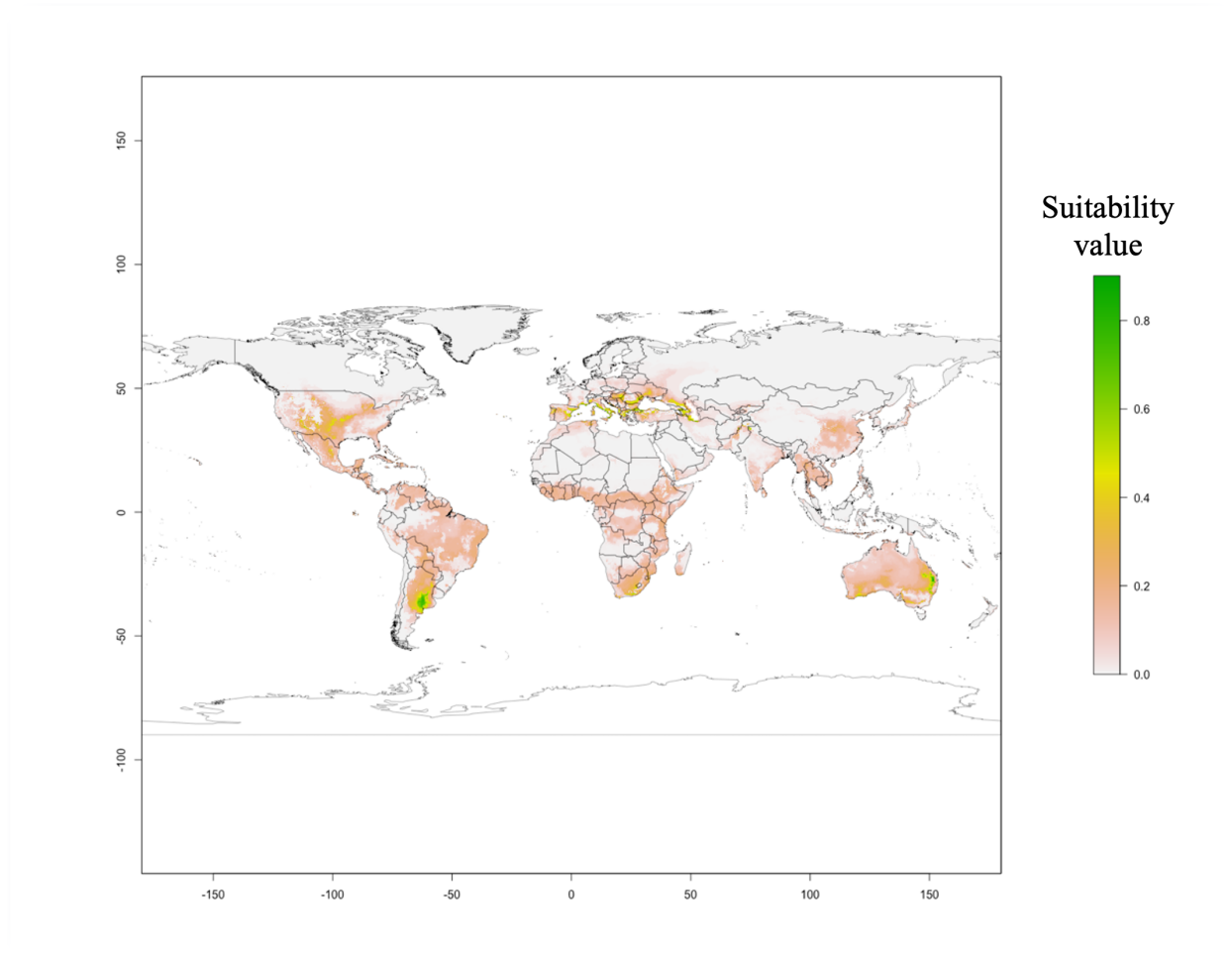


Figure 4.1 BIOCLIM global climatic suitability model for *Macrophomina phaseolina*. BIOCLIM algorithm and presence-background data records were used. A suitability value of 1 (green) indicates a location with high suitability and a value of 0 is given for locations predicted as unsuitable. Climatic variables used as predictors in the model were BIO18 = Precipitation of Warmest Quarter, BIO15 = Precipitation Seasonality (Coefficient of Variation), BIO17 = Precipitation of Driest Quarter, BIO10 = Mean Temperature of Warmest Quarter and BIO4 = Temperature Seasonality (standard deviation *100).

($r > 0.7$) (Supplementary figure A.3). The correlation decreases rapidly at distances greater than approximately 2500 km between points. At distances 3000 km, for most distance classes, the correlation is negative. Most climatic variables show a similar pattern of aggregation at points within 2000 km of each other, decreasing near to zero rapidly and shifting to negative correlations at greater distances. An exception is BIO10, mean temperature of warmest quarter, which showed peaks of positive correlation at greater distances (Supplementary figure A.3).

Based on these results, we can reject the null hypothesis that geographic and climatic distances are

uncorrelated with $p = 0.001$ for BIO18, BIO17 and BIO4. No significant correlation was found for BIO15 ($p=0.883$) and for BIO10 ($p=0.605$). The observed correlations for BIO18, BIO17 and BIO4 were $r=0.07$, 0.10 , 0.55 respectively indicate that points that are closer to each other have more similar climatic values than points that are far from each other.

4.4 Discussion

In this study, we developed a correlative BIOCLIM model for *M. phaseolina* to project the climate suitability of *M. phaseolina* and identify localities at risk of charcoal rot and other diseases caused by this pathogen at a global scale. Previous studies have reported the current *M. phaseolina* global distribution and its association to climate at a continental or biome resolution (Batista et al., 2021). This model constitutes the first attempt to predict the distribution of *M. phaseolina* at a resolution of approximately 20 km. Importantly, by using global records we provide an examination of temperature and precipitation variables that are predicted to be highly suitable for the occurrence of *M. phaseolina*.

