Winter 12-15-2019

Phylogenetic and Genomic Characterization of the Host-Pathogen Arms Race Between Bacterial Pathogens and Gossypium hirsutum

Anne Zimmerman Phillips
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences
Plant and Microbial Biosciences

Dissertation Examination Committee:
Rebecca Bart, Chair
Joshua Blodgett
Barbara Kunkel
Kenneth Olsen
Lucia Strader

Phylogenetic and Genomic Characterization of the Host-Pathogen Arms Race Between Bacterial Pathogens and *Gossypium hirsutum*

by

Anne Z. Phillips

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

December 2019
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

I am grateful to Washington University, the Donald Danforth Plant Science Center, the National Science Foundation, the William H. Danforth Plant Science Fellowship, and Cotton Inc. for funding and supporting my thesis research.

I would like to thank all of the scientific mentors without whom this thesis would not have been possible. I would like to thank Dr. Terry Wheeler for her insight into the world of cotton and for showing me how important collaboration is. I would like to thank Dr. Jim Carrington for instilling in me a passion for scientific communication and for his career advice. I would also like to thank my committee: Dr. Joshua Blodgett, Dr. Barbara Kunkel, Dr. Kenneth Olsen, and Dr. Lucia Strader for their dedication towards my thesis. Finally, I would like to express my sincere gratitude to Dr. Rebecca Bart for her guidance and mentorship throughout my PhD journey. Dr. Bart enabled me to explore my scientific interests both inside and outside of the lab, which has allowed me to grow as a scientist and find an exciting career path. I hope to emulate these mentorship skills throughout my career.

I also could not have completed this thesis without the friendship of my classmates and lab mates. Thank you Erica and Yunci for your Friday night cathartic dinners and Kira and Maggie for your advice and friendship. I of course would also like to thank Mark for his love, support, and humor throughout my thesis and thesis writing process.

Finally, I would also like to thank my parents and my sister for their unwavering love and support, even from afar. I am here because you encouraged my love of learning.

Anne Z. Phillips

Washington University in St. Louis

December 2019

Dedicated to my parents.

ABSTRACT OF THE DISSERTATION

Phylogenetic and Genomic Characterization of the Host-Pathogen Arms Race Between Bacterial Pathogens and *Gossypium hirsutum*

by

Anne Z. Phillips

Doctor of Philosophy in Biology and Biomedical Sciences

Plant and Microbial Biosciences

Washington University in St. Louis, 2019

Professor Rebecca Bart, Chair

Hosts and pathogens are eternally intertwined in an evolutionary arms race. When a pathogen causes a disease outbreak, scientists must identify resistance strategies that can durably tilt the arms race in favor of the host. This requires a deep understanding of both the genetic and environmental contexts in which the outbreak occurs. In this thesis I investigate the bacterial pathogens *Xanthomonas citri* pv. *malvacearum* (*Xcm*) and *Pseudomonas syringae* that caused disease outbreaks on *Gossypium hirsutum* from 2011-2017. I use pathogen genomics and host transcriptomics to develop hypotheses for how these pathogens emerged and how they cause disease. Phylogenetics and virulence factor analysis reveal few differences between contemporary and historical isolates. Data on agricultural practices point to changing germplasm dynamics as a reason for disease re-emergence. These data led to a RNA-Seq experiment that identified several new candidate susceptibility genes including four SWEET sugar transporters and two Mildew Locus-O homologs that have been used to confer resistance to bacterial pathogens in other systems. As an alternative approach, a diversity panel of 52 cotton varieties was screened to reveal seven sources of resistance that can be used to prevent *Xcm* outbreaks in

the future. While no resistance strategies have been identified for *P. syringae*, several virulence factors such as the type IV pilus and filamentous hemagglutinin were identified, which show evolutionary markers of a function in the host-pathogen arms race. These data are pivotal for determining the best strategies for developing durable host resistance strategies. Future work will focus on developing resistant varieties of cotton and performing reverse genetics to confirm the identities of virulence factors that contribute to pathogen fitness.

Chapter 1: The Host-Pathogen Arms Race

1.1 Abstract

Hosts and pathogens are continuously evolving new attack and defense strategies. This type of coevolution, termed the host-pathogen arms race, has been a topic of research for decades. Pathogen virulence factors and host resistance strategies, as well as the evolutionary mechanisms that influence them, determine who gains the upper hand. Furthermore, when the environment is factored in to this equation, hypotheses can be made for how disease outbreaks occur and strategies can be established for developing durable resistance. Here I describe the status of the field of the host-pathogen arms race in the context of the plant pathogens *Xanthomonas* and *Pseudomonas*. Both bacterial pathogen genomes encode many virulence factors that trigger both pattern-triggered immunity and effector-triggered immunity in host plants, making them ideal models for investigating the host-pathogen arms race. The data presented here inform the scientific questions and experimental designs that are applied in this thesis where I investigate the arms races between the bacterial pathogens *Xanthomonas citri* pv. *malvacearum* (*Xcm*) and *Pseudomonas syringae* and the host cotton (*Gossypium hirsutum*).

1.2 Introduction

When disease outbreaks occur in an agricultural environment, scientists and farmers work together to tilt the host-pathogen arms race in favor of the host. This requires understanding both the genetic mechanisms underlying the host-pathogen interaction as well as the environmental context in which it is occurring. In my thesis I will provide evidence for how each of the three sides of the disease triangle (the pathogen, host, and environment) contribute to disease by investigating recent bacterial disease outbreaks on cotton in the US. Here I describe the current

knowledge of 1) Pathogen virulence mechanisms and how they evolve, 2) Host resistance strategies and how they are developed, and 3) How the environment can help or hurt host-pathogen interactions.

Hosts and pathogens are continuously intertwined in a co-evolutionary arms race[1,2]. This arms race begins with pattern triggered immunity (PTI), the initial recognition of pathogens by host extracellular receptors [3]. Host extracellular receptors evolve to recognize conserved pathogen molecular patterns and the pathogen evolves to evade this recognition [4]. If the pathogen successfully evolves around this line of defense and is able to inject effectors into a host cell, it will then encounter the second line of defense: effector triggered immunity (ETI) [5,6]. Similar to PTI, in ETI, host intercellular receptors evolve to recognize pathogen effectors, which then evolve around this recognition in order to once again trigger disease [7,8]. While many recent studies cast doubt on the explicit delineation between PTI and ETI, it is nearly universally agreed upon that both levels of the host-pathogen arms race lead to reciprocal selective pressure [9–11]. The cyclical, continuous nature of this interaction has been named the Red Queen Hypothesis in which "it takes all the running you can do to keep in the same place" [12,13].

This model of pathogen-host interactions is termed the "Zig-Zag" model [3]. It emphasizes the back-and-forth interactions between hosts and pathogens as well as how ETI can trigger a stronger resistance response than PTI. However, PTI and ETI are not always strictly separated. For example, PTI can sometimes trigger a hypersensitive response (HR) and some effectors can be recognized in the apoplastic space rather than in the cytoplasm [14,15]. Recent articles have attempted to develop a new model called the "Spatial Invasion" model [9,10]. This model takes into consideration the complexities and exceptions to the PTI/ETI delineation. It also groups together all bacterial proteins recognized by the host immune system as invasion patterns and all host

receptors that recognize these patterns as invasion pattern receptors. Both models work on the underlying principle that there is a constant arms race happening between host and pathogen and whoever has the better arsenal "wins".

Understanding this battle and the elements that determine the winner are pivotal for developing durable resistance strategies. The evolution of new pathogen virulence factors occurs through three main mechanisms: horizontal gene transfer (HGT), recombination, and pathoadaptation. Horizontal gene transfer is the mechanism by which genetic material moves from one organism to another through transformation, transduction through phages, or conjugation [16]. Recombination involves exchange of genetic material between organisms or within the same organism, causing pathogenicity island movement, multiplication, or excision [17]. Pathoadaptation, derived from the term pathogenicity-adaptive, represents the small single nucleotide polymorphism (SNP) and insertion and deletion (INDEL) mutations acquired vertically over time that enhance bacterial virulence [18,19]. On the host side, the same mechanisms are used to generate new resistance genes and shut down susceptibility genes, but to a lesser extent. For example, host nucleotide-binding and leucine-rich repeat (NLR) resistance genes frequently undergo recombination and smaller SNP/INDEL changes, but horizontal gene transfer is less frequent [20,21] .

In nature disease is the exception not the rule. Disease requires a virulent pathogen, a susceptible host, and a conducive environment. This phenomenon, called the disease triangle, is the framework around which I investigate the mechanisms of disease emergence and contribute to the development of durable resistance strategies.

1.3 Players in the plant-pathogen arms race and how to pick the winning side

1.3.1 Plant pathogens and their virulence mechanisms: *Xanthomonas and Pseudomonas*

Xanthomonas and *Pseudomonas* plant pathogens

In order to develop durable resistance, you first need to know the identity of the pathogen and the conserved virulence factors being deployed. Two of the most widespread and widely studied plant pathogen genera are *Xanthomonas* and *Pseudomonas*. Both genera consist of gram negative rod-shaped Gammaproteobacteria that infect a wide range of plants that span from monocots such as rice and wheat to dicots such as kale, tomatoes, and beans [22,23]. Many species also survive as environmental isolates in the water system and as plant epiphytes [23–25]. Others survive as opportunistic human pathogens [26,27]. The plasticity of niches that these pathogens inhabit makes them particularly difficult to eradicate from an environment.

Traditionally, *Pseudomonas* and *Xanthomonas* pathovars are characterized using physical characteristics tests, pathogenicity tests on suspected hosts, and hypersensitive response tests on tobacco [28]. However, modern sequencing techniques have allowed for *Pseudomonas* and *Xanthomonas* pathogens to be divided into phylogroups based on Multi Locus Sequence Typing analysis [29–33]. This method uses concatenated regions of neutrally evolving housekeeping genes to develop phylogenetic trees, which allows for faster, more accurate identification of pathogens.

Pathovar host ranges are generally reported to be narrow, limited to one species or a subset of varieties within a species [22,34]. Exceptions include *P. syringae* pathovars in phylogroup 2 that infect hosts ranging from monocots to dicots and *X. campestris* pv. *campestris* isolates that infect the full range of brassica species [35,36]. Understanding the host range of a pathogen can aid

in the identification of pathogen reservoirs and ultimately aid in the development of the best host resistance strategies.

Due to the ubiquity of these pathogens and their importance in agriculture, there is a great depth of knowledge of their mechanisms of virulence. However, there are still many gaps in knowledge centered around how pathogens of non-model hosts evolve and how new outbreaks occur.

Virulence factors

The potential of a pathogen to cause disease is determined by its genetically encoded virulence factors. Virulence factors are the weapons that pathogens deploy in the host-pathogen arms race. Understanding the identities of these weapons as well as the diversity of weapons used across the pathogen population helps scientists decide which resistance strategies to use to fight back. Xanthomonads and Pseudomonads deploy many different tools that range from small systems such as ice nucleators to large biosynthetic pathways that produce phytotoxins and hormone mimics to massive, multi-part secretion systems that inject multiple effectors into host cells. In most bacterial pathogens, virulence factors are organized in pathogenicity islands. Here I use the definition of pathogenicity island described by Hacker et al. as a genomic region that: i) Contains many virulence genes, ii) Is present in pathogenic strains and less common in less-pathogenic strains, iii) Has a different GC content than the surrounding genome, iv) Occupies a large genomic region v) Has compact genes and is flanked by direct repeats, vi) Has tRNA genes and/or insertion sequence elements at its boundaries, vii) Is present near mobile genes such as integrases or transposases, and viii) Is not stable[48]. These characteristics make it easy to identify which virulence factors are in each genome and allow us to hypothesize the evolutionary history of the virulence island.

In plant-pathogen interactions, and particularly in *Xanthomonas/Pseudomonas* disease interactions, the most well studied virulence mechanism is the type three secretion system (T3SS) [37,38]. This system consists of a needle-like structure that can pierce through the host cell membrane. The needle delivers type three effectors (T3E), which promote pathogen growth and disease progression in many ways. The ability of these pathogens to cause disease on plants is directly related to the activity of the HR and pathogenicity (hrp) and HR and conserved (hrc) genes that encode the type three secretion system and its effectors. These systems are highly conserved. T3SS are present in most if not all gram negative plant pathogens such as *Pseudomonas, Xanthomonas, Erwinia,* and *Ralstonia* as well as animal pathogens such as *Pseudomonas aeruginosa*, *Salmonella, Yersinia*, *Shigella* [37,39]. Interestingly, the beneficial plant microbes Rhizobia and *Pseudomonas fluorescens* also have T3SS. Therefore, the presence of a T3SS does not automatically deem a bacterium a pathogen [40–42].

The mechanisms regulating the expression of the T3SS determine the behavior of the microbes in the presence of a host. For many plant pathogens, quorum sensing, as well as environmental stimuli such as pH, carbon source, micronutrients availability, osmolarity, and temperature work together to coordinate the expression of the T3SS [5]. This allows for a direct and organized attack against the host. In *Pseudomonas syringae,* T3Es are up regulated in unison by HrpL, an alternative sigma factor [43–45]. In other pathogens such as Xanthomonads, the transcription factors that up regulate T3SS components are more varied, but they are often triggered by the same environmental clues [46–48].

Within pathogens, the number and identity of T3Es varies greatly, even within the same species. The model pathogen *Pseudomonas syringae* pv. *tomato* DC3000 has approximately 30 T3Es, while one strain of *Pseudomonas syringae* pv. *japonica* only encodes nine [49].

Xanthomonads are also well studied for their use of a wide range of effectors that target host systems such as sugar export, apoplast ion balance, the regulation of PTI and ETI, and the cytoskeletal network[23,37,50–52]. Many of these effector families are shared with Pseudomonads and other bacterial pathogens, illustrating their importance in pathogen virulence [53]. However, knocking out individual effectors rarely affects the disease phenotype. This indicates that many effectors may have functional redundancy or work together to generate a phenotype [54–57]. Alternatively, some effectors may only be needed during certain environmental conditions or hosts [34]. For example, the human pathogen *Salmonella enterica* encodes two functioning T3SSs with corresponding sets of effectors [8]. The *Salmonella* Pathogenicity Island 1 (SPI-1) T3SS and effectors are deployed during the initial invasion of intestinal epithelial cells [58,59]. Then, once the bacteria are engulfed by macrophages, the second T3SS (SPI-2) is induced to promote survival within the macrophage [60]. Therefore, understanding the identity and regulation of virulence factors are important for combatting multi-leveled arms races.

In addition to the T3SS, Pseudomonads and Xanthomonads encode a variety of tools to enhance disease, including the production of phytohormones, extracellular polysaccharides, toxins, cell wall degrading enzymes, and effectors secreted by additional secretion systems. Therefore, while the T3SS and effectors often determine pathogen host range, other virulence factors can also contribute to pathogenicity and interactions with the environment. This information is pivotal for understanding how the pathogen fights back in the host-pathogen evolutionary arms race.

Pathogen evolution

Understanding the identity of virulence factors present in a pathogen helps breeders to identify which resistance strategies to deploy. However, this information is useless unless they

understand the variability in the pathogen population and how virulence factors evolve. Otherwise they run the risk of deploying a resistance strategy that can be rapidly overcome by pathogen evolution. For example, in may systems disrupting the interaction between a *Xanthomonas* Transcription Activator-Like effector and target host susceptibility gene confer resistance to the host [50,61]. However, further analysis revealed that field isolates of *Xanthomonas* deploy a suite of rapidly evolving TAL effectors that target many different susceptibility genes [62,63]. Therefore, disrupting one effector-host gene interaction would not confer resistance to the many *Xanthomonas* effectors present in the pathogen population. In this section I illustrate how virulence factors evolve and the consequences for host immunity.

Rapid gain in pathogenicity through HGT of a pathogenicity island

One hypothesis for how novel pathogens evolve is through the acquisition of the T3SS and effectors by non-pathogenic strains through a large, single horizontal gene transfer event. This hypothesis has been bolstered by evidence such as the divergent GC content and evolutionary relationships of T3SSs that indicate horizontal gene transfer, as well as their frequent positioning on transmissible plasmids [41]. However, at least one group has found that single HGT events may not be enough for pathogenicity to develop, and in fact they may be counter-productive [25].

Mobile effectors

Smaller, single gene presence/absence changes in virulence repertoires are more common and can have major effects on bacterial pathogenicity. The mechanisms of these effector profile changes are diverse. When a bacterium gains an effector from horizontal gene transfer, it can rapidly expand its host range and result in devastating crop losses [64–66]. This is because the

effector has already evolved to trigger disease and win the arms race against the host. Therefore, presence of the effector in a new pathogen can instantly confer pathogenicity to this host.

Novel effectors are also created through within-genome rearrangements. Due to the modular nature of T3Es (N-terminal T3SS domain and C-terminal functional domains), new effectors can be generated through the shuffling of effector domains [67]. This process, called terminal reassortment, is supported by the distinct phylogenetic lineages of many effector domains as well as the ability of some effectors to be post-translationally processed into two distinct proteins with different functions [67]. For example, SipA, a T3E of *Salmonella enterica*, activates a caspase that cleaves SipA [68]. This cleavage of SipA releases the C-terminal domain that aids in bacterial entry into epithelial cells and a N-terminal domain that extracellularly induces disease progression, both of which are necessary for pathogenicity [68].

Horizontal gene transfer and recombination are catalyzed by the positioning of virulence factors within prophage sequences and next to mobile elements [69–72]. These mobile pathogenicity islands can influence the pathogenicity of the bacterium, its race, and even its host range [64,71]. Therefore, when assessing a pathogenic population, horizontal gene transfers and recombinations are investigated as potential markers of virulence factor evolution. These data have led to the practice of pyramiding resistance genes in host germplasm and within a field [73,74]. This often makes germplasm durable against the target pathogen for many more years. However, this custom is not practiced in all host-pathogen systems [73].

Selective pressure/within gene evolution

Evidence of the host-pathogen arms race is not only seen with effector presence/absence changes or effector rearrangements, but also in smaller effector sequence variations. These types of changes rarely generate a pathogenicity change as quickly as horizontal gene transfer or

recombination because they are limited by the rate of random mutation within the host. However, they play an important role in the ability of a pathogen to evade detection in the ever continuing arms race [75]. This evolutionary process, deemed pathoadaptation, is caused by small changes (SNPs and INDELs) generated through mutation in vertically inherited effectors [19]. The location, frequency, and types of these changes act as signatures of selection caused by the host-pathogen arms race.

Selective pressure from the arms race can be steep for T3Es due to their often direct interactions with host components. The T3SS is often under functional constraint, leading to purifying selection and low nucleotide diversity [76,77]. However, the effectors and the parts of the T3SS that physically interact with the host are often under positive selection and diversifying selection due to their intimate interactions with host defenses [7,76,77]. Positive selection is defined by an advantageous allele sweeping a population. This is represented by an increase in the number of nonsynonymous mutations compared to synonymous mutations at a location. In the plant-pathogen arms race, positive selection often occurs when an allele or set of alleles is being selected for due to its benefits to pathogen fitness [7,18,78]. On the other hand, diversifying selection is defined by an increase in the accumulation of mutations at a region of interest. This can indicate that the presence of multiple alleles in a population may confer evolutionary fitness over a single allele. In this case, diversifying selection is often a marker of the arms race itself because multiple alleles make it difficult for the host to evolve pathogen recognition and trigger a defense response [49,76,77]. Both selective pressures are found within the T3SS as well as other virulence factors that directly interact with host proteins. The outcome of this selection depends on the balance between the selective pressure against retention of the virulence factor due to host

recognition and the selective pressure for retention of the virulence factor due to its contribution to pathogen survival.

Each mechanism of pathogen evolution has the potential to tilt the arms race in favor of the pathogen, causing a disease outbreak. While this evolutionary arms race cannot be stopped, understanding which virulence factors evolve quickly can inform which resistance strategies to choose for a pathosystem.

1.3.2 Using the arms race to our advantage: The host side of the disease triangle

Host susceptibility factors

Many mechanisms that effectors use to induce susceptibility can be leveraged against the pathogen. For example, susceptibility genes are host genes that are up regulated by pathogens during disease and confer fitness to the pathogen. Once identified, the interaction between pathogen and susceptibility gene can be disrupted in order to tilt the host-pathogen arms race back in favor of the host.

For example, the promoters of host SWEET sugar transporter genes are bound to and up regulated by *Xanthomonas* Transcription Activator-Like (TAL) effectors in both rice and cassava [50,61]. When these interactions are disrupted, disease symptoms are greatly diminished. However, in order for this to be a successful strategy in the field, the plants still have to produce as much yield as their susceptible counterparts. Unfortunately, there are often adverse effects on plant growth and flowering when sugar transporters are knocked out [79–82]. Therefore, many groups are now working on strategies to specifically interrupt the binding of the TAL effectors to susceptibility gene promoters using CRISPR [83]. This would allow for the native expression of the susceptibility gene to remain intact while making it inaccessible to the pathogen.

Another canonical susceptibility gene is the Mildew Locus O (MLO) gene. This gene was first identified in 1942 when a barley X-ray mutagenesis screen identified a mutant with increased resistance to powdery mildew, caused by the fungus *Blumeria graminis* f. sp. *hordei*[84]. While the biochemical mechanism of the MLO protein is still unknown, the broad-spectrum resistance is tied to the inhibition of pathogen penetration into host cells [85,86]. Since its discovery, mutations in this gene family have been found to confer resistance to many diseases, including *Xanthomonas* [85,87]. However, like SWEET genes, many mutants show decreased vigor and therefore cannot be used as resistance strategies in the field. Further research is required to understand why MLO-mediated disease resistance is so powerful and why some species show pleiotropic effects and others do not.

Both of these families of susceptibility genes can be mutated to confer broad-spectrum resistance against pathogens. Therefore they are ideal candidates for the development of new resistant varieties as long as two conditions are met: 1) Little to no pleiotropic effects are found and 2) These genes are targeted by all known isolates of a pathogen.

Host resistance strategies

The most common strategy for preventing disease progression in a field is the deployment of resistance genes. In contrast to susceptibility genes, many canonical resistance genes confer evolutionarily weak resistance against only certain races of a pathogen [88,89]. In fact, resistance genes categorize pathogen isolates into races according to which resistance genes an isolate can or cannot overcome [90]. However, despite their lack of durability, these canonical resistance genes are widely used throughout many agricultural systems due to the strength of the resistance incurred [89].

For example, at least 20 resistance loci have been identified in *Gossypium hirsutum* (upland cotton) [91,92]. In cotton and in other species, pyramiding resistance genes can make resistances strategies more broad and durable [93]. However, this does not always cause the resistance response to be stronger [93]. Despite progress in both the public and private sector, none of the causal genes have been cloned [94]. This is likely due to the large, tetraploid nature of the genome as well as the low genetic diversity among cultivars.

On a molecular level, these resistance genes are triggered via recognition of the T3SS and effectors. In a gene-for-gene interaction, a host protein can either directly or indirectly recognize the presence of a pathogen protein inside of the cell and trigger a rapid cell death (hypersensitive response)[90]. This restricts the movement of the pathogen and its ability to multiply in the host [95]. In nature this response is microscopic. However, large-scale infiltrations allow this response to be viewed with the naked eye. This is often the cause of host-range restrictions that separate pathovars (strains that cause disease on different host species) as well as races (strains that cause disease on different panels of host cultivars within one species)[90].

A classic example that illustrates how interconnected gene-for-gene interactions can be is the interaction between *Arabidopsis thaliana* and the *P. syringae* AvrRpt2 effector. The effector AvrRpm1 triggers a hypersensitive response when its modification of host protein RIN4 is recognized by resistance gene RPM1 [96]. However, many *P. syringae* strains have evolved around this resistance with the gain of effector AvrRpt2, which blocks the hypersensitive response and promotes disease [9796]. This kind of complex interaction demonstrates how the host-pathogen arms race is constantly in motion and can select for multi-layered virulence strategies.

Elucidating the identities of resistance genes and their genomic context will aid breeding practices and gene editing techniques. While the interactions between host resistance genes and

pathogen virulence genes may be complicated and prone to breaking down, they are the most common form of resistance in germplasm and are durable when multiple genes are pyramided together.

1.3.3 Leveraging the third side of the disease triangle (The environment) to reduce disease

The third side of the disease triangle, the environment, is a complex variable in agricultural ecosystems. It encompasses not only the weather but also the nearby flora and fauna, microbial community, and us! Farming practices, seed movement, and irrigation all contribute to the equation of whether or not disease will occur (and who wins the arms race).

The epidemiology of plant pathogens is still a burgeoning field that must synthesize genetic, supply chain, weather, and farming information from across the globe [98,99]. Current outbreaks of wheat blast disease and bacterial canker of tomato have been linked to the dissemination of contaminated seed [100–102]. While this form of disease transmission can be restricted through the methods of seed lot testing and regulation of seed movement, restriction of other mechanisms of transmission such as wind and insect dispersal are less feasible [103,104]. As Maraite et al explain, it is extremely important to understand the source of disease outbreaks because only then can you form effective control practices. For instance, if disease on sugar beet, wheat, and other crops was caused by the same ubiquitous bacteria, it would minimize the importance of seed sequestration and sterilization and emphasize the importance of proper crop rotation [105]. In addition, the revelation that many *Xanthomonas* and *Pseudomonas* strains thrive in water systems has led to changes in irrigation practices [36,106,107]

The agricultural practice of applying fungicides and nematocides to crops has greatly improved yields in many agricultural systems [108]. However, concerns over fungicide resistance and environmental health and safety are growing [109]. In addition, there are very few chemical

applications that have been approved to control bacterial disease. The only known EPA approved antibiotics for agricultural use were recently approved for controlling huanglongbing, the destructive bacterial disease that has been decimating orange groves in Florida. However, there is very little scientific evidence of the safety, efficacy, or environmental effects of this treatment [110]. There is some evidence that agricultural practices such as acid delinting cotton seeds may help control the spread of seed borne *Xcm* [111]. However, other groups suggest that pathogen overwintering on weeds may also contribute to *Xcm* outbreaks [112]. Therefore, while many farming practices can prevent the spread of bacterial pathogens and limit its severity, genetic resistance is the most durable and commonly used mechanism for preventing outbreaks.

1.4 Model diseases investigated in this thesis

1.4.1 Cotton and its diseases

Cotton is the world's largest fiber crop. It is grown in 171 countries and territories and generated nearly $59 billion dollars in exports in 2018 [113]. The US is the world's second largest exporter, responsible for $8.4 billion dollars in exports on its own.

Many different pathogens infect cotton, ranging from fungi such as *Fusarium oxysporum* and *Verticillium dahliae* to viruses such as cotton blue disease (Polerovirus) and leaf curl virus disease (Begomovirus) to the bacterial blight pathogen *Xanthomonas citri* pv. *malvacearum* (*Xcm*). These diseases equated to approximately 900 million dollars in losses in 2017 [114]. Mechanisms of fighting these diseases include monetarily and environmentally expensive fungicide treatments as well as crop rotation, equipment sterilization, and tailored irrigation techniques such as drip irrigation [106,115]. However, the most effective mechanism for fighting against disease is breeding for resistance.

1.4.2 *Xanthomonas* and *Pseudomonas* diseases of cotton

Cotton bacterial blight, caused by *Xanthomonas citri* pv. *malvacearum* (*Xcm*), was first identified in the US over 100 years ago [116]. Since then, multiple sources of resistance have been identified [91]. Cotton bacterial blight often starts as angular water soaking spots on leaves, bordered by veins [117]. It then can become a systemic disease, spreading to the petioles and stems through the vasculature, causing black arm rot [117]. It triggers measurable yield losses when lodging (breaking), cotton boll rot, and/or overall fitness reduction occurs. Once a field becomes infected with this pathogen, there is little to nothing that a farmer can do to protect the crop from yield loss. This emphasizes the importance of 1) breeding known resistance loci into farmer-preferred varieties and 2) identifying novel resistance strategies. In the US, genetic resistance was effective against all present races of the pathogen for decades [91]. However, beginning in 2011, extension scientists and farmers began identifying this disease in their fields again [63]. It was initially unclear whether this new outbreak was triggered by a change in the host, pathogen, or environmental side of the disease triangle.

In 2016, pathogenicity screens of bacteria isolated from cotton resulted in the identification of an isolate that produced symptoms dramatically different from *Xcm.* Koch's postulates and MLST analysis identified this pathogen as *Pseudomonas syringae* [118]. Unlike *Xcm,* almost nothing is known about this pathogen. Reports of a *P. syringae* pathogen of cotton are sparse and until recently only published in bulletins and conference proceedings [118–120]. It was initially unclear whether this new *Pseudomonas* outbreak stemmed from historical outbreaks or another source.

1.5 Chapter summary, significance, and scope

The host-pathogen arms race is a complex and multi-faceted system. In order to tilt the arms race in favor of the host, scientists must piece together 1) Which pathogen virulence factors

are being deployed, 2) Which host resistance strategies are available, and 3) how the environment/farming practices will influence this interaction. Furthermore, to make the puzzle more complicated, scientists must predict how the evolutionary arms race will affect the host-pathogen interaction in years to come. While much is known about host-pathogen interactions in model systems such as the interaction between *Pseudomonas syringae* pv. *tomato* DC3000 and *Arabidopsis thaliana*, relatively little is known about how this information can be applied in an agricultural setting. Disease outbreaks represent a unique opportunity for combining new techniques such as real-time long-read sequencing and RNA-Seq with classical techniques such as disease assays and resistance screening in order to determine which side of the disease triangle broke down in order to allow disease to occur.

In my thesis I provide explanations for disease emergence by investigating the host-pathogen arms race between upland cotton (*Gossypium hirsutum*) and two bacterial pathogens: *Xanthomonas citri* pv. *malvacearum* (*Xcm*) and *Pseudomonas syringae*. I achieve this by 1) Performing phylogenetics to illuminate the identity, evolutionary history and diversity of the pathogen, 2) Using comparative genomics to determine pathogen virulence factors that correlate with disease phenotypes, 3) Screening host germplasm for resistance and putative susceptibility factors, and 4) Identifying markers of selection in the pathogen that represent the ongoing host-pathogen arms race.

The results of this thesis contribute important information for understanding how emerging and re-emerging pathogens causes disease on cotton and will lay the groundwork for future research that ultimately aims to identify durable genetic resistance.

1.6 References

1. Abrams PA. Is predator-prey coevolutlon an arms race? *Trends Ecol Evol*. 1986;1(4):108-110. doi:10.1016/0169-5347(86)90037-6

2. Dawkins Richard, Krebs John Richard. Arms races between and within species. *Proc R Soc Lond B Biol Sci*. 1979;205(1161):489-511. doi:10.1098/rspb.1979.0081

3. Jones JDG, Dangl JL. The plant immune system. *Nature*. 2006;444(7117):323-329. doi:10.1038/nature05286

4. Boutrot F, Zipfel C. Function, Discovery, and Exploitation of Plant Pattern Recognition Receptors for Broad-Spectrum Disease Resistance. *Annu Rev Phytopathol*. 2017;55(1):257-286. doi:10.1146/annurev-phyto-080614-120106

5. Francis M. Regulation of type III secretion systems. *Curr Opin Microbiol*. 2002;5(2):166-172. doi:10.1016/S1369-5274(02)00301-6

6. Lindeberg M, Cunnac S, Collmer A. Pseudomonas syringae type III effector repertoires: last words in endless arguments. *Trends Microbiol*. 2012;20(4):199-208. doi:10.1016/j.tim.2012.01.003

7. Dillon MM, Almeida RND, Laflamme B, et al. Molecular Evolution of Pseudomonas syringae Type III Secreted Effector Proteins. *Front Plant Sci*. 2019;10. doi:10.3389/fpls.2019.00418

8. Stavrinides J, McCann HC, Guttman DS. Host–pathogen interplay and the evolution of bacterial effectors. *Cell Microbiol*. 2008;10(2):285-292. doi:10.1111/j.1462-5822.2007.01078.x

9. Cook DE, Mesarich CH, Thomma BPHJ. Understanding Plant Immunity as a Surveillance System to Detect Invasion. *Annu Rev Phytopathol*. 2015;53(1):541-563. doi:10.1146/annurev-phyto-080614-120114

10. Kanyuka K, Rudd JJ. Cell surface immune receptors: the guardians of the plant's extracellular spaces. *Curr Opin Plant Biol*. 2019;50:1-8. doi:10.1016/j.pbi.2019.02.005

11. Thomma BPHJ, Nürnberger T, Joosten MHAJ. Of PAMPs and Effectors: The Blurred PTI-ETI Dichotomy. *Plant Cell*. 2011;23(1):4-15. doi:10.1105/tpc.110.082602

12. Carroll L. *Through the Looking-Glass and What Alice Found There*. Macmillan; 1871.

13. Lee VanValen. A new evolutionary law. *Evol Theory*. 1973;1:1-30.

14. Kooman-Gersmann M, Honee G, Bonnema G, Wit PD. A High-Affinity Binding Site for the AVR9 Peptide Elicitor of Cladosporium fulvum Is Present on Plasma Membranes of Tomato and Other Solanaceous Plants. *Plant Cell*. 1996;8(5):929-938. doi:10.1105/tpc.8.5.929

15. Naito K, Ishiga Y, Toyoda K, Shiraishi T, Ichinose Y. N-terminal domain including conserved flg22 is required for flagellin-induced hypersensitive cell death in Arabidopsis thaliana. *J Gen Plant Pathol*. 2007;73(4):281-285. doi:10.1007/s10327-007-0017-9

16. Thomas CM, Nielsen KM. Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria. *Nat Rev Microbiol*. 2005;3(9):711-721. doi:10.1038/nrmicro1234

17. Sheppard SK, Guttman DS, Fitzgerald JR. Population genomics of bacterial host adaptation. *Nat Rev Genet*. 2018;19(9):549. doi:10.1038/s41576-018-0032-z

18. Ma W, Dong FFT, Stavrinides J, Guttman DS. Type III Effector Diversification via Both Pathoadaptation and Horizontal Transfer in Response to a Coevolutionary Arms Race. *PLOS Genet*. 2006;2(12):e209. doi:10.1371/journal.pgen.0020209

19. Sokurenko EV, Hasty DL, Dykhuizen DE. Pathoadaptive mutations: gene loss and variation in bacterial pathogens. *Trends Microbiol*. 1999;7(5):191-195. doi:10.1016/S0966-842X(99)01493-6

20. Baggs E, Dagdas G, Krasileva K. NLR diversity, helpers and integrated domains: making sense of the NLR IDentity. *Curr Opin Plant Biol*. 2017;38:59-67. doi:10.1016/j.pbi.2017.04.012

21. Keeling PJ, Palmer JD. Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet*. 2008;9(8):605-618. doi:10.1038/nrg2386

22. Jacques M-A, Arlat M, Boulanger A, et al. Using Ecology, Physiology, and Genomics to Understand Host Specificity in Xanthomonas. *Annu Rev Phytopathol*. 2016;54(1):163-187. doi:10.1146/annurev-phyto-080615-100147

23. Xin X-F, Kvitko B, He SY. Pseudomonas syringae: what it takes to be a pathogen. *Nat Rev Microbiol*. 2018;16(5):316-328. doi:10.1038/nrmicro.2018.17

24. da Silva ACR, Ferro JA, Reinach FC, et al. Comparison of the genomes of two Xanthomonas pathogens with differing host specificities. *Nature*. 2002;417(6887):459-463. doi:10.1038/417459a

25. Meline V, Delage W, Brin C, et al. Role of the acquisition of a type 3 secretion system in the emergence of novel pathogenic strains of Xanthomonas. *Mol Plant Pathol*. 2019;20(1):33-50. doi:10.1111/mpp.12737

26. Bodey GP, Bolivar R, Fainstein V, Jadeja L. Infections Caused by Pseudomonas aeruginosa. *Rev Infect Dis*. 1983;5(2):279-313. doi:10.1093/clinids/5.2.279

27. Nguyen MH, Muder RR. Meningitis Due to Xanthomonas maltophilia: Case Report and Review. *Clin Infect Dis*. 1994;19(2):325-326. doi:10.1093/clinids/19.2.325

28. Lelliott RA, Billing E, Hayward AC. A Determinative Scheme for the Fluorescent Plant Pathogenic Pseudomonads. *J Appl Bacteriol*. 1966;29(3):470-489. doi:10.1111/j.1365-2672.1966.tb03499.x

29. Almeida NF, Yan S, Cai R, et al. PAMDB, A Multilocus Sequence Typing and Analysis Database and Website for Plant-Associated Microbes. *Phytopathology*. 2010;100(3):208-215. doi:10.1094/PHYTO-100-3-0208

30. Bartoli C, Berge O, Monteil CL, et al. The Pseudomonas viridiflava phylogroups in the P. syringae species complex are characterized by genetic variability and phenotypic plasticity of pathogenicity-related traits. *Environ Microbiol*. 2014;16(7):2301-2315. doi:10.1111/1462-2920.12433

31. Berge O, Monteil CL, Bartoli C, et al. A User's Guide to a Data Base of the Diversity of Pseudomonas syringae and Its Application to Classifying Strains in This Phylogenetic Complex. *PLOS ONE*. 2014;9(9):e105547. doi:10.1371/journal.pone.0105547

32. Morris CE, Sands DC, Vanneste JL, et al. Inferring the Evolutionary History of the Plant Pathogen Pseudomonas syringae from Its Biogeography in Headwaters of Rivers in North America, Europe, and New Zealand. Lindow S, ed. *mBio*. 2010;1(3). doi:10.1128/mBio.00107-10

33. Parkinson N, Bryant R, Bew J, Elphinstone J. Rapid phylogenetic identification of members of the Pseudomonas syringae species complex using the rpoD locus: Identification of the Pseudomonas syringae complex. *Plant Pathol*. 2011;60(2):338-344. doi:10.1111/j.1365-3059.2010.02366.x

34. Baltrus DA, McCann HC, Guttman DS. Evolution, genomics and epidemiology of Pseudomonas syringae. Molecular Plant Pathology. doi:10.1111/mpp.12506

35. Morris CE, Lamichhane JR, Nikolić I, Stanković S, Moury B. The overlapping continuum of host range among strains in the Pseudomonas syringae complex. *Phytopathol Res*. 2019;1(1):4. doi:10.1186/s42483-018-0010-6

36. Vicente JG, Holub EB. Xanthomonas campestris pv. campestris (cause of black rot of crucifers) in the genomic era is still a worldwide threat to brassica crops. *Mol Plant Pathol*. 2013;14(1):2-18. doi:10.1111/j.1364-3703.2012.00833.x

37. Buttner D. Protein Export According to Schedule: Architecture, Assembly, and Regulation of Type III Secretion Systems from Plant- and Animal-Pathogenic Bacteria. *Microbiol Mol Biol Rev*. 2012;76(2):262-310. doi:10.1128/MMBR.05017-11

38. Deng W, Marshall NC, Rowland JL, et al. Assembly, structure, function and regulation of type III secretion systems. *Nat Rev Microbiol*. 2017;15(6):323-337. doi:10.1038/nrmicro.2017.20

39. Green ER, Mecsas J. Bacterial Secretion Systems: An Overview. *Microbiol Spectr*. 2016;4(1). doi:10.1128/microbiolspec.VMBF-0012-2015

40. Deakin WJ, Broughton WJ. Symbiotic use of pathogenic strategies: rhizobial protein secretion systems. *Nat Rev Microbiol*. 2009;7(4):312-320. doi:10.1038/nrmicro2091

41. Melnyk RA, Hossain SS, Haney CH. Convergent gain and loss of genomic islands drive lifestyle changes in plant-associated Pseudomonas. *ISME J*. 2019;13(6):1575. doi:10.1038/s41396-019-0372-5

42. Soto MJ, Sanjuan J, Olivares J. Rhizobia and plant-pathogenic bacteria: common infection weapons. *Microbiology*. 2006;152(11):3167-3174. doi:10.1099/mic.0.29112-0

43. Sineva E, Savkina M, Ades SE. Themes and variations in gene regulation by extracytoplasmic function (ECF) sigma factors. *Curr Opin Microbiol*. 2017;36:128-137. doi:10.1016/j.mib.2017.05.004

44. Xiao Y, Heu S, Yi J. Identification of a Putative Alternate Sigma Factor and Characterization of a Multicomponent Regulatory Cascade Controlling the Expression of Pseudomonas syringae pv. syringae Pss6l hrp and hrmA Genes. 1994;176:12.