The current distribution and disease dynamics of *M. phaseolina* are heavily influenced by climatic factors such as high temperature and low soil water availability (Sexton et al., 2016; Batista et al., 2021; Marquez et al., 2021; Cohen et al., 2022). The model was consistent with charcoal rot reports in areas with high mean temperature of warmest quarter (BIO10) and low precipitation of driest quarter (BIO17) and precipitation of warmest quarter (BIO18) around the world. In the US, areas projected as most suitable are in states with reported highest soybean yield losses due to charcoal rot (Allen et al., 2017; Bradley et al., 2021). Although, these reports are highest in the warmest and southernmost states, charcoal rot is a consistent threat to soybean grown in the northern US regions as well (Bradley et al., 2021; Roth et al., 2021). Our results of *M. phaseolina* potential distribution indicated by areas of intermediate suitability along the east north central and west north central regions suggest potential for further expansion of charcoal rot occurrences to these regions in the US.

Globally, our results projected the north-eastern region of Argentina as one of the largest areas with high suitability. The provinces of Buenos Aires, Tucuman and other northern provinces, have already reported charcoal rot epidemics in soybean, strawberry and canola (Gaetán et al., 2006; Bains et al., 2011; Viejobueno et al., 2017; Reznikov et al., 2018). Likewise, the model projected areas in the Eastern Cape

and Free State provinces of South Africa as intermediate to highly suitable. Charcoal rot in soybean and sunflower has been reported affecting fields in Free state province, one of the major producing regions for these crops in South Africa (Jordaan et al., 2019b). In Australia, the model is congruent with charcoal rot reports on olives, strawberry and other field and horticultural crops grown in eastern regions (Sergeeva et al., 2005; Hutton et al., 2013; Poudel et al., 2021). These observations suggest that our model can predict suitability of *M. phaseolina* in regions for which data points were not included but with reported presence or disease caused by this pathogen. Thus, we suggest this model may be used as an indicator for the potential risk of disease development. Similar models have used climatic suitability as proxy for disease caused by fungal and oomycete pathogens (Burgess et al., 2017; Hernández-Lambrano et al., 2018; Yonow et al., 2019).

Accurate predictions on the effects of climate on species occurrences face several challenges (Phillips et al., 2006; Franklin, 2013). For presence-background models, one of such challenges is that the accuracy of predictions is highly dependent on methods for background data selection (Phillips et al., 2006; Hijmans et al., 2017). Model performance as assessed by AUC in presence-background models tend to increase with larger spatial extents from which background points are sampled. To address this, we sampled background points within a radius of 100 km from the presence records (VanDerWal et al., 2009). This, although appropriate for our data, contributed to the relatively low observed AUC value (Phillips et al., 2006). In addition, presence-only and presence-background models using environment-only data have been identified as least accurate as compared to true-absence models in which additional factors related to the biology or epidemiology of organisms are accounted for to environment (Phillips et al., 2006). Thus, a limitation of our model is the relatively low number of records used to build the model, as compared to climatic suitability models developed at a global scale for other pathogens (Burgess et al., 2017; Hernández-Lambrano et al., 2018; Yonow et al., 2019) and the use of climatic only data.

To address the lack of biological data in our model and to provide an estimation of the effects of evolutionary processes into *M. phaseolina* distribution we used a complementary approach to model within-species evolutionary factors by estimating an adaptive index. This index was estimated using previously identified candidate loci for climate adaptation in data set of 95 *M. phaseolina* collected across

the US, Colombia, and Puerto Rico (Ortiz et al., under review). This data set encompasses isolates collected across a wide range of climates, making it suited for studying within-species adaptation to climate. The adaptive index estimated for the data set of 95 *M. phaseolina* isolates suggest that even in locations where similar temperatures are observed, isolates may differentially respond depending on the presence of adaptive loci. Climatic gradients have now been reported in plant species (Steane et al., 2014; Capblancq and Forester, 2021), however this the first time it has been used to predict the adaptive landscape in a plant pathogen. In summary, we provided a first species distribution model that serve as a basis for future more comprehensive predictive SDMs. Our model will also be useful for local adaptation that constitute the first step towards assessing the adaptive response of this fungal pathogen under climate change. Further improvements of the model will involve including larger data sets and the use of semi-mechanistic models (e.g., MAXENT) that allow the incorporation of biological parameters (Phillips et al., 2006), for example growth rates at different temperatures in fungal plant pathogens. Given the increasing impact of *M. phaseolina* on agroecosystems globally, the modelling of its distribution offers an important preliminary tool for monitoring and development of management strategies incorporating eco-evolutionary projections. Further, regional distribution models would provide a better assessment of charcoal rot risk in different crops. From a practical standpoint, of particular interest are crops and locations for which disease assessments data over time is available such as is the case for charcoal rot of soybean in the US (Bradley et al., 2021). A major need remains for *M. phaseolina* and other plant pathogens to examine the incorporation of disease risk assessments into management strategies.

4.5 Materials and methods

4.5.1 Study area and distribution data

Distribution data was obtained from two sources, records of *M. phaseolina* occurrences retrieved from the global biodiversity information facility (GBIF) database (GBIF.org) and a dataset on a collection of isolates throughout the US, Puerto Rico, and Colombia for which genomic data is available (Ortiz et al., under review). A total of 471 records for “*Macrophomina phaseolina*” were obtained from GBIF using R v4.0.0 (R Core Team 2019). After filtering for missing data and cleaning for potential georeferentiation mistakes, 231 records were maintained. For additional 66 records without longitude and latitude

information, coordinates based on location description were retrieved using the ‘geocode’ function as implemented in R v4.0.0 (R Core Team 2019). The longitude and latitude information of a collection of 95 isolates of *M. phaseolina* isolates as well as genomic data and adaptive candidate loci available from a previous population genomics study were used (Ortiz et al., under review). In brief, these isolates were collected mainly from soybean and dry bean in the US, Colombia and Puerto Rico from commercial fields and experimental stations. The entire data set covered occurrence records originating from plant tissues or soil in every continent, but Antarctica (Figure 4.2), consistent with the current reported distribution of *M. phaseolina* (Batista et al., 2021).