45. Xiao Y, Hutcheson SW. A Single Promoter Sequence Recognized by a Newly Identified Alternate Sigma Factor Directs Expression of Pathogenicity and Host Range Determinants in Pseudomonas syringae. *J Bacteriol*. 1994;176(10):3089-3091.

46. Huang D-L, Tang D-J, Liao Q, et al. The Zur of *Xanthomonas campestris* Is Involved in Hypersensitive Response and Positively Regulates the Expression of the *hrp* Cluster Via *hrpX* But Not *hrpG*. *Mol Plant Microbe Interact*. 2009;22(3):321-329. doi:10.1094/MPMI-22-3-0321

47. Ikawa Y, Tsuge S. The quantitative regulation of the *hrp* regulator HrpX is involved in sugar-source-dependent *hrp* gene expression in *Xanthomonas oryzae* pv. *oryzae*. Mongkolsuk S, ed. *FEMS Microbiol Lett*. 2016;363(10):fnw071. doi:10.1093/femsle/fnw071

48. Wei K, Tang D-J, He Y-Q, et al. hpaR, a Putative marR Family Transcriptional Regulator, Is Positively Controlled by HrpG and HrpX and Involved in the Pathogenesis, Hypersensitive Response, and Extracellular Protease Production of Xanthomonas campestris Pathovar campestris. *J Bacteriol*. 2007;189(5):2055-2062. doi:10.1128/JB.01331-06

49. Baltrus DA, Nishimura MT, Romanchuk A, et al. Dynamic Evolution of Pathogenicity Revealed by Sequencing and Comparative Genomics of 19 Pseudomonas syringae Isolates. Guttman DS, ed. *PLoS Pathog*. 2011;7(7):e1002132. doi:10.1371/journal.ppat.1002132

50. Cohn M, Bart RS, Shybut M, et al. Xanthomonas axonopodis Virulence Is Promoted by a Transcription Activator-Like Effector–Mediated Induction of a SWEET Sugar Transporter in Cassava. *Mol Plant Microbe Interact*. 2014;27(11):1186-1198. doi:10.1094/MPMI-06-14-0161-R

51. Medina CA, Reyes PA, Trujillo CA, et al. The role of type III effectors from Xanthomonas axonopodis pv. manihotis in virulence and suppression of plant immunity. *Mol Plant Pathol*. 2018;19(3):593-606. doi:10.1111/mpp.12545

52. Xin X-F, Nomura K, Aung K, et al. Bacteria establish an aqueous living space in plants crucial for virulence. *Nature*. 2016;539(7630):524-529. doi:10.1038/nature20166

53. Kay S, Bonas U. How Xanthomonas type III effectors manipulate the host plant. *Curr Opin Microbiol*. 2009;12(1):37-43. doi:10.1016/j.mib.2008.12.006

54. Almeida NF, Yan S, Lindeberg M, et al. A Draft Genome Sequence of Pseudomonas syringae pv. tomato T1 Reveals a Type III Effector Repertoire Significantly Divergent from That of Pseudomonas syringae pv. tomato DC3000. *Mol Plant Microbe Interact*. 2008;22(1):52-62. doi:10.1094/MPMI-22-1-0052

55. Cunnac S, Chakravarthy S, Kvitko BH, Russell AB, Martin GB, Collmer A. Genetic disassembly and combinatorial reassembly identify a minimal functional repertoire of type III effectors in Pseudomonas syringae. *Proc Natl Acad Sci*. 2011;108(7):2975-2980. doi:10.1073/pnas.1013031108

56. Kvitko BH, Park DH, Velásquez AC, et al. Deletions in the Repertoire of Pseudomonas syringae pv. tomato DC3000 Type III Secretion Effector Genes Reveal Functional Overlap among Effectors. *PLOS Pathog*. 2009;5(4):e1000388. doi:10.1371/journal.ppat.1000388

57. Popov G, Fraiture M, Brunner F, Sessa G. Multiple Xanthomonas euvesicatoria Type III Effectors Inhibit flg22-Triggered Immunity. *Mol Plant Microbe Interact*. 2016;29(8):651-660. doi:10.1094/MPMI-07-16-0137-R

58. Galán JE, Curtiss R. Cloning and molecular characterization of genes whose products allow Salmonella typhimurium to penetrate tissue culture cells. *Proc Natl Acad Sci*. 1989;86(16):6383-6387. doi:10.1073/pnas.86.16.6383

59. Penheiter KL, Mathur N, Giles D, Fahlen T, Jones BD. Non-invasive Salmonella typhimurium mutants are avirulent because of an inability to enter and destroy M cells of ileal Peyer's patches. *Mol Microbiol*. 1997;24(4):697-709. doi:10.1046/j.1365-2958.1997.3741745.x

60. Hensel M, Shea JE, Gleeson C, Jones MD, Dalton E, Holden DW. Simultaneous identification of bacterial virulence genes by negative selection. *Science*. 1995;269(5222):400-403. doi:10.1126/science.7618105

61. Streubel J, Pesce C, Hutin M, Koebnik R, Boch J, Szurek B. Five phylogenetically close rice SWEET genes confer TAL effector-mediated susceptibility to Xanthomonas oryzae pv. oryzae. *New Phytol*. 2014:808-819. doi:10.1111/nph.12411@10.1002/(ISSN)1469-8137(CAT)VirtualIssues(VI)Phytopathogeneffectorproteins

62. Cox KL, Meng F, Wilkins KE, et al. TAL effector driven induction of a *SWEET* gene confers susceptibility to bacterial blight of cotton. *Nat Commun*. 2017;8:15588. doi:10.1038/ncomms15588

63. Phillips AZ, Berry JC, Wilson MC, et al. Genomics-enabled analysis of the emergent disease cotton bacterial blight. *PLOS Genet*. 2017;13(9):23.

64. Merda D, Bonneau S, Guimbaud J-F, et al. Recombination-prone bacterial strains form a reservoir from which epidemic clones emerge in agroecosystems: Recombinant strains as a reservoir for epidemics. *Environ Microbiol Rep*. 2016;8(5):572-581. doi:10.1111/1758-2229.12397

65. O'Brien HE, Thakur S, Gong Y, et al. Extensive remodeling of the Pseudomonas syringae pv. avellanae type III secretome associated with two independent host shifts onto hazelnut. *BMC Microbiol*. 2012;12(1):141. doi:10.1186/1471-2180-12-141

66. Neale HC, Laister R, Payne J, Preston G, Jackson RW, Arnold DL. A low frequency persistent reservoir of a genomic island in a pathogen population ensures island survival and improves pathogen fitness in a susceptible host: Reservoir of a genomic island in a pathogen. *Environ Microbiol*. 2016;18(11):4144-4152. doi:10.1111/1462-2920.13482

67. Stavrinides J, Ma W, Guttman DS. Terminal Reassortment Drives the Quantum Evolution of Type III Effectors in Bacterial Pathogens. *PLOS Pathog*. 2006;2(10):e104. doi:10.1371/journal.ppat.0020104

68. Srikanth CV, Wall DM, Maldonado-Contreras A, et al. Salmonella Pathogenesis and Processing of Secreted Effectors by Caspase-3. *Science*. 2010;330(6002):390-393. doi:10.1126/science.1194598

69. Alfano JR, Charkowski AO, Deng W-L, et al. The Pseudomonas syringae Hrp pathogenicity island has a tripartite mosaic structure composed of a cluster of type III secretion genes bounded by exchangeable effector and conserved effector loci that contribute to parasitic fitness and pathogenicity in plants. *Proc Natl Acad Sci U S A*. 2000;97(9):4856-4861.

70. Chen J, Carpena N, Quiles-Puchalt N, Ram G, Novick RP, Penadés JR. Intra- and inter-generic transfer of pathogenicity island-encoded virulence genes by *cos* phages. *ISME J*. 2015;9(5):1260-1263. doi:10.1038/ismej.2014.187

71. Hulin MT, Armitage AD, Vicente JG, et al. Comparative genomics of Pseudomonas syringae reveals convergent gene gain and loss associated with specialization onto cherry (Prunus avium). *New Phytol*. 2018;219(2):672-696. doi:10.1111/nph.15182

72. Kim EJ, Yu HJ, Lee JH, et al. Replication of Vibrio cholerae classical CTX phage. *Proc Natl Acad Sci*. 2017;114(9):2343-2348. doi:10.1073/pnas.1701335114

73. Djidjou-Demasse R, Moury B, Fabre F. Mosaics often outperform pyramids: insights from a model comparing strategies for the deployment of plant resistance genes against viruses in agricultural landscapes. *New Phytol*. 2017;216(1):239-253. doi:10.1111/nph.14701

74. Fuchs M. Pyramiding resistance-conferring gene sequences in crops. *Curr Opin Virol*. 2017;26:36-42. doi:10.1016/j.coviro.2017.07.004

75. Kunkeaw S, Tan S, Coaker G. Molecular and Evolutionary Analyses of *Pseudomonas syringae* pv. *tomato* Race 1. *Mol Plant Microbe Interact*. 2010;23(4):415-424. doi:10.1094/MPMI-23-4-0415

76. Guttman DS, Gropp SJ, Morgan RL, Wang PW. Diversifying Selection Drives the Evolution of the Type III Secretion System Pilus of Pseudomonas syringae. *Mol Biol Evol*. 2006;23(12):2342-2354. doi:10.1093/molbev/msl103

77. Rohmer L, Guttman DS, Dangl JL. Diverse Evolutionary Mechanisms Shape the Type III Effector Virulence Factor Repertoire in the Plant Pathogen Pseudomonas syringae. *Genetics*. 2004;167(3):1341-1360. doi:10.1534/genetics.103.019638

78. Weber E, Koebnik R. Positive Selection of the Hrp Pilin HrpE of the Plant Pathogen Xanthomonas. *J Bacteriol*. 2006;188(4):1405-1410. doi:10.1128/JB.188.4.1405-1410.2006

79. Chen L-Q, Qu X-Q, Hou B-H, et al. Sucrose Efflux Mediated by SWEET Proteins as a Key Step for Phloem Transport. *Science*. 2012;335(6065):207-211. doi:10.1126/science.1213351

80. Eom J-S, Chen L-Q, Sosso D, et al. SWEETs, transporters for intracellular and intercellular sugar translocation. *Curr Opin Plant Biol*. 2015;25:53-62. doi:10.1016/j.pbi.2015.04.005

81. Lin IW, Sosso D, Chen L-Q, et al. Nectar secretion requires sucrose phosphate synthases and the sugar transporter SWEET9. *Nature*. 2014;508(7497):546-549. doi:10.1038/nature13082

82. Yang B, Sugio A, White FF. Os8N3 is a host disease-susceptibility gene for bacterial blight of rice. *Proc Natl Acad Sci*. 2006;103(27):10503-10508. doi:10.1073/pnas.0604088103

83. Jiang W, Zhou H, Bi H, Fromm M, Yang B, Weeks DP. Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic Acids Res*. 2013;41(20):e188-e188. doi:10.1093/nar/gkt780

84. Jørgensen IH. Discovery, characterization and exploitation of Mlo powdery mildew resistance in barley. *Euphytica*. 1992;63(1):141-152. doi:10.1007/BF00023919

85. Kusch S, Panstruga R. mlo-Based Resistance: An Apparently Universal "Weapon" to Defeat Powdery Mildew Disease. *Mol Plant Microbe Interact*. 2017;30(3):179-189. doi:10.1094/MPMI-12-16-0255-CR

86. Miklis M, Consonni C, Bhat RA, Lipka V, Schulze-Lefert P, Panstruga R. Barley MLO Modulates Actin-Dependent and Actin-Independent Antifungal Defense Pathways at the Cell Periphery. *Plant Physiol*. 2007;144(2):1132-1143. doi:10.1104/pp.107.098897

87. Kim DS, Hwang BK. The pepper MLO gene, CaMLO2, is involved in the susceptibility cell-death response and bacterial and oomycete proliferation. *Plant J*. 2012;72(5):843-855. doi:10.1111/tpj.12003

88. Bent AF, Mackey D. Elicitors, Effectors, and R Genes: The New Paradigm and a Lifetime Supply of Questions. *Annu Rev Phytopathol*. 2007;45(1):399-436. doi:10.1146/annurev.phyto.45.062806.094427

89. Jones JDG. Putting knowledge of plant disease resistance genes to work. *Curr Opin Plant Biol*. 2001;4(4):281-287. doi:10.1016/S1369-5266(00)00174-6

90. Flor HH. Inheritance of pathogenicity in a cross between physiologic races 22 and 24 of Melampsora lini. *Phytopathology*. 1942;32(5):653-669.

91. Delannoy E, Lyon BR, Marmey P, et al. Resistance of Cotton Towards Xanthomonas campestris pv. malvacearum. *Annu Rev Phytopathol*. 2005;43(1):63-82. doi:10.1146/annurev.phyto.43.040204.140251

92. Xiao J, Fang DD, Bhatti M, Hendrix B, Cantrell R. A SNP haplotype associated with a gene resistant to Xanthomonas axonopodis pv. malvacearum in upland cotton (Gossypium hirsutum L.). *Mol Breed*. 2010;25(4):593-602. doi:10.1007/s11032-009-9355-y

93. Essenberg M, Bayles MB, Pierce ML, Verhalen LM. Pyramiding B Genes in Cotton Achieves Broader But Not Always Higher Resistance to Bacterial Blight. *Phytopathology*. 2014;104(10):1088-1097. doi:10.1094/PHYTO-06-13-0167-R

94. Cox KL, Babilonia K, Wheeler T, He P, Shan L. Return of old foes — recurrence of bacterial blight and Fusarium wilt of cotton. *Curr Opin Plant Biol*. 2019;50:95-103. doi:10.1016/j.pbi.2019.03.012

95. Heath MC. Hypersensitive response-related death. In: Lam E, Fukuda H, Greenberg J, eds. *Programmed Cell Death in Higher Plants*. Dordrecht: Springer Netherlands; 2000:77-90. doi:10.1007/978-94-010-0934-8_6

96. Kim H-S, Desveaux D, Singer AU, Patel P, Sondek J, Dangl JL. The Pseudomonas syringae effector AvrRpt2 cleaves its C-terminally acylated target, RIN4, from Arabidopsis membranes to block RPM1 activation. *Proc Natl Acad Sci*. 2005;102(18):6496-6501. doi:10.1073/pnas.0500792102

97. Eschen-Lippold L, Jiang X, Elmore JM, et al. Bacterial AvrRpt2-Like Cysteine Proteases Block Activation of the Arabidopsis Mitogen-Activated Protein Kinases, MPK4 and MPK11. *Plant Physiol*. 2016;171(3):2223-2238. doi:10.1104/pp.16.00336

98. Gardy JL, Loman NJ. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat Rev Genet*. 2018;19(1):9-20. doi:10.1038/nrg.2017.88

99. Lamichhane JR, Messéan A, Morris CE. Insights into epidemiology and control of diseases of annual plants caused by the Pseudomonas syringae species complex. *J Gen Plant Pathol*. 2015;81(5):331-350. doi:10.1007/s10327-015-0605-z

100. Barret M, Guimbaud J-F, Darrasse A, Jacques M-A. Plant microbiota affects seed transmission of phytopathogenic microorganisms: Plant microbiota affects seed transmission. *Mol Plant Pathol*. 2016;17(6):791-795. doi:10.1111/mpp.12382

101. Ceresini PC, Castroagudín VL, Rodrigues FÁ, et al. Wheat blast: from its origins in South America to its emergence as a global threat. *Mol Plant Pathol*. 2019;20(2):155-172. doi:10.1111/mpp.12747

102. Ialacci GM, Bella P, Licciardello G, et al. Clonal populations of Clavibacter michiganensis subsp. michiganensis are responsible for the outbreaks of bacterial canker in greenhouse tomatoes in Italy. *Plant Pathol*. 2016;65(3):484-495. doi:10.1111/ppa.12424

103. Jacobson AL, Duffy S, Sseruwagi P. Whitefly-transmitted viruses threatening cassava production in Africa. *Curr Opin Virol*. 2018;33:167-176. doi:10.1016/j.coviro.2018.08.016

104. Schiro G, Verch G, Grimm V, Müller MEH. Alternaria and Fusarium Fungi: Differences in Distribution and Spore Deposition in a Topographically Heterogeneous Wheat Field. *J Fungi*. 2018;4(2):63. doi:10.3390/jof4020063

105. Maraite H, Weyns J. Pseudomonas Syringae pv. aptata and pv. atrofaciens, Specific Pathovars or Members of pv. Syringae? In: Rudolph K, Burr TJ, Mansfield JW, Stead D, Vivian A, von Kietzell J, eds. *Pseudomonas Syringae Pathovars and Related Pathogens*. Vol 9. Dordrecht: Springer Netherlands; 1997:515-520. doi:10.1007/978-94-011-5472-7_94

106. King CJ, Brinkerhoff LA. The dissemination of Xanthomonas malvacearum by irrigation water. *Phytopathology*. 1949;39:88-90.

107. Morris CE, Sands DC, Vinatzer BA, et al. The life history of the plant pathogen *Pseudomonas syringae* is linked to the water cycle. *ISME J*. 2008;2(3):321-334. doi:10.1038/ismej.2007.113

108. Leadbeater A. Recent developments and challenges in chemical disease control – a review. *Plant Prot Sci*. 2016;51(No. 4):163-169. doi:10.17221/83/2015-PPS

109. Russell PE. A century of fungicide evolution. *J Agric Sci*. 2005;143(1):11-25. doi:10.1017/S0021859605004971

110. McKenna M. Antibiotics set to flood Florida's troubled orange orchards. *Nature*. 2019;567:302. doi:10.1038/d41586-019-00878-4

111. Alexander AS, Woodward JE, Boman RK, Wheeler TA, Hopper NW. Effect of the Easiflo Cottonseed Processing Method on Recovery of Xanthomonas axonopodis pv. malvacearum. *Tex J Agric Nat Resour*. 2016;25(0):13-23.

112. Koczan J, Albers DW, Louis S, Gholston K. 2017: IDENTIFICATION OF AN ALTERNATE SOURCE OF INOCULUM CAUSING BACTERIAL BLIGHT IN COTTON. 2017:2.

113. Workman D. Cotton Exports by Country. World's Top Exports. http://www.worldstopexports.com/cotton-exports-by-country/. Published April 1, 2019. Accessed June 18, 2019.

114. Pest Management: National Cotton Council. https://www.cotton.org/tech/pest/. Accessed June 18, 2019.

115. Wheeler TA. Identification of Factors that Influence Screening for Bacterial Blight Resistance. 2007;11(1):7.

116. Verma JP. *Bacterial Blight of Cotton*. CRC Press; 1986.

117. Hillocks RJ. Bacterial blight. *Cotton Dis*. 1992:39-85.

118. Phillips AZ, Wheeler T, Woodward J, Bart RS. *Pseudomonas syringae* Pathogen Causes Foliar Disease of Upland Cotton in Texas. *Plant Dis*. 2018;102(6):1171-1171. doi:10.1094/PDIS-11-17-1700-PDN

119. Bird LS, Hefner JJ, Blackmon CW, Pore RS. Pseudomonas Wilt of Cotton. *Agric Mech Coll Tex Tex Agric Exp Stn*. 1960;MP-471:1-8.

120. Mertely J, Gannaway J, Kaufman H. Pathogens associated with cotton seedling diseases on the southern high plains of Texas. *Proc Beltwide Cotton Dis Counc*. 1994:246.

Chapter 2: Genomics-Enabled Analysis of the Emergent Disease Cotton Bacterial Blight

This chapter was previously published in the journal PLoS Genetics as:

Phillips AZ, Berry JC*, Wilson MC*, Vijayaraghavan A, Burke J, Bunn JI, Allen TW, Wheeler T, and Bart RS. (2017) Genomics-enabled analysis of the emergent disease cotton bacterial blight. PLoS Genet 13(9): e1007003. https://doi.org/10.1371/journal.pgen.1007003

2.1 Personal contributions

This manuscript was the product of a collaboration among several talented scientists. As first author, I coordinated the project as a whole and contributed in the form of experimental design and execution, data collection, data analysis, figure generation, and writing. My contributions to experimental design and data analysis included identifying which *Xanthomonas* strains would be sequenced and investigated to determine if a race or host shift had occurred. I then generated a multi locus sequence typing phylogeny that demonstrated that a host shift had not occurred. I also curated an effector table and performed disease assays that showed that a host shift did not occur. Subsequently I assembled disease incidence reports and cotton planting statistics from extension scientists and federal databases to determine how fast the disease spread over time and to demonstrate that disease progression correlated with the proportion of susceptible varieties planted. I collected additional data through taking representative photographs of disease symptoms from the field as well as symptoms generated in the laboratory in order to characterize the many symptoms of the disease and the range of symptom severity that individual pathogens cause. This aided in the identification of two phenotypically and phylogenetically distinct *Xcm* strains that I further genetically characterized through genome alignments and Transcription Activator-Like (TAL) effector characterization. This allowed me to

identify TAL binding sites in the promoters of putative host susceptibility genes. I performed TAL effector westerns to support these data and demonstrate that many TAL effectors are expressed. I then helped design an RNA-Seq experiment with these two strains and two genetically divergent cotton varieties in order to identify susceptibility genes. My analysis of the resulting gene expression data resulted in the identification of several putative SWEET and Mildew Locus-O susceptibility genes. At the culmination of this project, I wrote the manuscript and generated and assembled all figures with the exception of the Circos plot in Fig. 4, the venn diagram in Figure 6c, and the plots in Figure 7, where images were generated to my specifications.

2.2 Abstract

Cotton bacterial blight (CBB), an important disease of (*Gossypium hirsutum*) in the early 20th century, had been controlled by resistant germplasm for over half a century. Recently, CBB re-emerged as an agronomic problem in the United States. Here, we report analysis of cotton variety planting statistics that indicate a steady increase in the percentage of susceptible cotton varieties grown each year since 2009. Phylogenetic analysis revealed that strains from the current outbreak cluster with race 18 *Xanthomonas citri* pv. *malvacearum* (*Xcm*) strains. Illumina based draft genomes were generated for thirteen *Xcm* isolates and analyzed along with 4 previously published *Xcm* genomes. These genomes encode 24 conserved and nine variable type three effectors. Strains in the race 18 clade contain 3 to 5 more effectors than other *Xcm* strains. SMRT sequencing of two geographically and temporally diverse strains of *Xcm* yielded circular chromosomes and accompanying plasmids. These genomes encode eight and thirteen distinct transcription activator-like effector genes. RNA-sequencing revealed 52 genes induced within two cotton cultivars by both tested *Xcm* strains. This gene list includes a homeologous pair of

genes, with homology to the known susceptibility gene, MLO. In contrast, the two strains of *Xcm* induce different clade III SWEET sugar transporters. Subsequent genome wide analysis revealed patterns in the overall expression of homeologous gene pairs in cotton after inoculation by *Xcm*. These data reveal novel insights into the *Xcm-G. hirsutum* disease complex and strategies for future development of resistant cultivars.

2.3 Author summary

Cotton bacterial blight (CBB), caused by *Xanthomonas citri* pv. *malvacearum* (*Xcm*), significantly limited cotton yields in the early 20th century but has been controlled by classical resistance genes for more than 50 years. In 2011, the pathogen re-emerged with a vengeance. In this study, we compare diverse pathogen isolates and cotton varieties to further understand the virulence mechanisms employed by *Xcm* and to identify promising resistance strategies. We generate fully contiguous genome assemblies for two diverse *Xcm* strains and identify pathogen proteins used to modulate host transcription and promote susceptibility. RNA-sequencing of infected cotton reveals novel putative gene targets for the development of durable *Xcm* resistance. Together, the data presented reveal contributing factors for CBB re-emergence in the U.S. and highlight several promising routes towards the development of durable resistance including classical resistance genes and potential manipulation of susceptibility targets.

2.4 Introduction

Upland cotton (*Gossypium hirsutum* L.) is the world's leading natural fiber crop. Cotton is commercially grown in over 84 countries, and in the United States, is responsible for $74 billion annually[1,2]. Numerous foliar diseases affect cotton throughout the world's cotton growing regions. Historically, one of the most significant foliar diseases has been bacterial blight, caused by *Xanthomonas citri* pv. *malvacearum*. Cotton bacterial blight significantly limited cotton yield

in the late 20th century. In the 1940's and 1950's, breeders identified and introgressed multiple resistance loci into elite germplasm[3-5]. This strategy proved durable for over half a century. In 2011, cotton bacterial blight (CBB) returned and caused significant losses to farmers in the southern United States, including in Arkansas and Mississippi. Nonetheless, CBB has received little research focus during the last several decades because, prior to 2011, losses from this disease were not substantial. Modern molecular and genomic technologies can now be employed expeditiously to deduce the underlying cause of the disease re-emergence and pinpoint optimized routes towards the development of durable resistance.

CBB is caused by *X. citri pv. malvacearum* (*Xcm*); however, the pathogen has previously been placed within other species groupings[6-9]. The *Xcm* pathovar can be further divided into at least 19 races according to virulence phenotypes on a panel of historical cotton cultivars: Acala-44, Stoneville 2B-S9, Stoneville 20, Mebane B-1, 1-10B, 20-3, and 101-102.B[10,11]. Historically, the most common race observed in the U.S. has been race 18, which was first isolated in 1973[12]. This race is highly virulent, causing disease on all cultivars in the panel except for 101-102.B. However, this diagnostic panel of cotton varieties used to race type strains is no longer available from the USDA/ARS, Germplasm Resources Information Network (GRIN).

CBB can occur at any stage in the plant's life cycle and on any aerial organ. Typical symptoms include seedling blight as either pre- or post-emergent damping-off, black arm on petioles and stems, water-soaked spots on leaves and bracts, and most importantly boll rot[10]. The most commonly observed symptoms are the angular-shaped lesions on leaves that can coalesce and result in a systemic infection. Disease at each of these stages can cause yield losses either by injury to the plant or direct damage to the boll. No effective chemical treatments for the disease have been released to date. Methods to reduce yield loss as a result of CBB include acid delinting

cotton seed prior to planting, field cultivation practices to reduce sources of overwintering inoculum and planting cultivars with known sources of resistance[3,4,8,13,14].

Xanthomonads assemble the type three secretion system (T3SS), a needle-like structure, to inject diverse type three effectors (T3Es) into the plant cell to suppress immunity and promote disease[15-19]. For example, transcription activator-like (TAL) effectors influence the expression levels of host genes by binding directly to promoters in a sequence-specific way[20]. Up-regulated host genes that contribute to pathogen virulence are termed susceptibility genes and may be modified through genome editing for the development of resistant crop varieties[21].

Plants have specialized immune receptors, collectively known as nucleotide-binding leucine rich repeat receptors that recognize, either directly or indirectly, the pathogen effector molecules[22,23]. Historically, this host-pathogen interaction has been termed the 'gene-for-gene' model of immunity, wherein a single gene from the host and a single gene from the pathogen are responsible for recognition[24]. Recognition triggers a strong immune response that often includes a localized hypersensitive response (HR) in which programmed cell death occurs around the infection site[25]. Nineteen CBB resistance loci have been reported in *Gossypium hirsutum* breeding programs; however, none have been molecularly identified[8,13].

Here we combine comparative genomics of the pathogen *Xcm* with transcriptomics of the host to identify molecular determinants of Cotton Bacterial Blight. This will inform the development of durable resistance strategies.

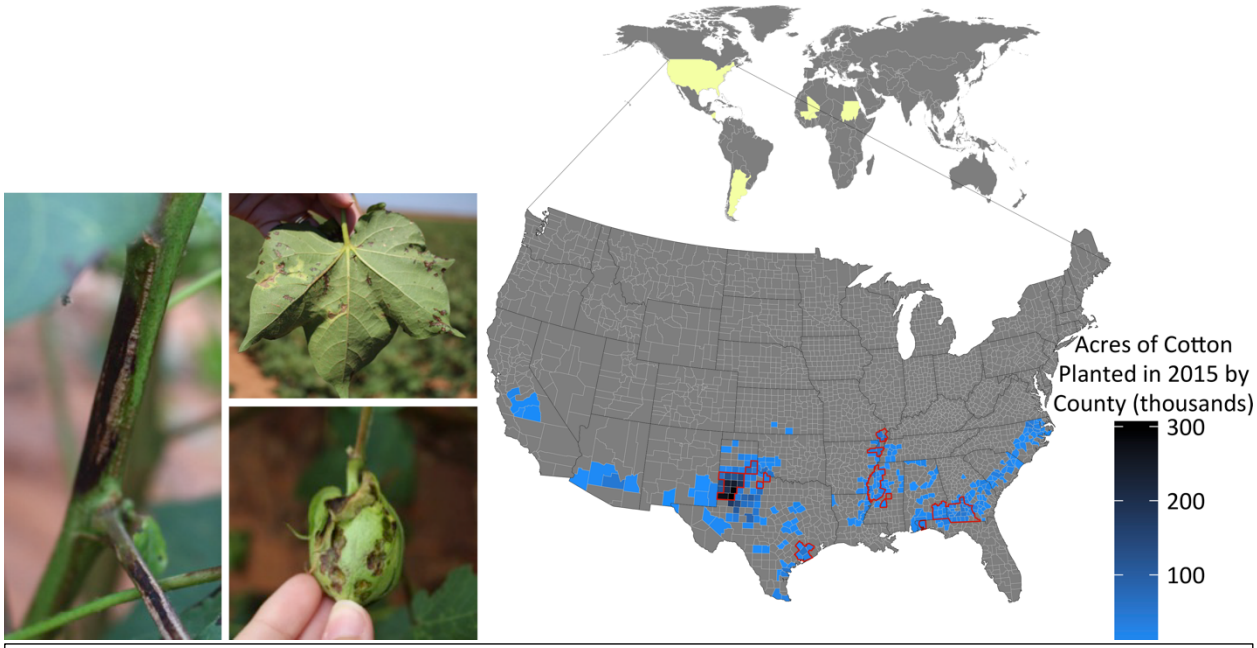


Figure 1: Cotton Bacterial Blight (CBB) symptoms and reemergence across the southern United States. (Left) Typical CBB symptoms present in cotton fields near Lubbock, TX during the 2015 growing season include angular leaf spots, boll rot, and black arm rot. Yellow shading within world map (top) indicates origin of strains included in this study. Acres of cotton planted per county in the United States in 2015 (blue) and counties with confirmed CBB in 2015 (red outline). Statistics on the area of cotton planted in the U.S. were acquired from the USDA. CBB was reported by extension agents, extension specialists, and certified crop advisers in their respective states.

2.5 Results

2.5.1 CBB reemergence in the US

In 2011, farmers, extension specialists, and certified crop advisers in Missouri, Mississippi, and Arkansas observed cotton plants exhibiting symptoms of CBB. Widespread infected plant material was observed throughout much of the production area, but appeared to be centered around Clarksdale, Mississippi. In figure 1, we collate reports from this outbreak and overlay these data with US cotton planting statistics to reveal that this disease has spread through much of the cotton belt in the southern U.S. (Figs 1 and S1, Table S1). Since 2016, CBB has been reported from at least eight out of the sixteen states that grow cotton (Fig 1). In 2014, we collected diseased cotton leaves from two sites across Mississippi and demonstrated pathogen causality following Koch's postulates[26]. In addition, PCR amplification of the 16S rRNA gene

confirmed that the causal agent was a member of the *Xanthomonas* genus. Multi locus sequence type (MLST) analysis and maximum-likelihood analysis were performed using concatenated sections of the *gltA*, *lepA*, *lacF*, *gyrB*, *fusA* and *gap-1* loci for increased phylogenetic resolution (Fig 2a). The newly sequenced strains were named MS14002 and MS14003 and were compared to four previously published *Xcm* genomes and thirty-six additional *Xanthomonas* genomes representing thirteen species (Tables 1, S2). MS14002 and MS14003 grouped with the previously published *Xcm* strains as a single unresolved clade, further confirming that the current disease outbreak is CBB and is caused by *Xcm*. The species designation reported here is consistent with previous reports[6,7].

2.5.2 Contemporary U.S. *Xcm* strains cluster phylogenetically with historical race 18 strains.

Race groups have been described for *Xcm* strains by analyzing compatible (susceptible) and incompatible (resistant) interactions on a panel of seven cotton cultivars. Different geographies often harbor different pathogen races[7]. Consequently, one possible explanation for the recent outbreak of CBB would be the introduction of a new race of *Xcm* capable of overcoming existing genetic resistance. Only 2 varieties of the original cotton panel plus three related cultivars, were available and these cultivars were not sufficient to determine whether a new race had established within the U.S. Thirteen *Xcm* strains were sequenced using Illumina technology to determine the phylogenetic relationship between recent isolates of *Xcm* and historical isolates. Isolates designated as race 1, race 2, race 3, race 12 and race 18 have been maintained at Mississippi State University with these designations. Additional isolates were obtained from the Collection Française de Bactéries associées aux Plantes (CFBP) culture collection. Together, these isolates include nine strains from the US, three from Africa, and one from South America and span collection dates ranging from 1958 through 2014 (Fig 1, Table 1).

Illumina reads were mapped to the *Xanthomonas citri* subsp. *citri* strain Aw12879 (Genbank assembly accession: GCA_000349225.1) using Bowtie2 and single nucleotide polymorphisms (SNPs) were identified using Samtools[27,28]. Only regions of the genome with at least 10x coverage for all genomes were considered. This approach identified 17,853 sites that were polymorphic in at least one genome. Nucleotides were concatenated and used to build a neighbor-joining tree (Fig 2b). This analysis revealed that recent U.S. *Xcm* isolates grouped with the race 18 clade. Notably, the race 18 clade is phylogenetically distant from the other *Xcm* isolates.

2.5.3 Contemporary US *Xcm* strains have conserved type three virulence protein arsenals and disease phenotypes with historical race 18 strains.

Xanthomonads deploy many classes of virulence factors to promote disease. Type three effectors (T3E) are of particular interest for their role in determining race designations. T3E profiles from sixteen *Xcm* isolates were compared to determine whether a change in the virulence protein arsenal of the newly isolated strains could explain the re-emergence of CBB. Genomes from 13 *Xcm* isolates were *de novo* assembled with SPAdes and annotated with Prokka based on annotations from the *X. euvesicatoria* (aka. *X. campestris* pv. *vesicatoria*) 85-10 genome (NCBI accession: NC_007508.1). T3Es pose a particular challenge for reference-based annotation as no bacterial genome contains all effectors. Consequently, an additional protein file containing known T3Es from our previous work was included within the Prokka annotation pipeline[15,29]. This analysis revealed 24 conserved and 9 variable *Xcm* T3Es (Fig 3a). Race 18 clade isolates contain more effectors than other isolates that were sequenced. The recent *Xcm* isolates (MS14002 and MS14003) were not distinguishable from the historical race 18 isolate, with the exception of XcmNI86 isolated from Nicaragua in 1986, which contains mutations in XopE2 and XopP.

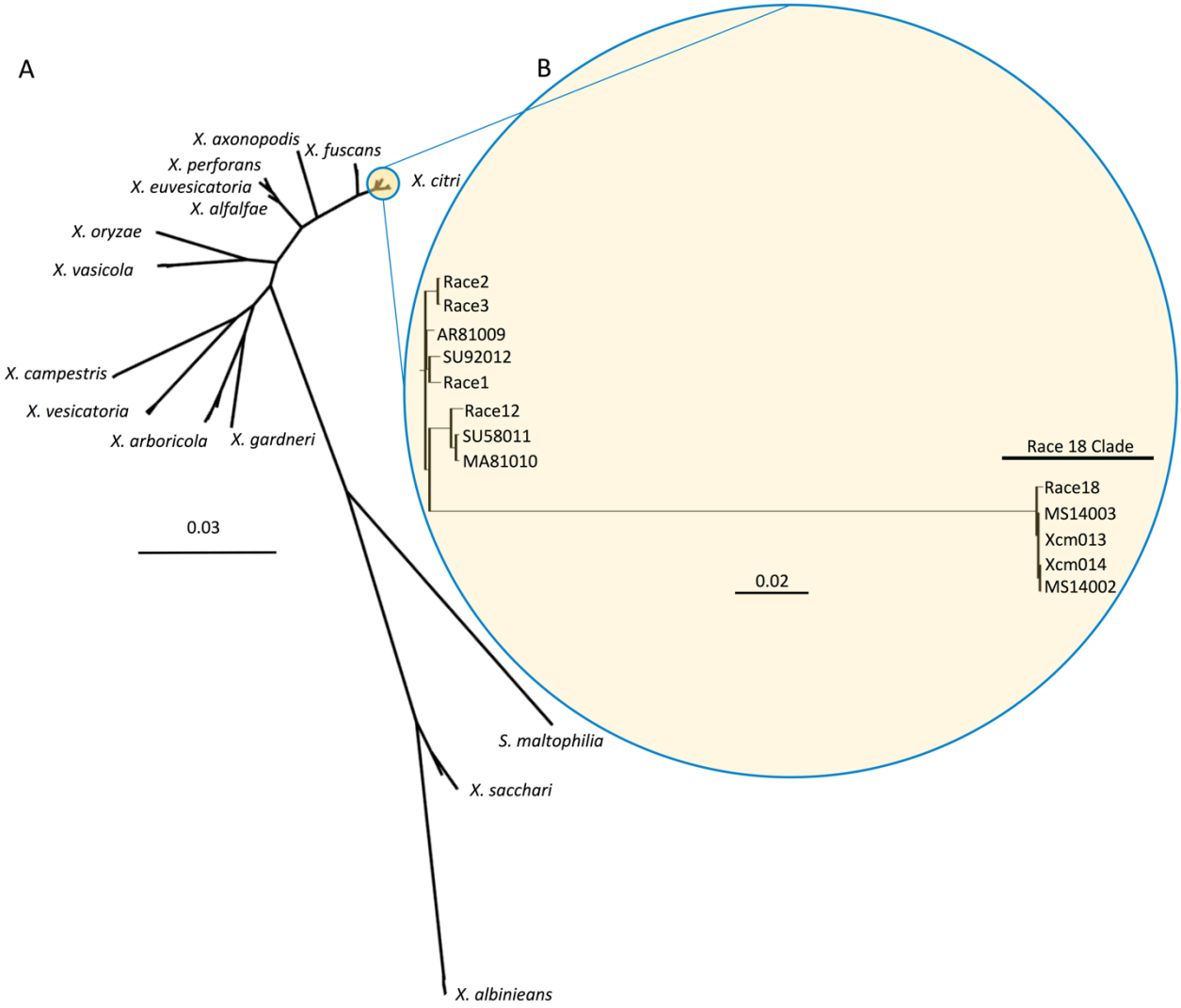


Figure 2: Phylogenetic analysis of *Xcm* isolates and 13 species of *Xanthomonas* A) MLST (Multi Locus Sequence Typing) and maximum likelihood analysis of 13 Illumina sequenced *Xcm* isolates (this paper) and 40 other Xanthomonads using concatenated sections of the gltA, lepA, lacF, gyrB, fusA and gap-1 loci. B) SNP based neighbor-joining tree generated from 17,853 variable loci between 13 *Xcm* isolates and the reference genome *Xanthomonas citri* subsp. *citri* strain Aw12879. The tree was made using the Simple Phylogeny tool from ClustalW2.

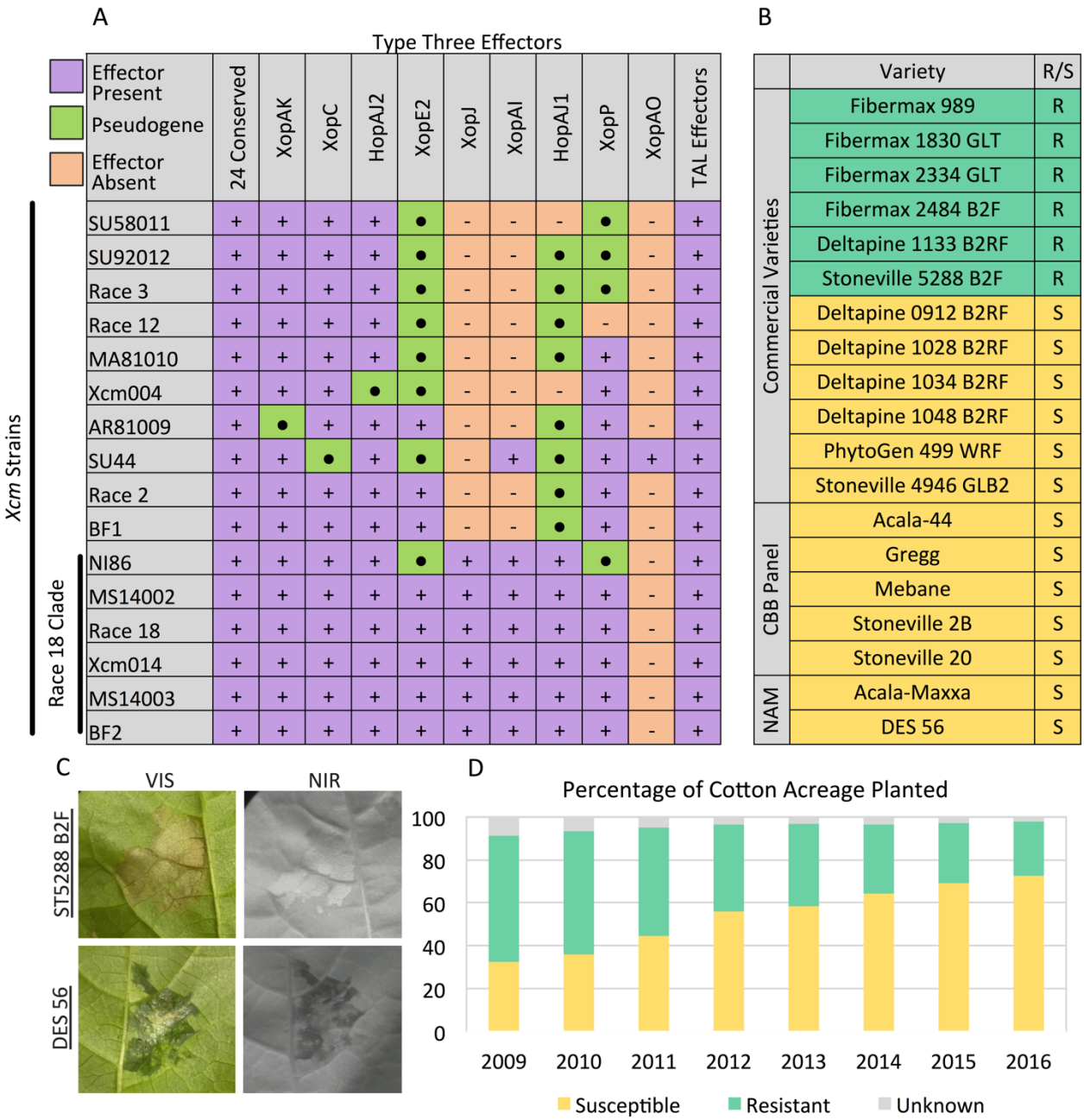


Figure 3: Molecular and phenotypic analysis of *Xcm* and *G. hirsutum* interactions. A) Type three effector profiles of *Xcm* isolates were deduced from *de novo*, Illumina based genome assemblies. Effector presence or absence was determined based on homology to known type three effectors using the program Prokka. B) Commercial and public *G. hirsutum* cultivars were inoculated with 13 *Xcm* isolates. Susceptible (S) indicates water soaking symptoms. Resistant (R) indicates a visible hypersensitive response. Plants were screened with a range of inoculum concentration from OD_{600} = 0.001-0.5. C) Disease symptoms on *G. hirsutum* cultivars Stoneville 5288 B2F and DES 56 after inoculation with *Xcm* strain AR81009 (OD_{600} = 0.05). Symptoms are visualized under visible (VIS) and near infrared (NIR) light. D) The proportion of US fields planted with susceptible and resistant cultivars of *G. hirsutum* was determined using planting acreage statistics from the USDA-AMA and disease phenotypes based on previous reports for common cultivars[34-36].

Analysis of the genomic sequence of T3Es revealed presence/absence differences, frameshifts and premature stop codons. However, this analysis does not preclude potential allelic or expression differences among the virulence proteins that could be contributing factors to the re-emergence of CBB. Therefore, newly isolated strains may harbor subtle genomic changes that have allowed them to overcome existing resistance phenotypes. Many commercial cultivars of cotton are reported to be resistant to CBB[30-32]. Based on these previous reports, we selected commercial cultivars resistant and susceptible (6 of each) to CBB. In addition, we included 5 available varieties that are related to the historical panel as well as 2 parents from a nested association mapping (NAM) population currently under development[33]. All varieties inoculated with the newly isolated *Xcm* strains exhibited inoculation phenotypes consistent with previous reports (Figs 3b,c). In these assays, bright field and near infrared (NIR) imaging were used to distinguish water-soaked disease symptoms from rapid cell death (HR) that is indicative of an immune response. These data confirm that existing resistance genes present within cotton germplasm are able to recognize the newly isolated *Xcm* strains and trigger a hypersensitive response. Together, the phylogenetic analysis, effector profile conservation and cotton inoculation phenotypes, confirm that the recent outbreak of *Xcm* in the US represents a re-emergence of a race 18 clade *Xcm* and is not the result of a dramatic shift in the pathogen.

Table 1: Illumina and SMRT sequenced *Xcm* genomes described in this paper.

Strain Name	Identifier	Country	Year	Platform	Contig #	Avg Contig Len	Total Bases	n50
MS14002		US	2014	Illumina	545	9443.27	5146580	62542
MS14003		US	2014	Illumina	2577	1511.35	3894744	2209
Race1		US		Illumina	523	10127.35	5296606	48599
Race2		US		Illumina	387	13402.57	5186796	54804
Race3		US		Illumina	725	7207.34	5225324	28344
Race12		US		Illumina	632	8134.35	5140911	21428
Race18		US		Illumina	369	13924.03	5137968	112543
AR81009	CFBP2035	Argentina	1981	Illumina	306	17182.59	5257872	86594
MA81010	CFBP2036	Mali	1981	Illumina	584	9033.09	5275326	23323
SU58011	CFBP2530	Sudan	1958	Illumina	1134	4563.33	5174819	9682
SU9012	CFBP5637	Sudan	1992	Illumina	377	13919.54	5247665	88522
Xcm013	MSCT4	US		Illumina	2169	2151.5	4666607	3869
Xcm014	MSCT8	US		Illumina	580	8929.58	5179156	88255
MS14003		US	2014	SMRT	4	1286176.5	5144706	5029617
AR81009	CFBP2035	Argentina	1981	SMRT	4	1352212	5408848	5267057

The USDA Agricultural Marketing Service (AMS) releases reports on the percentage of upland cotton cultivars planted in the U.S. each year (www.ams.usda.gov/mnreports/cnavar.pdf). Most of these varieties are screened for resistance or susceptibility to multiple strains of *Xcm* by extension scientists and published in news bulletins[30,31,34-38]. These distinct datasets were cross-referenced to reveal that only 25% of the total cotton acreage was planted with resistant cultivars in 2016 (Fig 3d, Table S3). This is part of a larger downward trend in which the acreage of resistant cultivars has fallen each year since at least 2009 when the percentage of acreage planted with resistant varieties was at 75%.

2.5.4 Comparative genome analysis for two *Xcm* strains

Differences in virulence were observed among *Xcm* strains at the molecular and phenotypic level. In order to gain insight into these differences, we selected two strains from our collection that differed in T3E content, virulence level, geography of origin and isolation date. AR81009 was isolated in Argentina in 1981 and is one of the most virulent strains investigated in

this study; MS14003 was isolated in Mississippi in 2014 and is a representative strain of the race 18 clade (Fig S2). The latter strain causes comparatively slower and diminished leaf symptoms; however, both strains are able to multiply and cause disease on susceptible varieties of cotton (Fig S3). Full genome sequences were generated with Single Molecule Real-Time (SMRT) sequencing. Genomes were assembled using the PacBio Falcon assembler, which yielded circular 5Mb genomes and associated plasmids. Genic synteny between the two strains was observed with the exception of two 1.05 Mb inversions (Fig 4). Regions of high and low GC content, indicative of horizontal gene transfer, were identified in both genomes. In particular, a 120kb insertion with low GC content was observed in AR81009. This region contains one T3E as well as two annotated type four secretion system-related genes, two conjugal transfer proteins, and two multi drug resistant genes (Fig 4 insert). MS14003 contains three plasmids (52.4, 47.4, and 15.3kb) while AR81009 contains two plasmids (92.6 and 22.9kb). Analysis of homologous regions among the plasmids was performed using progressiveMauve[39]. This identified four homologous regions greater than 1kb that were shared among multiple plasmids (Fig 4).

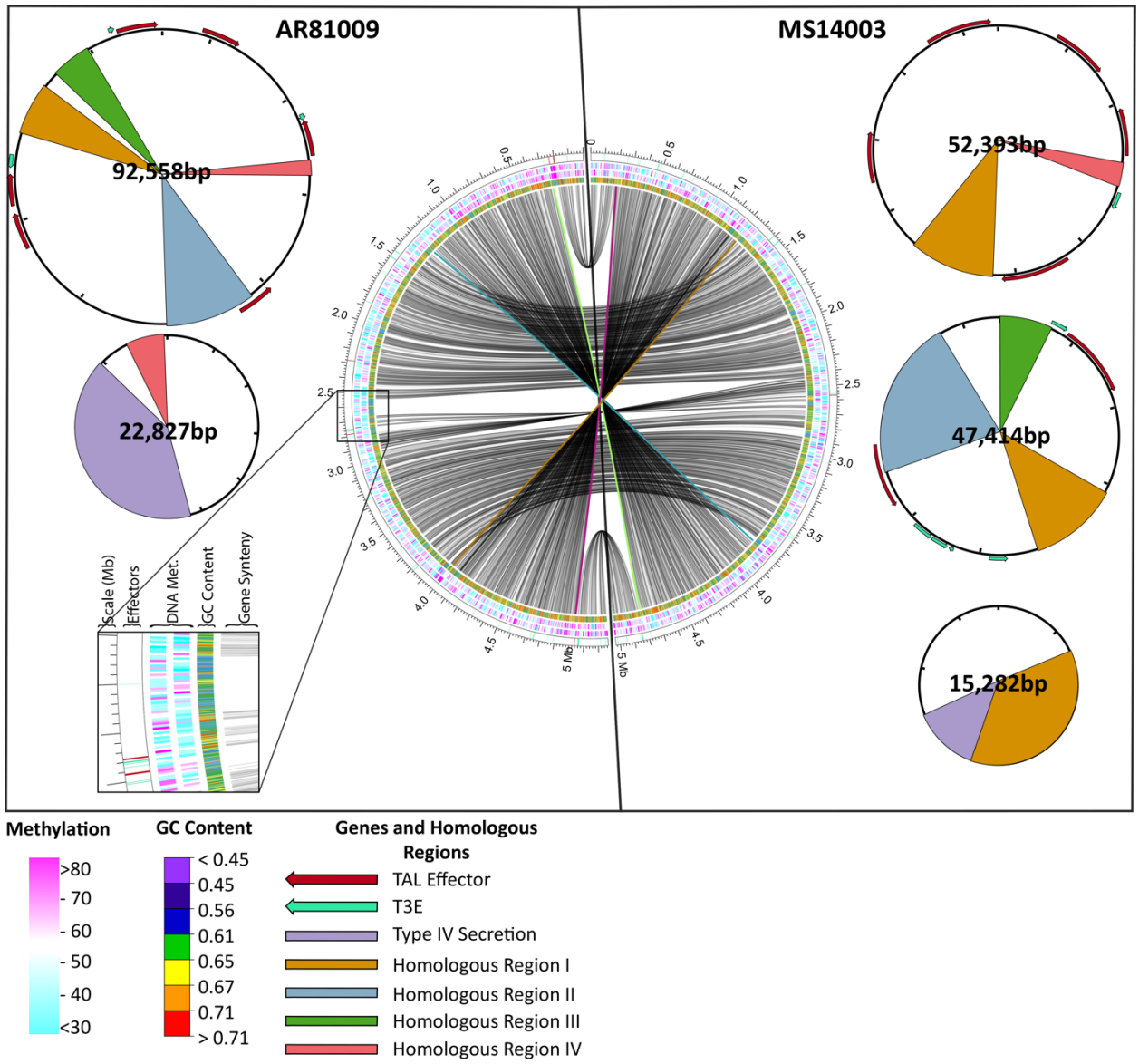


Figure 4: SMRT sequencing of two phenotypically and geographically diverse *Xcm* isolates: MS14003 and AR81009. Circos plot visualization of two circular *Xcm* genomes. Tracks are as follows from inside to outside: synteny of gene models; GC Content; DNA Methylation on + and – strands; location of type three effectors (teal) and TAL effectors (red), and position. On each side, accompanying plasmids are cartooned. Type three effector repertoires and the type IV secretion systems were annotated using Prokka. Homologous regions greater than 1kb were identified using MAUVE, and TAL effectors were annotated using AnnoTALE.

Both strains express TAL effector proteins as demonstrated through western blot analysis using a TAL effector specific polyclonal antibody (Fig 5)[40]. However, the complexity of TAL effector repertoires within these strains prevented complete resolution of each individual TAL

effector using Illumina sequencing. In contrast, the long reads obtained from SMRT sequencing are able to span whole TAL effectors, allowing for full assemblies of the TAL effectors in each strain. The AR81009 genome encodes twelve TAL effectors that range in size from twelve to twenty three repeat lengths, six of which reside on plasmids. The MS14003 genome encodes eight TAL effectors that range in size from fourteen to twenty eight repeat lengths, seven of which reside on plasmids (Fig 5). Three partial TAL effector-like coding sequences were also identified within these genomes and are presumed to be non-functional. A 1-repeat gene with reduced 5' and 3' regions was identified in both strains directly upstream of a complete TAL effector. In addition, a large 4kb TAL effector was identified in AR81009 with a 1.5 kb insertion and 10 complete repeat sequences. The tool AnnoTALE was used to annotate and group TAL effectors based on the identities of the repeat variable diresidues (RVDs) in each gene[41]. Little homology was identified among TAL RVD sequences within and between strains; only two TAL effectors were determined to be within the same TAL class between strains (TAL19b of AR81009 and TAL19 of MS14003) and two within strain MS14003 (TAL14b and TAL16).

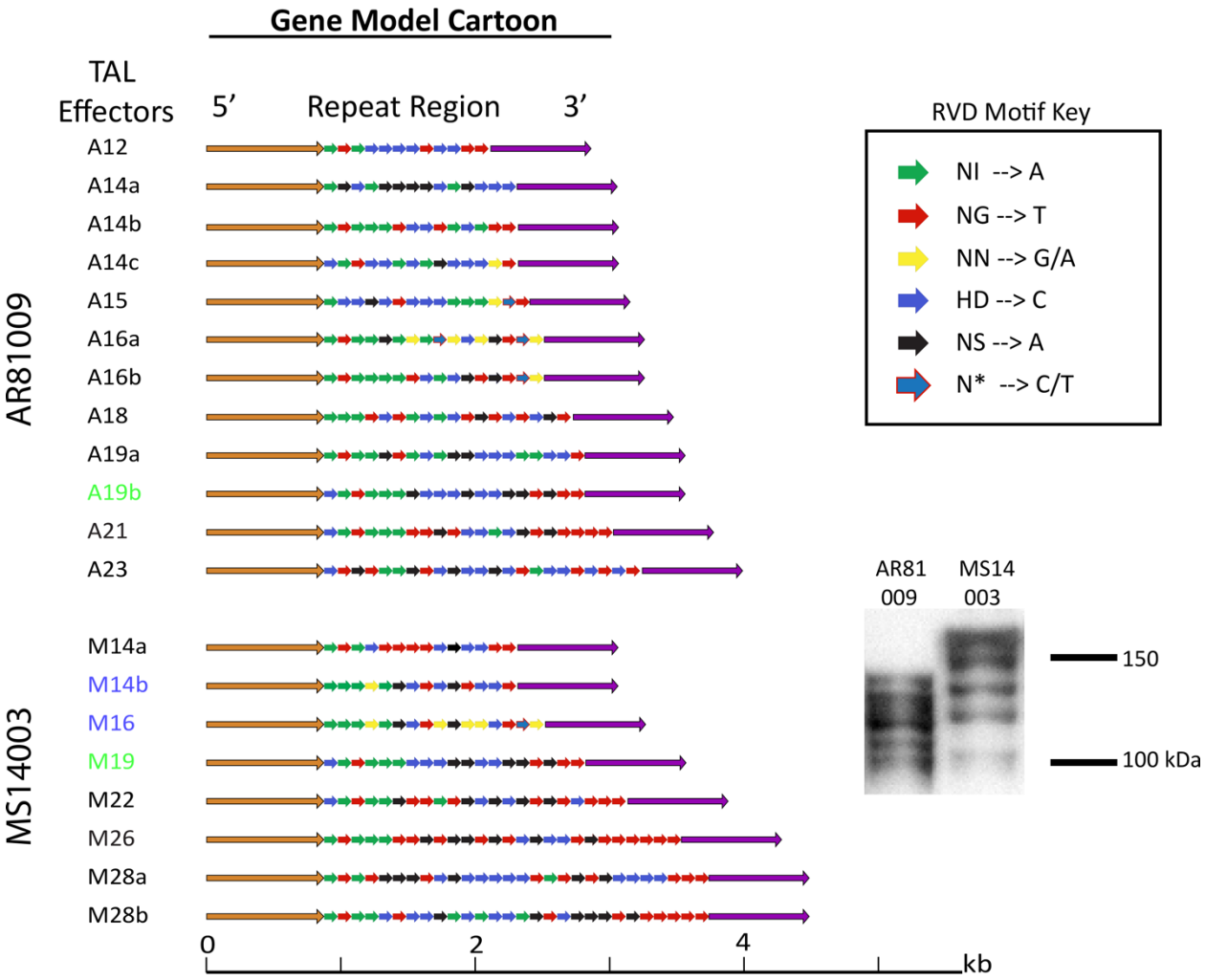


Figure 5: SMRT sequencing and western blot reveal diverse TAL effector repertoires between *Xcm* strains MS14003 and AR81009. Western Blot of TAL effectors using polyclonal TAL-specific antibody and gene models of TAL effectors identified by AnnoTALE. Blue and Green highlighted gene models represent TALs grouped in the same clade by repeat variable di-residue (RVD) sequence using AnnoTALE.

2.5.5 Transcriptome changes are induced by *Xcm* in *G. hirsutum*.

An RNA-sequencing experiment was designed to determine whether AR81009 and MS14003 incite different host responses during infection (Fig 6a, b). Isolates were inoculated into the phylogenetically diverse *G. hirsutum* cultivars Acala Maxxa and DES 56[33]. Infected and mock-treated tissue was collected at 24 and 48 hours post inoculation. First, we considered global transcriptome patterns of gene expression. Fifty-two genes were determined to be induced

in all *Xcm*-*G. hirsutum* interactions at 48 hours (Fig 6c, Table S4). Of note among this list is a homeologous pair of genes with homology to the known susceptibility target MLO[42-45]. Gene induction by a single strain was also observed; AR81009 and MS14003 uniquely induced 127 and 16 *G. hirsutum* genes, respectively (Fig 6c). In contrast, the average magnitude of gene induction between the two strains was not significantly different (Fig S4). Both *Xcm* strains caused more genes to be differentially expressed in DES 56 than in Acala Maxxa. Among the 52 genes significantly induced by both strains, sixteen conserved targets are homeologous pairs, whereas seventeen and fifteen genes are encoded by the A and D sub-genomes, respectively (Tables 2 and S4). It has been previously reported that homeologous genes encoded on the *G. hirsutum* A and D sub-genomes are differentially regulated during abiotic stress[46]. A set of approximately 10,000 homeologous gene pairs was selected and differential gene expression was assessed (Fig 7). For each pairwise comparison of *Xcm* strain and *G. hirsutum* cultivar, a similar number of genes were differentially expressed in each of the A and D subgenomes. However, some homeologous pairs were up- or down-regulated differentially in response to disease, indicating a level of sub-genome specific responses to disease. For example, SWEET sugar transporter gene Gh_D12G1898 in the D genome is induced over fourfold during infection with *Xcm* strain AR81009, while the homeolog Gh_A12G1747 in the A genome is induced to a much smaller extent.

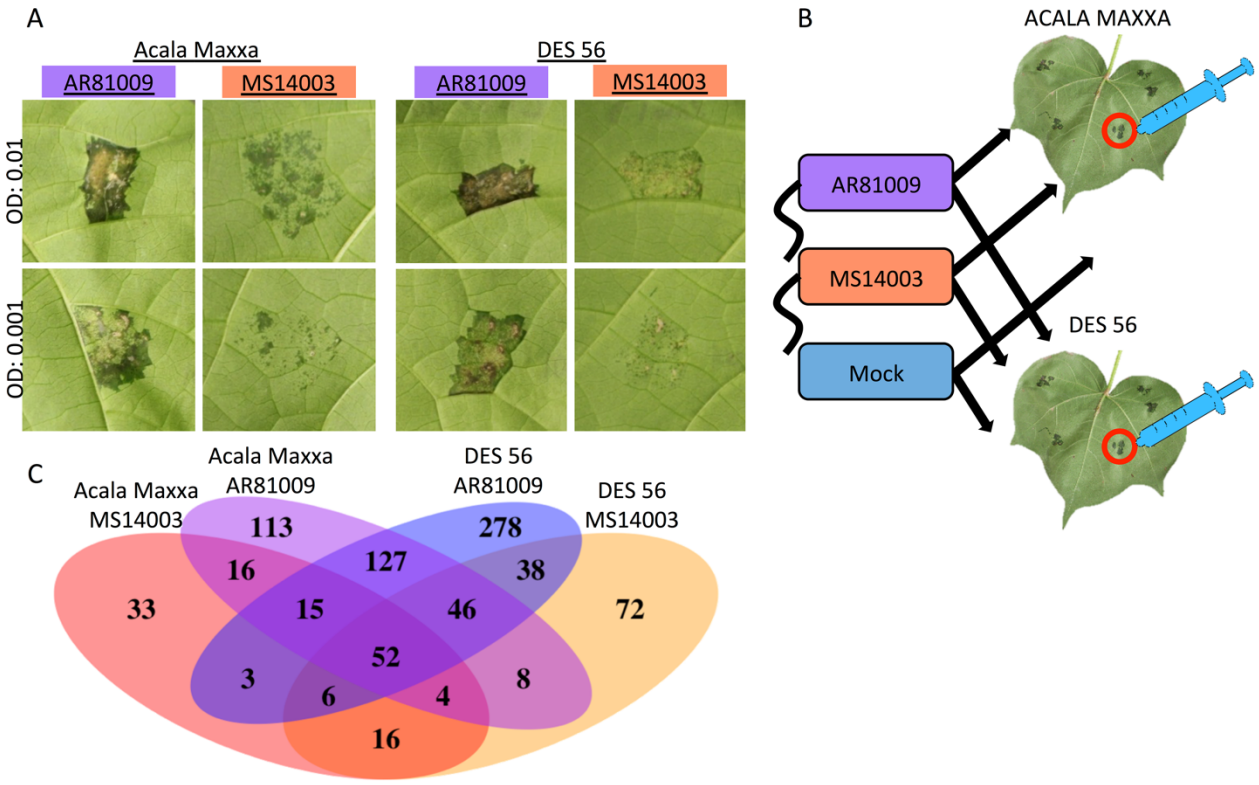


Figure 6: RNA-Sequencing analysis of infected *G. hirsutum* tissue demonstrates transcriptional changes during CBB. A) Disease phenotypes of *Xcm* strains MS14003 and AR81009 on *G. hirsutum* cultivars Acala Maxxa and DES 56, 7 days post inoculation. B) Acala Maxxa and DES 56 were inoculated with *Xcm* strains MS14003 and AR81009 at an OD of 0.5 and a mock treatment of 10mM $MgCl_2$. Inoculated leaf tissue was collected at 24 and 48 hpi (before disease symptoms emerged). C) Venn diagram of up regulated *G. hirsutum* genes (Log2(fold change in FPKM) ≥ 2 and p value ≤ 0.05) in response to *Xcm* inoculation. Venn diagram was created using the VennDiagram package in R.

2.5.6 Different strains of *Xcm* target distinct SWEET transporters in *G. hirsutum.*

SWEET sugar transporter genes have been reported to be targets of and up regulated by *Xanthomonas* TAL effectors in *Manihot esculenta*, *Oryza sativa*, and *Citrus sinensis*[21,40,47,48]. In rice and cassava, the SWEET genes are confirmed susceptibility genes that contribute to disease symptoms. The previously reported susceptibility genes and the SWEETs identified here, are clade III sugar transporters (Fig S5). The NBI *Gossypium hirsutum* genome encodes 54 putative SWEET sugar transporter genes. Of these 54 genes, three were up regulated greater than fourfold in response to inoculation by one of the two *Xcm* strains (Fig 8). Predicted TAL effector binding

sites were identified using the program TALEnt[49]. MS14003 significantly induces the homeologs Gh_A04G0861 and Gh_D04G1360 and contains the TAL effectors M14b, M28a, and M28b, which are predicted to bind within the 300bp promoter sequences of at least one of these genes. Of note is TAL M28a, which is predicted to bind both homeologs (Fig S6a). In contrast, AR81009 induces Gh_D12G1898 to a greater extent than its homeolog Gh_A12G1747. TAL effectors A14c and A16b from AR81009 are predicted to bind to the Gh_D12G1898 and Gh_A12G1747 promoters; however, TAL A14a is predicted to bind only the Gh_D12G1898 promoter (Fig S6b). We note that while Gh_A12G1747 did not pass the fourfold cut off for gene induction, this gene is slightly induced compared to mock inoculation.

Table 2: Eight homeologous pairs of *Gossypium hirsutum* genes are up regulated in both Acala Maxxa and DES 56 varieties 48 hours post inoculation with *Xanthomonas citri* pv. *malvacearum* strains MS14003 and AR81009.

A Genome	D Genome	Gene Annotation
Gh_A02G0615	Gh_D02G0670	Seven transmembrane MLO family protein
Gh_A03G0560	Gh_D03G0971	Pectate lyase family protein
Gh_A05G2012	Gh_D05G2256	Protein of unknown function DUF688
Gh_A06G0439	Gh_D06G0479	basic chitinase
Gh_A07G1129	Gh_D07G1229	Protein of unknown function (DUF1278)
Gh_A10G0257	Gh_D10G0257	Protein E6
Gh_A10G1075	Gh_D10G1437	Pectin lyase-like superfamily protein
Gh_A13G1467	Gh_D13G1816	pathogenesis-related 4

2.6 Discussion

Cotton Bacterial Blight was considered controlled in the U.S. until an outbreak was observed during the 2011 growing season in Missouri, Mississippi and Arkansas[50]. Until 2011, seed sterilization, breeding for resistant varieties, and farming techniques such as crop rotation and sterilizing equipment prevented the disease from becoming an economic concern[51]. The number of counties reporting incidence of CBB has increased from 17 counties in 2011 to 77 counties in 2015[38,52,53]. This paper investigates the root of the re-emergence and identifies several routes towards control of the disease.

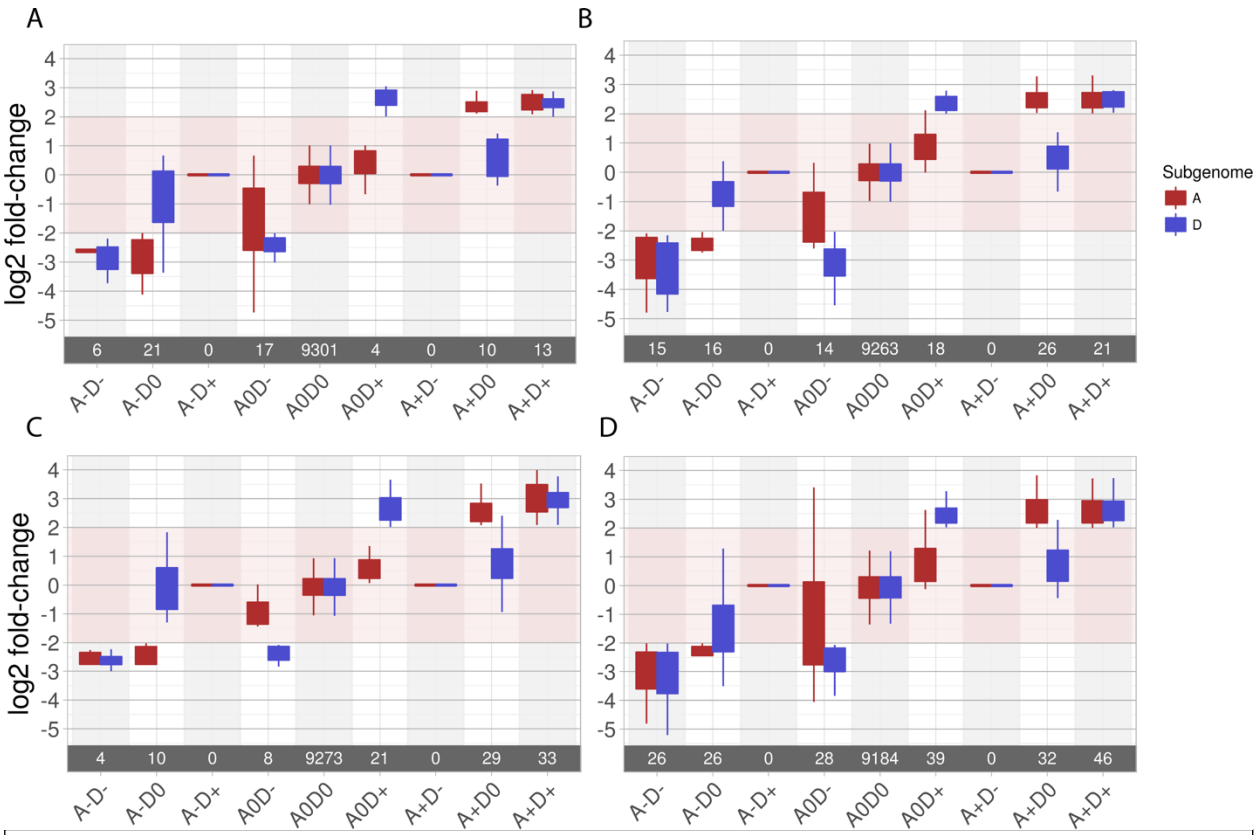


Figure 7: Expression of homeologous pairs across the A and D *G. hirsutum* genomes in response to *Xcm* inoculation. Genes are considered up or down regulated if the absolute value of gene expression change after inoculation as compared to mock treatment was Log2(fold change in FPKM) ≥ 2 and p value ≤ 0.05. By these criteria, pink shading indicates no significant gene expression change. A-D-: both members of the homeologous gene pair are down regulated; A-D0: only the 'A' sub-genome homeolog is down regulated; A-D+: 'A' sub-genome homeolog is down regulated, 'D' sub-genome homeolog is up regulated; etc. Number of gene pairs (n) meeting each expression pattern is indicated within the grey bar. For all genes meeting each expression pattern, the distribution of expression patterns is displayed as a box plot. Rectangles indicate the interquartile range and the whiskers show 1.5 times the interquartile range. A) Acala Maxxa inoculated with MS14003 B) DES 56 inoculated with MS14003 C) Acala Maxxa inoculated with AR 81009 D) DES 56 inoculated with AR81009.

When CBB was first recognized as re-emerging, several possible explanations were proposed including: (1) A highly virulent race of the pathogen that had been introduced to the U.S.; (2) Historical strains of *Xcm* that had evolved to overcome existing resistance (e.g. an effector gene change or host shift); and (3) Environmental conditions over the last several years that had been particularly conducive to the disease. Here, we present evidence that the re-emergence of CBB is not due to a large genetic change or race shift in the pathogen. Rather, the

re-emergence of the disease is likely due to agricultural factors such as large areas of susceptible cultivars being planted. The presented data do not rule out potential environmental conditions that may also have contributed to the re-emergence. In this context, environmental conditions include disease conducive temperature and humidity as well as potentially contaminated seed or other agronomic practices that may have perpetuated spread of the disease outbreaks. Importantly, the presented data confirm that the presence of resistance loci could be deployed to prevent further spread of this disease. However, since many of the most popular farmer preferred varieties lack these resistance traits, additional breeding or biotechnology strategies will be needed to maximize utility. Notably, the current *Xcm* isolates characterized in this study all originate from Mississippi cotton fields in 2014. During the 2015 and 2016 growing seasons, resistant cotton cultivars were observed in Texas with symptoms indicative of bacterial infection distinct from CBB. Additional work is underway to identify and characterize the causal agent(s) of these disease symptoms.

However, races are not necessarily phylogenetically distinct clades. Race 18 isolates have been reported overseas, indicating that there may be independent origins of the race or cross-continent movement of this pathogen. Phenotypic race delineations were created before modern genetic and phylogenetic techniques were developed. However, modern genetics presents the opportunity to begin classifying strains based upon phylogenetic and effector profiles rather than phenotypes on a limited range of host varieties. Here, we identify all known and putative race 18 isolates as phylogenetically grouped into a single clade and distinct from other *Xcm* isolates. Future efforts can further explore phylogenetic relatedness among diverse isolates.

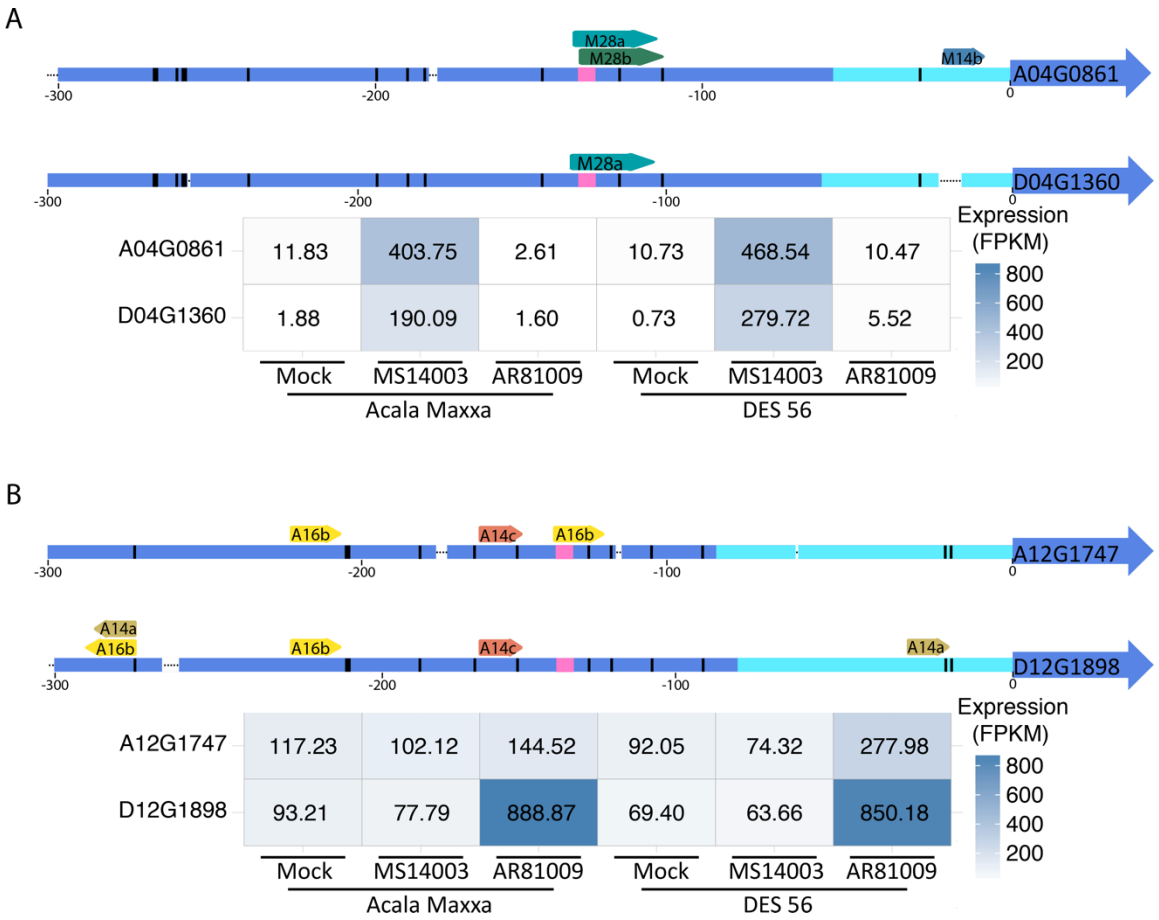


Figure 8: Three candidate *G. hirsutum* susceptibility genes are targeted by two different *Xcm* strains. Recent work on CBB in the US has focused on the most prevalent US *Xcm* race: race 18. A) The homeologous pair of SWEET genes A04_G0861 and D04_G1360 are up regulated in the presence of *Xcm* strain MS14003. (top) Cartoon summary of 300bp promoters of A04_G0861 and D04_G1360. (bottom) Heat-map of the expressions of A04_G0861 and D04_G1360 48 hours after mock or *Xcm* inoculation. B) The SWEET gene D12_G1898 is up regulated in the presence of *Xcm* strain AR81009. (top) Cartoon summary of 300bp promoters of D12_G1898 and A12_G1747. (bottom) Heat-map of the expressions of A12_G1747 and D12_G1898 48 hours after mock or *Xcm* inoculation. TAL effector binding sites were predicted with TALEsf using a quality score cutoff of 4. Gene promoter cartoon legend: Arrow: TAL effector binding site; Black dot: Deletion; Black bar: SNP; Pink bar: TATA box; Teal section: 5'UTR.