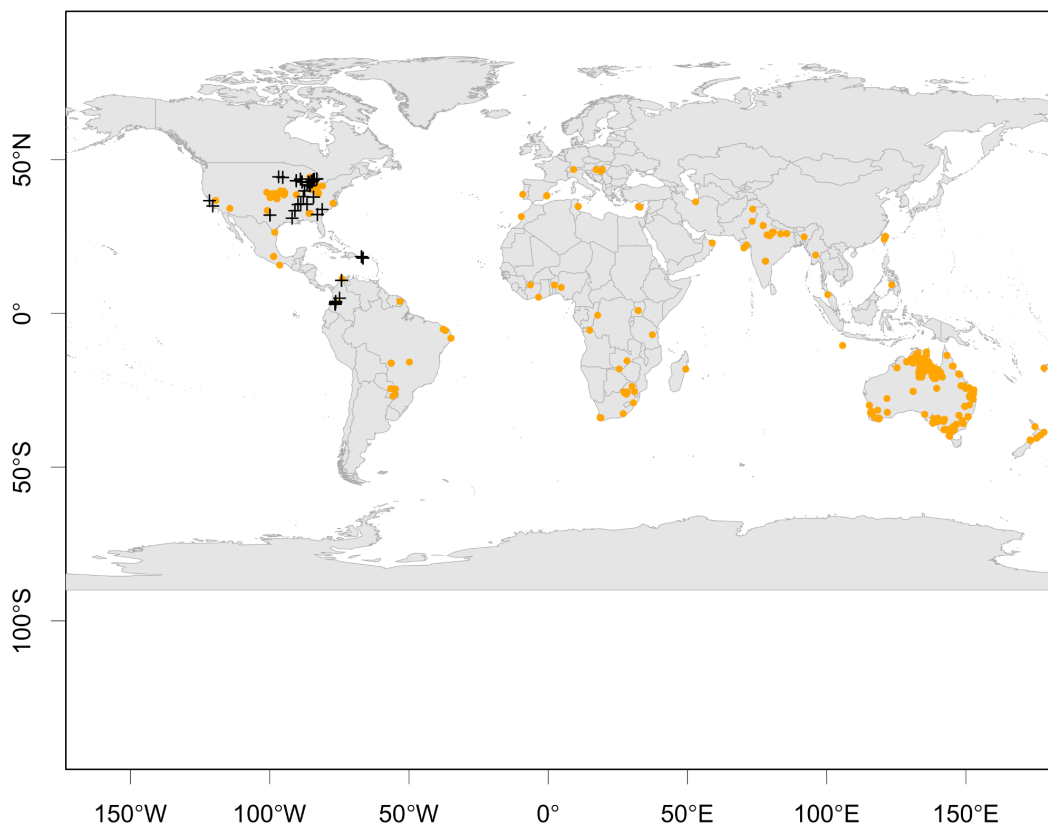


Figure 4.2 Geographic locations of *Macrophomina phaseolina* records included in the BIOCLIM model. Records obtained from the global biodiversity information facility (GBIF) are shown in orange circles. Isolate collection sites of 95 *M. phaseolina* isolates collected in the US, Puerto Rico and Colombia are depicted in black.

4.5.2 Distribution model

Five bioclimatic variables previously selected from the 19 standard bioclimatic variables (WorldClim v2) as described previously were used (Ortiz et al., under review). In summary, the bioclimatic variables are the average for the years 1970 to 2000 and were obtained at a resolution of 2.5 min (21.5 km²) which correspond with that of the data for the isolate collection, recorded at a field to municipality scale. This set of climatic variables was selected based on ecological relevance and after removing correlated variables ($|r| > 0.7$). The selected variables were: BIO18 = Precipitation of Warmest Quarter, BIO15 = Precipitation Seasonality (Coefficient of Variation), BIO17 = Precipitation of Driest Quarter, BIO10 = Mean Temperature of Warmest Quarter and BIO4 = Temperature Seasonality (standard deviation *100).

The species distribution model (SDM) was built using BIOCLIM as implemented in ‘dismo’ R package (Hijmans et al., 2017). We used BIOCLIM algorithm with presence-background data. The algorithm creates percentile distributions for the climatic data values at the locations of species occurrence (“training sites”). The values for each climatic variable are compared to the percentile distribution of the training sites providing a measure of similarity between locations. Since one-tailed percentile distributions are used (10th percentile is treated as equivalent to 90th percentile), the closer to the 50th percentile (the median), the more suitable a location is. Here, we used the ‘dismo’ implementation in which the suitability values are scaled, thus resulting in values between 0 and 1. The value of 1 is given for a location that would have the median values of the training data for all the variables considered, while 0 will be given for cells with climatic values outside of the range of the training data for at least one of the variables. The final BIOCLIM model was fitted with all presence records from the GBIF cleaned dataset and 95 records from the previously published *M. phaseolina* isolate collection using the five selected climatic variables as predictors.

Since we used a presence-background species distribution modeling approach, we selected background data for model parameterization (Hijmans and Elith, 2017). Background localities were generated at random within a radius of 100 km from the presence records (VanDerWal et al., 2009). The models were assessed and compared according to their discrimination capacity of suitable versus unsuitable areas for *M. phaseolina* using the area under the receiver operator curve (AUC) in the ‘dismo’ implementation.