While resistant cotton cultivars were identified for all strains in this study, variability in symptom severity was observed for different strains when inoculated into susceptible cultivars. Two strains in particular, MS14003 and AR81009, have different effector profiles as well as different disease phenotypes. Comparative genomic analysis of the two pathogens revealed many

differences that may contribute to the relative disease severity phenotypes. Similarly, transcriptomic analysis of two cultivars of *G. hirsutum* inoculated with these strains confirms that the genomic differences between the two strains result in a divergence in their molecular targets in the host.

Over the past decade, susceptibility genes have become targets for developing disease tolerant plants[54,55]. These genes are typically highly induced during infection[56]. Therefore, RNA-Seq of infected plants has become a preferred way to identify candidate susceptibility genes. Once identified, genome editing can be used to block induction of these genes[57]. We report a homeologous pair of genes that are homologs of the MLO gene as targeted by both *Xcm* strains in both cotton cultivars. These genes are excellent candidates for future biotechnology efforts. Because the potential importance of these genes in cotton biology is unknown, their role in cotton physiology must first be explored. Knock-out mutations of MLO genes in other systems has led to durable resistance against powdery mildew as well as oomycetes and bacteria such as *Xanthomonas*[42,45]. The dual purpose of host susceptibility genes has been observed previously. For example, the rice *Xa13* (aka. *Os8N3* and *OsSWEET11*) gene is required for pollen development but also targeted by a rice pathogen during infection[58]. Xa13 is a member of the clade III SWEET sugar transporters implicated in many pathosystems. In this case, the induction of *Xa13* for pathogen susceptibility is mediated by a TAL effector. Of the 54 SWEET genes in the *G. hirsutum* genome, at least three are significantly up regulated during *Xcm* infection. In contrast to MLO, no single SWEET gene was induced by both pathogen strains in both hosts.

Analysis of SWEET gene expression after inoculation revealed a context for polyploidy in the *G. hirsutum*-*Xcm* pathosystem. This relatively unexplored area of plant-microbe interactions arose from our observation of a potential difference in induction magnitude between

the homeologous Gh_A12G1747 and Gh_D12G1898 SWEET genes. Further analysis revealed many examples of preferentially induced or down-regulated homeologs in response to *Xcm* infection. Characterization of sub-genome specialization may lead to new insights regarding durability of resistance and susceptibility loci in polyploid crops. Future research may investigate the diploid ancestors of tetraploid cotton to further explore the evolution of host and pathogen in the context of ploidy events[59].

Multiple putative TAL effector binding sites were identified within each up-regulated SWEET promoter. These observations suggest that TAL M28a from MS14003 may induce the homeologs Gh_A04G0861 and Gh_D04G1360. Further, TAL effector A14a from AR81009 is likely responsible for the up regulation of Gh_D12G1898. Whether additional TAL effectors are involved in these responses is not clear. Genome organization in the host, such as histone modifications or other epigenetic regulations may also be affecting these interactions. Future research will investigate these mechanisms further.

Collectively, the data presented here suggest that the widespread planting of CBB-susceptible cultivars has contributed to the re-emergence of CBB in the southern U.S. It is possible that a reservoir of race 18 *Xcm* was maintained in cotton fields below the level of detection due to resistant cultivars planted in the 1990s and early 2000s. Alternatively, the pathogen may have persisted on an alternate host or was re-introduced by contaminated seed[9,10]. Regardless of the cause of the re-emergence, the genomic comparisons among pathogen races and host cultivars have identified several possible routes towards resistance. These include the use of existing effective resistance loci as well as the potential disruption of the induction of susceptibility genes through genome editing. The latter is an attractive strategy in part because of recent progress in genome editing[60,61]. In summary, within a relatively short time frame, through

the deployment of modern molecular and genomic techniques, we were able to identify factors that likely contribute to the re-emergence of cotton bacterial blight and generate data that can now be rapidly translated to effective disease control strategies.

2.7 Materials & methods

2.7.1 *Xcm* strain isolation and manipulation

New *Xcm* strains were isolated from infected cotton leaves by grinding tissue in 10mM $MgCl_2$ and culturing bacteria on NYGA media. The most abundant colony type was selected, single colony purified and then 16S sequencing was used to confirm the bacterial genus as previously described[62]. In addition, single colony purified strains were re-inoculated into cotton leaves and the appearance of water soaked symptoms indicative of CBB infection was confirmed. Both newly isolated strains as well as strains received from collaborators were used to generate a rifampicin resistance version of each strain. Wild-type strains were grown on NYGA, then transferred to NYGA containing 100µg/ml rifampicin. After approximately 4-5 days, single colonies emerged. These were single colony purified and stored at -80C. The rifampicin resistant version of each *Xcm* strain was used in all subsequent experiments reported in this manuscript unless otherwise noted.

2.7.2 Plant inoculations

Cotton varieties from the original cotton panel for determining *Xcm* race designations were obtained from the USDA/ARS, Germplasm Resources Information Network (GRIN). Varieties included in the *G. hirsutum* NAM population were provided by Vasu Kuraparthy[33]. Other commercial varieties were obtained from Terry Wheeler and Tom Allen. Disease assays were conducted in a growth chamber set at 30°C and 80% humidity. *Xcm* strains were grown on NYGA plates containing 100µg/ml rifampicin at 30°C for two days before inoculations were

performed. Inoculations were conducted by infiltrating a fully expanded leaf with a bacterial solution in 10mM $MgCl_2$ (OD_{600} specified within each assay).

The field tests were conducted as follows: Cotton cultivars are planted in two row plots (10 – 11 m in length, 1 m row spacing), in a randomized complete block design with four replications. Approximately 60 to 80 days after planting, *Xcm* was applied to the test area similar to that described in Wheeler et al. (2007)[37]. Briefly, Xcm is grown in trypticase soy broth (30 g/L) for 1 ½ days and then 19 L of the concentrated bacterial solution (10^8 cfu/ml) are diluted into 189 L of water (resulting in 10^6 cfu/ml) . The surfactant Silwet L-77 (polyalkyleneoxide modified heptamethyltrisiloxane, Loveland Industries, Greely, CO) is added at 0.2% v/v. The suspension of bacteria is sprayed over the top of the cotton at a pressure of 83 kpa and rate of 470 L/ha. The nozzles used were TeeJet 8008. Symptoms were typically visible 14 days after application and plots were rated for incidence of symptoms 17-21 days after application[34-37].

2.7.3 Cotton cultivar statistics

Area of cotton planted per county in the United States in 2015 was obtained from the USDA National Agricultural Statistics Service: www.nass.usda.gov/Statistics_by_Subject/result.php?7061F36A-A4C6-3C65-BD7F-129B702CFBA2&sector=CROPS&group=FIELD%20CROPS&comm=COTTONUSDA. Estimated percentage of upland cotton planted for each variety was obtained from the Agricultural Marketing Service (AMS): www.ams.usda.gov/mnreports/canvar.pdf.

2.7.4 Bacterial sequencing and phylogenetics

Illumina based genomic datasets were generated as previously described[29]. Paired-end Illumina reads were trimmed using Trimmomatic v0.32 (ILLUMINACLIP:TruSeq3-

PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36)[63]. Genome assemblies were generated using the SPAdes *de novo* genome assembler[64]. Strain information is reported in Supplemental Table 1. Similar to our previously published methods[29], the program Prokka was used in conjunction with a T3E database to identify type three effector repertoires for each of the 12 *Xcm* isolates as well as four *Xcm* genomes previously deposited on NCBI (S2Table)[65].

Multi-locus sequence analysis was conducted by concatenating sequences of the *gltA*, *lepA*, *lacF*, *gyrB*, *fusA* and *gap-1* loci obtained from the Plant-Associated Microbes Database (PAMDB) for each strain as previously described[66]. A maximum-likelihood tree using these concatenated sequences was generated using CLC Genomics 7.5.

2.7.5 Variant based phylogeny

A variant based dendrogram was created by comparing 12 Illumina sequenced *Xcm* genomes to the complete *Xanthomonas citri* subsp. *citri* strain Aw12879 reference genome (Genbank assembly accession: GCA_000349225.1) on NCBI. Read pairs were aligned to the reference genome using Bowtie2 v2.2.9 with default alignment parameters[27]. From these alignments, single nucleotide polymorphisms (SNPs) were identified using samtools mpileup v1.3 and the bcftools call v1.3.1 multi-allelic caller[28]. Using Python v2.7, the output from samtools mpileup was used to identify loci in the *X. citri* subsp. *citri* reference genome with a minimum coverage of 10 reads in each *Xcm* genome used Python version 2.7 available at http://www.python.org. Vcftools v0.1.14 and bedtools v2.25.0 were used in combination to remove sites marked as insertions or deletions, low quality, or heterozygous in any of the genomes[67,68]. Remaining loci were concatenated to create a FASTA alignment of confident loci. Reference loci were used where SNP's were not detected in a genome. The resulting FASTA

alignment contained 17853 loci per strain. This alignment was loaded into the online Simple Phylogeny Tool from the ClustalW2 package to create a neighbor joining tree of the assessed strains[69,70]. Trees were visualized using FigTree v1.4.2.

2.7.6 Genome assembly

Single Molecule, Real Time (SMRT) sequencing of *Xcm* strains MS14003 and AR81009 was obtained from DNA prepped using a standard CTAB DNA preparation. Blue Pippin size selection and library preparation was done at the University of Delaware Sequencing Facility. The genomes were assembled using FALCON-Integrate (https://github.com/PacificBiosciences/FALCON-integrate/commit/cd9e93)[71]. The following parameters were used: Assembly parameters for MS14003: length_cutoff = 7000; length_cutoff_pr = 7000; pa_HPCdaligner_option = -v -dal8 -t16 -e.70 -l2000 -s240 -M10; ovlp_HPCdaligner_option = -v -dal8 -t32 -h60 -e.96 -l2000 -s240 -M10; falcon_sense_option = --output_multi --min_idt 0.70 --min_cov 5 --local_match_count_threshold 2 --max_n_read 300 --n_core 6; overlap_filtering_setting = --max_diff 80 --max_cov 160 --min_cov 5 --bestn 10; Assembly parameters for AR81009: length_cutoff = 8000; length_cutoff_pr = 8000; pa_HPCdaligner_option = -v -dal8 -t16 -e.72 -l2000 -s240 -M10; ovlp_HPCdaligner_option = -v -dal8 -t32 -h60 -e.96 -l2000 -s240 -M10; falcon_sense_option = --output_multi --min_idt 0.72 --min_cov 4 --local_match_count_threshold 2 --max_n_read 320 --n_core 6; overlap_filtering_setting = --max_diff 90 --max_cov 300 --min_cov 10 --bestn 10. Assemblies were polished using iterations of pbalign and quiver, which can be found at https://github.com/PacificBiosciences/pbalign/commit/cda7abb and https://github.com/PacificBiosciences/GenomicConsensus/commit/43775fa. Two iterations were run for *Xcm* strain MS14003 and 3 iterations for AR81009. Chromosomes were then reoriented

to the DnaA gene and plasmids were reoriented to ParA. The assemblies were checked for overlap using BLAST, and trimmed to circularize the sequences[72]. TAL effectors were annotated and grouped by RVD sequences using AnnoTALE[41]. Homologous regions among plasmids that are greater than 1 kb were determined using progressiveMauve[39]. Genomic comparisons between the MS14003 and AR81009 chromosomes were visualized using Circos[73]. Single-copy genes on each of the chromosomes were identified and joined using their annotated gene IDs. Lines connecting the two chromosomes represent these common genes and their respective positions in each genome. A sliding window of 1KB was used to determine the average GC content. Methylation was determined using the Base Modification and Motif Analysis workflow from pbsmrtpipe v0.42.0 at https://github.com/PacificBiosciences/pbsmrtpipe.

2.7.7 Western blot analysis

Western Blot analysis of Transcription Activator-Like (TAL) effectors was performed using a polyclonal TAL specific antibody[40]. Briefly, bacteria were suspended in 5.4 pH minimal media for 4.5 hours to induce effector production and secretion. Bacteria were pelleted and then suspended in laemmli buffer and incubated at 95 degrees Celsius for three minutes to lyse the cells. Freshly boiled samples were loaded onto a 4-6% gradient gel and run for several hours to ensure sufficient separation of the different sized TAL effectors.

2.7.8 Gene expression analysis

Susceptible cotton were inoculated with *Xcm* using a needleless syringe at an OD_{600} of 0.5. Infected and mock-treated tissue were collected and flash frozen at 24 and 48 hours post inoculation. RNA was extracted using the Sigma tRNA kit. RNA-sequencing libraries were generated as previously described[74].

Raw reads were trimmed using Trimmomatic[63]. The Tuxedo Suite was used for mapping reads to the TM-1 NBI *Gossypium hirsutum* genome[75], assembling transcripts, and quantifying differential expression[27].

Read mapping identified several mis-annotated SWEET genes that skewed differential expression results. The annotations of SWEET genes Gh_A12G1747, Gh_D07G0487, and Gh_D12G1898 were shortened to exclude 20-30kb introns. Two exons were added to Gh_D05G1488. The 2.7kb scaffold named Scaffold013374 was also removed from analysis because its gene Gh_Sca013374G01 has exact sequence homology to Gh_A12G1747 and created multi-mapped reads that interfered with expression analysis.

Homeologous pairs were identified based on syntenic regions with MCScan[76]. A syntenic region was defined as a region with a minimum of five genes with an average intergenic distance of two and within extended distance of 40. All other values were set to the default. Comparisons between homeologs was performed by examining cuffdiff differential expression and classifying them according to the sub-genome expression pattern. Genes considered up or down regulated meet both differential expression from mock significance of q-value < 0.05 and the absolute value of the log2 fold change is greater than 2.

2.7.9 TAL binding sites

Bioinformatic prediction of TAL effector binding sites on the *G. hirsutum* promoterome was performed using the TAL Effector-Nucleotide Targeter (TALEnt)[50]. In short, the regions of the genome that were within 300 base pairs of annotated genes were queried with the RVD's of MS14003 and AR81009 using a cutoff score of 4. Promiscuously binding TALs 16 from MS14003 and 16a from AR81009 were removed from analysis.

Acknowledgements

The authors would like to acknowledge Dr. Robert Nichols for useful discussions throughout the presented research and preparation of this manuscript.

2.8 References

1. Cotton: World Markets and Trade [Internet]. 2016 [cited August 2016]. Available from: http://apps.fas.usda.gov/psdonline/circulars/cotton.pdf.

2. World Agricultural Supply and Demand Estimates [Internet]. 2016 [cited August 12, 2016]. Available from: http://usda.mannlib.cornell.edu/usda/current/wasde/wasde-08-12-2016.pdf.

3. Blank LM, Bird LS. Breeding Strains of Cotton Resistant to Bacterial Blight. Texas: Texas Agricultural Experiment Station; 1951. p. 1-25.

4. Green JM, Brinkerhoff LA. Inheritance of Three Genes for Bacterial Blight Resistance in Upland Cotton1. Agronomy Journal. 1956;48(11):481-5. doi: 10.2134/agronj1956.00021962004800110001x.

5. Simpson D, Weindling R. Bacterial blight resistance in a strain of Stoneville cotton. Journal of the American Society of Agronomy. 1946.

6. Huang X, Zhai J, Luo Y, Rudolph K. Identification of a highly virulent strain of Xanthomonas axonopodis pv. malvacearum. European Journal of Plant Pathology. 2008;122(4):461-9. doi: 10.1007/s10658-008-9312-3.

7. Cunnac S, Bolot S, Forero Serna N, Ortiz E, Szurek B, Noel LD, et al. High-Quality Draft Genome Sequences of Two Xanthomonas citri pv. malvacearum Strains. Genome Announc. 2013;1(4). doi: 10.1128/genomeA.00674-13. PubMed PMID: 23990578; PubMed Central PMCID: PMCPMC3757453.

8. Delannoy E, Lyon BR, Marmey P, Jalloul A, Daniel JF, Montillet JL, et al. Resistance of cotton towards Xanthomonas campestris pv. malvacearum. Annu Rev Phytopathol. 2005;43:63-82. doi: 10.1146/annurev.phyto.43.040204.140251. PubMed PMID: 16078877.

9. Brinkerhoff LA. Variation in Xanthomonas malvacearum and its relation to control. Annu Rev Phytopathol. 1970;8:85-110.

10. Verma JP. Bacterial Blight of Cotton. CRC Press1986.

11. Hunter RE, Brinkerhoff LA, Bird LS. The Development of a Set of Upland Cotton Lines for Differentiating Races of Xanthomonas malvacearum. Phytopathology. 1968;58:830-2.

12. Baldwin Jr C, editor Report of the Bacterial Blight Committee--1975 [Cotton, USA]. Proceedings Beltwide Cotton Production Research Conferences; 1976.

13. Essenberg M, Bayles MB, Pierce ML, Verhalen LM. Pyramiding B genes in cotton achieves broader but not always higher resistance to bacterial blight. Phytopathology. 2014;104(10):1088-97. doi: 10.1094/PHYTO-06-13-0167-R. PubMed PMID: 24655289.

14. Gabriel DW, Burges A, Lazo GR. Gene-for-gene interactions of five cloned avirulence genes from Xanthomonas campestris pv. malvacearum with specific resistance genes in cotton. Proceedings of the National Academy of Sciences. 1986;83:6415-9.

15. Schwartz AR, Potnis N, Timilsina S, Wilson M, Patané J, Martins J, et al. Phylogenomics of Xanthomonas field strains infecting pepper and tomato reveals diversity in effector repertoires and identifies determinants of host specificity. Frontiers in Microbiology. 2015;6. doi: 10.3389/fmicb.2015.00535.

16. Jacques MA, Arlat M, Boulanger A, Boureau T, Carrere S, Cesbron S, et al. Using Ecology, Physiology, and Genomics to Understand Host Specificity in Xanthomonas. Annu Rev Phytopathol. 2016;54:163-87. doi: 10.1146/annurev-phyto-080615-100147. PubMed PMID: 27296145.

17. Macho AP, Zipfel C. Targeting of plant pattern recognition receptor-triggered immunity by bacterial type-III secretion system effectors. Current Opinion in Microbiology. 2015;23:14-22. doi: http://dx.doi.org/10.1016/j.mib.2014.10.009.

18. Kim J-G, Stork W, Mudgett Mary B. Xanthomonas Type III Effector XopD Desumoylates Tomato Transcription Factor SlERF4 to Suppress Ethylene Responses and Promote Pathogen Growth. Cell Host & Microbe. 2013;13(2):143-54. doi: http://dx.doi.org/10.1016/j.chom.2013.01.006.

19. Wang S, Sun J, Fan F, Tan Z, Zou Y, Lu D. A Xanthomonas oryzae pv. oryzae effector, XopR, associates with receptor-like cytoplasmic kinases and suppresses PAMP-triggered stomatal closure. Science China Life Sciences. 2016:1-8. doi: 10.1007/s11427-016-5106-6.

20. Doyle EL, Stoddard BL, Voytas DF, Bogdanove AJ. TAL effectors: highly adaptable phytobacterial virulence factors and readily engineered DNA-targeting proteins. Trends Cell Biol. 2013;23(8):390-8. doi: 10.1016/j.tcb.2013.04.003. PubMed PMID: 23707478; PubMed Central PMCID: PMCPMC3729746.

21. Streubel J, Pesce C, Hutin M, Koebnik R, Boch J, Szurek B. Five phylogenetically close rice SWEET genes confer TAL effector-mediated susceptibility to Xanthomonas oryzae pv. oryzae. New Phytol. 2013;200(3):808-19. doi: 10.1111/nph.12411. PubMed PMID: 23879865.

22. Schreiber KJ, Baudin M, Hassan JA, Lewis JD. Die another day: Molecular mechanisms of effector-triggered immunity elicited by type III secreted effector proteins. Seminars in Cell & Developmental Biology. 2016;56:124-33. doi: http://dx.doi.org/10.1016/j.semcdb.2016.05.001.

23. Lee HA, Yeom SI. Plant NB-LRR proteins: tightly regulated sensors in a complex manner. Brief Funct Genomics. 2015;14(4):233-42. doi: 10.1093/bfgp/elv012. PubMed PMID: 25825425.

24. Flor HH. Current Status of the Gene-for-Gene Concept. Annu Rev Phytopathol. 1971;9:275-96.

25. Cui H, Tsuda K, Parker JE. Effector-Triggered Immunity: From Pathogen Perception to Robust Defense. Annual Review of Plant Biology. 2015;66(1):487-511. doi: 10.1146/annurev-arplant-050213-040012.

26. Koch R. Untersuchungen ueber Bakterien V. Die Aetiologie der Milzbrand-Krankheit, begruendent auf die Entwicklungsgeschichte des Bacillus Anthracis. Beitr. z. Biol. D. Pflanzen 2: 277-310. Milestones in Microbiology. 1876;1556.

27. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012;7(3):562-78. doi: 10.1038/nprot.2012.016. PubMed PMID: 22383036; PubMed Central PMCID: PMCPMC3334321.

28. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9. doi: 10.1093/bioinformatics/btp352.

29. Bart R, Cohn M, Kassen A, McCallum EJ, Shybut M, Petriello A, et al. High-throughput genomic sequencing of cassava bacterial blight strains identifies conserved effectors to target for durable resistance. Proceedings of the National Academy of Sciences. 2012;109(32):13130-. doi: 10.1073/pnas.1211014109.

30. Dever JK, Wheeler TA, Boman RK, Kerns D, Foster M, Nesmith D, et al. Cotton Performance Tests in the Texas High Plains and Trans Pecos Areas of Texas --2009. Texas AgriLife Research and Extension Service Center. 2010;10(2):1-74.

31. Dever JK, Morgan V, Kelly CM, Wheeler TA, Elkins H, Mendoza V, et al. Cotton Performance Tests in the Texas High Plains, 2014. Texas AgriLife Research and Extension Service Center. 2015;15(1).

32. Organization CSaIR, inventorCotton Variety Fibermax 9892004 March 3, 2004.

33. Tyagi P, Gore MA, Bowman DT, Campbell BT, Udall JA, Kuraparthy V. Genetic diversity and population structure in the US Upland cotton (Gossypium hirsutum L.).

Theor Appl Genet. 2014;127(2):283-95. doi: 10.1007/s00122-013-2217-3. PubMed PMID: 24170350.

34. Wheeler TA, Woodward JE. Response of cotton varieties to bacterial blight race 18 ain 2016. Texas AgriLife Research and Extension Service Center. 2016. doi: Lubbock.tamu.edu/files/2016/12/Bacterial-blight-trials-2016-2.pdf.

35. Wheeler TA, Woodward JE. Response of cotton varieties to diseases on the Texas High Plains. Texas AgriLife Research and Extension Service Center. 2012. doi: Lubbock.tamu.edu/files/2013/02/2012-cotton-disease-ratings.pdf.

36. Wheeler TA, Woodward JE. Response of cotton varieties to diseases on the Southern High Plains of Texas, 2010. Texas AgriLife Research and Extension Service Center. 2010. doi: Lubbock.tamu.edu/files/2011/11/DiseaseRecommendations.pdf.

37. Wheeler TA, Sagram US, Schuster GL, Gannaway JR. Identification of Factors that Influence Screening for Bacterial Blight Resistance. The Journal of Cotton Science. 2007;11(2):91-7. doi: http://www.cotton.org/journal/2007-11/2/upload/jcs11-91.pdf

38. Woodward JE, Wheeler TA. Bacterial Blight Update for Cotton on the High Plains 2015. Available from: http://plantclinic.tamu.edu/files/2015/08/WoodwardWheeler2015.pdf.

39. Darling AE, Mau B, Perna NT. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. PLOS ONE. 2010;5(6):e11147. doi: 10.1371/journal.pone.0011147.

40. Cohn M, Bart RS, Shybut M, Dahlbeck D, Gomez M, Morbitzer R, et al. Xanthomonas axonopodis virulence is promoted by a transcription activator-like effector-mediated induction of a SWEET sugar transporter in cassava. Mol Plant Microbe Interact. 2014;27(11):1186-98. doi: 10.1094/MPMI-06-14-0161-R. PubMed PMID: 25083909.

41. Grau J, Reschke M, Erkes A, Streubel J, Morgan RD, Wilson GG, et al. AnnoTALE: bioinformatics tools for identification, annotation, and nomenclature of TALEs from Xanthomonas genomic sequences. Sci Rep. 2016;6:21077. doi: 10.1038/srep21077. PubMed PMID: 26876161; PubMed Central PMCID: PMCPMC4753510.

42. Büschges R, Hollricher K, Panstruga R, Simons G, Wolter M, Frijters A, et al. The Barley Mlo Gene: A Novel Control Element of Plant Pathogen Resistance. Cell. 1997;88(5):695-705. doi: http://dx.doi.org/10.1016/S0092-8674(00)81912-1.

43. Jarosch B, Kogel K-H, Schaffrath U. The Ambivalence of the Barley Mlo Locus: Mutations Conferring Resistance Against Powdery Mildew (Blumeria graminis f. sp. hordei) Enhance Susceptibility to the Rice Blast Fungus Magnaporthe grisea. Molecular Plant-Microbe Interactions. 1999;12(6):508-14. doi: 10.1094/MPMI.1999.12.6.508.

44. Jørgensen IH. Discovery, characterization and exploitation of Mlo powdery mildew resistance in barley. Euphytica. 1992;63(1):141-52. doi: 10.1007/BF00023919.

45. Kim DS, Hwang BK. The pepper MLO gene, CaMLO2, is involved in the susceptibility cell-death response and bacterial and oomycete proliferation. The Plant Journal. 2012;72(5):843-55. doi: 10.1111/tpj.12003.

46. Dong S, Adams KL. Differential contributions to the transcriptome of duplicated genes in response to abiotic stresses in natural and synthetic polyploids. New Phytol. 2011;190(4):1045-57. doi: 10.1111/j.1469-8137.2011.03650.x. PubMed PMID: 21361962.

47. Chen LQ. SWEET sugar transporters for phloem transport and pathogen nutrition. New Phytol. 2014;201(4):1150-5. PubMed PMID: 24649486.

48. Verdier V, Triplett LR, Hummel AW, Corral R, Cernadas RA, Schmidt CL, et al. Transcription activator-like (TAL) effectors targeting OsSWEET genes enhance virulence on diverse rice (Oryza sativa) varieties when expressed individually in a TAL effector-deficient strain of Xanthomonas oryzae. The New phytologist. 2012;196(4):1197-207. doi: 10.1111/j.1469-8137.2012.04367.x. PubMed PMID: 23078195.

49. Doyle EL, Booher NJ, Standage DS, Voytas DF, Brendel VP, VanDyk JK, et al. TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. Nucleic Acids Research. 2012;40(W1):W117-W22.

50. Allen T. ALERT Bacterial Blight of Cotton 2011 [updated July 15, 2011]. Available from: http://www.mississippi-crops.com/2011/07/15/alert-bacterial-blight-of-cotton/.

51. Aaron S. Alexander JEW, Randal K. Boman, Terry A. Wheeler, Norman W. Hopper. Effect of the Easiflo Cottonseed Processing Method on Recovery of Xanthomonas axonopodis pv. malvacearum. The Texas Journal of Agriculture and Natural Resource. 2012;25:13-23.

52. Allen TW. Epidemiology of Bacterial Blight of Cotton: 2011 to 2015. 2016 Cotton Incorporated - Bacterial Blight Review. 2016.

53. Kemerait B. Cotton bacterial blight is back and this is what you need to know 2016 [updated August 10, 2016]. Available from: http://southeastfarmpress.com/cotton/cotton-bacterial-blight-back-and-what-you-need-know.

54. Li T, Liu B, Spalding MH, Weeks DP, Yang B. High-efficiency TALEN-based gene editing produces disease-resistant rice. Nature Biotechnology. 2012;30(5):390-2. doi: 10.1038/nbt.2199. PubMed PMID: 22565958.

55. Jia H, Orbovic V, Jones JB, Wang N. Modification of the PthA4 effector binding elements in Type I CsLOB1 promoter using Cas9/sgRNA to produce transgenic Duncan grapefruit alleviating XccDeltapthA4:dCsLOB1.3 infection. Plant Biotechnol J. 2016;14(5):1291-301. doi: 10.1111/pbi.12495. PubMed PMID: 27071672.

56. Strauß T, van Poecke RMP, Strauß A, Römer P, Minsavage GV, Singh S, et al. RNA-seq pinpoints a Xanthomonas TAL-effector activated resistance gene in a large-crop genome. Proceedings of the National Academy of Sciences. 2012;109(47):19480-5.

57. Yang B, Sugio A, White FF. Os8N3 is a host disease-susceptibility gene for bacterial blight of rice. Proceedings of the National Academy of Sciences. 2006;103(27):10503-8.

58. Chu Z, Yuan M, Yao J, Ge X, Yuan B, Xu C, et al. Promoter mutations of an essential gene for pollen development result in disease resistance in rice. Genes & Development. 2006;20(10):1250-5. doi: 10.1101/gad.1416306.

59. Yoo M-J, Wendel JF. Comparative Evolutionary and Developmental Dynamics of the Cotton (*Gossypium hirsutum*) Fiber Transcriptome. PLoS Genet. 2014;10(1):e1004073. doi: 10.1371/journal.pgen.1004073.

60. Janga MR, Campbell LM, Rathore KS. CRISPR/Cas9-mediated targeted mutagenesis in upland cotton (Gossypium hirsutum L.). Plant Molecular Biology. 2017:1-12. doi: 10.1007/s11103-017-0599-3.

61. Li C, Unver T, Zhang B. A high-efficiency CRISPR/Cas9 system for targeted mutagenesis in Cotton (Gossypium hirsutum L.). Sci Rep. 2017;7:43902. doi: 10.1038/srep43902. PubMed PMID: 28256588; PubMed Central PMCID: PMCPMC5335549.

62. Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 16S ribosomal DNA amplification for phylogenetic study. Journal of bacteriology. 1991;173(2):697-703. PubMed PMID: 1987160.

63. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20. doi: 10.1093/bioinformatics/btu170. PubMed PMID: 24695404; PubMed Central PMCID: PMCPMC4103590.

64. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology. 2012;19(5):455-77. doi: 10.1089/cmb.2012.0021.

65. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068-9.

66. Almeida NF, Yan S, Cai R, Clarke CR, Morris CE, Schaad NW, et al. PAMDB, a multilocus sequence typing and analysis database and website for plant-associated microbes. Phytopathology. 2010;100(3):208-15. doi: 10.1094/PHYTO-100-3-0208. PubMed PMID: 20128693.

67. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156-8. doi: 10.1093/bioinformatics/btr330.

68. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841-2. doi: 10.1093/bioinformatics/btq033.

69. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23(21):2947-8. doi: 10.1093/bioinformatics/btm404.

70. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution. 1987;4(4):406-25. doi: 10.1093/oxfordjournals.molbev.a040454.

71. Chin C-S, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, et al. Phased diploid genome assembly with single-molecule real-time sequencing. Nat Meth. 2016;13(12):1050-4. doi: 10.1038/nmeth.4035

72. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421. doi: 10.1186/1471-2105-10-421. PubMed PMID: 20003500; PubMed Central PMCID: PMCPMC2803857.

73. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19(9):1639-45. doi: 10.1101/gr.092759.109. PubMed PMID: 19541911; PubMed Central PMCID: PMCPMC2752132.

74. Wilson MC, Mutka AM, Hummel AW, Berry J, Chauhan RD, Vijayaraghavan A, et al. Gene expression atlas for the food security crop cassava. New Phytol. 2017;213(4). doi: 10.1111/nph.14443. PubMed PMID: 28116755.

75. Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, et al. Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement. Nat Biotechnol. 2015;33(5):531-7. doi: 10.1038/nbt.3207. PubMed PMID: 25893781.

76. Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. Synteny and Collinearity in Plant Genomes. Science. 2008;320(5875):486.

Chapter 3a: *Pseudomonas syringae* Pathogen Causes Foliar Disease of Upland Cotton in Texas.

This chapter was previously published in the journal Plant Disease as:

Phillips, A. Z.; Wheeler, T.; Woodward, J.; Bart, R. S. *Pseudomonas Syringae* Pathogen Causes Foliar Disease of Upland Cotton in Texas. *Plant Dis.* 2018, *102* (6), 1171–1171. https://doi.org/10.1094/PDIS-11-17-1700-PDN.

3a.1 Personal contributions

I isolated the first known sample of this pathogen since 1994. I then characterized this pathogen and alerted the scientific community of its presence in Texas. This involved purifying the pathogen from contaminating *Xcm* and performing Koch's postulates to determine that it was the causal agent of necrotic disease symptoms. I used 16S sequencing to identify the pathogen as a Pseudomonad. I then characterized disease symptoms over time with *Xcm* and alone at varying concentrations. Subsequently, I alerted several extension scientists who began isolating this pathogen from fields and characterizing its pathogenicity on many cotton varieties. Genome sequencing of the original isolate allowed me to generate a multi locus sequence typing phylogeny and determine that the pathogen is most closely related to the *Pseudomonas syringae* pathovars *atrofaciens*, *aptata*, and *pisi.*

3a.2 Disease note

Cotton bacterial blight (CBB), caused by *Xanthomonas citri* pv. *malvacearum* (*Xcm*), can cause significant yield losses on susceptible varieties of upland cotton (*Gossypium hirsutum* L.)[1]. CBB has re-emerged in the United States since 2011[2]. In 2015, cotton fields near Plains, TX,

exhibited symptoms of an unknown foliar disease on cotton cultivars that were reported to be resistant to CBB. In June 2016, bacteria were isolated from Fibermax 2007GLT (CBB Resistant) leaves exhibiting CBB-like symptoms. Culture on nutrient-rich NYGA medium with 50 μM cycloheximide revealed two predominant colony morphologies, yellow and white. The yellow colonies were confirmed to be *Xcm*. Koch's postulates were used to determine that the white bacteria caused necrotic foliar lesions when infiltrated into cotton leaves alone or in combination with *Xcm*. Sequencing of the 16s rRNA gene identified this bacterium as *Pseudomonas* sp. with 99% sequence identity to the *Pseudomonas syringae* pv. *syringae* B728a (Gbk 230265-9). Further multilocus sequence analysis using concatenated regions of the *gyrB*, *rpoD*, *gap1*, and *gltA* on the PAMDB database grouped this pathogen with *P. syringae* pv. *atrofaciens*, *P. syringae* pv. *apata*, *P. syringae* pv. *pisi*, *P. syringae* pv. *syringae,* and *P. syringae* pv. *aceris* (http://genome.ppws.vt.edu/cgi-bin/MLST/home.pl). Identification of a Pseudomonad cotton pathogen has occurred at least once before in cotton seedlings near Lubbock, TX. This was recorded in the 1994 Cotton Beltwide Conferences Proceedings (ATCC 51506). No other reference to a pseudomonad pathogen of cotton has been found. However, evidence of a *Pseudomonas-Xanthomonas* disease complex has been found at least once in leafy crucifers[3]. Inoculations with a needleless syringe resulted in symptoms that initially appeared as a cell death phenotype but quickly progressed to a spreading necrotic lesion. Foliar disease symptoms were observed after inoculation with a bacterial suspension at OD_{600} of 0.0001 to 0.01. Symptoms appeared as early as 1 day after inoculation. To date, this pathogen has been isolated from 11 different fields in six counties and always in association with *Xcm*. Additional isolates were identified from symptomatic leaf tissue of at least eight different cultivars at various locations. Four isolates from Donley County, TX, and four cotton cultivars (DP 1454NRB2RF, FM

1320GL, FM 1830GLT, and FM 2484B2F) were screened using the scratch method to identify whether variability existed in the disease response of cultivars. The cultivars were arranged in a split-plot design, with isolate as the main factor and cultivar as the subfactor. Both the main factor and the subfactors were randomized. There were four replications for each isolate/cultivar combination. The test was repeated once. Disease symptoms were less severe on FM 2484B2F ($P < 0.05$) than the other three cultivars, suggesting that variation exists among cotton germplasm for tolerance to this pathogen. Future phylogenetic analysis will focus on the origin of virulence of this pathogen and its distribution within the cotton production regions in the United States.

3a.3 e-Xtra

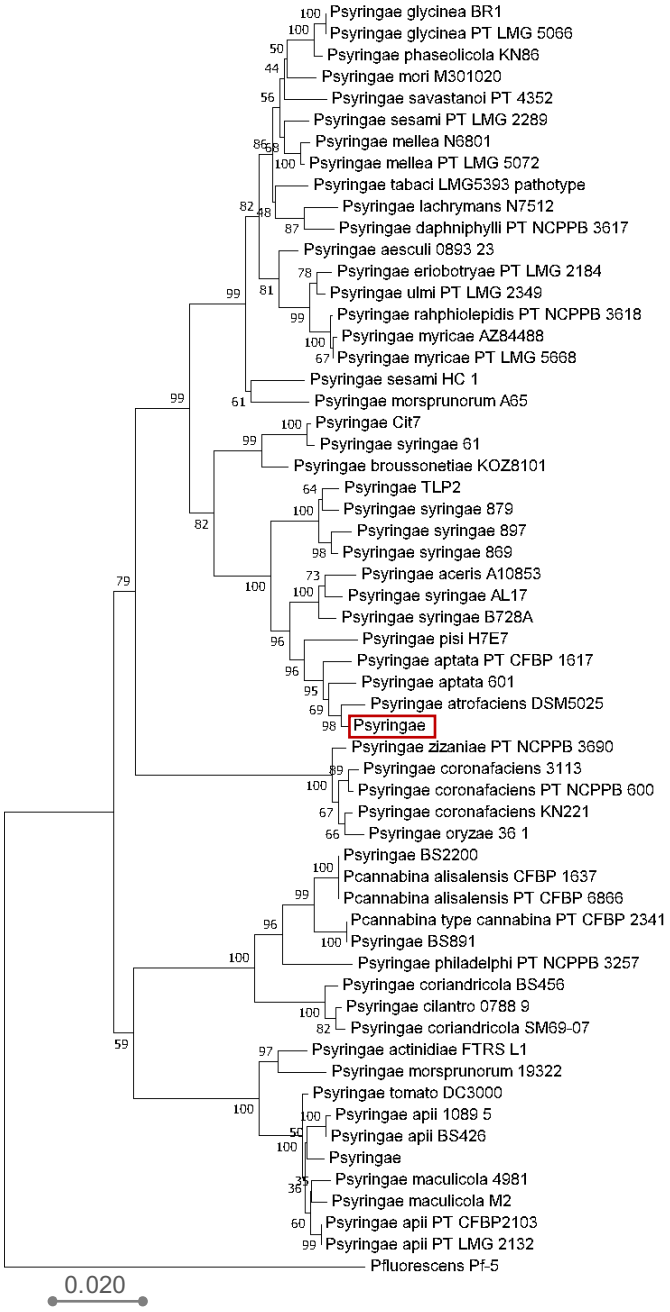


Figure 1: Evolutionary relationship of the *Pseudomonas syringae* cotton pathogen (red box) to *Pseudomonas syringae* species. The evolutionary history was inferred using the Neighbor-Joining method using 500 bootstraps. Branch labels: Percentage of bootstraps associated with the branching. The evolutionary distances were computed using the Maximum Composite Likelihood method using the number of base substitutions per site.

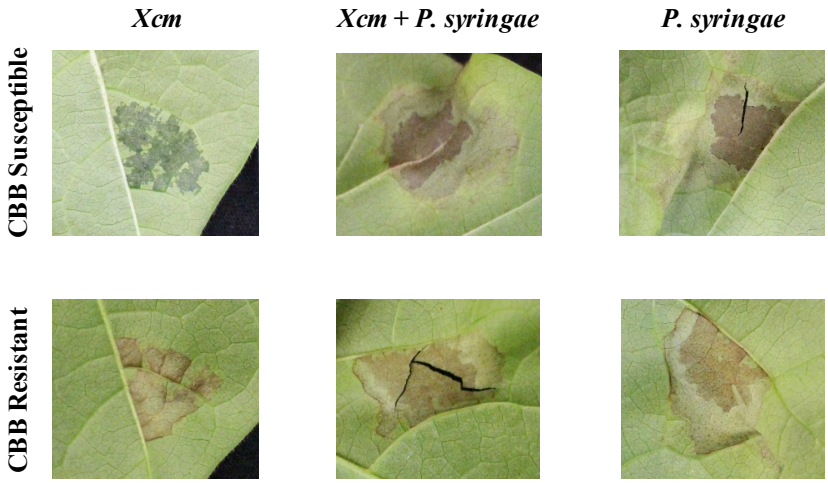


Figure 2: *Xcm* and *P. syringae* phenotypes on CBB susceptible (Phy499WRF) and CBB resistant (ST5288B2RF) *G. hirsutum* varieties. Bacteria were re-suspended in $MgCl_2$ at an OD600 = 0.05 and then introduced into cotton leaves using a needleless syringe. Symptoms were evaluated 6 days post inoculation.

3a.4 References

1. Verma, J. P. *Bacterial Blight of Cotton*; CRC Press, 1986.

2. Phillips, A. Z.; Berry, J. C.; Wilson, M. C.; Vijayaraghavan, A.; Burke, J.; Bunn, J. I.; Allen, T. W.; Wheeler, T.; Bart, R. S. Genomics-Enabled Analysis of the Emergent Disease Cotton Bacterial Blight. *PLOS Genet.* **2017**, *13* (9), 23.

3. Zhao, Y; Damicone J. P.; Demezas, D. H.; Bender, C. L. Bacterial Leaf Spot Diseases of Leafy Crucifers in Oklahoma Caused by Pathovars of *Xanthomonas campestris*. *Plant Dis.* **2000**, *84* (9), 1008-1014. https://doi.org/10.1094/PDIS.2000.84.9.1008

Chapter 3b: Evolutionary Context of the *Pseudomonas syringae* Cotton Pathogen and Mechanisms of Virulence

3b.1 Abstract

Rapid long-read sequencing provides a unique opportunity for investigating new pathogen outbreaks as they occur. This technology enables in-depth analysis of the host-pathogen arms race and the discovery of mechanisms that allow the pathogen to gain the upper hand. The *Pseudomonas syringae* pathogen of cotton (*Gossypium hirsutum*) is an elusive pathogen that has only been identified twice before. Here we use long-read sequencing to identify the evolutionary origins of this pathogen, uncover putative virulence factors, and investigate how it is influenced by the host-pathogen arms race. Whole genome SNP analysis of seven circularized genomes identify all historical and contemporary *P. syringae* isolates as monophyletic phylogroup 2 pathogens. Genome annotation reveals multiple type III effectors as well as five putative toxin biosynthetic clusters that may contribute to the characteristic rapid necrosis phenotype of this pathogen. Whole genome alignments uncover several putative virulence factors such as the type IV pilus and Filamentous Hemagglutinin that are undergoing diversifying selection. Further phylogenetic analysis reveals several potential mechanisms underlying the diversifying selection including horizontal gene transfer, pathoadaptation, and recombination. The results of this paper contribute important information for understanding how this little-known pathogen causes disease on cotton and lays the groundwork for future research that ultimately aims to identify genetic resistance.

3b.2 Introduction

Hosts and pathogens are intertwined in a co-evolutionary arms race[1,2]. Pattern triggered immunity (PTI) is the initial recognition of pathogens by host extracellular receptors, collectively known as pattern recognition receptors (PRRs). PRRs recognize conserved microbe associated molecular patterns (also known as pathogen associated molecular patterns (PAMPs)) and trigger a basal immune response that controls pathogen proliferation[3,4]. Successful pathogens have evolved mechanisms to avoid or suppress PTI, often through specific effector proteins that enter the host cell. These effectors may trigger a second line of host defense known as effector triggered immunity (ETI)[5,6]. Similar to PTI, in ETI, host intercellular receptors, collectively known as nucleotide-binding leucine-rich repeat proteins (NLRs), have evolved to recognize the presence of pathogen effectors and trigger strong resistance. Within the co-evolutionary arms race, pathogens are under selective pressure to evolve around this recognition and overcome resistance[7,8]. Within this conceptual framework, there is overlap between PTI and ETI and importantly, evolution within the arms race is continuous[9–11]. Consequently, this interaction has been characterized as the Red Queen Hypothesis in which "it takes all the running you can do to keep in the same place"[12,13].

Genomic evolution occurs through multiple mechanisms including horizontal gene transfer, recombination, and the accumulation of smaller, vertically inherited mutations (single nucleotide polymorphisms (SNPs) and small insertions and deletions (INDELs))[14,15]. Genes that are involved in the host-pathogen arms race may display specific signatures of evolutionary selection[7,8]. For example, a pathogen gene that is required for virulence, may contain low nucleotide diversity compared to the rest of a neutrally evolving genome[16]. Conversely, regions of high nucleotide diversity may indicate diversifying selection and evasion of host recognition.

Similarly, this could indicate an advantage associated with maintaining multiple alleles within a population[17,18].

The recent development of long read sequencing has enabled investigations of pathogen outbreaks. This technology can be used to track the movement of entire virulence islands, plasmids, and other structural changes through the use of publicly available assembly and annotation software. Furthermore, because it is relatively cheap and fast, nanopore sequencing has made long read sequencing available to many more scientists. While this technology is still improving, it is already revolutionizing the identification of pathogen outbreaks both in agricultural and medical settings[19–22].

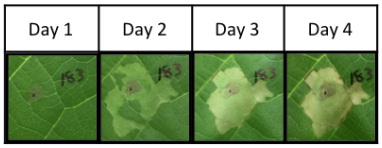
Pseudomonas species are ubiquitous in the environment and are commonly found in rain water, soil, and in association with plants[14,23]. This genus encompasses many economically damaging plant pathogens that cause diseases such as bacterial speck on tomato, bacterial blight of soybean, and bleeding canker on horse chestnut[24–26]. These pathogens are genetically and pathologically diverse, spanning five major phylogroups. The most intensively studied Pseudomonad plant pathosystem is *P. syringae* pv. *tomato* DC3000 and *Arabidopsis thaliana*[27–29]. Research on this model system has dissected the molecular mechanisms governing pathogen virulence and host immunity[28].

Pst DC3000 belongs to phylogroup 1[30]. In contrast, less is known about the other four phylogroups. Of particular interest are the phylogroup 2 *P. syringae* pathogens. Unlike many of their more well studied relatives, these pathogens often have extremely broad host ranges[31,32]. This characteristic has been hypothesized to contribute to rapid spread of the pathogen as well as the difficulty of controlling outbreaks[31,32]. At least 25 new First Reports of *P. syringae* outbreaks between January 2015-July 2018 were published in the journal Plant Disease[32]. This observation

highlights the urgent need to understand how phylogroup 2 and other *P. syringae* outbreaks occur.

Cotton is particularly vulnerable to emerging diseases due to its low genetic diversity compared to other staple crops in the US such as corn and alfalfa[33–35]. This low genetic diversity may facilitate pathogen evolution around host genetic defenses and may have contributed to the recent outbreaks of *Xanthomonas citri* pv. *malvacearum* (*Xcm*), *Fusarium oxysporum* f. sp. *vasinfectum* Race 4 (FOV4), and cotton leafroll dwarf virus (CLRDV)[36–38]. In this manuscript we investigate a recent outbreak of *Pseudomonas syringae* on cotton in the US. Here we present the first sequenced genomes of this pathogen as well as analysis that gives evolutionary context to this previously understudied *Pseudomonas* pathogen. The identification of putative virulence factors and markers of selection provide insights into how this elusive pathogen causes disease and participates in the plant-pathogen arms race.

A.



C.

B.

Strain	Location	Variety	Year	Disease
O183	Donley Co, TX	NexGen 3406B2XF	2016	
A203	Donley Co, TX	FM1911 GLT	2016	
A234	Brownfield, TX	--	2017	
P.s. aptata	USA	*Beta vulgaris*	1959	
A236	Lubbock, TX	--	1994	
A238	Lubbock, TX	--	1994	
A248	College Station, TX	Croplan	2017	
X480	Plains, TX	FM2007-GLT	2015	

Figure 1. *Pseudomonas syringae* causes disease on *Gossypium hirsutum*. A) Ps183 disease symptoms 1 to 4 days post inoculation (dpi). B) *P. syringae* strains investigated in this study. Images were taken 4 dpi. C) Map of Texas where *P. syringae* strains were originally isolated. Strains are color coded as in B. All bacterial concentrations are $1x10^6$.

3b.3 Results

3b.3.1 The *Pseudomonas syringae* cotton pathogen

We previously reported that *Xanthomonas citri* pv. *malvacearum* (*Xcm*) re-emerged as a major disease of *Gossypium hirsutum* (upland cotton) from 2011 to 2017[38]. Recently, we reported a second disease of cotton and identified the causal agent as *Pseudomonas syringae*[39]. When infiltrated into cotton leaves, this pathogen caused a rapid spreading necrosis phenotype. Disease symptoms began one day post inoculation (DPI) as a mild necrosis at the site of inoculation (Fig. 1a). By 2 DPI, rapid cell collapse occurred outside of the site of inoculation and continued to expand at 3 and 4 DPI. Four additional *P.s.* isolates were collected from infected fields in the Texas high plains and near College Station from 2016-2017 (Ps183, Ps203, Ps234, Ps248, and Ps480). We obtained further additional strains from culture collections. From ATCC we received a *P. syringae* culture originally isolated from cotton in 1994 (*P. syringae* ATCC 51506). This culture showed two distinct colony types and when purified, both caused disease on cotton. Further, through sequencing the 16s rRNA gene, we confirmed both bacteria belong to the *P. syringae* species. Hence forth, these strains will be referred to as Ps236 and Ps238. We also received a closely related strain, *P. syringae* pv. *aptata* CFBP 1617 from the CIRM-CFBP culture collection that was isolated from *Beta vulgaris* (sugar beet) and will be referred to as Ps418 in this manuscript. All *P. s.* isolates triggered spreading lesions with the exception of Ps203, which only induced symptoms at the site of inoculation (Fig. 1b). To further assess relative virulence among these strains, bacterial growth after inoculation was assessed for strains Ps418, Ps183 and Ps203. All three strains grew to similarly high populations by 4 and 7 days post inoculation. Therefore, we conclude that these strains are pathogenic on cotton. The data

also suggest that the magnitude of symptoms does not necessarily correlate with pathogen population in this pathosystem (Sup. Fig 1).

3b.3.2 Long read sequencing and genome assembly

To uncover the virulence mechanisms deployed by this pathogen, we generated full genome sequences for each isolate using long read sequencing technology (Sup. Table 1). The first isolate, Ps480, was sequenced using PacBio. We obtained 31,554 reads with a mean read length of 18,652bp. Reads were assembled using Falcon and polished with Quiver[40,41]. The seven remaining isolates were first sequenced using Nanopore technology. We obtained 214,343-586,120 reads per isolate with a mean read length of 8,474-18,652bp. These genomes were assembled using Canu and then polished with Nanopolish[42,43].

For all isolates, genome assembly yielded 6MB circularized genomes (Sup. Table 1). We also identified a 68kb plasmid in four of the cotton pathogen genomes. The *P. syringae* pv. *aptata* CFBP 1617 (Ps418) genome contained a shorter 58kb plasmid that aligned with the 68kb plasmids for over 50% of its length (Sup. Fig 2). The genome quality was assessed using Benchmarking Universal Single-Copy Ortholog (BUSCO) scores[44]. Genomes assembled with nanopore reads alone resulted in BUSCO scores less than 85%. Therefore, the genomes were polished again with 100x Illumina reads using Pilon[45]. This was able to correct for assembly errors inherent to error-prone Nanopore reads. Each resulting genome received BUSCO scores over 98%, similar to Illumina-only genome assemblies (Sup. Fig 3). Genomes were annotated using Prokka[46].

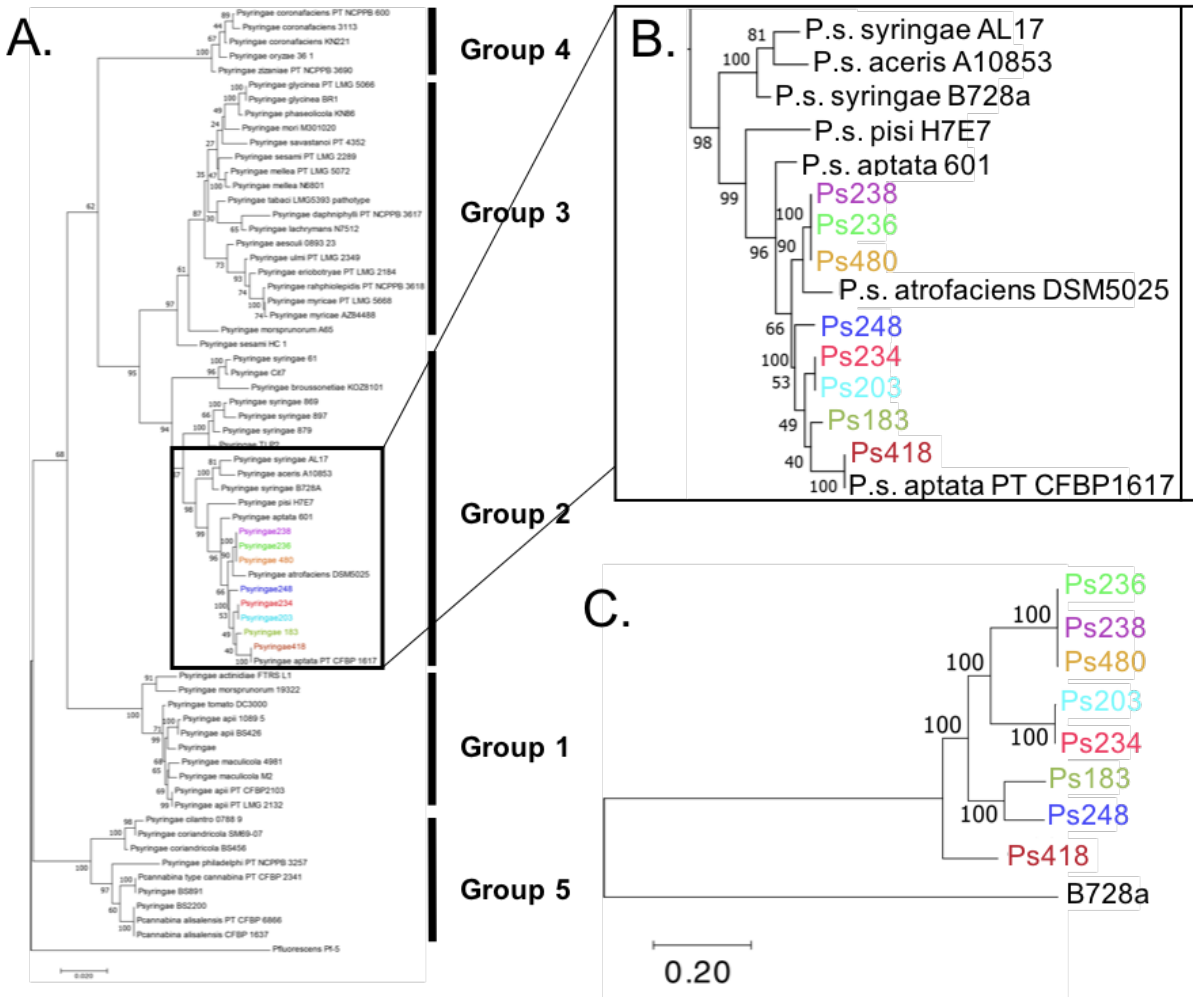


Figure 2. Phylogenetic context of the *P. syringae* cotton pathogen. A) Multi Locus Sequence Typing (MLST) phylogeny of 66 *Pseudomonas syringae* plant pathogens based on Maximum Likelihood phylogenetic analysis of concatenated regions of the house keeping genes: *gyrB*, *rpoD*, *gap1*, and *gltA*. Colored names represent isolates sequenced in this study. B) Zoom region of phylogroup 2 with *P. syringae* cotton pathogens. C) Whole genome Maximum Likelihood phylogeny of cotton pathogens in addition to Ps418, a sugar beet pathogen and B728a as an out group. Whole genomes were aligned using Mauve. 331,898 SNPs were concatenated and used to generate at tree. For both phylogenies, branch labels are bootstrap percentage values based on 500 bootstraps. Branch lengths measured in the number of substitutions per site. All positions containing gaps and missing data were eliminated.

3b.3.3 Phylogenetic relatedness among *Pseudomonas syringae* cotton pathogens

Concatenated regions of the housekeeping genes *gyrB*, *rpoD*, *gap1*, and *gltA* were taken from the complete *P. syringae* genomes and used to generate a Multi Locus Sequence Typing (MLST) phylogeny (Fig. 2a, b). The eight newly sequenced strains were compared against 57 *P. syringae* isolates that span all 5 major phylogroups. All *P. syringae* isolates collected from cotton formed a monophyletic group within the *P. syringae* phylogroup 2, supporting the previous placement of the pathogen within this group[39]. Phylogroup 2 encompasses many *P.s.* pathovars including *atrofaciens, aptata,* and *pisi,* pathogens of wheat, sugar beet, and peas, respectively. Two strains, *P. syringae* pv. *aptata* PT CFBP1617 and *P. syringae atrofaciens* DSM5025 co-located within the clade of cotton pathogens. These data, along with the finding that *P. syringae* pv. *aptata* CFBP 1617 (Ps418) can cause disease on cotton, supports previous reports of wide host ranges within phylogroup 2.

To further dissect the evolutionary relationships within phylogroup 2 we pursued whole genome phylogenetic analysis. The complete chromosomes of seven *P. syringae* cotton isolates, *P. syringae* pv. *aptata* (strain Ps418), and the previously published *P. syringae* pv. *syringae* B728a chromosome, were aligned using Mauve[47]. We identified 331,898 SNPs within colinear blocks. These were concatenated and used to generate a Maximum Likelihood tree (Fig. 2c). Because the 68kb plasmid was not found in all isolates it was not included in the SNP phylogeny. Ps480, collected in 2015, clustered tightly with the historical strains Ps236 and Ps238, both collected in 1994. Only 5,259 SNPs differentiated these three chromosomes.

Similarly, Ps248 and Ps183 were collected from distant locations in Texas yet were phylogenetically clustered. Isolates Ps203 and Ps183 did not fall within the same clade, despite the fact that they were isolated from the same field. Thus, among the strains included in this study, evolutionary relationships are distinct from temporal and spatial patterns.

Phylogenetic relationships also did not correlate with relative virulence on cotton. For example, Ps234 and Ps203 clustered in the MLST and SNP based phylogenies. However, the former induces strong spreading lesions while the latter triggers more mild symptoms (Fig. 1). Therefore, we hypothesize that the variable virulence levels among these strains is related to small differences in virulence factor repertoires, not large phylogenetic changes.

3b.3.4 Virulence factors present in *P. syringae* cotton pathogens

We investigated the identities of conserved virulence factors present in each genome due to their putative importance for pathogenicity on cotton. In addition, we investigated the identities of non-conserved virulence factors because they may contribute to the variability in pathogenicity among strains. Genome sequencing and subsequent annotation of 7 *P. syringae* cotton pathogens revealed several instances of pathogenicity islands with homology to Type Three Secretion Systems (T3SS)[48]. Unsurprisingly, a full tripartite pathogenicity island (T-PAI) containing the T3SS and many effectors was identified in all isolates based on synteny with the *P. syringae* pv. *tomato* DC3000 canonical T-PAI (Fig. 3). In addition, three isolates also contained a Rhizobial pathogenicity island (R-PAI) T3SS, based on homology to the *P. syringae* pv. *phaseolicola* 1448a R-PAI (Sup. Fig.4). This additional T3SS is common among many *P. syringae* phylogroups, though less common in Phylogroup 2. The function of this secretion system in pathogenicity has not yet been determined[14,49]. In addition to the T3SS, several strains were found to contain 68kb plasmids that encodes a complete Type Four Secretion System

(TIVSS) (*virB1-virB10*). However, because the secretion system is absent from many virulent strains we conclude that it is not essential for disease.

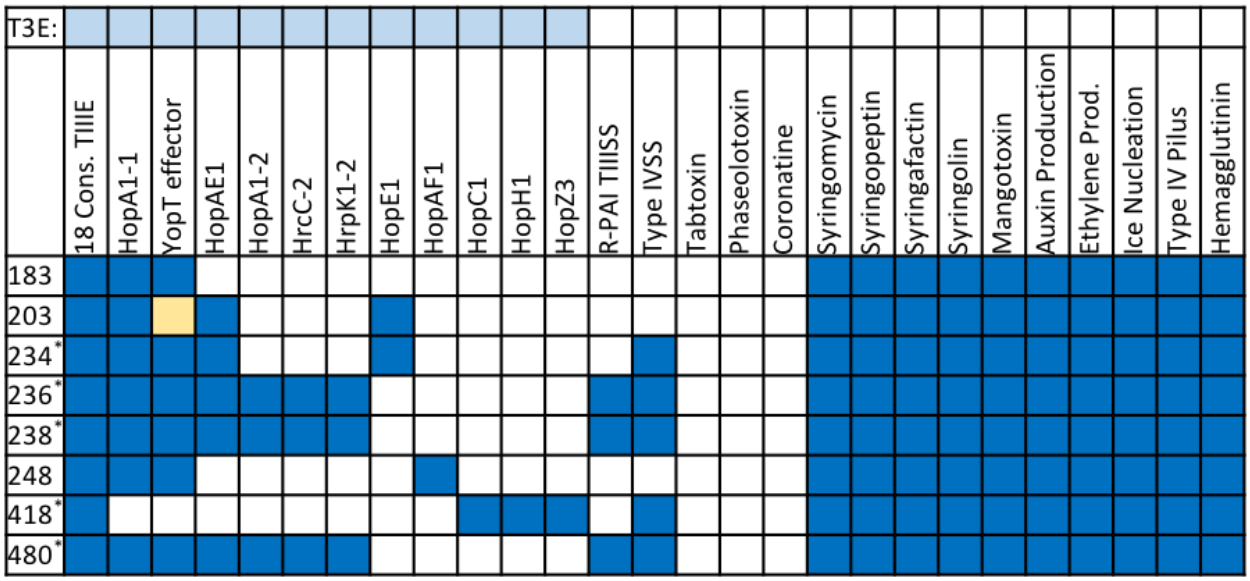


Figure 3. Virulence factors present in *P. syringae* cotton pathogens. Blue boxes represent the presence and white boxes represent the absence of a gene or biosynthetic cluster. Yellow boxes represent genes that are disrupted but still identifiable via BLAST searches. (At top) light blue boxes represent variable Type III effectors. *indicates a genome that includes a 67kb plasmid.

P. syringae pathovars are well known for their utilization of effectors, phytotoxins, and other virulence factors to trigger disease. Therefore, we used Prokka annotations, synteny, and tBLASTn to compare putative virulence factors and biosynthetic clusters in the cotton pathogen genomes to previously studied virulence factors and biosynthetic clusters in other pathovars. A tBLASTn e-value cutoff of $1x10^{-6}$ was used for significance, consistent with the annotation software Prokka [46]. However, it should be noted that this cutoff does not ensure mechanistic conservation. Additional characterization would need to be conducted to reveal the functions of these genes in this pathosystem.

The variability in pathogenicity within and between *P. syringae* pathovars is often attributed to the variable T3E profiles present in each isolate. We identified 18 conserved and 11 variable T3E genes within the 8 newly sequenced genomes based on the Prokka annotations and

BLAST analysis. This is comparable to other studies that found eight to sixteen T3Es within the genomes of phylogroup 2 *P. syringae* pathogens[50]. Ps203 and Ps418 induced the weakest disease symptoms, correlating with the absence of a YopT-like effector, suggesting that this effector may be important for virulence (Fig 3).

Phylogroup 2 *Pseudomonas syringae* strains often specifically use lipopeptide phytotoxins to trigger disease. Evidence of a conserved lipopeptide biosynthetic cluster was identified, including many large non-ribosomal peptide synthases. Within this biosynthetic cluster we found regions with 72.1% and 82.8% identity to the syringomycin and syringopeptin biosynthetic clusters in *P. s.* pv. *syringae* B728a, respectively (Fig. 3, Sup. Table 2). It also contained a region with 97.4% identity to the *P. s.* pv. *syringae* B64 syringafactin biosynthetic cluster. Additional biosynthetic clusters were found with 96.0% identity to *P. s.* pv. *syringae* B301 D-R syringolin A biosynthetic cluster and 97.4% identity to the *P. s.* pv. *syringae* CFBP3388 mangotoxin biosynthetic cluster. Therefore, it is possible that multiple toxins are working in unison to promote pathogenicity on cotton and generate the rapidly spreading necrotic lesions that are characteristic of this disease.

P. s. BR2 produces tabtoxin and *P. s.* pv. *phaseolicola* NPS3121 produces phaseolotoxin[51,52]. We used these reference isolates to search for homologous genes in the *P.s.* cotton pathogens (Fig 3, Sup. Table 2). BLAST hits were returned for tabtoxin genes *tabP*, *tabD*, *tblA*, *tblS*, *tblC*, *tblD*, *tblF*, and *tblR*, but not *tabA*, *tabB*, *tabC*, and *tblE.* Similarly, of the phaseolotoxin genes *ptx1-ptx22*, only genes *ptx7*, *ptx10* (*desA*), *ptx13*, *ptx18*, and *ptx19* produced positive BLAST hits. Prokka annotated several putative genes in the coronafacic acid (*cfa1*, *2*, *6*, and *7*) and coronamic acid (*cmaB*, *E*, *U*, and *X*) pathways in these genomes. However, these genes were not found in the canonical CMA and CFA biosynthetic clusters and several

coronatine biosynthesis genes (*cfa4*, *PSPTO_4688*, *cmaC*, *cmaD*, and *alanyl tRNA synthetase*) were missing. Therefore, we find it unlikely that these Pseudomonads synthesize tabtoxin, phaseolotoxin or coronatine.

We observed evidence for genes corresponding to ethylene production (*efe* from *P. s.* pv. *cannabina*) and auxin biosynthesis (*iaaM* and *iaaH* from *P. s.* pv. *syringae* Y30, e-values $1.34x10^{-55}$ and $6.13x10^{31}$) but not for auxin inactivation (*iaaL* from *P. s.* pv. *tomato* DC3000) (Sup. Table 3). We also did not find a homolog of the syringolide gene *avrD* from *P. s.* pv. *tomato* PT23. filamentous hemagglutinin and the ice nucleation gene, *inaZ*, were annotated in each genome (Fig 3, Sup. Table 2).

The type IVa pilus associated with gamma proteobacteria mobility is usually encoded by 5 to 6 operons located across the genome. Several genes within these operons were annotated by prokka, including the *ponA-aroB* operon (*ponA*, *pilM*, *pilN*, *pilO*, *pilP*, *pilQ*, *aroK*, *aroB*) and the *pilB-yacG* operon (*pilB*, *pilC*, *pilD*, *coaE*, *yacG*) (Sup. Table 2). The type IVa pilus of *Pseudomonas aeruginosa* is the best characterized type IV pilus in the *Pseudomonas* genus due to its role in human pathogenesis[53]. tBLASTn against *P. aeruginosa* PA96 type IV pilus operons provided evidence for the *pilU-yggS* operon (*pilU*, *pilT*, *yggS*), the *ispG-yfgB* operon (*ispG*, *yfgA*, *pilF*, *yfgB*), and part of the *pilE-fimT* operon (*pilE*, *pilY2*, *pilY1*, *pilV*, *fimU*), and *pilA*. Therefore, it is likely that this pilus system is functional in the genomes sequenced in this study (Fig. 3).

3b.3.5 Nucleotide diversity among putative virulence factors

Regions of a genome that display a high frequency of SNPs may indicate diversifying selection and a role in the host-pathogen arms race. SNPs across a whole genome alignment of the *P. syringae* genomes assembled in this study were identified by Mauve (Fig. 4a) and then converted into a variant call format (VCF)[47]. Pairwise nucleotide diversity (π) was assessed in

sliding 100bp windows overlapping by 50bp using VCFtools[54]. This resulted in the identification of eleven loci scattered across the genome that had a nucleotide diversity of over 0.25 (Fig. 4b). These regions were re-aligned and checked for local misalignment as well as surrounding gene annotations. Notably, no T3E sequences demonstrated a nucleotide diversity distinguishable from the neutrally evolving genome (Sup. Fig 5).

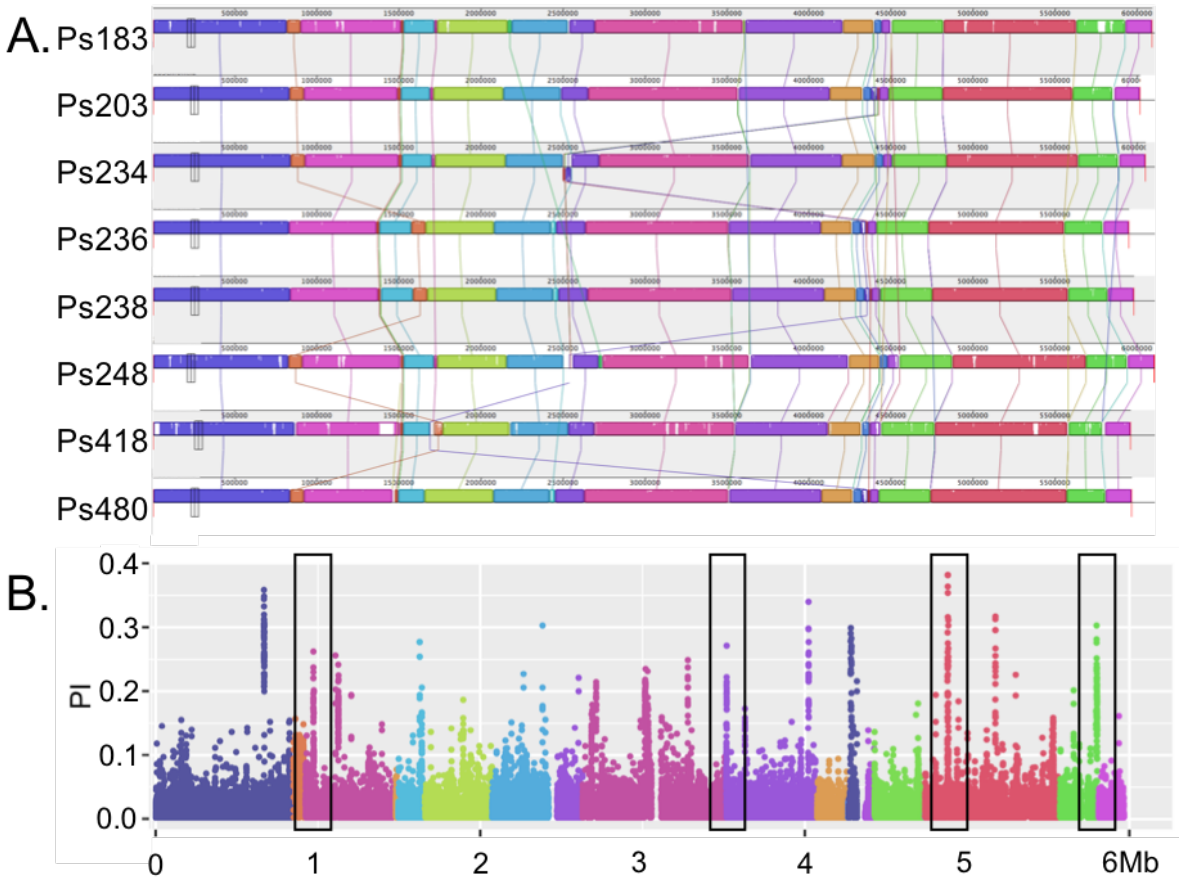


Figure 4. Genetic Diversity Among *P. syringae* Cotton Pathogens. A) Whole genome alignments of *P. syringae* chromosomes generated by Mauve. B) Pairwise Nucleotide Diversity (PI) Across Genome. SNP outputs from the Mauve alignment were converted into a VCF file using the Ps480 genome as a reference. Pi was assessed in 100bp windows staggered every 50 bp across each colinear block.

Four regions with notably high pairwise nucleotide diversity included an operon encoding PilB, PilC, and PilD of the type IV pilus and genes encoding an Insecticidal Toxin, Filamentous Hemagglutinin, and the Rhs element Vgr protein (Fig. 4b, Sup. Fig6). The latter

three genes are members of the class of toxins called polymorphic toxins. These genes often undergo recombination due to their modular and/or highly repetitive nature[55,56]. While these genes are most often studied in the context of microbe-microbe interactions, the Type IV pilus and Filamentous Hemagglutinin have also been implicated in host-microbe interactions and therefore were investigated further[57–63].

As a complementary approach, we investigated the phylogenetic relationships of type IV pilus and filamentous hemagglutinin genes among the *P. syringae* cotton pathogens (Fig. 5). As expected, the nucleotide *pilB*, *pilC*, and *pilD* phylogenetic trees displayed greater depths due to greater genetic differences than the MLST tree based on housekeeping genes. Next, for each gene tree, we considered the phylogenetic relationships between isolates. The *pilB*-, *pilC*-, and, *pilD*-based trees were mostly similar except for the placement of isolates Ps183 and Ps248, which vary in respect to each other as well as the other isolates in each phylogeny. We investigated the alignments of these genes and found SNPs dispersed throughout the length of each *pilB*, *pilC*, and *pilD* gene in each genome, which may indicate either horizontal gene transfer or variable levels of selective pressure on each gene (Sup. Fig. 7, 8, 9, 10). We then investigated the GC content across this region as an indicator of horizontal gene transfer. No dramatic changes in GC content were observed across the *pilBCD* operon and surrounding region suggesting that no horizontal gene transfer has occurred (Sup. Fig. 11). However, the possibility of a historical horizontal gene transfer event or a horizontal gene transfer between closely related species could not be ruled out.

The filamentous hemagglutinin phylogenetic tree was also found to be deeper compared to the MLST phylogenetic tree, indicating that more divergence has occurred among these genes than in the neutrally evolving genome as a whole (Fig. 5a and e). This high level of diversity is

likely due to the highly repetitive and modular nature of this gene, which encourages local recombination events and transcription errors[55,64]. We identified between 24-40 hemagglutinin repeats in each filamentous hemagglutinin gene within the *Pseudomonas* genomes, as assigned by PFAM (Sup. Fig 12). Evidence of recombination was supported by the identification of two large 310aa and 595 INDELs and several additional 9-18aa INDELs as well as the presence of variable numbers of smaller filamentous hemagglutinin related genes at the 5' end of each gene (Sup. Fig 13). This trend in phylogenetic depth was also found at the amino acid level, indicating that more divergence occurred among the Type IV pilus genes than in the neutrally evolving genome as a whole (Sup. Fig. 14). This analysis also revealed potential unequal levels of diversifying selection across the type IV pilus genes. The test statistic dN-dS was evaluated per codon to further investigate the effects of selection on the *pilBCD* operon (Sup. Fig. 15). As shown in both the dN-dS data and amino acid phylogenies, most of the SNPs confer synonymous mutations that do not affect the amino acid sequences. Therefore positive selection is not likely to be occurring on the operon as a whole or on a codon-by-codon basis.

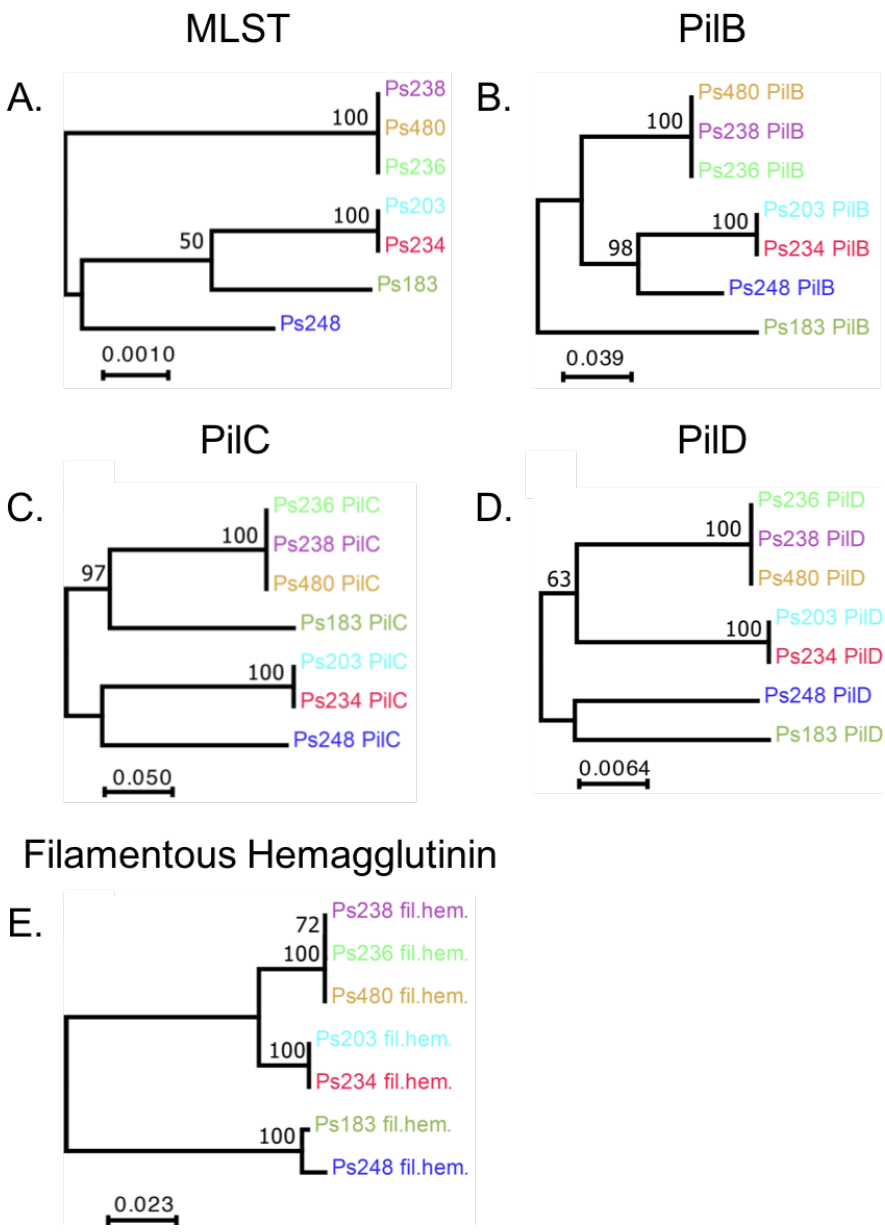


Figure 5. MLST and gene phylogenies of type IV pilus genes and filamentous hemagglutinin within *P. syringae* cotton pathogens genomes. Maximum Likelihood phylogenetic tree based on nucleotide sequences of A) Concatenated regions of housekeeping genes: *gyrB*, *rpoD*, *gap1*, and *gltA*. B) *pilB,* C) *pilC,* D) *pilD,* and E) Filamentous hemagglutinin within *P. syringae* cotton pathogens. Branch labels are bootstrap percentage values based on 100 bootstraps.