Two additional classification assessment indices were used (Fielding Bell, 1997): sensitivity (true positive rate i.e., the proportion of correctly classified presences) and specificity (true negative rate, i.e., the proportion of correctly classified absences). We divided the presence data in training and testing sets via cross-validation with k-fold (k=5) data partitioning. The background data was only used for model testing and was not partitioned. The mean AUC of the five cross-validation runs was reported as well as the maximum of the sum of the sensitivity (true positive rate) and specificity (true negative rate) (Hijmans et al., 2017).

4.5.3 Spatial autocorrelation

Spatial autocorrelation was tested using BIOCLIM in 'dismo' R package (Hijmans et al., 2017). A subset of 81 records out of the 95 isolate collection records for which the resolution was at least to the municipality level. Similarly, the GBIF records with exact longitude and latitude coordinates as recorded in the GBIF dataset were used (i.e., records that georeferenced using geocode were excluded) for spatial autocorrelation analysis. The associated climatic data values for the five variables for each record, as it was retrieved for the SDM analysis, was used. A geographic distance matrix was computed using longitude and latitude coordinates as well as distance matrices for each of the five environmental predictors. Correlograms for each of the climatic variables were performed using 20 distance classes with 1000 km distance increments. A Mantel test was run for each of the environmental predictors between the geographic distance matrix and the environmental distance matrix for each climatic predictor to test for significant autocorrelation at each distance class.

BIBLIOGRAPHY

Allen TW, Bradley CA, Sisson AJ, Byamukama E, Chilvers MI, Coker CM, Collins AA, Damicone JP, Dorrance AE, Dufault NS, et al (2017) Soybean yield loss estimates due to diseases in the United States and Ontario, Canada, from 2010 to 2014. *Plant Heal Prog* 18: 19–27

Altizer S, Ostfeld RS, Johnson PTJ, Kutz S, Harvell CD (2013) Climate Change and Infectious Diseases: From Evidence to a Predictive Framework. *Science* (80-) 341: 514–519

Baino OM, Salazar SM, Ramallo AC, Kirschbaum DS (2011) First Report of *Macrophomina phaseolina* Causing Strawberry Crown and Root Rot in Northwestern Argentina . *Plant Dis* 95: 1477–1477

Batista E, Lopes A, Alves A (2021) What do we know about Botryosphaeriaceae? An overview of a worldwide cured dataset. *Forests*. doi: 10.3390/f12030313

Bradley CA, Allen TW, Sisson AJ, Bergstrom GC, Bissonnette KM, Bond J, Byamukama E, Chilvers MI, Collins AA, Damicone JP, et al (2021) Soybean Yield Loss Estimates Due to Diseases in the United States and Ontario, Canada, from 2015 to 2019. *Plant Heal Prog* 22: 483–495

Burgess TI, Scott JK, McDougall KL, Stukely MJC, Crane C, Dunstan WA, Brigg F, Andjic V, White D, Rudman T, et al (2017) Current and projected global distribution of *Phytophthora cinnamomi*, one of the world's worst plant pathogens. *Glob Chang Biol* 23: 1661–1674

Capblancq T, Forester BR (2021) Redundancy analysis: A Swiss Army Knife for landscape genomics. *Methods Ecol Evol* 12: 2298–2309

Casano S, Cotan AH, Delgado MM, García-Tejero IF, Saavedra OG, Puig AA, de los Santos B (2018) First report of charcoal rot caused by *Macrophomina phaseolina* on hemp (*Cannabis sativa*) varieties cultivated in southern Spain. *Plant Dis* 102: 1665

Cohen R, Elkabetz M, Paris HS, Gur A, Dai N, Rabinovitz O, Freeman S (2022) Occurrence of *Macrophomina phaseolina* in Israel: Challenges for Disease Management and Crop Germplasm Enhancement. *Plant Dis* 106: 15–25

Croll D, McDonald BA (2016) The genetic basis of local adaptation for pathogenic fungi in agricultural ecosystems. *Mol Ecol*. doi: 10.1111/mec.13870

Dhingra OD, Sinclair JB (1974) Survival of *Macrophomina phaseolina* Sclerotia in Soil: Effects of Soil Moisture, Carbon:Nitrogen Ratios, Carbon Sources, and Nitrogen Concentrations. *Phytopathology* 65: 236–240

Dudney J, Willing CE, Das AJ, Latimer AM, Nesmith JCB, Battles JJ (2021) Nonlinear shifts in infectious rust disease due to climate change. *Nat Commun*. doi: 10.1038/s41467-021-25182-6

Elith J, Leathwick JR (2009) Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annu Rev Ecol Evol Syst* 40: 677–697

Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, Gurr SJ (2012) Emerging fungal threats to animal, plant and ecosystem health. *Nature* 484: 186–194

Franklin J (2013) Species distribution models in conservation biogeography: Developments and challenges. *Divers Distrib* 19: 1217–1223

Gaetán SA, Fernandez L, Madia M (2006) Occurrence of Charcoal Rot Caused by *Macrophomina phaseolina* on Canola in Argentina. *Plant Dis* 90: 524

Gerin D, Dongiovanni C, De Miccolis Angelini RM, Pollastro S, Faretra F (2018) First Report of *Macrophomina phaseolina* Causing Crown and Root Rot on Strawberry in Italy. *Plant Dis* 102: 1857

Hernández-Lambrano RE, González-Moreno P, Sánchez-Agudo JÁ (2018) Environmental factors associated with the spatial distribution of invasive plant pathogens in the Iberian Peninsula: The case of *Phytophthora cinnamomi* Rands. *For Ecol Manage* 419–420: 101–109

Hijmans RJ, Elith J (2017) Species Distribution Modeling with R. *Ref Modul Life Sci*. doi: 10.1016/b978-0-12-809633-8.02390-6

Hijmans RJ, Phillips S, Leathwick J, Maintainer JE (2017) Package “dismo” Species Distribution Modeling.