3b.3.6 The diversification of type IV pilus genes across the *Pseudomonas syringae* species

In order to investigate whether the observed genomic diversity was unique to the *P. syringae* cotton pathogen or common among the *P. syringae* species as a whole, we considered representatives of each of the five major *P. syringae* phylogroups. Unfortunately, we were not able to include filamentous hemagglutinin in this analysis because most of these genomes were assembled using Illumina data, which is not conducive to assembling these large,

repetitive genes. Therefore, we focused on the Type IV pilus operon. As before, the *pilC* and to a lesser extent *pilB* phylogenetic trees had a greater depth than the MLST or *pilD* trees (Fig. 6). In contrast to Figure 5, Ps248 *pilB*, *pilC*, and *pilD* genes all remained within the Phylogroup 2 clade (Fig 6). This suggests that other evolutionary forces such as varying levels of pathoadaptation, not horizontal gene transfer, may be the cause of the genetic diversity of these genes within Ps248. In contrast, the *pilB*- and *pilC*- based trees placed Ps183 in a separate clade with *P. syringae* pv. *tabaci* LMG5393, a Phylogroup 3 pathogen (Fig. 6). This trend is maintained in the amino acid sequences (Sup. Fig. 16). This suggests that the *pilB* and *pilC* genes of isolate Ps183 may have undergone horizontal gene transfer from a Phylogroup 3 pathogen. This would also explain why there was no large change in GC content within the region of the *pilBCD* operon because a horizontal gene transfer among closely related strains would not show this marker (Sup. Fig. 11).

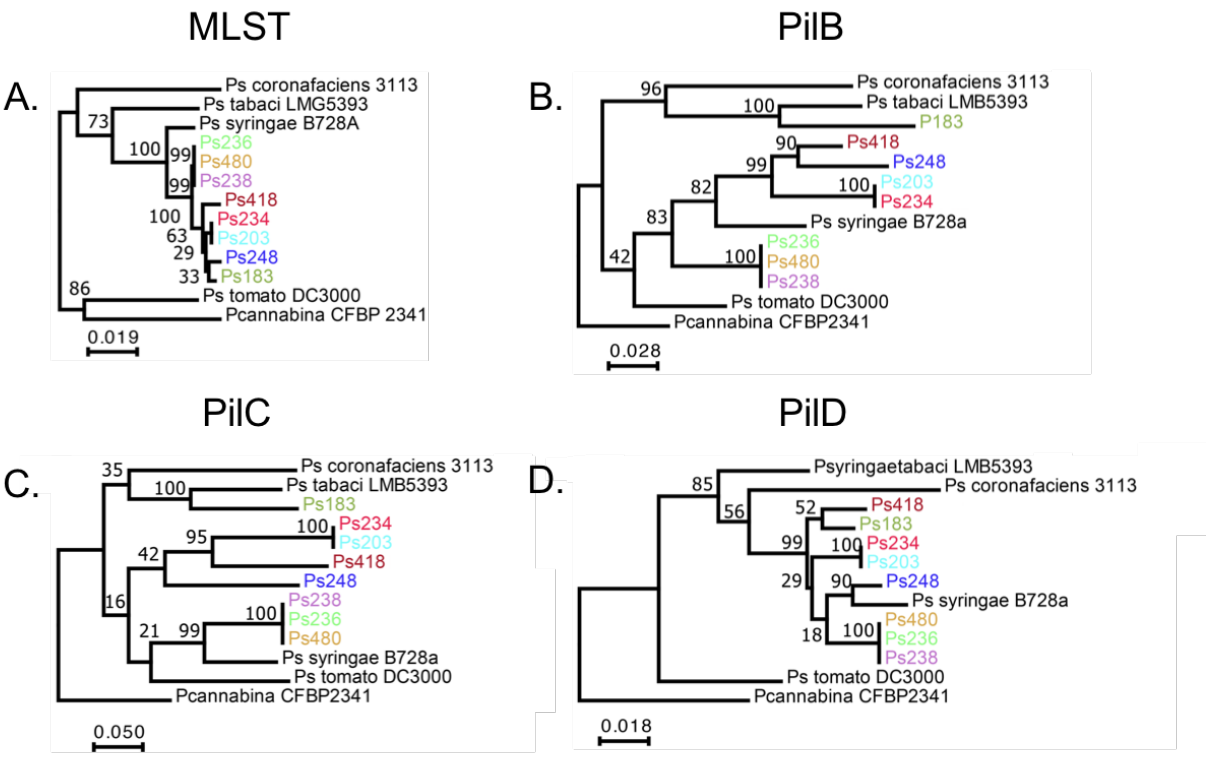


Figure 6. MLST and gene phylogenies of Type IV Pilus Genes spanning 5 *P. syringae* phylogroups. Maximum Likelihood phylogenetic tree based on nucleotide sequences of A) concatenated regions of housekeeping genes: *gyrB*, *rpoD*, *gap1*, and *gltA*. B) *pilB* C) *pilC* D) *pilD* within *P. syringae* pathogens across 5 phylogroups. Branch labels are bootstrap percentage values based on 100 bootstraps.

To further understand how the pairwise nucleotide diversity of the *pilBCD* operon compares to other genes and other pathosystems, we calculated pairwise nucleotide diversity of housekeeping and virulence related genes from *P. syringae* pathogens of tomato, kiwi, and cherry (Sup. Table 4)[65–67]. Each of these datasets provided a unique view of *P. syringae* host-pathogen evolution. The *P. syringae* pv. *tomato* T1 pathogens were previously determined to be "monomorphic" with low genetic diversity[68]. In contrast, the *P. syringae* cherry pathogens were previously placed into three different phylogroups and likely have separate evolutionary origins[65]. The kiwi pathogens hypothesized to have recently emerged from a single origin and therefore have an intermediate amount of genetic diversity[67]. The commonly used MLST genes

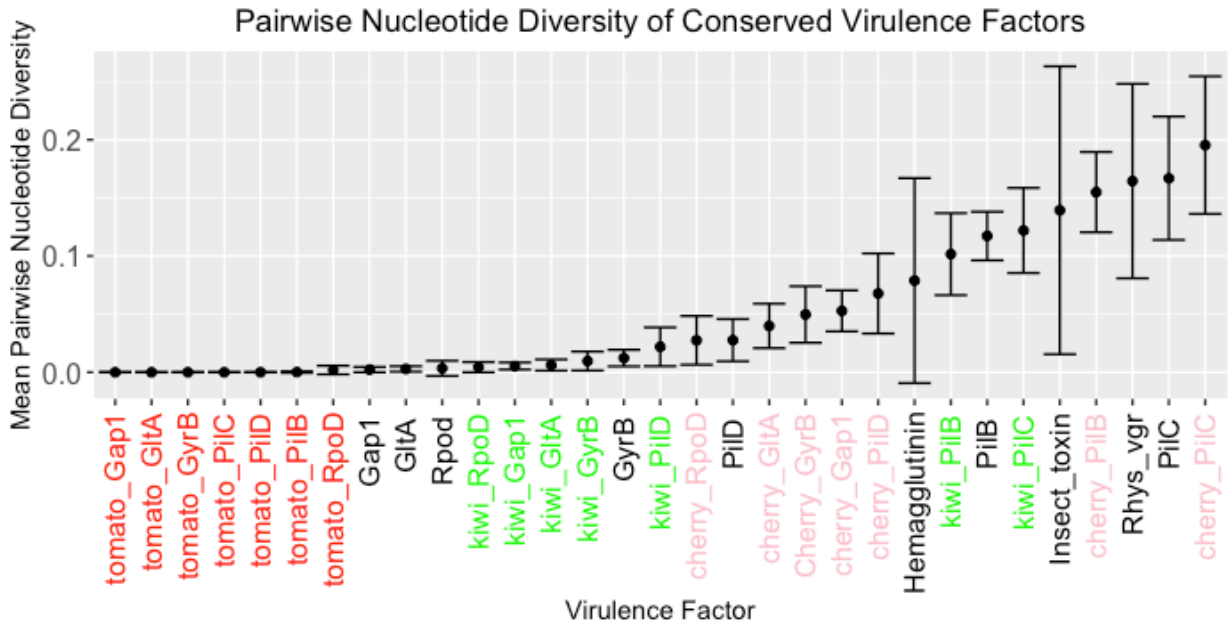


Figure 7. Pairwise nucleotide diversity among conserved virulence factors and housekeeping genes within 4 *P. syringae* pathovars. Pairwise Nucleotide Diversity was assessed in 100bp windows across nucleotide clustal alignments of housekeeping genes and virulence factors that are conserved in *P. syringae* cotton pathogen genomes (black) As well as housekeeping and *pilB*, *pilC*, *pilD* genes from T1 tomato pathogens (red), kiwi pathogens (green), and cherry pathogens (pink). Error bars: Standard Deviation.

gap1, *gltA*, *gyrB*, and *rpoD*, were used as a proxy for overall genomic nucleotide diversity within each dataset. As expected, the tomato T1 housekeeping genes had the lowest diversity, followed by kiwi and then cherry (Fig. 7). In fact, only *rpoD* had any SNPs differentiating tomato T1 strains. The same trend held for *pilB*, *pilC*, and *pilD* genes. *pilC* was the most diverse gene in the operon in kiwi, cherry, and cotton datasets, followed by *pilB* and then *pilD*, which was only slightly more diverse than the housekeeping genes. Among the *pilB*, *pilC*, and *pilD* genes in the tomato T1 isolates, only *pilB* displayed any SNPs. The *pilB* and *pilC* genes of the cherry pathogens displayed the highest levels of nucleotide diversity among all pathovars. This was likely due to the basal levels of diversity among these pathogens, as shown by the high diversity within housekeeping genes (Fig. 7). The low levels of diversity among the housekeeping genes

of cotton pathogens compared to the high levels of diversity within *pilB* and *pilC* suggests that cotton pathogens may be undergoing high levels of diversifying selection at this locus (Fig. 7).

Table 1: Population Genetic Analyses of Putative Virulence Factors within *P. syringae* cotton pathogens. Statistics were calculated using whole gene nucleotide alignments using MEGA7. Genes are ordered in ascending order of Tajima's D value.

Gene	Positions in Final Dataset (n)	Segregating Sites (S)	S/n (P_s)	P_s/a1 (Θ)	π	Tajima's D
HopAC1-1	1929	54	0.028	0.011	0.009	-0.986
HopA1	1149	17	0.015	0.006	0.005	-0.753
HopAN1	1290	26	0.020	0.008	0.008	-0.455
AvrE1	4863	175	0.360	0.015	0.014	-0.242
HopAJ2	1338	24	0.018	0.007	0.007	-0.212
HopAH1	1284	22	0.017	0.007	0.007	-0.017
HopJ1	342	13	0.038	0.016	0.016	0.126
HopAC1-2	6069	150	0.025	0.010	0.010	0.143
HopAK1	1617	77	0.048	0.019	0.020	0.159
HopM1	2145	60	0.028	0.011	0.012	0.425
PilD	873	55	0.063	0.026	0.028	0.621
PilB	1695	440	0.260	0.106	0.118	0.644
MLST	2010	27	0.013	0.005	0.006	0.653
HopAG1	2109	46	0.022	0.009	0.010	0.669
HopAA1	1452	42	0.029	0.012	0.014	0.960
HrcC	2106	21	0.010	0.004	0.005	1.000
HopAH2	1143	15	0.013	0.005	0.006	1.093
Fil. Hemagglutinin	14222	2329	0.164	0.067	0.080	1.188
Rhs Element Vgr	1841	634	0.344	0.141	0.169	1.209
HrpE	582	10	0.017	0.007	0.009	1.270
HrpK1	2307	48	0.021	0.008	0.010	1.356
HopAI1	804	29	0.036	0.015	0.018	1.363
PilC	1218	422	0.346	0.141	0.178	1.528
Insecticidal Toxin	7344	1866	0.254	0.104	0.135	1.794
HrpA	327	15	0.046	0.019	0.025	1.867

However, high levels of nucleotide diversity alone cannot be used as evidence of diversifying selection. Tajima's D compares the pairwise nucleotide diversity to the expected

level of variation, which is calculated from the sum of segregating sites. This allows for the identification of the presence of an overabundance of rare alleles, indicative of diversifying selection. Using this statistic, *pilC* has a positive value (1.528), confirming that it is under diversifying selection (Table 1). This value is much higher than the neutrally evolving MLST regions of *gap1*, *gltA*, gyrB, and *rpoD* (0.653). Interestingly, two other putative virulence factors have even higher Tajima's D values than *pilC*: an insecticidal toxin encoding gene and *hrpA*. Therefore diversifying selection may be acting on multiple *P. syringae* virulence factors. In contrast, several other T3E have negative Tajima's D values indicating that they are under purifying selection. This suggests that these genes are likely important for pathogen fitness. The particular importance of PilC in the *P. syringae*- cotton pathosystem is supported by the larger differences in tree depths between the *pilBCD* operon trees and the MLST reference tree in figure 5 than in figure 6 where *P. syringae* strains from across the species are included. Taken together, these data suggest that higher diversifying selection may be acting on cotton pathogen genomes than on the genomes of other *P. syringae* pathovars. This further suggests that PilB and PilC may be PTI targets of cotton and therefore the type IV pilus may play an important role in the host-pathogen arms race in this pathosystem.

3b.4 Materials and Methods

3b.4.1 Isolate collection

Pseudomonas syringae isolates were collected from diseased leaf tissue originating from cotton fields in Texas. Leaf tissue was macerated in 10mM $MgCl_2$ and plated on either NYG or KB agar and incubated for two days at 30 degrees Celsius. Single colonies were isolated and tested for Koch's postulates and 16S identity as previously described[69]. Five isolates were

identified in this way from material collected from 2015 to 2017 (Ps183, Ps203, Ps234, Ps248, and Ps480). Two additional Isolates were obtained from culture collections: *P. syringae* pv. *aptata* CFBP 1617 (Ps418) and *P. syringae* ATCC 51506 (Ps236, Ps238) isolated from cotton in 1994. *P. syringae* ATCC 51506 was split into two samples because the original culture contained colonies of two morphologies: white-Ps236 and yellow-Ps238. Cultures of both cell types were able to cause similar amounts of disease, therefore both samples were included in this study. In total, this resulted in 8 *P. syrinage* isolates.

3b.4.2 Inoculations

Cotton variety DES56 was grown under greenhouse conditions until five fully expanded leaves appeared. Fully expanded leaves were inoculated with $1x10^6$ CFU/ml of bacteria suspended in 10mM $MgCl_2$ with a needleless syringe. Disease assays were conducted in a growth chamber set at 30 degrees Celsius with 80% humidity and 14 hour days. Images were taken four days post inoculation.

3b.4.3 Genome sequencing, assembly, and annotation

DNA with high molecular weight was extracted using a standard CTAB DNA preparation. DNA was sequenced using a nanopore MinION R9 flow cell and SQK-RAD004 Rapid Sequencing kit. Genomes were assembled with Canu and polished with Nanopolish[42,43]. These genomes were circularized and chromosomes were reoriented to DnaA and plasmids were reoriented to RepA.

100x Shotgun Illumina MiSeq library prep, 2x250 paired-end sequencing, and trimming was performed at the Roy J. Carver Biotechnology Center at the University of Illinois at Urbana-Champaign. Bacterial genomes were then polished again with Pilon using the paired-end reads[45].

In total, two rounds of Nanopolish and three rounds of Pilon were performed. This resulted in genomes with gammaproteobacteria BUSCO scores >98% (Sup. Fig. 6)[44]. Chromosome sizes range from 5,936,430bp to 6,087,715bp. Genomes Ps234, Ps236, Ps238, and Ps480 contained a 68kb plasmid and Ps418 contained a homologous 58kb plasmid. Genomes were annotated using prokka and a database of T3Es, as described previously[70] as well as the database of *P. syringae* pv. *tomato* DC3000 effectors[46,71].

Sample Ps480 was sequenced separately using the SMRT PacBio platform. This genome was assembled with Falcon and polished with Quiver[40,41].

3b.4.4 Phylogenetics

Multi locus sequence typing analysis was performed using concatenated regions of the *gyrB*, *rpoD*, *gap1*, and *gltA* genes. The concatenated regions of the newly sequenced strains were aligned with corresponding regions of 57 *Pseudomonas syringae* isolates using the PAMDB database and Mega7 and a Maximum Likelihood tree was generated[72,73]. Bootstrap values represent the percentage of trees out of 500 that support that branch. All alignment gaps were removed from the analysis, leaving 2001 base pairs in the final dataset. Branch lengths are numbers of substitutions per site.

In order to generate a more finely tuned tree of the *P. syringae* cotton pathogens, whole chromosomes were aligned using default settings on Mauve[47]. This tool identifies rearrangements and possible HGT events through the detection of colinear blocks. This allows for phylogenetic analysis to be performed on data that has been corrected for these events that may skew the results. 331,898 SNPs among all colinear blocks were identified by Mauve. The SNPs from the Mauve SNP file were concatenated and a Maximum Likelihood tree was

generated using Mega7. Bootstrap values represent the percentage of trees out of 500 that support that branch. All alignment gaps were removed from the analysis.

3b.4.5 Virulence Factor Identification

Prokka along with two databases of effectors were used to identify the virulence factors in each genome. The type III effector annotations were manually inspected. Any effector annotated in some but not all genomes was against the other genomes using BLAST in order to determine if a frame shift mutation or early stop codon caused this effector to not be annotated.

The biochemical pathways of *P. syringae* toxins were first searched for in Prokka annotations. If no Prokka annotations were found, putative biosynthetic clusters were identified by comparing functionally annotated toxin pathways against the newly sequenced genomes using synteny and BLAST. A tBLASTn e-value cutoff of $1x10^{-6}$ was used for significance, as used by the annotation software prokka[46].

3b.4.6 Analysis of Nucleotide Diversity and Selection

The Mauve SNP file was converted to a VCF file using snp-sites, using ps480 as the reference[74]. Then, VCFtools was used to calculate pairwise nucleotide diversity (pi) across the genome within the LCB's using a 100bp sliding window staggered every 50bp[54]. Gaps were removed from the analysis.

Peaks with pairwise nucleotide diversity greater than 0.25 were manually inspected for alignment errors. Genomic regions that did not have alignment errors were identified in each genome and aligned to assess nucleotide diversity across 100bp non-overlapping windows using VCFtools[54]. Maximum Likelihood phylogenetic trees, population genetics, and dN-dS analyses

of these genes were generated using Mega7. Bootstrap values represent the percentage of trees out of 100 that support that branch. All alignment gaps were removed from the analysis.

3b.5 Discussion

This manuscript represents the first report of the genetic diversity of the *Pseudomonas syringae* pathogen of cotton. Our thorough investigation of the phylogenetics, genomics, and virulence factor arsenals of these isolates has revealed many insights into how this pathogen causes disease and how it participates in the host-pathogen arms race. However, it continues to be a mystery as to why so few reports of this disease have occurred. Phylogenetic analysis suggests that this is not a new disease to cotton. Rather, it has likely laid dormant or unobserved since it was last identified in 1994. It is possible that changes in the environment or agricultural practices may have allowed it to re-emerge. For example, an increase in rainfall or a change in the methods of acid delinting cotton seed may have helped this pathogen proliferate in cotton fields. Alternatively, it is possible that resistant germplasm was unknowingly being used until recently, similar to how *Xanthomonas citri* pv. *malvacearum* emerged on cotton in 2011(Anne Z Phillips et al., 2017). Notably, no qualitative genetic resistance to this pathogen has been identified in contemporary *G. hirsutum* varieties. Future work will focus on screening cotton diversity panels for sources of resistance.

The most intriguing hypothesis for the intermittent emergence of this pathogen is that it is an opportunistic pathogen. Much like opportunistic pathogens in a hospital setting, it may only infect hosts in an agricultural setting that are pre-disposed or weakened due to a pre-existing disease[75–77]. This hypothesis is supported by the fact that every pathogenic *P. syringae* that we isolated also came from a plant that was infected with *Xcm.* This suggests the intriguing

hypothesis that *P. syringae* and *Xcm* cause a disease complex on cotton. Bacterial-bacterial disease complexes are rare across all pathosystems but they are not unheard of[78–80]. Therefore, future research will focus on investigating this system as a possible disease complex.

An alternative approach for investigating the emergence of this disease is by exploring the possibility that it is a member of a *P. syringae* phylogroup 2 generalist pathovar. Phylogroup 2 *P. syringae* pathogens are well known for their wide host ranges[32]. Our finding that a sugar beet pathogen (Ps418) can cause disease on cotton supports this hypothesis. If this pathogen is indeed a generalist, its presence in agricultural fields may ebb and flow depending on the contact that cotton has with environmental sources of contamination such as weeds, shared farm equipment, and rotation crops. Virulence assays on multiple hosts will illuminate whether this pathogen is indeed a generalist or if it has been more specifically co-evolving with cotton for millennia. These experiments will also determine the pathovar designation of these isolates. As of now, the necessary disease assays required to designate these isolates as the member of an existing or new pathovar have not been performed. However, given the recent studies demonstrating an "overlapping continuum" of host ranges within *P. syringae*, perhaps the practice of designating strains into pathovars is antiquated[32].

It has been previously suggested that phylogroup 2 pathogens are generalists due to the many toxins that they often secrete[32,50]. A necrotic phenotype is common in phylogroup 2 pathogens and often obscures the difference between resistance and susceptible responses without the use of CFU growth assays[31]. The characteristic rapid spreading necrosis of the cotton pathogens described in this study suggests that they too secrete toxins during pathogenesis. We outline several potential culprits including syringomycin, syringopeptin, syringolin, and mangotoxin, which are all known to cause necrotic lesions on hosts[49,81–83]. Syringafactin is better

known as a surfactant than a toxin, but its potential contributions to this phenotype cannot be ruled out[84–86]. Reverse genetics, diffusion, erythrocyte lysis, and surface tension assays as well as LC-MS will be needed to definitively identify which families of lipopeptides are produced and whether they have toxic properties and/or are involved in pathogen virulence on cotton.

When a new or relatively unknown disease emerges, it is now common practice to first identify the arsenal of virulence factors present in the pathogen genome[26,38,65,67]. The motivation for this analysis is the hypothesis that presence/absence changes in virulence factors such as T3E can trigger a host shift or new pathogen race[14,87]. While this method is valid and has revealed many fascinating examples of pathogen evolution[88–91], we suggest that going one step further and characterizing the evolutionary forces underlying virulence factor variation can not only explain how a host change occurs, but also predict how a host change might occur in the future. The method of studying the evolutionary forces acting on virulence factors is not new. In fact, numerous studies have revealed mechanisms such as HGT, recombination, and pathoadaptation that alter the sequences of T3E and therefore tilt the arms race in the pathogen's favor[7,17,18,50]. Interestingly, while the *P. syringae* cotton pathogens have many presence/absence changes in their T3E repertoire, all conserved effectors have low pairwise nucleotide diversity, comparable to the rest of the genome and their Tajima's D values vary greatly from gene to gene. HrpA, which encodes the subunits of the T3SS needle, displays the highest Tajima's D value of all virulence factors tested. This supports previous analysis that the gene is undergoing high levels of diversifying selection across the *P. syringae* species[17]. In contrast, many other T3Es have negative Tajima's D values suggesting that they are undergoing purifying selection. This indicates that diversifying selection is not acting on each of these genes equally.

Furthermore, we apply these statistics to identify two relatively unknown virulence factors in *P. syringae*: the type IV pilus and Filamentous Hemagglutinin. We find that these genes are undergoing strong diversifying selection, suggesting their involvement in the host-pathogen arms race. These virulence factors are less well studied in *Pseudmonas syringae* than T3Es, perhaps due to the lack of known gene-for-gene interactions in this system. However, both genes are prevalent in both pathogenic and non-pathogenic bacteria and have been implicated as virulence factors and immunity targets in human pathogens such as *Neisseria gonorrhoeae* and *Bordetella pertussis*[60,92,93]. The predominant function of both systems is cell-cell adhesion, but the type IV pilus has also been implicated in biofilm formation, swarming, and twitching motility, all functions important for virulence[53,56,63]. Furthermore, both virulence factors are present on the outside of the pathogen cell, making them likely targets for host pattern triggered immunity (PTI).

The function of the type IV pilus in plant-pathogen systems is less clear than in mammalian-pathogen systems. Their importance in pathogenicity, movement, and seed transmission has been implicated[94–97], though not in all systems[63,97]. For example, abolishing the type IV pilus in *Xylella* decreased its ability to move upstream of water movement, but actually increased biofilm formation[96]. Filamentous hemagglutinin has also been found to be a virulence factor of the plant pathogens *Xanthomonas axonopodis* pv. *citri*, *Xylella fastidosa*, and necrotrophic pathogens such as *Erwinia chrysanthemi*[59,98,99]. However, it is also upregulated during biofilm formation of the plant growth promoting rhizobacteria *Enterobacter*[100]. Therefore, the function of this protein varies from bacteria to bacteria.

Our identification of high Tajima's D values among *pilC* and filamentous hemagglutinin genes in closely related *P. syringae* isolates supports the hypothesis that the host-pathogen arms

race is acting on these genes through diversifying selection. This is further supported by the juxtaposition against the relatively low rates of nucleotide diversity in these genes within tomato, cherry, and kiwi pathogens. Interestingly, the markers of diversifying selection are quite different between the filamentous hemagglutinin gene and the type IV pilus *pilBCD* operon. While the *pilBCD* operon shows equally distributed SNPs indicative of pathoadaptation as well as evidence of horizontal gene transfer in one genome, filamentous hemagglutinin is repetitive and modular in nature which may have allowed for high rates of recombination, representative of its common status as a polymorphic toxin[18,55]. High levels of sequence variation within both type IV pilus and filamentous hemagglutinin genes has also been reported within other pathosystems[99,101–103]. This suggests that the function of these genes in the molecular arms-race may extend past the *P. syringae – G. hirsutum* pathosystem.

The functions of the type IV pilus genes *pilB* and *pilC*, are inherently interconnected. PilB is a cytoplasmic pilus polymerization ATPase[104]. Its N-terminal domain physically interacts with and is coordinated by the inner membrane platform protein PilC[104,105]. Therefore, it is a logical conclusion that if diversifying selection (driven by host PTI) were to act on *pilB*, reciprocal mutations would accumulate in *pilC* and vice versa. PilD is a peptidase that processes prepilins into mature pilins that can be incorporated into the pilus structure[106]. Therefore, this role may require more functional constraint in order for the structural integrity of the pilus to be preserved. Future research will focus on reverse genetics to identify the roles that these genes play in pathogenicity on cotton as well as which aspects of host PTI recognize these virulence factors. Further, we will determine if the *pilB* and *pilC* genes are interchangeable among *P. syringae* cotton pathogens or if they have diversified so much as to only be functional within their original Pseudomonad genome.

Collectively, the use of long-read sequencing has allowed us to rapidly characterize an emerging disease, propose hypotheses for its emergence, and investigate how it interacts in the host-pathogen arms race. Through phylogenetics and comparative genomics, we determine that this rare but re-emerging *P. syringae* pathogen of cotton is the same phylogroup 2 pathogen that was last identified in 1994. We also identify several putative virulence factors including two T3SSs, many T3Es, and five lipopeptide toxin biosynthetic clusters that may contribute to the characteristic rapid spreading necrosis of this pathogen. Further, we probed the host-pathogen arms race to identify two putative virulence factors that may be involved in PTI. Ultimately these data lay the foundation for understanding how this pathogen emerged and how we can develop durable resistance strategies against it.

3b.6 References

1. Abrams PA. Is predator-prey coevolutlon an arms race? *Trends Ecol Evol*. 1986;1(4):108-110. doi:10.1016/0169-5347(86)90037-6

2. Dawkins Richard, Krebs John Richard. Arms races between and within species. *Proc R Soc Lond B Biol Sci*. 1979;205(1161):489-511. doi:10.1098/rspb.1979.0081

3. Boutrot F, Zipfel C. Function, Discovery, and Exploitation of Plant Pattern Recognition Receptors for Broad-Spectrum Disease Resistance. *Annu Rev Phytopathol*. 2017;55(1):257-286. doi:10.1146/annurev-phyto-080614-120106

4. Jones JDG, Dangl JL. The plant immune system. *Nature*. 2006;444(7117):323-329. doi:10.1038/nature05286

5. Francis M. Regulation of type III secretion systems. *Curr Opin Microbiol*. 2002;5(2):166-172. doi:10.1016/S1369-5274(02)00301-6

6. Lindeberg M, Cunnac S, Collmer A. Pseudomonas syringae type III effector repertoires: last words in endless arguments. *Trends Microbiol*. 2012;20(4):199-208. doi:10.1016/j.tim.2012.01.003

7. Dillon MM, Almeida RND, Laflamme B, et al. Molecular Evolution of Pseudomonas syringae Type III Secreted Effector Proteins. *Front Plant Sci*. 2019;10. doi:10.3389/fpls.2019.00418

8. Stavrinides J, McCann HC, Guttman DS. Host–pathogen interplay and the evolution of bacterial effectors. *Cell Microbiol*. 2008;10(2):285-292. doi:10.1111/j.1462-5822.2007.01078.x

9. Cook DE, Mesarich CH, Thomma BPHJ. Understanding Plant Immunity as a Surveillance System to Detect Invasion. *Annu Rev Phytopathol*. 2015;53(1):541-563. doi:10.1146/annurev-phyto-080614-120114

10. Kanyuka K, Rudd JJ. Cell surface immune receptors: the guardians of the plant's extracellular spaces. *Curr Opin Plant Biol*. 2019;50:1-8. doi:10.1016/j.pbi.2019.02.005

11. Thomma BPHJ, Nürnberger T, Joosten MHAJ. Of PAMPs and Effectors: The Blurred PTI-ETI Dichotomy. *Plant Cell*. 2011;23(1):4-15. doi:10.1105/tpc.110.082602

12. Carroll L. *Through the Looking-Glass and What Alice Found There*. Macmillan; 1871.

13. Lee VanValen. A new evolutionary law. *Evol Theory*. 1973;1:1-30.

14. Baltrus DA, McCann HC, Guttman DS. Evolution, genomics and epidemiology of Pseudomonas syringae. Molecular Plant Pathology. doi:10.1111/mpp.12506

15. Sokurenko EV, Hasty DL, Dykhuizen DE. Pathoadaptive mutations: gene loss and variation in bacterial pathogens. *Trends Microbiol*. 1999;7(5):191-195. doi:10.1016/S0966-842X(99)01493-6

16. Rohmer L, Guttman DS, Dangl JL. Diverse Evolutionary Mechanisms Shape the Type III Effector Virulence Factor Repertoire in the Plant Pathogen Pseudomonas syringae. *Genetics*. 2004;167(3):1341-1360. doi:10.1534/genetics.103.019638

17. Guttman DS, Gropp SJ, Morgan RL, Wang PW. Diversifying Selection Drives the Evolution of the Type III Secretion System Pilus of Pseudomonas syringae. *Mol Biol Evol*. 2006;23(12):2342-2354. doi:10.1093/molbev/msl103

18. Ma W, Dong FFT, Stavrinides J, Guttman DS. Type III Effector Diversification via Both Pathoadaptation and Horizontal Transfer in Response to a Coevolutionary Arms Race. *PLOS Genet*. 2006;2(12):e209. doi:10.1371/journal.pgen.0020209

19. Boykin LM, Sseruwagi P, Alicai T, et al. Tree Lab: Portable genomics for Early Detection of Plant Viruses and Pests in Sub-Saharan Africa. *Genes*. 2019;10(9):632. doi:10.3390/genes10090632

20. Gardy JL, Loman NJ. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat Rev Genet*. 2018;19(1):9-20. doi:10.1038/nrg.2017.88

21. Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol*. 2016;17(1). doi:10.1186/s13059-016-1103-0

22. Quick J, Grubaugh ND, Pullan ST, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc*. 2017;12(6):1261-1276. doi:10.1038/nprot.2017.066

23. Lamichhane JR, Messéan A, Morris CE. Insights into epidemiology and control of diseases of annual plants caused by the Pseudomonas syringae species complex. *J Gen Plant Pathol*. 2015;81(5):331-350. doi:10.1007/s10327-015-0605-z

24. Giesler LJ. Bacterial Diseases of Soybean. *NebGuide*. 2011:2.

25. Preston GM. Pseudomonas syringae pv. tomato: the right pathogen, of the right plant, at the right time. *Mol Plant Pathol*. 2000;1(5):263-275. doi:10.1046/j.1364-3703.2000.00036.x

26. Scortichini M, Marcelletti S, Ferrante P, Petriccione M, Firrao G. Pseudomonas syringae pv. actinidiae: a re-emerging, multi-faceted, pandemic pathogen. *Mol Plant Pathol*. 2012;13(7):631-640. doi:10.1111/j.1364-3703.2012.00788.x

27. Buell CR, Joardar V, Lindeberg M, et al. The complete genome sequence of the Arabidopsis and tomato pathogen Pseudomonas syringae pv. tomato DC3000. *Proc Natl Acad Sci*. 2003;100(18):10181-10186. doi:10.1073/pnas.1731982100

28. Xin X-F, He SY. Pseudomonas syringae pv. tomato DC3000: A Model Pathogen for Probing Disease Susceptibility and Hormone Signaling in Plants. *Annu Rev Phytopathol*. 2013;51(1):473-498. doi:10.1146/annurev-phyto-082712-102321

29. Xin X-F, Kvitko B, He SY. Pseudomonas syringae: what it takes to be a pathogen. *Nat Rev Microbiol*. 2018;16(5):316-328. doi:10.1038/nrmicro.2018.17

30. Berge O, Monteil CL, Bartoli C, et al. A User's Guide to a Data Base of the Diversity of Pseudomonas syringae and Its Application to Classifying Strains in This Phylogenetic Complex. *PLOS ONE*. 2014;9(9):e105547. doi:10.1371/journal.pone.0105547

31. Maraite H, Weyns J. Pseudomonas Syringae pv. aptata and pv. atrofaciens, Specific Pathovars or Members of pv. Syringae? In: Rudolph K, Burr TJ, Mansfield JW, Stead D, Vivian A, von Kietzell J, eds. *Pseudomonas Syringae Pathovars and Related Pathogens*. Vol 9. Dordrecht: Springer Netherlands; 1997:515-520. doi:10.1007/978-94-011-5472-7_94

32. Morris CE, Lamichhane JR, Nikolić I, Stanković S, Moury B. The overlapping continuum of host range among strains in the Pseudomonas syringae complex. *Phytopathol Res*. 2019;1(1):4. doi:10.1186/s42483-018-0010-6

33. Flajoulot S, Ronfort J, Baudouin P, et al. Genetic diversity among alfalfa (Medicago sativa) cultivars coming from a breeding program, using SSR markers. *Theor Appl Genet*. 2005;111(7):1420-1429. doi:10.1007/s00122-005-0074-4

34. Liu K, Goodman M, Muse S, Smith JS, Buckler E, Doebley J. Genetic Structure and Diversity Among Maize Inbred Lines as Inferred From DNA Microsatellites. *Genetics*. 2003;165(4):2117-2128.

35. Tyagi P, Gore MA, Bowman DT, Campbell BT, Udall JA, Kuraparthy V. Genetic diversity and population structure in the US Upland cotton (Gossypium hirsutum L.). *Theor Appl Genet*. 2014;127(2):283-295. doi:10.1007/s00122-013-2217-3

36. Aboughanem-Sabanadzovic N, Allen TW, Wilkerson TH, et al. First Report of Cotton Leafroll Dwarf Virus in Upland Cotton (Gossypium hirsutum) in Mississippi. *Plant Dis*. 2019;103(7):1798. doi:10.1094/PDIS-01-19-0017-PDN

37. Cox KL, Babilonia K, Wheeler T, He P, Shan L. Return of old foes — recurrence of bacterial blight and Fusarium wilt of cotton. *Curr Opin Plant Biol*. 2019;50:95-103. doi:10.1016/j.pbi.2019.03.012

38. Phillips AZ, Berry JC, Wilson MC, et al. Genomics-enabled analysis of the emergent disease cotton bacterial blight. *PLOS Genet*. 2017;13(9):23.