Hutton DG, Gomez AO, Mattner SW (2013) *Macrophomina phaseolina* and Its Association with Strawberry Crown Rot in Australia. *Int J Fruit Sci* 13: 149–155

Hyder S, Gondal AS, Ahmed R, Sahi ST, Rehman A, Hannan A (2018) First Report of Charcoal Rot in Tomato Caused by *Macrophomina phaseolina* (Tassi) Goid. *From Pakistan*. *Plant Dis* 102: 1459

Ireland KB, Kriticos DJ (2019) Why are plant pathogens under-represented in eco-climatic niche modelling? *Int J Pest Manag* 65: 207–216

Jay F, Manel S, Alvarez N, Durand EY, Thuiller W, Holderegger R, Taberlet P, François O (2012) Forecasting changes in population genetic structure of alpine plants in response to global warming. *Mol Ecol* 21: 2354–2368

Jordaan E, van der Waals JE, McLaren NW (2019a) Effect of irrigation on charcoal rot severity, yield loss and colonization of soybean and sunflower. *Crop Prot* 122: 63–69

Jordaan E, van der Waals JE, McLaren NW (2019b) Effect of irrigation on charcoal rot severity, yield loss and colonization of soybean and sunflower. *Crop Prot* 122: 63–69

Juroszek P, Von Tiedemann A (2013) Plant pathogens, insect pests and weeds in a changing global climate: A review of approaches, challenges, research gaps, key studies and concepts. *J Agric Sci* 151: 163–188

Kendig SR, Rupe JC, Scott HD (2000) Effect of irrigation and soil water stress on densities of *Macrophomina phaseolina* in soil and roots of two soybean cultivars. *Plant Dis* 84: 895–900

Koehler AM, Shew HD (2018) First Report of Charcoal Rot of Stevia Caused by *Macrophomina phaseolina* in North Carolina. *Plant Dis* 102: 241–241

Marquez N, Giachero ML, Declerck S, Ducasse DA (2021) *Macrophomina phaseolina*: General Characteristics of Pathogenicity and Methods of Control. *Front Plant Sci*. doi: 10.3389/fpls.2021.634397

- Meena BR, Nagendran K, Tripathi AN, Kumari S, Sagar V, Gupta N, Rai AB, Singh B (2018) First report of charcoal rot caused by *Macrophomina phaseolina* in *Basella alba* in India. Plant Dis 102: 1669
- Mengistu A, Arelli PA, Bond JP, Shannon GJ, Wrather AJ, Rupe JB, Chen P, Little CR, Canaday CH, Newman MA, et al (2011a) Evaluation of soybean genotypes for resistance to charcoal rot. Plant Heal Prog. doi: 10.1094/PHP-2010-0926-01-RS
- Mengistu A, Smith JR, Ray JD, Bellaloui N (2011b) Seasonal progress of charcoal rot and its impact on soybean productivity. Plant Dis 95: 1159–1166
- Meyer W., Sinclair JB (1974) Factors Affecting Charcoal Rot of Soybean Seedlings. Phytopathology 64: 845–849
- Nishad I, Srivastava AK, Saroj A, Babu BK, Samad A (2018) First Report of Root Rot of *Nepeta cataria* Caused by *Macrophomina phaseolina* in India. Plant Dis 102: 2380
- Nouri MT, Zhuang G, Culumber CM, Trouillas F (2018) First Report of *Macrophomina phaseolina* Causing Trunk and Cordon Canker Disease of Grapevine in the United States. Plant Dis PDIS-07-18-1189-PDN
- Paini DR, Sheppard AW, Cook DC, De Barro PJ, Worner SP, Thomas MB (2016) Global threat to agriculture from invasive species. Proc Natl Acad Sci 113: 7575–7579
- Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Modell 190: 231–259
- Poudel B, Shivas RG, Adorada DL, Barbetti MJ, Bithell SL, Kelly LA, Moore N, Sparks AH, Tan YP, Thomas G, et al (2021) Hidden diversity of *Macrophomina* associated with broadacre and horticultural crops in Australia. Eur J Plant Pathol 161: 1–23
- Reznikov S, Vellicce GR, Mengistu A, Arias RS, Gonzalez V, Lisi V De, Garcia MG, Rocha Carla ML, Pardo EM, Castagnaro AP, Ploper LD (2018) Disease incidence of charcoal rot (*Macrophomina phaseolina*) on soybean in north-western Argentina and genetic characteristics of the pathogen. Can J Plant Pathol 40: 423–433
- Roth MG, Webster RW, Mueller DS, Chilvers MI, Faske TR, Mathew FM, Bradley CA, Damicone JP, Kabbage M, Smith DL (2021) Integrated Management of Important Soybean Pathogens of the United States in Changing Climate. J Integr Pest Manag. doi: 10.1093/jipm/pmaa013
- Savary S, Willocquet L, Pethybridge SJ, Esker P, McRoberts N, Nelson A (2019) The global burden of pathogens and pests on major food crops. Nat Ecol Evol 3: 430–439
- Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. Nat Rev Genet 14: 807–820
- Savolainen O, Pyhäjärvi T, Knürr T, Pyhäjärvi T, Knürr T (2007) Gene Flow and Local Adaptation in Trees. Source Annu Rev Ecol Evol Syst 38: 595–619
- Sergeeva V, Tesoriero L, Spooner-Hart A R, Nair N (2005) First report of *Macrophomina phaseolina* on olives (*Olea europaea*) in Australia. Australas Plant Pathol 34: 273–274

Sexton ZF, Hughes TJ, Wise KA (2016) Analyzing isolate variability of *Macrophomina phaseolina* from a regional perspective. Crop Prot 81: 9–13