39. Phillips AZ, Wheeler T, Woodward J, Bart RS. *Pseudomonas syringae* Pathogen Causes Foliar Disease of Upland Cotton in Texas. *Plant Dis*. 2018;102(6):1171-1171. doi:10.1094/PDIS-11-17-1700-PDN

40. Chin C-S, Alexander DH, Marks P, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods*. 2013;10(6):563-569. doi:10.1038/nmeth.2474

41. Chin C-S, Peluso P, Sedlazeck FJ, et al. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods*. 2016;13(12):1050-1054. doi:10.1038/nmeth.4035

42. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive *k* -mer weighting and repeat separation. *Genome Res*. 2017;27(5):722-736. doi:10.1101/gr.215087.116

43. Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled *de novo* using only nanopore sequencing data. *Nat Methods*. 2015;12(8):733-735. doi:10.1038/nmeth.3444

44. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31(19):3210-3212. doi:10.1093/bioinformatics/btv351

45. Walker BJ, Abeel T, Shea T, et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLOS ONE*. 2014;9(11):e112963. doi:10.1371/journal.pone.0112963

46. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30(14):2068-2069. doi:10.1093/bioinformatics/btu153

47. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. *Genome Res*. 2004;14(7):1394-1403. doi:10.1101/gr.2289704

48. Hacker J, Blum-Oehler G, Mühldorfer I, Tschäpe H. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol Microbiol*. 1997;23(6):1089-1097. doi:10.1046/j.1365-2958.1997.3101672.x

49. Martínez-García PM, Rodríguez-Palenzuela P, Arrebola E, et al. Bioinformatics Analysis of the Complete Genome Sequence of the Mango Tree Pathogen Pseudomonas syringae pv. syringae UMAF0158 Reveals Traits Relevant to Virulence and Epiphytic Lifestyle. *PLOS ONE*. 2015;10(8):e0136101. doi:10.1371/journal.pone.0136101

50. Baltrus DA, Nishimura MT, Romanchuk A, et al. Dynamic Evolution of Pathogenicity Revealed by Sequencing and Comparative Genomics of 19 Pseudomonas syringae Isolates. Guttman DS, ed. *PLoS Pathog*. 2011;7(7):e1002132. doi:10.1371/journal.ppat.1002132

51. Aguilera S, Lopez-Lopez K, Nieto Y, et al. Functional Characterization of the Gene Cluster from Pseudomonas syringae pv. phaseolicola NPS3121 Involved in Synthesis of Phaseolotoxin. *J Bacteriol*. 2007;189(7):2834-2843. doi:10.1128/JB.01845-06

52. Kinscherf TG, Willis DK. The Biosynthetic Gene Cluster for the β-Lactam Antibiotic Tabtoxin in Pseudomonas syringae. *J Antibiot (Tokyo)*. 2005;58(12):817. doi:10.1038/ja.2005.109

53. Pelicic V. Type IV pili: e pluribus unum? *Mol Microbiol*. 2008;68(4):827-837. doi:10.1111/j.1365-2958.2008.06197.x

54. Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156-2158. doi:10.1093/bioinformatics/btr330

55. Jamet A, Nassif X. New Players in the Toxin Field: Polymorphic Toxin Systems in Bacteria. *mBio*. 2015;6(3). doi:10.1128/mBio.00285-15

56. Zhang D, Souza RF de, Anantharaman V, Iyer LM, Aravind L. Polymorphic toxin systems: Comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. *Biol Direct*. 2012;7(1):18. doi:10.1186/1745-6150-7-18

57. Choi HW, Kim DS, Kim NH, Jung HW, Ham JH, Hwang BK. Xanthomonas Filamentous Hemagglutinin-Like Protein Fha1 Interacts with Pepper Hypersensitive-Induced Reaction Protein CaHIR1 and Functions as a Virulence Factor in Host Plants. *Mol Plant Microbe Interact*. 2013;26(12):1441-1454. doi:10.1094/MPMI-07-13-0204-R

58. Dunger G, Llontop E, Guzzo CR, Farah CS. The Xanthomonas type IV pilus. *Curr Opin Microbiol*. 2016;30:88-97. doi:10.1016/j.mib.2016.01.007

59. Guilhabert MR, Kirkpatrick BC. Identification of Xylella fastidiosa Antivirulence Genes: Hemagglutinin Adhesins Contribute to X. fastidiosa Biofilm Maturation and Colonization and Attenuate Virulence. *Mol Plant Microbe Interact*. 2005;18(8):856-868. doi:10.1094/MPMI-18-0856

60. Henderson MW, Inatsuka CS, Sheets AJ, et al. Contribution of Bordetella Filamentous Hemagglutinin and Adenylate Cyclase Toxin to Suppression and Evasion of Interleukin-17-Mediated Inflammation. *Infect Immun*. 2012;80(6):2061-2075. doi:10.1128/IAI.00148-12

61. Narulita E, Addy HS, Kawasaki T, Fujie M, Yamada T. The involvement of the PilQ secretin of type IV pili in phage infection in Ralstonia solanacearum. *Biochem Biophys Res Commun*. 2016;469(4):868-872. doi:10.1016/j.bbrc.2015.12.071

62. Nguyen LC, Taguchi F, Tran QM, et al. Type IV pilin is glycosylated in Pseudomonas syringae pv. tabaci 6605 and is required for surface motility and virulence. *Mol Plant Pathol*. 2012;13(7):764-774. doi:10.1111/j.1364-3703.2012.00789.x

63. Roine E, Raineri DM, Romantschuk M, Wilson M, Nunn DN. Characterization of Type IV Pilus Genes in *Pseudomonas syringae* pv. *tomato* DC3000. *Mol Plant Microbe Interact*. 1998;11(11):1048-1056. doi:10.1094/MPMI.1998.11.11.1048

64. Stavrinides J, Ma W, Guttman DS. Terminal Reassortment Drives the Quantum Evolution of Type III Effectors in Bacterial Pathogens. *PLOS Pathog*. 2006;2(10):e104. doi:10.1371/journal.ppat.0020104

65. Hulin MT, Armitage AD, Vicente JG, et al. Comparative genomics of Pseudomonas syringae reveals convergent gene gain and loss associated with specialization onto cherry (Prunus avium). *New Phytol*. 2018;219(2):672-696. doi:10.1111/nph.15182

66. Kunkeaw S, Tan S, Coaker G. Molecular and Evolutionary Analyses of *Pseudomonas syringae* pv. *tomato* Race 1. *Mol Plant Microbe Interact*. 2010;23(4):415-424. doi:10.1094/MPMI-23-4-0415

67. McCann HC, Rikkerink EHA, Bertels F, et al. Genomic Analysis of the Kiwifruit Pathogen Pseudomonas syringae pv. actinidiae Provides Insight into the Origins of an Emergent Plant Disease. *PLOS Pathog*. 2013;9(7):e1003503. doi:10.1371/journal.ppat.1003503

68. Cai R, Lewis J, Yan S, et al. The Plant Pathogen Pseudomonas syringae pv. tomato Is Genetically Monomorphic and under Strong Selection to Evade Tomato Immunity. *PLOS Pathog*. 2011;7(8):e1002130. doi:10.1371/journal.ppat.1002130

69. Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol*. 1991;173(2):697-703. doi:10.1128/jb.173.2.697-703.1991

70. Bart R, Cohn M, Kassen A, et al. High-throughput genomic sequencing of cassava bacterial blight strains identifies conserved effectors to target for durable resistance. *Proc Natl Acad Sci*. 2012;109(28):E1972-E1979. doi:10.1073/pnas.1208003109

71. Lindeberg M, Stavrinides J, Chang JH, et al. Proposed guidelines for a unified nomenclature and phylogenetic analysis of type III Hop effector proteins in the plant pathogen Pseudomonas syringae. *Mol Plant-Microbe Interact MPMI*. 2005;18(4):275-282. doi:10.1094/MPMI-18-0275

72. Almeida NF, Yan S, Cai R, et al. PAMDB, A Multilocus Sequence Typing and Analysis Database and Website for Plant-Associated Microbes. *Phytopathology*. 2010;100(3):208-215. doi:10.1094/PHYTO-100-3-0208

73. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol*. 2016;33(7):1870-1874. doi:10.1093/molbev/msw054

74. Page AJ, Taylor B, Delaney AJ, et al. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb Genomics*. 2016;2(4). doi:10.1099/mgen.0.000056

75. Hunter P. Co-infection: when whole can be greater than the sum. *EMBO Rep*. 2018;19(8):e46601. doi:10.15252/embr.201846601

76. Kim S, Covington A, Pamer EG. The intestinal microbiota: Antibiotics, colonization resistance, and enteric pathogens. *Immunol Rev*. 2017;279(1):90-105. doi:10.1111/imr.12563

77. Suleyman G, Alangaden G, Bardossy AC. The Role of Environmental Contamination in the Transmission of Nosocomial Pathogens and Healthcare-Associated Infections. *Curr Infect Dis Rep*. 2018;20(6):12. doi:10.1007/s11908-018-0620-2

78. Lamichhane JR, Venturi V. Synergisms between microbial pathogens in plant disease complexes: a growing trend. *Front Plant Sci*. 2015;06. doi:10.3389/fpls.2015.00385

79. O'Keeffe KR, Carbone I, Jones CD, Mitchell CE. Plastic potential: how the phenotypes and adaptations of pathogens are influenced by microbial interactions within plants. *Curr Opin Plant Biol*. 2017;38:78-83. doi:10.1016/j.pbi.2017.04.014

80. Tollenaere C, Susi H, Laine A-L. Evolutionary and Epidemiological Implications of Multiple Infection in Plants. *Trends Plant Sci*. 2016;21(1):80-90. doi:10.1016/j.tplants.2015.10.014

81. Bender CL, Alarcón-Chaidez F, Gross DC. Pseudomonas syringae Phytotoxins: Mode of Action, Regulation, and Biosynthesis by Peptide and Polyketide Synthetases. *Microbiol Mol Biol Rev*. 1999;63(2):266-292.

82. Bender CL, Scholz-Schroeder BK. New Insights Into the Biosynthesis, Mode of Action, and Regulation of Syringomycin, syringopeptin, and Coronatine. In: Ramos J-L, ed. *Pseudomonas*. Boston, MA: Springer US; 2004:125-158. doi:10.1007/978-1-4419-9084-6_4

83. Schellenberg B, Ramel C, Dudler R. Pseudomonas syringae Virulence Factor Syringolin A Counteracts Stomatal Immunity by Proteasome Inhibition. *Mol Plant Microbe Interact*. 2010;23(10):1287-1293. doi:10.1094/MPMI-04-10-0094

84. Burch AY, Zeisler V, Yokota K, Schreiber L, Lindow SE. The hygroscopic biosurfactant syringafactin produced by *P seudomonas syringae* enhances fitness on leaf surfaces during fluctuating humidity: Hygroscopic biosurfactants in the phyllosphere. *Environ Microbiol*. 2014;16(7):2086-2098. doi:10.1111/1462-2920.12437

85. Geissler M, Heravi KM, Henkel M, Hausmann R. Chapter 6 - Lipopeptide Biosurfactants From Bacillus Species. In: Hayes DG, Solaiman DKY, Ashby RD, eds. *Biobased Surfactants (Second Edition)*. AOCS Press; 2019:205-240. doi:10.1016/B978-0-12-812705-6.00006-X

86. Zouari O, Lecouturier D, Rochex A, et al. Bio-emulsifying and biodegradation activities of syringafactin producing Pseudomonas spp. strains isolated from oil contaminated soils. *Biodegradation*. November 2018. doi:10.1007/s10532-018-9861-x

87. Jacques M-A, Arlat M, Boulanger A, et al. Using Ecology, Physiology, and Genomics to Understand Host Specificity in Xanthomonas. *Annu Rev Phytopathol*. 2016;54(1):163-187. doi:10.1146/annurev-phyto-080615-100147

88. Kousik CS (North CSU, Ritchie DF. Race shift in Xanthomonas campestris pv. vesicatoria within a season in field-grown pepper. *Phytopathol USA*. 1996. http://agris.fao.org/agris-search/search.do?recordID=US9744228. Accessed July 20, 2019.

89. Merda D, Bonneau S, Guimbaud J-F, et al. Recombination-prone bacterial strains form a reservoir from which epidemic clones emerge in agroecosystems: Recombinant strains as a reservoir for epidemics. *Environ Microbiol Rep*. 2016;8(5):572-581. doi:10.1111/1758-2229.12397

90. Neale HC, Laister R, Payne J, Preston G, Jackson RW, Arnold DL. A low frequency persistent reservoir of a genomic island in a pathogen population ensures island survival and improves pathogen fitness in a susceptible host: Reservoir of a genomic island in a pathogen. *Environ Microbiol*. 2016;18(11):4144-4152. doi:10.1111/1462-2920.13482

91. O'Brien HE, Thakur S, Gong Y, et al. Extensive remodeling of the Pseudomonas syringae pv. avellanae type III secretome associated with two independent host shifts onto hazelnut. *BMC Microbiol*. 2012;12(1):141. doi:10.1186/1471-2180-12-141

92. Craig L, Pique ME, Tainer JA. Type IV pilus structure and bacterial pathogenicity. *Nat Rev Microbiol*. 2004;2(5):363. doi:10.1038/nrmicro885

93. Shahin RD, Amsbaugh DF, Leef MF. Mucosal immunization with filamentous hemagglutinin protects against Bordetella pertussis respiratory infection. *Infect Immun*. 1992;60(4):1482-1488.

94. Büttner D, Bonas U. Regulation and secretion of Xanthomonas virulence factors. *FEMS Microbiol Rev*. 2010;34(2):107-133. doi:10.1111/j.1574-6976.2009.00192.x

95. Darsonval A, Darrasse A, Durand K, Bureau C, Cesbron S, Jacques M-A. Adhesion and Fitness in the Bean Phyllosphere and Transmission to Seed of *Xanthomonas fuscans* subsp. *fuscans*. *Mol Plant Microbe Interact*. 2009;22(6):747-757. doi:10.1094/MPMI-22-6-0747

96. Meng Y, Li Y, Galvani CD, et al. Upstream Migration of Xylella fastidiosa via Pilus-Driven Twitching Motility. *J Bacteriol*. 2005;187(16):5560-5567. doi:10.1128/JB.187.16.5560-5567.2005

97. Ojanen-Reuhs T, Kalkkinen N, Westerlund-Wikström B, et al. Characterization of the fimA gene encoding bundle-forming fimbriae of the plant pathogen Xanthomonas campestris pv. vesicatoria. *J Bacteriol*. 1997;179(4):1280-1290. doi:10.1128/jb.179.4.1280-1290.1997

98. Gottig N, Garavaglia BS, Garofalo CG, Orellano EG, Ottado J. A Filamentous Hemagglutinin-Like Protein of Xanthomonas axonopodis pv. citri, the Phytopathogen Responsible for Citrus Canker, Is Involved in Bacterial Virulence. *PLOS ONE*. 2009;4(2):e4358. doi:10.1371/journal.pone.0004358

99. Rojas CM, Ham JH, Deng W-L, Doyle JJ, Collmer A. HecA, a member of a class of adhesins produced by diverse pathogenic bacteria, contributes to the attachment, aggregation, epidermal cell killing, and virulence phenotypes of Erwinia chrysanthemi EC16 on Nicotiana clevelandii seedlings. *Proc Natl Acad Sci*. 2002;99(20):13142-13147. doi:10.1073/pnas.202358699

100. Taghavi S, Wu X, Ouyang L, et al. Transcriptional Responses to Sucrose Mimic the Plant-Associated Life Style of the Plant Growth Promoting Endophyte Enterobacter sp. 638. *PLOS ONE*. 2015;10(1):e0115455. doi:10.1371/journal.pone.0115455

101. da Silva ACR, Ferro JA, Reinach FC, et al. Comparison of the genomes of two Xanthomonas pathogens with differing host specificities. *Nature*. 2002;417(6887):459-463. doi:10.1038/417459a

102. Green S, Studholme DJ, Laue BE, et al. Comparative Genome Analysis Provides Insights into the Evolution and Adaptation of Pseudomonas syringae pv. aesculi on Aesculus hippocastanum. *PLOS ONE*. 2010;5(4):e10224. doi:10.1371/journal.pone.0010224

103. Qian W, Jia Y, Ren S-X, et al. Comparative and functional genomic analyses of the pathogenicity of phytopathogen Xanthomonas campestris pv. campestris. *Genome Res*. 2005;15(6):757-767. doi:10.1101/gr.3378705

104. Chiang P, Sampaleanu LM, Ayers M, Pahuta M, Howell PL, Burrows LL. Functional role of conserved residues in the characteristic secretion NTPase motifs of the Pseudomonas aeruginosa type IV pilus motor proteins PilB, PilT and PilU. *Microbiology*. 2008;154(1):114-126. doi:10.1099/mic.0.2007/011320-0

105. Takhar HK, Kemp K, Kim M, Howell PL, Burrows LL. The Platform Protein Is Essential for Type IV Pilus Biogenesis. *J Biol Chem*. 2013;288(14):9721-9728. doi:10.1074/jbc.M113.453506

106. Strom MS, Nunn DN, Lory S. A single bifunctional enzyme, PilD, catalyzes cleavage and N-methylation of proteins belonging to the type IV pilin family. *Proc Natl Acad Sci*. 1993;90(6):2404-2408. doi:10.1073/pnas.90.6.2404

Chapter 4: Discussion and Future Directions

4.1 Abstract

Disease outbreaks are caused by diverse triggers ranging from pathogen evolution to changes in weather systems to human behavior. Because of this, it is a great scientific challenge to determine the exact causes of a disease outbreak and how the host or pathogen wins the evolutionary arms race. In this thesis I identified several aspects of the host, pathogen, and environment sides of the disease triangle that contributed to the outbreaks of *Xanthomonas citri* pv. *malvacearum* (*Xcm*) and *Pseudomonas syringae* on cotton (*Gossypium hirsutum*). While these data contribute significant information towards the field of pathology, they also revealed many more remaining questions. Here I highlight several key findings from this thesis and the scientific questions they reveal: Phylogenetic analysis of contemporary and historical isolates revealed *Xcm* and *P. syringae* as re-emergent pathogens on cotton, but the contributions of mixed-infections remains unknown. Genome sequencing and virulence factor identification contributed to the basic understanding of these pathogens and their contributions to the host-pathogen arms race; however, the mechanisms of putative virulence factors remains to be deciphered. Finally, host susceptibility genes and sources of resistance were identified that can lead to a new generation of resistant cotton; however, the mechanisms of suspected susceptibility genes and the genetic basis for *G. hirsutum* resistance to *Xcm* remains to be identified. Over all, each of the findings outlined in this thesis as well as their resulting questions contributes to the understanding of *Gossypium hirsutum* pathosystems and the greater field of pathology as a whole.

4.2 Introduction

Cotton (*Gossypium hirsutum*) is the most widely used natural fiber in the world. This industry is threatened by many diseases every year. I focused my thesis on two such pathogens: *Xanthomonas citri* pv. *malvacearum* (*Xcm*) and *Pseudomonas syringae.* In 2011, the historical cotton pathogen *Xcm* re-emerged in the US despite the presence of several sources of genetic resistance [1]. Investigations of this disease revealed that a second pathogen, *P. syringae,* was infecting the same US fields. To our knowledge, this pathogen has only been identified twice before, and only in Lubbock, Texas. When I began my thesis, very little was known about the genetic diversity, mechanisms of virulence, or evolution of either of these pathogens because they were not investigated as threats during the era of modern sequencing techniques. In this thesis I use phylogenetics to confirm the status of *Xcm* and *P. syringae* as re-emergent diseases on cotton, comparative genomics to identify putative virulence factors and the evolutionary forces acting on them, and RNA-Seq and virulence assays to identify new resistance strategies all in the context of the never ending host-pathogen arms race. Here I highlight the contributions of these findings to the field as well as the gaps in knowledge that they expose.

4.3 Disease emergence and pathogen-pathogen interactions

Disease outbreaks and the underlying host-pathogen arms race are the result of the combination of pathogen genetics, host genetics, and the environment. Previous studies on plant pathogen outbreaks have identified the movement of propagation material, seeds, insect vectors, and wind patterns as sources of outbreaks [2–6]. Others have found that genetic changes in pathogen virulence factors allowed them to cause disease on previously resistant hosts [7–9]. In this thesis I investigated whether the *Xcm* outbreak on *G. hirsutum* was caused by genetic changes in *Xcm* that allowed it to overcome resistance. I also investigated whether *P. syringae* underwent

any genetic changes that would explain its re-emergence on cotton. In summary, the genetic and phylogenetic analysis described in this thesis allowed me to conclude that both outbreaks were not likely to have been caused by a host shift or race shift. Instead, we report analysis of cotton variety planting statistics that indicate this outbreak was likely exacerbated by a steady increase in the percentage of susceptible cotton varieties grown each year since 2009.

It is still a mystery as to why the *P. syringae* cotton pathogen re-emerged and why its emergence is so rare. It is possible that the necrotic lesions caused by this pathogen are being incorrectly identified as a different pathogen. Alternatively, genetic resistance to this disease may have been unknowingly deployed in fields until recently, similar to how *Xcm* re-emerged [1]. It is also possible that this pathogen is not particularly virulent in cotton fields and therefore only occurs during rare weather patterns. Likely many of these factors, in conjunction with many more factors yet to be known, work together to cause this pathogen to re-emerge once every few decades.

The hypothesis for *P. syringae* emergence that is currently being investigated is whether or not it is a coincidence that *P. syringae* re-emerged during an *Xcm* outbreak. To our knowledge, every time *P. syringae* was isolated from a cotton plant this decade, *Xcm* was isolated along with it. This may be caused by a sampling bias due to extension scientists visiting *Xcm* infected fields more frequently than healthy fields. Alternatively, these two pathogens could have a commensal or even mutualistic relationship. Many pathogens are opportunistic and require for the host to be already weakened before disease can occur [10–12]. A commensal relationship would suggest that *Xcm* infections facilitate *P. syringae* disease progression. It is also possible that *Xcm* benefits from the presence of *P.* syrinage in the form of a mutualistic disease complex. Disease complexes occur when two pathogens cause a greater amount of

disease together than they would apart [13–15]. Disease complexes are difficult to identify due to the many environmental variables present within an agricultural field that may influence the distribution of pathogens and the expression of disease symptoms. However, if *Xcm* and *P. syringae* do cause a disease complex, it would become an ideal model system for understanding how disease complexes occur. This is because both pathogens are culturable and easily genetically manipulated and we can study their infections in real time in the field.

The presence of a disease complex could be confirmed if bacterial concentrations and/or disease symptoms are found to be stronger when the two pathogens are inoculated together than when they are inoculated separately. Inoculations with a needleless syringe are typical for quantifying bacterial colony forming units within the leaf due to the highly controlled levels of inoculum that this method introduces into the plant. However, spray or dip inoculations better mimic the mode of disease progression within a field. Therefore both methods of inoculation should be used to determine whether the disease complex may be due to 1. One or both pathogens enabling greater bacterial entry into the host or 2. One or both pathogens enabling higher growth and/or disease symptoms once they have entered the host. Either of these methods of mutualistic behavior could be caused due to direct pathogen-pathogen interactions such as quorum signaling or through modulation of host defenses through hormone signaling.

Alternatively, a commensal interaction could be occurring if the interaction only benefits *P. syringae*. This interaction would explain why *P. syringae* is so rarely found in the fields and *Xcm* is a common pathogen. *P. syringae* may require high disease pressure from *Xcm* such as from the recent *Xcm* disease outbreak in order to successfully infect a host. This hypothesis would also be confirmed using the above methods and would result in higher disease

symptoms/bacterial concentrations of *P. syringae* in co-inoculations than when it is inoculated alone.

A third hypothesis is that these pathogens do not enhance each other's disease causing abilities. In fact, if the pathogens are not working together to cause disease in any way, it is likely that they are either directly or indirectly antagonistic. This is because the host environment contains a limited amount of resources that can support a finite number of bacterial cells. Co-inoculation experiments and more will be used to determine the nature of the interaction of these two bacterial pathogens of cotton.

4.4 Susceptibility genes

Susceptibility genes are host genes that are upregulated by pathogens during infection and promote disease. The presence of *G. hirusutum* susceptibility genes was investigated in this thesis in order to better understand the basic interaction between *Xcm* and *G. hirsutum* as well as to identify potential targets for generating new resistant varieties. Time course RNA-Seq data of *Xcm* inoculated and non-inoculated host tissue enabled us to identify several SWEET sugar transporter and mildew locus-O (MLO) genes as putative susceptibility genes.

Four candidate susceptibility genes that we identified are SWEET sugar transporters. In other systems, these genes are susceptibility genes that are upregulated by *Xanthomonas* Transcription Activator-Like (TAL) effectors [16,17]. This occurs through the unique ability of TAL effectors to bind to Effector Binding Elements (EBEs) in susceptibility gene promoters in a sequence-specific manner [18,19]. When SWEET genes are knocked down, or when the interaction between the TAL effector and SWEET promoter is disrupted, disease symptoms are diminished. The identification of several TAL effectors in each *Xcm* strain used in this study supports the hypothesis that these SWEET genes are susceptibility genes that are used by *Xcm* to promote

virulence. In fact, Cox et al. showed that at least one SWEET gene in *Gossypium hirsutum* can be disrupted and confer quantitative resistance to a race of *Xcm* [20]. Interestingly, as I show in this thesis, one set of homeologous SWEET genes is upregulated by one *Xcm* strain and the other set of SWEET genes is upregulated by the other *Xcm* strain that has a different repertoire of TAL effectors. Therefore the TAL effector-SWEET promoter binding specificity is likely to be highly variable from isolate to isolate. In the future this system could be probed to understand the contributions of the many SWEET homologs and homeologs to susceptibility in the *G. hirsutum*-*Xcm* pathosystem. Therefore, while the *Xcm*-*G. hirsutum* pathosystem was identified as a good model system for understanding TAL-SWEET gene interactions, it is not a great candidate system for the development of resistant varieties through disrupted EBE sites.

A more likely candidate gene for the development of a *Xcm* resistant *G. hirsutum* variety is Mildew Locus-O (MLO). Unlike the SWEET genes, this homeologous pair of genes was induced equally by two phylogenetically distinct *Xcm* strains. MLO has been deployed in many systems to confer resistance to its namesake powdery mildew, but some evidence suggests that this gene family may confer resistance to Xanthomonads [21]. The next step in determining whether this gene is a susceptibility gene is to disrupt the function of the gene and determine if the subsequent plants are resistant to *Xcm.* Preliminary experiments indicate that knocking down this gene using virus induced gene silencing (VIGS) may be difficult. Therefore, future efforts will focus on CRISPR knock-outs of this gene. Cotton has historically been a difficult gene to genetically manipulate due in part to low transformation rates and the tetraploid nature of its genome. However, recent breakthroughs in both transformation efficiency and CRISPR technology may make this knock-out possible [22–26].

These data reveal insights into the *Xcm-G. hirsutum* disease complex and strategies for future development of resistant cultivars. Susceptibility genes are particularly good candidates for developing resistant germplasm because they generally confer non-race specific resistance. In contrast to resistance genes, these genes often target a conserved host mechanism that the pathogen needs to proliferate and successfully cause disease [27,28]. However, due to the functional redundancy of SWEET genes in the *Gossypium hirsutum* genome, this broad-spectrum resistance is lost. In contrast, the MLO-based resistance may be more widely applicable than *Xcm* due to its uses against multiple pathogens in other systems [21,28]. However, one caveat to the use of many susceptibility gene mutants is that many of them cause pleiotropic effects. Whether or not pleiotropic effects occur is often species specific and the underlying causes are a current topic of research in the field. Therefore, future work on the contributions of these susceptibility genes to *Xcm* disease as well as their pleiotropic effects are still needed.

4.5 Resistance genes

In contrast to susceptibility gene mutants, resistance genes have been used as *Xcm* resistant germplasm for decades [29,30]. In cotton, many of these genes are incorporated into germplasm through traditional breeding practices [29–32]. However, we still do not know the identities of any of the 20 hypothesized resistance genes. My preliminary work, performed in collaboration with Vasu Kuraparthy, has identified nine sources of qualitative resistance to *Xcm* across a diverse panel of cotton varieties [31]. I identified one variety that has resistance tied to a single dominantly inherited trait, a common characteristic of resistance genes [33]. Other varieties are also likely to have this type of resistance, but were not screened. Many of these resistant varieties have now been crossed with susceptible varieties and self-fertilized until the F6 generation. Future research will focus on screening these populations for resistance and

ultimately map the location of these genes. This research will lead to more targeted breeding practices for developing resistance and may eventually lead to the incorporation of *Xcm* resistance into more diverse and farmer preferred varieties.

Due to the sparse research attention that the *Pseudomonas syringae* cotton pathogen has had, there have been no known previous attempts to identify genetic resistance to this pathogen. Alarmingly, our preliminary studies have not identified a cotton cultivar that is completely resistant to this pathogen. In this thesis I present the first characterization of the genetic and phenotypic diversity of the *P. syringae* cotton pathogen. These data will enable us to screen the cotton diversity panel for resistance to diverse *P. syringae* strains so that durable resistance strategies can be developed.

4.6 Virulence factor mechanisms

Bacterial pathogens use a plethora of virulence factors to promote growth and disease symptoms. These range from highly specialized effectors that bind to and upregulate host susceptibility genes to toxins that lyse any host membrane in their path.

Filamentous hemagglutinin and the PilBCD type IV pilus operon were singled out as potential players in the *P. syringae-G. hirsutum* arms race due to their high pairwise nucleotide diversity. However, it is still unclear what contributions these genes have towards pathogenicity. In other systems, these genes promote pathogenicity by facilitating adhesion to host cells and movement across the plant surface [34–37]. However, these contributions are highly variable from system to system and knock-outs of these genes do not always result in reduced pathogenicity [38–40]. Gene knock-outs in multiple strains of the *P. syringae* cotton pathogen are needed to confirm that they function as virulence factors in this system. In addition, the great variability in sequences from strain to strain leads to the question of whether these genes are able to

complement each other in different strains. In particular, I found high levels of nucleotide diversity in the pilB and pilC genes of the type IV pilus. In *Pseudomonas aerigunosa*, these genes physically interact to promote the extension of the pilus [41,42]. Therefore, it is possible that the separate evolution of these genes within strains has made them unable to function in different contexts. Knock-out and reciprocal complementation assays will illuminate the roles that these genes play in virulence as well as the impact of diversification on their functions.

Phylogroup 2 *P. syringae* pathovars often deploy phytotoxins that promote pathogenicity and trigger necrosis [43,44]. I identified several toxin biosynthetic clusters that may be the source of the rapid necrosis that is characteristic of the *P. syringae* pathogen of cotton. Each toxin, syringomycin, syringopeptin, syringolin, and mangotoxin, cause necrosis unlike coronatine, phaseolotoxin, and tabtoxin that cause chlorosis [45,46]. Interestingly, one isolate, Ps203, does not cause spreading necrosis and yet is still able to multiply to similar levels as strains that cause more dramatic symptoms. LC-MS will be needed to identify which toxins are secreted by each of these strains, if any. An alternative hypothesis for the spreading necrosis is that it is an unsuccessful resistance response. Essenberg et al. reported a "spreading collapse" in some cotton genotypes in response to spray inoculation of *Xcm* [30]. This response resulted in a decrease in growth of pathogenic bacteria compared to susceptible varieties and therefore was deemed a form of "runaway cell death" similar to the disease phenotype of *lsd1 Arabidopsis thaliana* mutants [47]. However, this did not occur until a week after inoculation with concentrations of 1 x 10^6 and did not occur with point inoculations or in field inoculations. Therefore, it is more likely that this phenotype is caused by a pathogen toxin than a host resistance response.

The discovery of each of these virulence factors has contributed to the basic understanding of bacterial cotton pathogens and how they compete in the plant-pathogen arms

race. Additional analyses of codon usage, gene expression levels, plasmid copy number, and metabolomics may shed further light on how finely tuned these virulence systems are their hosts. Genetic manipulation experiments will confirm the contributions that these factors have on virulence.

4.7 Conclusions of the thesis

Xcm and *P. syringae* are re-emergent pathogens on cotton. The data exhibited in this thesis provide insights into how these pathogens cause disease and how resistant varieties can be developed in the future. This was accomplished through long read sequencing and assembly of bacterial genomes that revealed phylogenetic and evolutionary explanations for disease emergence as well as evidence for putative virulence factors. Additionally, RNA-Sequencing and disease screening identified putative susceptibility genes and the presence of resistance genes that may be used in the future to develop new resistant varieties of cotton. These data provide important information for the understanding of these diseases and lay the groundwork for future genetic manipulation of both the host and pathogen.

4.8 References

1. Phillips AZ, Berry JC, Wilson MC, et al. Genomics-enabled analysis of the emergent disease cotton bacterial blight. *PLOS Genet*. 2017;13(9):23.

2. Giovanardi D, Biondi E, Ignjatov M, Jevtić R, Stefani E. Impact of bacterial spot outbreaks on the phytosanitary quality of tomato and pepper seeds. *Plant Pathol*. 2018;67(5):1168-1176. doi:10.1111/ppa.12839

3. Ialacci GM, Bella P, Licciardello G, et al. Clonal populations of Clavibacter michiganensis subsp. michiganensis are responsible for the outbreaks of bacterial canker in greenhouse tomatoes in Italy. *Plant Pathol*. 2016;65(3):484-495. doi:10.1111/ppa.12424

4. Jacobson AL, Duffy S, Sseruwagi P. Whitefly-transmitted viruses threatening cassava production in Africa. *Curr Opin Virol*. 2018;33:167-176. doi:10.1016/j.coviro.2018.08.016

5. Mazzaglia A, Studholme DJ, Taratufolo MC, et al. Pseudomonas syringae pv. actinidiae (PSA) Isolates from Recent Bacterial Canker of Kiwifruit Outbreaks Belong to the Same Genetic Lineage. *PLOS ONE*. 2012;7(5):e36518. doi:10.1371/journal.pone.0036518

6. Schiro G, Verch G, Grimm V, Müller MEH. Alternaria and Fusarium Fungi: Differences in Distribution and Spore Deposition in a Topographically Heterogeneous Wheat Field. *J Fungi*. 2018;4(2):63. doi:10.3390/jof4020063

7. Hulin MT, Armitage AD, Vicente JG, et al. Comparative genomics of Pseudomonas syringae reveals convergent gene gain and loss associated with specialization onto cherry (Prunus avium). *New Phytol*. 2018;219(2):672-696. doi:10.1111/nph.15182

8. Kousik CS (North CSU, Ritchie DF. Race shift in Xanthomonas campestris pv. vesicatoria within a season in field-grown pepper. *Phytopathol USA*. 1996. http://agris.fao.org/agris-search/search.do?recordID=US9744228. Accessed July 20, 2019.

9. O'Brien HE, Thakur S, Gong Y, et al. Extensive remodeling of the Pseudomonas syringae pv. avellanae type III secretome associated with two independent host shifts onto hazelnut. *BMC Microbiol*. 2012;12(1):141. doi:10.1186/1471-2180-12-141

10. Hunter P. Co-infection: when whole can be greater than the sum. *EMBO Rep*. 2018;19(8):e46601. doi:10.15252/embr.201846601

11. Kim S, Covington A, Pamer EG. The intestinal microbiota: Antibiotics, colonization resistance, and enteric pathogens. *Immunol Rev*. 2017;279(1):90-105. doi:10.1111/imr.12563

12. Suleyman G, Alangaden G, Bardossy AC. The Role of Environmental Contamination in the Transmission of Nosocomial Pathogens and Healthcare-Associated Infections. *Curr Infect Dis Rep*. 2018;20(6):12. doi:10.1007/s11908-018-0620-2

13. Lamichhane JR, Venturi V. Synergisms between microbial pathogens in plant disease complexes: a growing trend. *Front Plant Sci*. 2015;06. doi:10.3389/fpls.2015.00385

14. O'Keeffe KR, Carbone I, Jones CD, Mitchell CE. Plastic potential: how the phenotypes and adaptations of pathogens are influenced by microbial interactions within plants. *Curr Opin Plant Biol*. 2017;38:78-83. doi:10.1016/j.pbi.2017.04.014

15. Tollenaere C, Susi H, Laine A-L. Evolutionary and Epidemiological Implications of Multiple Infection in Plants. *Trends Plant Sci*. 2016;21(1):80-90. doi:10.1016/j.tplants.2015.10.014

16. Cohn M, Bart RS, Shybut M, et al. Xanthomonas axonopodis Virulence Is Promoted by a Transcription Activator-Like Effector–Mediated Induction of a SWEET Sugar Transporter in Cassava. *Mol Plant Microbe Interact*. 2014;27(11):1186-1198. doi:10.1094/MPMI-06-14-0161-R

17. Streubel J, Pesce C, Hutin M, Koebnik R, Boch J, Szurek B. Five phylogenetically close rice SWEET genes confer TAL effector-mediated susceptibility to Xanthomonas oryzae pv. oryzae. *New Phytol*. 2014:808-819. doi:10.1111/nph.12411@10.1002/(ISSN)1469-8137(CAT)VirtualIssues(VI)Phytopathogeneffectorproteins

18. Boch J, Scholze H, Schornack S, et al. Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science*. 2009;326(5959):1509-1512. doi:10.1126/science.1178811

19. Moscou MJ, Bogdanove AJ. A Simple Cipher Governs DNA Recognition by TAL Effectors. *Science*. 2009;326(5959):1501-1501. doi:10.1126/science.1178817

20. Cox KL, Meng F, Wilkins KE, et al. TAL effector driven induction of a *SWEET* gene confers susceptibility to bacterial blight of cotton. *Nat Commun*. 2017;8:15588. doi:10.1038/ncomms15588

21. Kim DS, Hwang BK. The pepper MLO gene, CaMLO2, is involved in the susceptibility cell-death response and bacterial and oomycete proliferation. *Plant J*. 2012;72(5):843-855. doi:10.1111/tpj.12003

22. Chen X, Lu X, Shu N, et al. Targeted mutagenesis in cotton (*Gossypium hirsutum* L.) using the CRISPR/Cas9 system. *Sci Rep*. 2017;7:44304. doi:10.1038/srep44304

23. Li B, Rui H, Li Y, et al. Robust CRISPR/Cpf1 (Cas12a)-mediated genome editing in allotetraploid cotton (*Gossypium hirsutum*). *Plant Biotechnol J*. June 2019. doi:10.1111/pbi.13147

24. Li C, Unver T, Zhang B. A high-efficiency CRISPR/Cas9 system for targeted mutagenesis in Cotton (*Gossypium hirsutum* L.). *Sci Rep*. 2017;7:43902. doi:10.1038/srep43902

25. Long L, Guo D-D, Gao W, et al. Optimization of CRISPR/Cas9 genome editing in cotton by improved sgRNA expression. *Plant Methods*. 2018;14(1). doi:10.1186/s13007-018-0353-0

26. Nekrasov V, Wang C, Win J, Lanz C, Weigel D, Kamoun S. Rapid generation of a transgene-free powdery mildew resistant tomato by genome deletion. *Sci Rep*. 2017;7(1):482. doi:10.1038/s41598-017-00578-x

27. Eom J-S, Chen L-Q, Sosso D, et al. SWEETs, transporters for intracellular and intercellular sugar translocation. *Curr Opin Plant Biol*. 2015;25:53-62. doi:10.1016/j.pbi.2015.04.005

28. Kusch S, Panstruga R. mlo-Based Resistance: An Apparently Universal "Weapon" to Defeat Powdery Mildew Disease. *Mol Plant Microbe Interact*. 2017;30(3):179-189. doi:10.1094/MPMI-12-16-0255-CR

29. Delannoy E, Lyon BR, Marmey P, et al. Resistance of Cotton Towards Xanthomonas campestris pv. malvacearum. *Annu Rev Phytopathol*. 2005;43(1):63-82. doi:10.1146/annurev.phyto.43.040204.140251

30. Essenberg M, Bayles MB, Pierce ML, Verhalen LM. Pyramiding B Genes in Cotton Achieves Broader But Not Always Higher Resistance to Bacterial Blight. *Phytopathology*. 2014;104(10):1088-1097. doi:10.1094/PHYTO-06-13-0167-R

31. Tyagi P, Gore MA, Bowman DT, Campbell BT, Udall JA, Kuraparthy V. Genetic diversity and population structure in the US Upland cotton (Gossypium hirsutum L.). *Theor Appl Genet*. 2014;127(2):283-295. doi:10.1007/s00122-013-2217-3

32. Xiao J, Fang DD, Bhatti M, Hendrix B, Cantrell R. A SNP haplotype associated with a gene resistant to Xanthomonas axonopodis pv. malvacearum in upland cotton (Gossypium hirsutum L.). *Mol Breed*. 2010;25(4):593-602. doi:10.1007/s11032-009-9355-y

33. Fuchs M. Pyramiding resistance-conferring gene sequences in crops. *Curr Opin Virol*. 2017;26:36-42. doi:10.1016/j.coviro.2017.07.004

34. Craig L, Pique ME, Tainer JA. Type IV pilus structure and bacterial pathogenicity. *Nat Rev Microbiol*. 2004;2(5):363. doi:10.1038/nrmicro885

35. Dunger G, Llontop E, Guzzo CR, Farah CS. The Xanthomonas type IV pilus. *Curr Opin Microbiol*. 2016;30:88-97. doi:10.1016/j.mib.2016.01.007

36. Guilhabert MR, Kirkpatrick BC. Identification of Xylella fastidiosa Antivirulence Genes: Hemagglutinin Adhesins Contribute to X. fastidiosa Biofilm Maturation and Colonization and Attenuate Virulence. *Mol Plant Microbe Interact*. 2005;18(8):856-868. doi:10.1094/MPMI-18-0856

37. Sun Y-Y, Chi H, Sun L. Pseudomonas fluorescens Filamentous Hemagglutinin, an Iron-Regulated Protein, Is an Important Virulence Factor that Modulates Bacterial Pathogenicity. *Front Microbiol*. 2016;7. doi:10.3389/fmicb.2016.01320

38. Ojanen-Reuhs T, Kalkkinen N, Westerlund-Wikström B, et al. Characterization of the fimA gene encoding bundle-forming fimbriae of the plant pathogen Xanthomonas campestris pv. vesicatoria. *J Bacteriol*. 1997;179(4):1280-1290. doi:10.1128/jb.179.4.1280-1290.1997

39. Roine E, Raineri DM, Romantschuk M, Wilson M, Nunn DN. Characterization of Type IV Pilus Genes in *Pseudomonas syringae* pv. *tomato* DC3000. *Mol Plant Microbe Interact*. 1998;11(11):1048-1056. doi:10.1094/MPMI.1998.11.11.1048

40. Taghavi S, Wu X, Ouyang L, et al. Transcriptional Responses to Sucrose Mimic the Plant-Associated Life Style of the Plant Growth Promoting Endophyte Enterobacter sp. 638. *PLOS ONE*. 2015;10(1):e0115455. doi:10.1371/journal.pone.0115455

41. Chiang P, Sampaleanu LM, Ayers M, Pahuta M, Howell PL, Burrows LL. Functional role of conserved residues in the characteristic secretion NTPase motifs of the Pseudomonas aeruginosa type IV pilus motor proteins PilB, PilT and PilU. *Microbiology*. 2008;154(1):114-126. doi:10.1099/mic.0.2007/011320-0

42. Takhar HK, Kemp K, Kim M, Howell PL, Burrows LL. The Platform Protein Is Essential for Type IV Pilus Biogenesis. *J Biol Chem*. 2013;288(14):9721-9728. doi:10.1074/jbc.M113.453506

43. Baltrus DA, Nishimura MT, Romanchuk A, et al. Dynamic Evolution of Pathogenicity Revealed by Sequencing and Comparative Genomics of 19 Pseudomonas syringae Isolates. Guttman DS, ed. *PLoS Pathog*. 2011;7(7):e1002132. doi:10.1371/journal.ppat.1002132

44. Xin X-F, Kvitko B, He SY. Pseudomonas syringae: what it takes to be a pathogen. *Nat Rev Microbiol*. 2018;16(5):316-328. doi:10.1038/nrmicro.2018.17

45. Arrebola E, Cazorla FM, Durán VE, et al. Mangotoxin: a novel antimetabolite toxin produced by Pseudomonas syringae inhibiting ornithine/arginine biosynthesis. *Physiol Mol Plant Pathol*. 2003;63(3):117-127. doi:10.1016/j.pmpp.2003.11.003

46. Bender CL, Alarcón-Chaidez F, Gross DC. Pseudomonas syringae Phytotoxins: Mode of Action, Regulation, and Biosynthesis by Peptide and Polyketide Synthetases. *Microbiol Mol Biol Rev*. 1999;63(2):266-292.