Shaw MW, Osborne TM (2011) Geographic distribution of plant pathogens in response to climate change. Plant Pathol 60: 31–43

Sparks AH, Forbes GA, Hijmans RJ, Garrett KA (2014) Climate change may have limited effect on global risk of potato late blight. Glob Chang Biol 20: 3621–3631

Steane DA, Potts BM, McLean E, Prober SM, Stock WD, Vaillancourt RE, Byrne M (2014) Genome-wide scans detect adaptation to aridity in a widespread forest tree species. Mol Ecol 23: 2500–2513

Sutherland WJ (2006) Predicting the ecological consequences of environmental change: A review of the methods. J Appl Ecol 43: 599–616

Tančić Živanov S, Dedić B, Dimitrijević A, Dušanić N, Mikić S, Jocić S, Miladinović D, Miklič V (2018) First Report of Charcoal Rot on Zebra Plant (*Aphelandra squarrosa*) Caused by *Macrophomina phaseolina*. Plant Dis 102: 2377

VanDerWal J, Shoo LP, Graham C, Williams SE (2009) Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? Ecol Modell 220: 589–594

Váry Z, Mullins E, Mcelwain JC, Doohan FM (2015) The severity of wheat diseases increases when plants and pathogens are acclimatized to elevated carbon dioxide. Glob Chang Biol 21: 2661–2669

Velásquez AC, Castroverde CDM, He SY (2018) Plant–Pathogen Warfare under Changing Climate Conditions. Curr Biol 28: R619–R634

Viejobueno J, Ramallo AC, Kirschbaum DS, Bains OM, Salazar SM (2017) Severe outbreaks of strawberry crown and root charcoal rot caused by *Macrophomina phaseolina* in Tucumán, Argentina. Rev agronómica del noroeste argentino 37: 111–114

Waldvogel A-M, Feldmeyer B, Rolshausen G, Exposito-Alonso M, Rellstab C, Kofler R, Mock T, Schmid K, Schmitt I, Bataillon T, et al (2020) Evolutionary genomics can improve prediction of species' responses to climate change. Evol Lett 4–18

Wang W, Liu Z, Wang W, Song X (2020) First Report of *Macrophomina phaseolina* Causing Stalk Rot of Sacha Inchi (*Plukenetia volubilis*) in China. Plant Dis 104: 570–570

Yang XB, Navi SS (2005) First Report of Charcoal Rot Epidemics Caused by *Macrophomina phaseolina* in Soybean in Iowa. Plant Dis 89: 526

Yonow T, Ramirez-Villegas J, Abadie C, Darnell RE, Ota N, Kriticos DJ (2019) Black Sigatoka in bananas: Ecoclimatic suitability and disease pressure assessments. PLoS One 14: e0220601

APPENDIX

SUPPLEMENTARY FIGURES

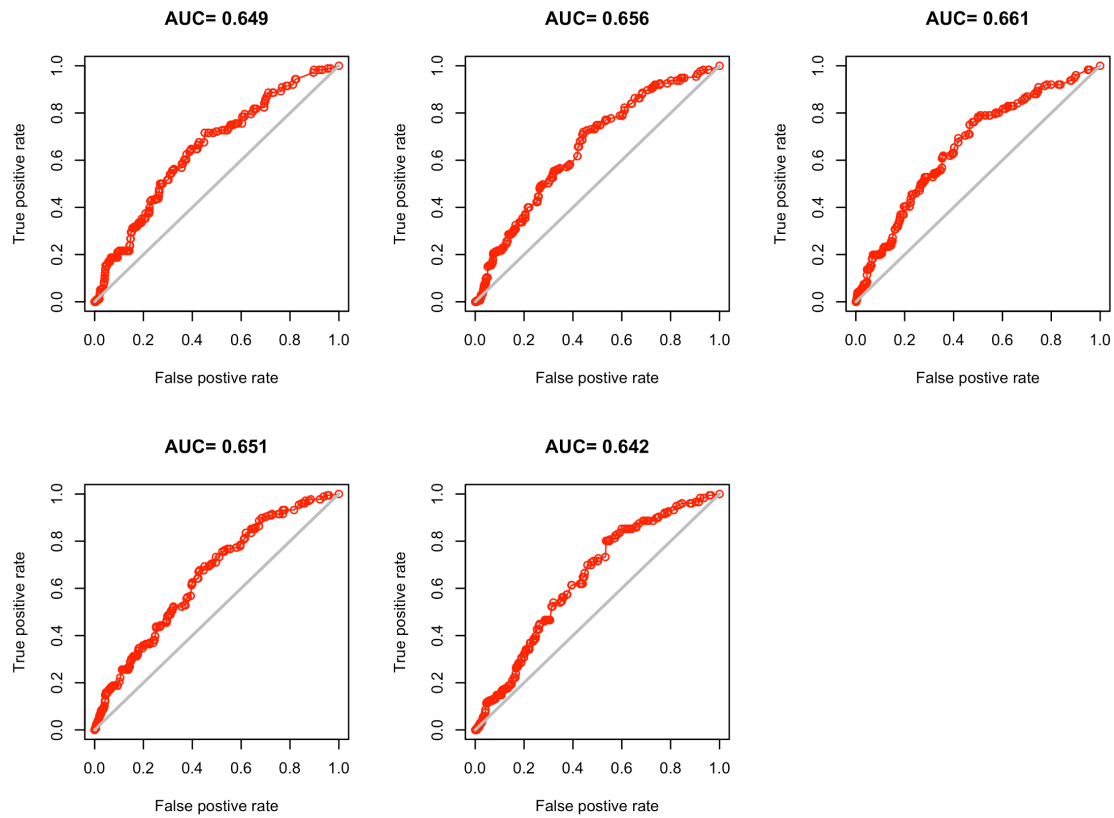


Figure A.1 Model performance AUC values for the BIOCLIM model. Area under the receiver operator curve (AUC) values for each of five cross-validations runs illustrating discrimination capacity of suitable versus unsuitable areas for *M. phaseolina*.

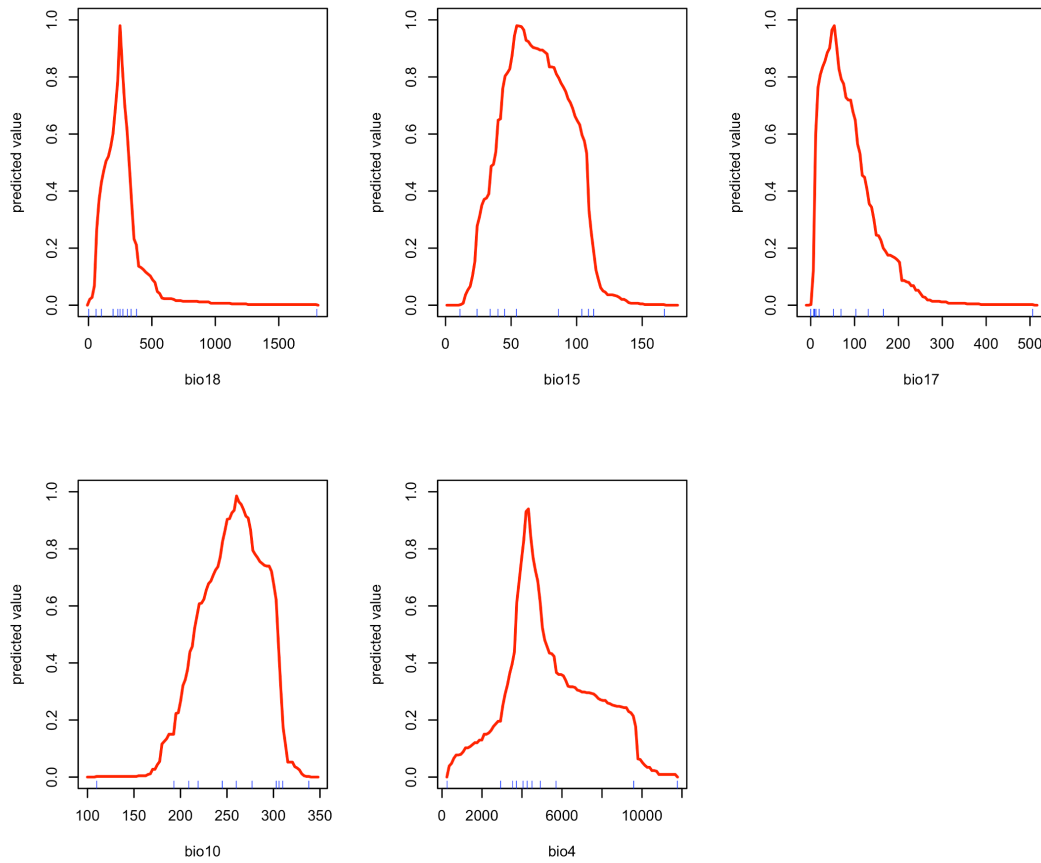


Figure A.2 Suitability values predicted for temperature and precipitation variables values found across locations of the complete data set of *M. phaseolina* occurrences. Predicted values are suitability values. A suitability value of 1 indicates a climatic value with predicted high suitability and a value of 0 is given for climatic values predicted as unsuitable. Climatic variables are BIO18 = Precipitation of Warmest Quarter, BIO15 = Precipitation Seasonality (Coefficient of Variation), BIO17 = Precipitation of Driest Quarter, BIO10 = Mean Temperature of Warmest Quarter and BIO4 = Temperature Seasonality (standard deviation * 100).

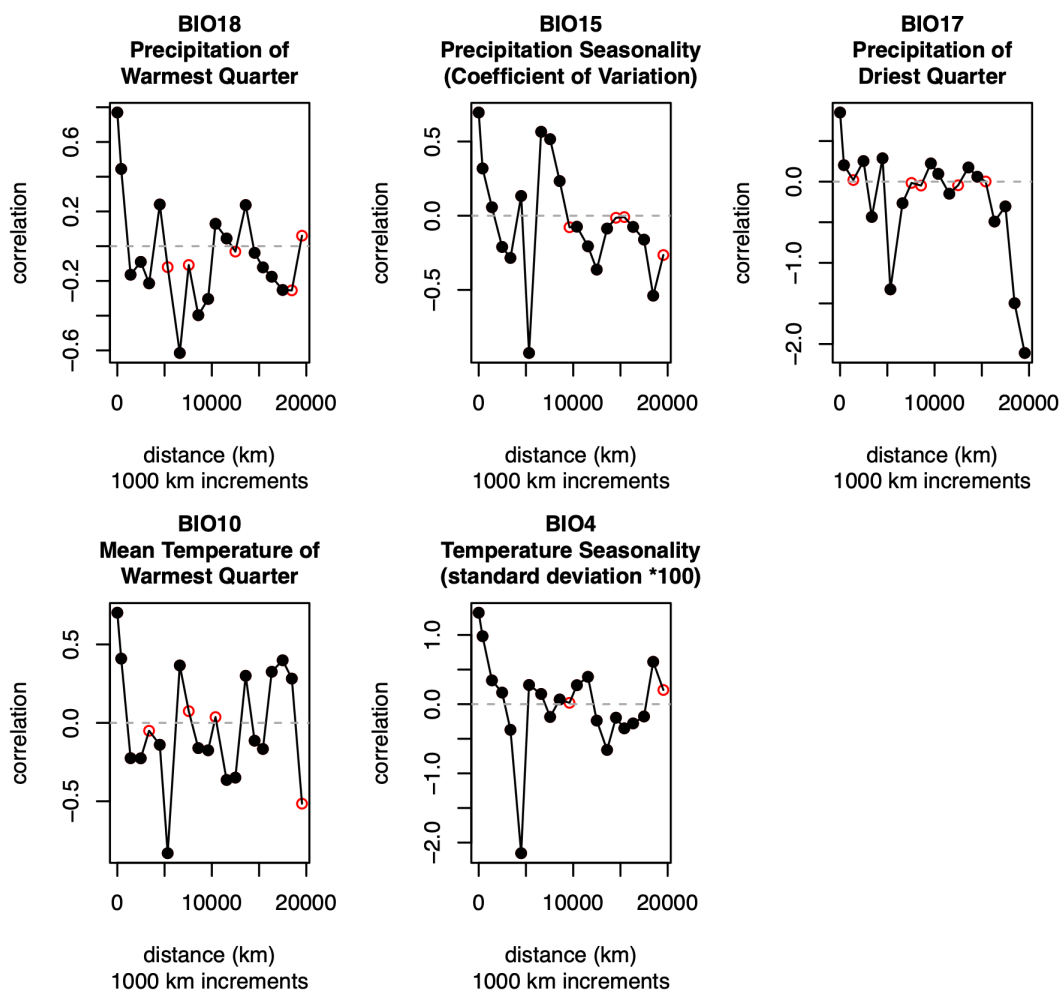


Figure A.3 Mantel's correlograms of climatic variables used in the BIOCLIM model. Correlation between climatic and geographic distances. Geographic distance classes are defined by 1000 km increments. Climatic variables are BIO18 = Precipitation of Warmest Quarter, BIO15 = Precipitation Seasonality (Coefficient of Variation), BIO17 = Precipitation of Driest Quarter, BIO10 = Mean Temperature of Warmest Quarter and BIO4 = Temperature Seasonality (standard deviation *100).

CHAPTER 5
CONCLUDING STATEMENT

Charcoal rot and diseases caused by *Macrophomina phaseolina* are a threat to agricultural production affecting many important economic and subsistence crops worldwide. Importantly, one of increasing concern under climate change. This research focused on understanding the genetic diversity and evolutionary potential of *M. phaseolina*, to inform and provide tools for improved charcoal rot management strategies.

Populations of *M. phaseolina* in the continental US, Puerto Rico and Colombia collected from soybean and dry bean fields were found to be structured in a hierarchical manner with subcontinental regional stability and instability at local scales consistent with a metapopulation dynamics perspective. These results are in line with a scenario of evolution after migration driven by divergence following clonal expansions. Additionally, this research identified the potential for anthropogenic influence in the movement of *M. phaseolina* to locations around the world. Climate was found to significantly contribute to genetic divergence in this pathogen and identified candidate genomic regions for adaptation. Putatively adaptive functions associated to these regions may benefit *M. phaseolina* in specific environments. This knowledge expands the impact that population genomics and genotype-environment associations can have on our ability to characterize adaptive potential in plant pathogens.

Effective chemical-control means are lacking for the management of charcoal rot. Therefore, the efficacy of active ingredients currently used in commercial fungicide formulations in crop production was investigated. Our results on the *in-vitro* efficacy of boscalid, iprodione and prothioconazole indicate that formulations with these active ingredients, may reduce *M. phaseolina* seedling infection originating from infected seeds or inoculum in the soil. Particularly, our results on the *in-vitro* efficacy of prothioconazole suggest that commercial formulations with this active ingredient may be of particular interest for charcoal rot management. Information regarding mutations in fungicide target genes was lacking for *M. phaseolina*. None of the point mutations found in our isolate collection were correlated with levels of fungicide sensitivity. Finally, in this study we developed a bioclimatic model for *M. phaseolina* to project the climate suitability of *M. phaseolina* at a global scale and identify localities at risk of charcoal rot and other diseases caused by this pathogen. The model projected high suitability for localities with low precipitation of driest quarter and precipitation of warmest quarter, and high mean temperature of warmest quarter.

Notably, areas of high suitability were projected in the southern US, north-eastern Argentina, eastern Australia, and southern Europe. These predictions correspond to an expected distribution of higher *M. phaseolina* suitability at warm and dry regions and with increased disease reports in these regions.

5.0.1 Future directions

Future studies investigating the adaptive potential of *M. phaseolina* will be needed to identify the degree to which global populations reflect their adaptation to host and climate. Such studies will benefit from comprehensive samplings schemes including diverse hosts and climates. In addition, long-read sequencing technologies will allow further characterization of the role of genomic variation, including structural variation, in *M. phaseolina* adaptation to host and the climatic environment.

Our data on the *in-vitro* efficacy of prothioconazole suggest that commercial formulations with this active ingredient may be of particular interest for future *in-vivo* efficacy testing in soybean and dry bean. Data on the *in-vivo* efficacy of prothioconazole in preventing seedling colonization and charcoal rot disease development is needed in order to determine its effectiveness in charcoal rot control. Additional charcoal rot management efforts should be directed at identifying novel effective fungicides and monitoring the potential development of resistance to fungicides use in modern crop production.

Given the increasing impact of *M. phaseolina* on agroecosystems globally, the modelling of its distribution constitutes an important tool for the monitoring and development of management strategies. More comprehensive predictive species distribution models, including ensemble models, should provide a better understanding of the adaptive response of this fungal pathogen under climate change. Further improvements of the model presented in this research, will involve the use of larger data sets and semi-mechanistic models. Similarly, regional distribution models would provide a better assessment of charcoal rot risk for major crop production regions. A major need remains to incorporate disease risk assessments and eco-evolutionary projections into charcoal rot management strategies. The characterization of adaptation in plant pathogens enabled by population genomics should become increasingly utilized for plant disease risk prediction models especially under climate change.