47. Jabs T, Dietrich RA, Dangl JL. Initiation of Runaway Cell Death in an Arabidopsis Mutant by Extracellular Superoxide. *Science*. 1996;273(5283):1853-1856. doi:10.1126/science.273.5283.1853

Appendix I: Supplemental Figures For: Genomics-Enabled Analysis of the Emergent Disease Cotton Bacterial Blight

Table AI.1(S1 Table): US Counties with reported CBB incidence from 2009 to 2016.

St.011	Co.2011	St.2012	Co.2012	St.2014	Co.2014	St.2015	Co.2015	St.2016	Co.2016
Arkansas	Chicot	Arkansas	Miss	Arkansas	Miss	Alabama	Henry	Alabama	Escambia
Arkansas	Craighead	Miss	Bolivar	Georgia	Irwin	Alabama	Houston	Alabama	Tallapoosa
Arkansas	Crittenden	Miss	Calhoun	Louisiana	Catahoula	Alabama	Geneva	Alabama	Houston
Arkansas	Lee	Miss	Coahoma	Louisiana	Natchitoches	Alabama	Barbour	Arkansas	Greene
Arkansas	Miss	Miss	Grenada	Louisiana	Pointe Coupee	Alabama	Coffee	Arkansas	Miss
Arkansas	St. Francis	Miss	Holmes	Louisiana	Red River	Alabama	Dale	Florida	Jackson
Miss	Bolivar	Miss	Humphreys	Louisiana	Tensas	Arkansas	Craighead	Florida	Santa Rosa
Miss	Coahoma	Miss	Issaquena	Missouri	New Madrid	Arkansas	Miss	Georgia	Colquitt
Miss	Calhoun	Miss	Lafayette	Missouri	Stoddard	Florida	Santa Rosa	Georgia	Decatur
Miss	Grenada	Miss	Leflore	Miss	Coahoma	Georgia	Irwin	Georgia	Early
Miss	Leflore	Miss	Lowndes	Miss	Chickasaw	Georgia	Seminole	Georgia	Miller
Miss	Monroe	Miss	Madison	Miss	Issaquena	Georgia	Decatur	Georgia	Seminole
Miss	Quitman	Miss	Monroe	Miss	Humphreys	Georgia	Grady	Georgia	Grady
Miss	Sunflower	Miss	Montgomery	Miss	Monroe	Georgia	Thomas	Georgia	Thomas
Miss	Tallahatchie	Miss	Noxubee	Miss	Noxubee	Georgia	Brooks	Georgia	Baker
Miss	Washington	Miss	Panola	Miss	Sharkey	Georgia	Colquitt	Georgia	Mitchell
Miss	Yalobusha	Miss	Quitman	Miss	Rankin	Georgia	Mitchell	Georgia	Cook
Missouri	Pemiscot	Miss	Rankin	Miss	Yazoo	Georgia	Miller	Georgia	Berrien
		Miss	Sharkey	Miss	Forrest	Georgia	Early	Georgia	Lanier
		Miss	Sunflower	Miss	Covington	Georgia	Baker	Georgia	Atkinson
		Miss	Tallahatchie	Miss	Lee	Georgia	Cook	Georgia	Ware
		Miss	Washington	Miss	Leake	Georgia	Lowndes	Georgia	Pierce
		Miss	Yalobusha	Miss	Quitman	Georgia	Echols	Georgia	Wayne
		Miss	Yazoo	Miss	Bolivar	Georgia	Lanier	Georgia	Appling
		Miss	Scott	Miss	Tallahatchie	Georgia	Berrien	Georgia	Bacon
		Miss	Desoto	Miss	Tunica	Georgia	Tift	Georgia	Jeff Davis
		Missouri	Pemiscot	Texas	Bailey	Georgia	Worth	Georgia	Coffee
		Tennessee	Lake	Texas	Collingsworth	Georgia	Irwin	Georgia	Irwin
				Texas	Crosby	Georgia	Turner	Georgia	Ben Hill
				Texas	Dawson	Georgia	Crisp	Georgia	Turner
				Texas	Floyd	Georgia	Dooly	Georgia	Tift
				Texas	Gaines	Georgia	Daugherty	Georgia	Clay

				Texas	Hale	Georgia	Calhoun	Georgia	Quitman
				Texas	Hall	Georgia	Clay	Georgia	Randolph
				Texas	Hockley	Georgia	Randolph	Georgia	Terrell
				Texas	Lamb	Georgia	Terrell	Georgia	Lee
				Texas	Lubbock	Georgia	Lee	Georgia	Dougherty
				Texas	Lynn	Louisiana	Cataoula	Georgia	Calhoun
				Texas	Knox	Louisiana	Natchitoches	Georgia	Brooks
				Texas	Terry	Louisiana	Pointe Coupee	Georgia	Worth
				Texas	Wheeler	Louisiana	Red River	Georgia	Wilcox
				Texas	Wilbarger	Louisiana	Tensas	Georgia	Telfair
				Texas	Colorado	Missouri	New Madrid	Georgia	Dodge
				Texas	Matagorda	Missouri	Stoddard	Georgia	Bleckley
				Texas	Wharton	Missouri	Scott	Georgia	Laurens
				Texas	Brazoria	Miss	Coahoma	Georgia	Johnson
				Texas	Jackson	Miss	Issaquena	Georgia	Webster
						Miss	Humphreys	Georgia	Sumter
						Miss	Sharkey	Georgia	Wilcox
						Miss	Rankin	Georgia	Lowndes
						Miss	Yazoo	Georgia	Toombs
						Miss	Tallahatchie	Louisiana	Rapides
						Miss	Tunica	Louisiana	Tensas
						Miss	Leflore	Louisiana	Morehouse
						Miss	Washington	Louisiana	West Carroll
						Miss	Panola	Louisiana	East Carroll
						Miss	Holmes	Louisiana	Richland
						Miss	Sunflower	Louisiana	Madison
						Texas	Bailey	Louisiana	Franklin
						Texas	Collingsworth	Louisiana	Grant
						Texas	Crosby	Louisiana	Concordia
						Texas	Dawson	Louisiana	Avoyelles
						Texas	Floyd	Louisiana	Point Coupee
						Texas	Gaines	Louisiana	Catahoula
						Texas	Hale	Missouri	New Madrid
						Texas	Hall	Miss	Madison
						Texas	Hockley	Miss	Rankin
						Texas	Lamb	Miss	Washington
						Texas	Lubbock	Miss	Yazoo
						Texas	Lynn	Miss	Humphreys
						Texas	Knox	Miss	Issaquena
						Texas	Terry	Miss	Coahoma
						Texas	Wheeler	Miss	Grenada

						Texas	Wilbarger	Miss	Lafayette
						Texas	Colorado	Miss	Montgomery
						Texas	Matagorda	Miss	Yalobusha
						Texas	Wharton	Miss	Leflore
						Texas	Brazoria	Miss	Carroll
						Texas	Jackson	Miss	Oktibbeha
								Miss	Tunica
								Miss	Forrest
								Miss	Perry
								Miss	Covington
								Miss	Desoto
								Miss	Scott
								Miss	Calhoun
								Miss	Clay
								Miss	Noxubee
								Miss	Tallahatchi
								Oklahoma	Jackson
								Oklahoma	Tillman
								Tennessee	Carroll
								Tennessee	Tipton
								Texas	Dawson
								Texas	Lubbock
								Texas	Medina
								Texas	Wharton
								Texas	Burleson
								Texas	Fort Bend
								Texas	Williamson
								Texas	Gaines
								Texas	Terry
								Texas	Yoakum
								Texas	Lynn
								Texas	Haskell
								Texas	Collingsworth
								Texas	Donley
								Texas	Wilbarger
								Texas	Hardeman
								Texas	Hall
								Texas	Baylor
								Texas	Childress
								Texas	Hale
								Texas	Swisher

								Texas	Childress
								Texas	Parmer
								Texas	Borden
								Texas	Martin

Table AI.2(S2 Table): *Xanthomonas* genomes previously deposited on NCBI that are referenced in this paper.

GenBank	Strain	Abbreviation
GCA_000072485	Stenotrophomonas maltophilia K279a	S. maltophilia
GCA_000087965	Xanthomonas albilineans GPE PC73	X. albilineans_1
GCA_000962915	Xanthomonas albilineans strain HVO082	X. albilineans_2
GCA_000962925	Xanthomonas albilineans strain PNG130	X. albilineans_3
GCA_000225915	Xanthomonas axonopodis pv. citrumelo F1 !Xaxon_pvCitrum_F1	X. alfalfa_2
GCA_000488955	Xanthomonas alfalfae subsp. alfalfae CFBP 3836	X. alfalfa_1
GCA_000306055	Xanthomonas arboricola pv. juglandis str. NCPPB 1447	X. arboricola_1
GCA_000355635	Xanthomonas arboricola pv. corylina str. NCCB 100457	X. arboricola_2
GCA_000521365	Xanthomonas arboricola pv. pruni MAFF 301420	X. arboricola_3
GCA_000259445	Xanthomonas axonopodis pv. Glycines str. 12-2	X. axonopodis_1
GCA_000265565	Xanthomonas axonopodis pv. Manihotis (Bart Lab)	X. axonopodis_2
GCA_000007145	Xanthomonas campestris pv. campestris str. ATCC 33913	X. campestris_1
GCA_000012105	Xanthomonas campestris pv. campestris str. 8004	X. campestris_2
GCA_000403575	Xanthomonas campestris pv. campestris str. CN15	X. campestris_3
GCA_000007165	Xanthomonas axonopodis pv. citri str. 306	X. citri_1
GCA_000263335	Xanthomonas citri pv. mangiferaeindicae LMG 941	X. citri_2
GCA_000349225	Xanthomonas citri subsp. citri Aw12879	X. citri_3
GCA_000009165	Xanthomonas campestris pv. Vesicatoria	X. euvesicatoria_1
GCA_000802325	Xanthomonas euvesicatoria strain 66b	X. euvesicatoria_2
GCA_000175135	Xanthomonas fuscans subsp. aurantifolii str. ICPB 11122	X. fuscans_1
GCA_000741885	Xanthomonas fuscans subsp. fuscans strain CFBP4884	X. fuscans_2
GCA_000817715	Xanthomonas fuscans subsp. fuscans strain X621	X. fuscans_3
GCA_000192065	Xanthomonas gardneri ATCC 19865	X. gardneri_1
GCA_000007385	Xanthomonas oryzae pv. oryzae KACC 10331	X. oryzae_1
GCA_000010025	Xanthomonas oryzae pv. oryzae MAFF 311018	X. oryzae_2
GCA_000019585	Xanthomonas oryzae pv. oryzae PXO99A	X. oryzae_3
GCA_000192045	Xanthomonas perforans 91-118	X. perforans_1
GCA_000800665	Xanthomonas perforans strain 4P1S2	X. perforans_2
GCA_000225975	Xanthomonas sacchari NCPPB 4393	X. sacchari_1

GCA_000815185	Xanthomonas sacchari strain R1	X. sacchari_2
GCA_000831625	Xanthomonas sacchari strain LMG 476	X. sacchari_3
GCA_000159795	Xanthomonas vasicola pv. vasculorum NCPPB 702	X. vasicola_1
GCA_000277995	Xanthomonas vasicola pv. vasculorum NCPPB 1326	X. vasicola_2
GCA_000772695	Xanthomonas vasicola strain NCPPB 1241	X. vasicola_3
GCA_000192025	Xanthomonas vesicatoria ATCC 35937	X. vesicatoria_1
GCA_000803145	Xanthomonas vesicatoria strain 53M X	X. vesicatoria_2
GCA_000803155	Xanthomonas vesicatoria strain 15b	X. vesicatoria_3
GCA_000454525	Xanthomonas citri pv. malvacearum X20 !Xcitri_malvX20	Xcm_BF_1
GCA_000454505	Xanthomonas citri pv. malvacearum X18 !Xcitri_malvX18	Xcm_BF_2
GCA_000309925	Xanthomonas axonopodis pv. malvacearum str. GSPB2388 !Xaxon_malv2388	Xcm_SU44
GCA_000309905	Xanthomonas axonopodis pv. malvacearum str. GSPB1386 !Xaxon_malv1386	Xcm_NI86

Table AI.3(S3 Table): Disease phenotypes and percent acreage of commercial *G. hirsutum* varieties planted in the US from 2009-2016.
<Attached>

Table AI.4(S4 Table): RNA-Seq analysis reveals that 52 genes are induced in all *Xcm-G. hirsutum* interactions at 48 hours ($p \leq 0.05$) with a Log2 (fold change in FPKM) ≥ 2).

A Genome	D Genome	Gene Annotation
Gh_A02G0615	Gh_D02G0670	Seven transmembrane MLO family protein
Gh_A03G0560	Gh_D03G0971	Pectate lyase family protein
Gh_A05G2012	Gh_D05G2256	Protein of unknown function DUF688
Gh_A06G0439	Gh_D06G0479	basic chitinase
Gh_A07G1129	Gh_D07G1229	Protein of unknown function (DUF1278)
Gh_A10G0257	Gh_D10G0257	Protein E6
Gh_A10G1075	Gh_D10G1437	Pectin lyase-like superfamily protein
Gh_A13G1467	Gh_D13G1816	pathogenesis-related 4
Gh_A01G0779		Predicted Protein
Gh_A01G1712		terpene synthase 21
Gh_A02G0972		glycosyl hydrolase 9B13
Gh_A03G0875		Protein of unknown function (DUF1666)
Gh_A04G0364		Cysteine proteinases superfamily protein
Gh_A04G0366		Cysteine proteinases superfamily protein
Gh_A05G1967		Predicted Protein
Gh_A07G0470		malate synthase
Gh_A08G1167		downstream target of AGL15 2
Gh_A09G0128		EXS (ERD1/XPR1/SYG1) family protein
Gh_A09G1148		Protein of unknown function, DUF642
Gh_A09G1803		Pectin lyase-like superfamily protein
Gh_A12G2323		PAR1 protein
Gh_A13G0185		expansin A4
Gh_A13G0205		Ypt/Rab-GAP domain of gyp1p superfamily protein
Gh_A13G0281		Subtilase family protein
Gh_A13G1662		Protein of unknown function (DUF1677)
	Gh_D01G1158	hydroxy methylglutaryl CoA reductase 1
	Gh_D02G1352	glutaredoxin-related
	Gh_D02G1437	Plant invertase/pectin methylesterase inhibitor superfamily protein
	Gh_D03G1462	osmotin 34

Gh_D05G2589	laccase 14
Gh_D06G0662	Nucleotide-diphospho-sugar transferases superfamily protein
Gh_D07G1960	Uncharacterized membrane protein
Gh_D07G1997	RAB GTPase homolog A5E
Gh_D08G0336	WUSCHEL related homeobox 13
Gh_D08G2134	Protein of unknown function (DUF1635)
Gh_D09G1130	beta-1,3-glucanase 3
Gh_D10G1861	expansin A8
Gh_D11G0279	chloroplast beta-amylase
Gh_D11G1628	reversibly glycosylated polypeptide 1
Gh_D12G2309	glycosyl hydrolase 9C2
Gh_Sca005130G01	photosystem II reaction center protein B
Gh_Sca005423G01	Leucine-rich receptor-like protein kinase family protein
Gh_Sca006797G01	TBP-ASSOCIATED FACTOR 6B

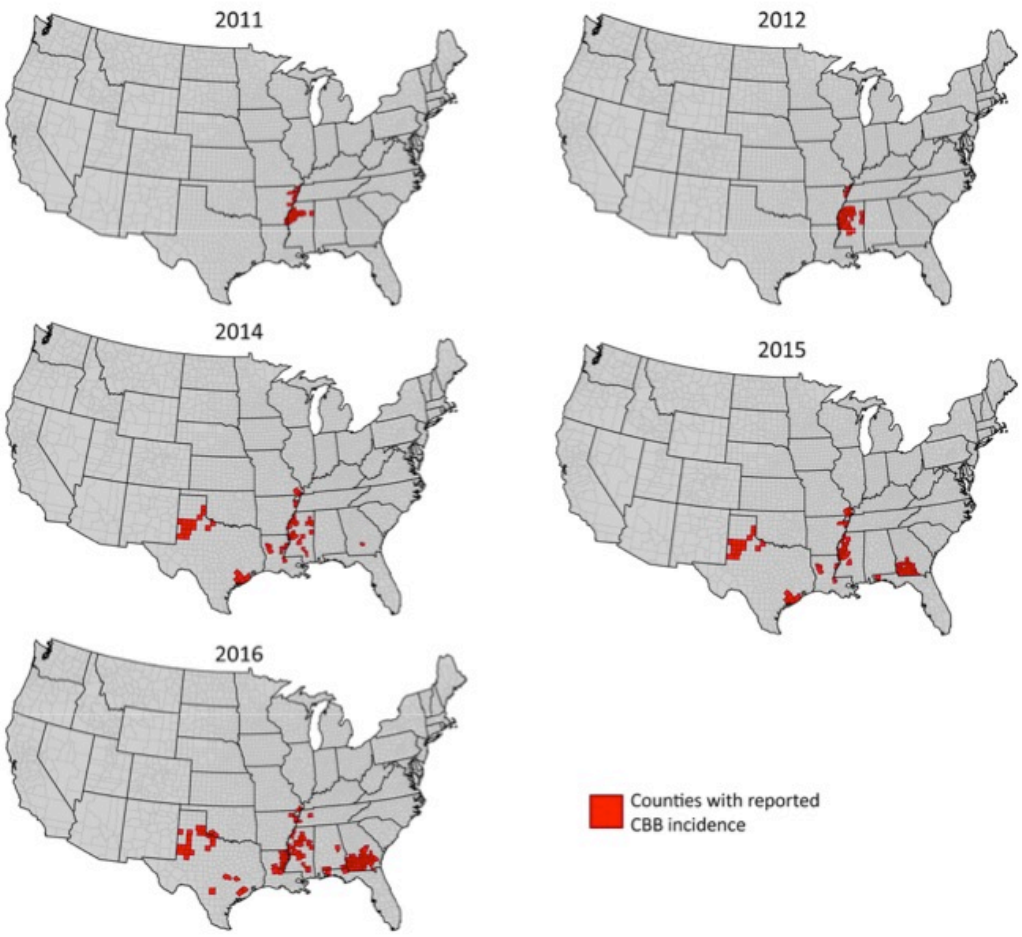
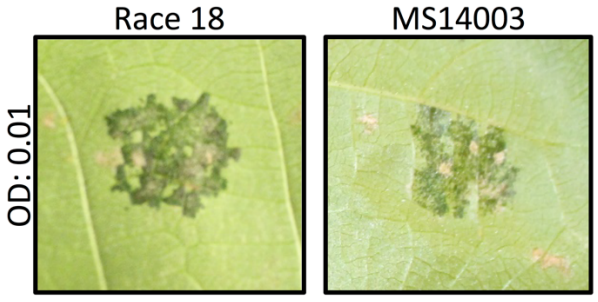


Figure AI.1(S1 Fig): Maps of CBB incidence in the US from 2011-2012 and 2014-2016. CBB incidence was reported by extension agents, extension specialists and certified crop advisers in their respective states for the years 2011-2012 and 2014-2016, and compiled by Tom Allen. CBB reports for 2013 were infrequent.



FigureAI.2(S2 Fig): Disease phenotypes of historical Race18 strain and MS14003 strain. *Xcm* strains Race18 and MS14003 were inoculated into *G. hirsutum* variety PHY499 WRF at an OD600 of 0.01 and imaged at 8 dpi.

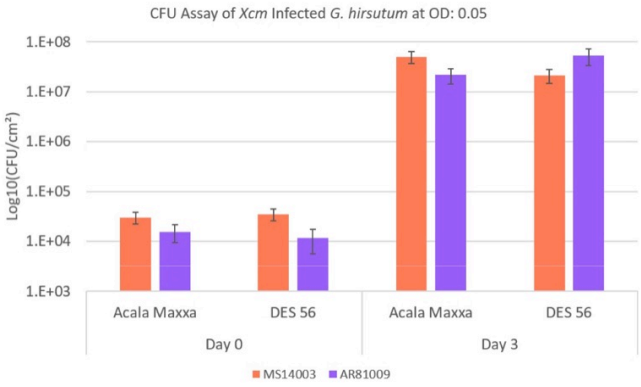


Figure AI.3(S3 Fig): Growth assay of MS14003 and AR81009 on cotton varieties Acala Maxxa and DES 56. *G. hirsutum* varieties were inoculated with *Xcm* at an OD600: 0.05. Tissue was collected at day 0 and day 3 and processed as described in materials and methods.

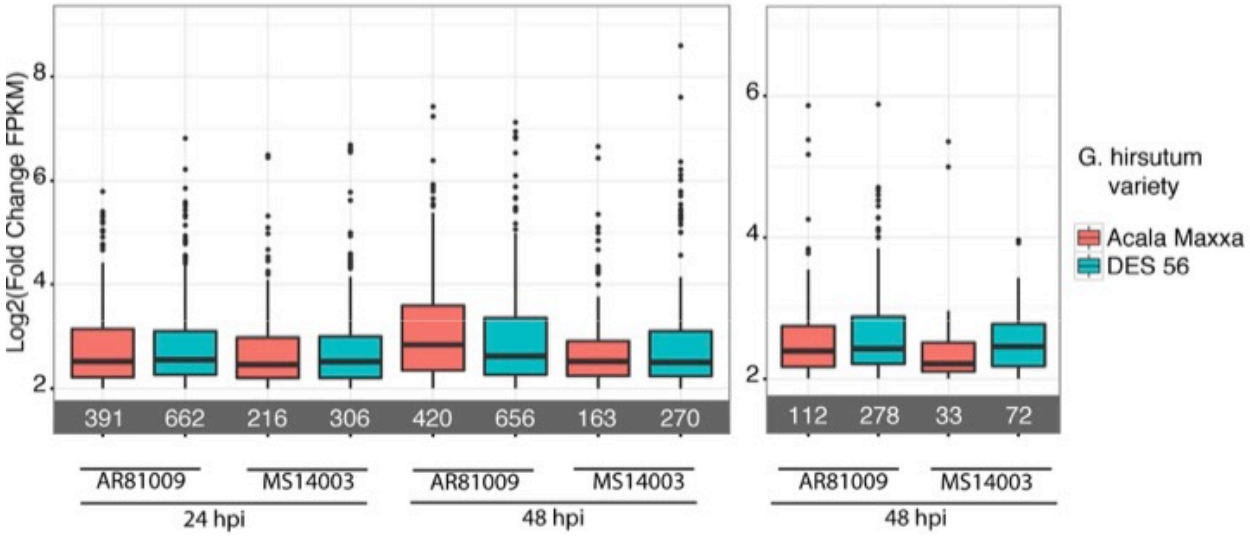


Figure AI.4(S4 Fig): Expression levels of significantly upregulated genes with a Log2 fold change of 2 in *G. hirsutum* A) All significantly upregulated genes with a Log2 fold change of 2 B) All significantly upregulated genes ($p \leq 0.05$) with a Log2 (fold change in FPKM) ≥ 2 that are unique to each cultivar/*Xcm* disease interaction in *G. hirsutum.* Numbers in grey bar indicate the total number of genes for each condition.

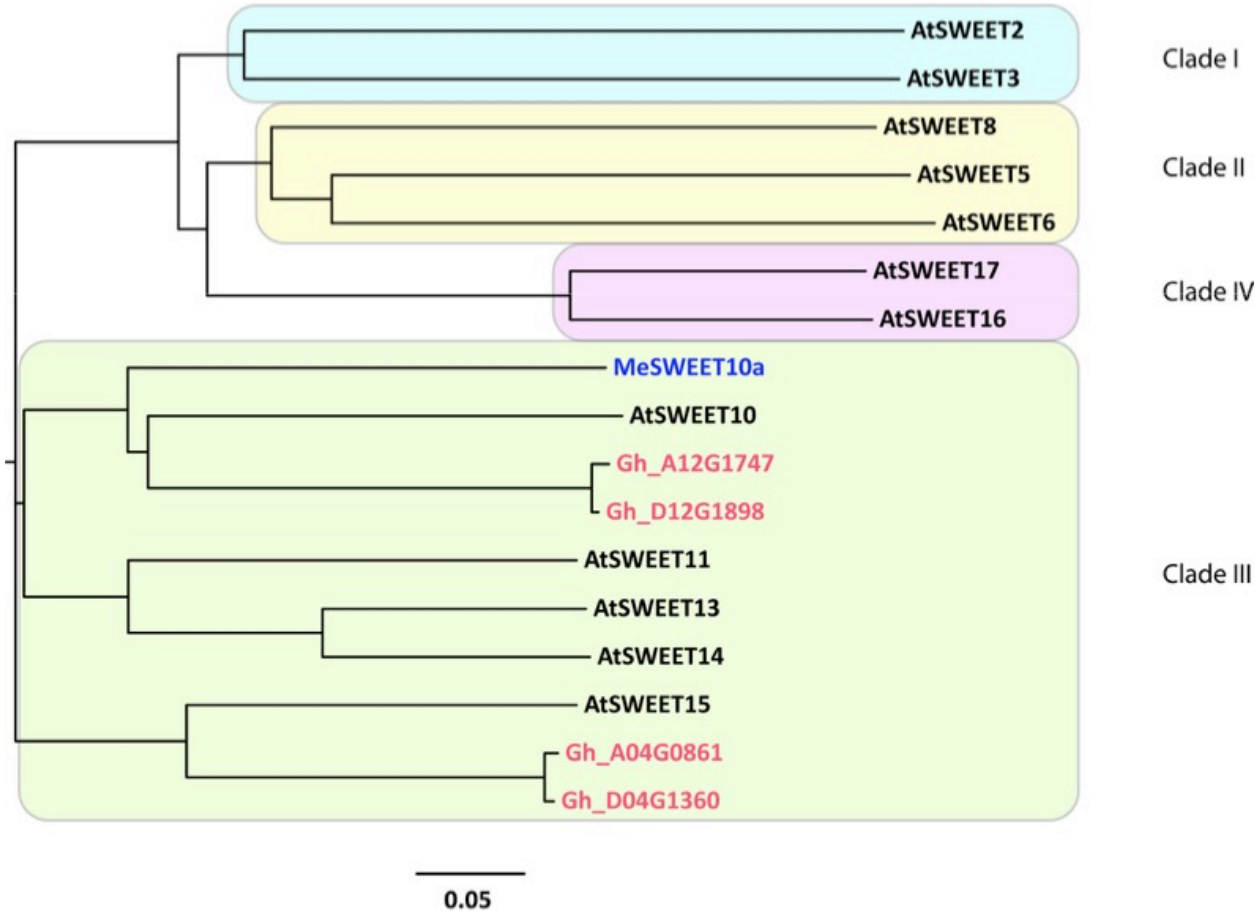


Figure AI.1(S5 Fig): Phylogeny of SWEET genes from *Gossypium hirsutum*, *Manihot esculenta,* and *Arabidopsis thaliana.* Four predicted *G. hirsutum* SWEET genes are compared to classified *A. thaliana* SWEET genes and the MeSWEET10a *M. esculenta* susceptibility gene. A protein alignment and phylogenetic tree was generated by Clustal Omega, and the tree was visualized using Figtree v1.4.2.

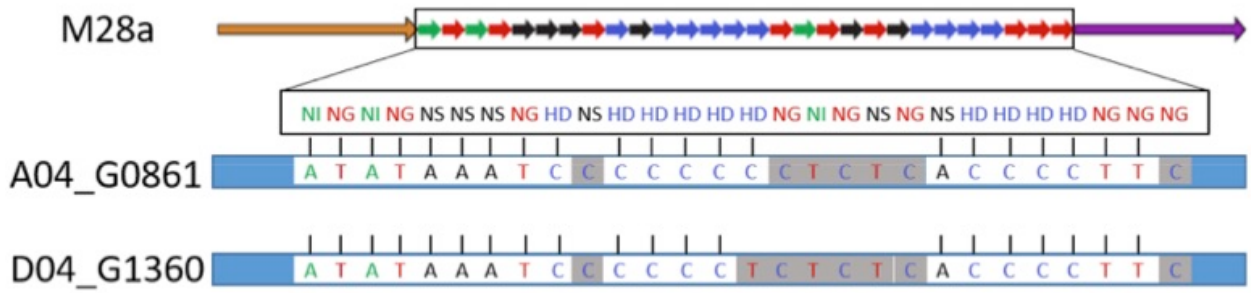


Figure AI.6(S6 Fig): Alignment of predicted TAL effector binding sites on induced *G. hirsutum* SWEET genes. A) TAL M28a is predicted to bind to and up-regulate the homeologous pair of SWEET genes: A04_G0861 and D04_G1360 in *G. hirsutum* varieties Acala Maxxa and DES56 after inoculation with *Xcm* strain MS14003. B) TAL A14a is predicted to bind to and up-regulate the SWEET gene D12_G1898 G1360 in *G. hirsutum* varieties Acala Maxxa and DES56 after inoculation with *Xcm* strain AR81009.

Appendix II: Supplemental Figures For: Evolutionary Context of the *Pseudomonas syringae* Cotton Pathogen and Mechanisms of Virulence

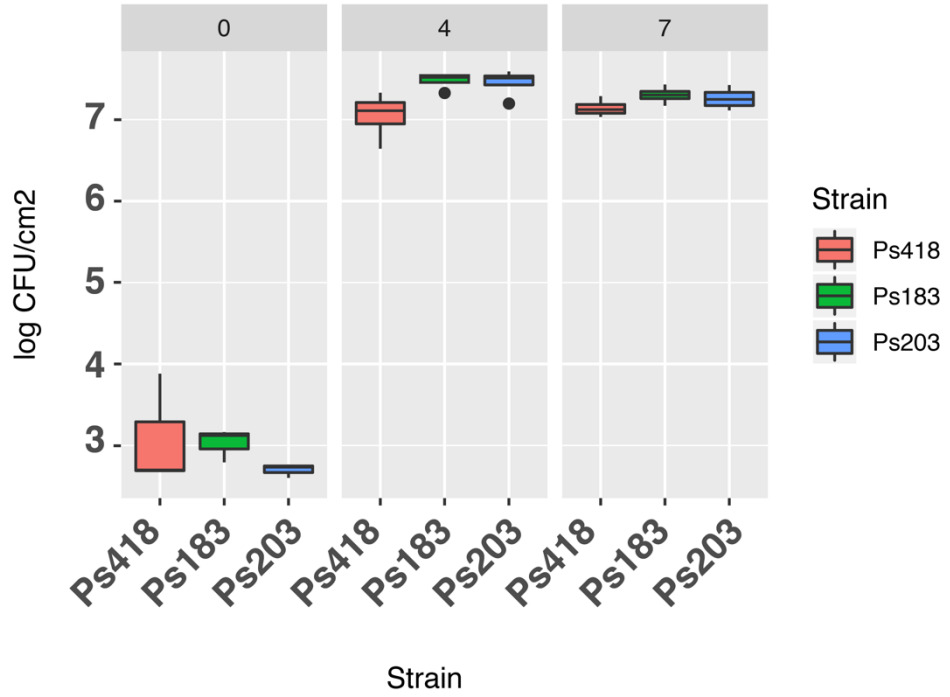


Figure AII.1(Sup. Fig. 1): Growth assay of *P. syringae* strains Ps418, Ps183, and Ps203 on *Gossypium hirsutum.* Bacterial concentrations: $1x10^7$.

Table AII.1(Sup. Table 1): Genome assembly statistics of *P. syringae* cotton pathogens. All genomes were sequenced with nanopore technology, assembled with Canu, polished with Nanopolish, circularized, and further polished with Illumina reads using Pilon, except Ps480, which was sequenced using PacBio technology, assembled with Falcon, polished with Quiver, and circularized.

Strain	# Raw Reads	Mean Read Length (bp)	Contig Type	Nanopore Reads	Nanopore Coverage	Contig Length (bp)	Illumina Coverage
Ps183	228,000	14,590	Chromosome	56,234	149.16	6,087,715	135
Ps203	214,343	10,402	Chromosome	55,554	95.14	5,994,796	125
Ps234	586,120	8,526	Chromosome	135,685	216.87	6,048,721	122
-	-	-	Plasmid	4,514	339.61	68,178	175
Ps236	244,313	8,474	Chromosome	20,607	48	5,936,430	141
-	-	-	Plasmid	164	46.37	67,923	132
Ps238	439,474	17,569	Chromosome	23,599	72.08	5,978,218	171
-	-	-	Plasmid	1,890	251.49	68,216	122
Ps248	318,254	12,583	Chromosome	85,189	198.26	6,080,853	117
Ps418	272,650	16,563	Chromosome	39,757	150.13	5,956,445	120
-	-	-	Plasmid	4,782	1413.06	58,518	242
Ps480	31,554	18,652	Chromosome	-	-	5,965,816	--
-	-	-	Plasmid	-	-	68,280	--

Figure AII.2(Sup. Fig. 2): Mauve Alignment of *P. syringae* plasmid sequences.

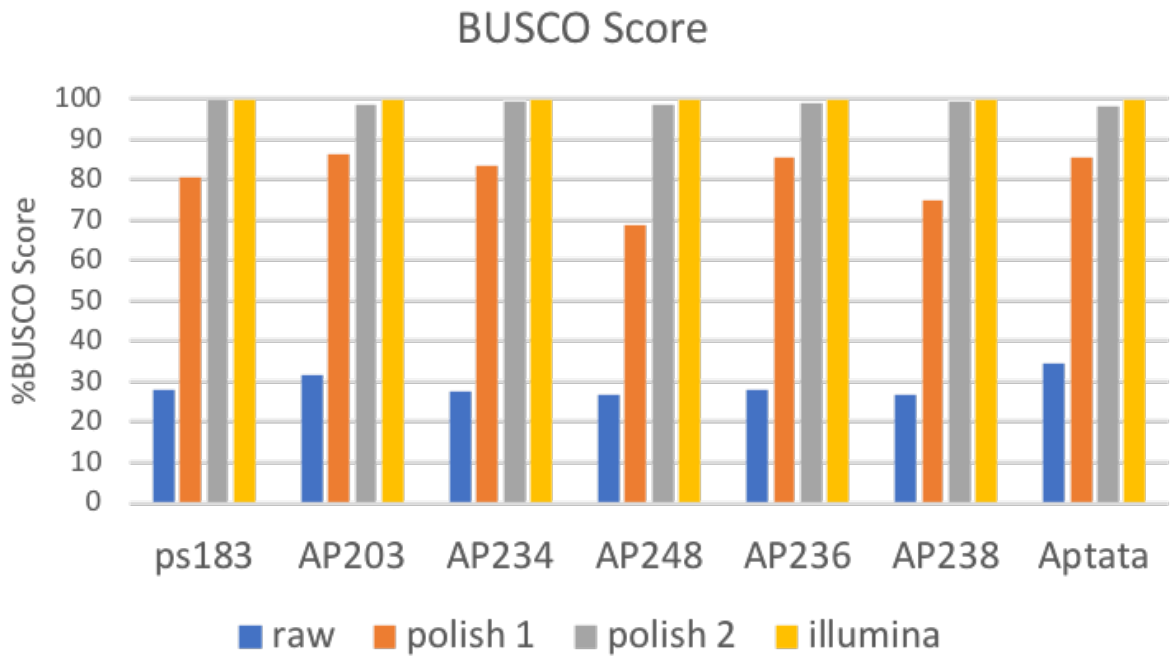


Figure AII.3(Sup. Fig. 3): Quality Assessment of *P. syringae* genomes: Genome Quality was assessed using BUSCO scores (Benchmarking Universal Single-Copy Orthologs). Raw: Initial Canu genome assembly; Polish 1: After polishing with Nanopolish using Nanopore reads; Polish 2: After polishing with Pilon using Illumina reads; Illumina: Illumina only genomes.

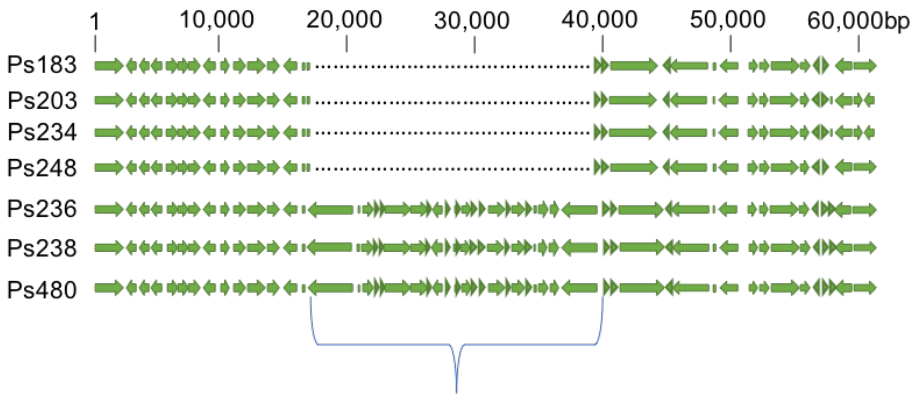


Figure AII.4(Sup. Fig. 4): Nucleotide alignment illustrating the Rhizobial pathogenicity island (R-PAI)T3SS present in 3 out of 7 *P. syringae* cotton pathogens sequenced.

Table AII.2(Table S2): Virulence factor protein sequences used to BLAST P. syringae genomes

Virulence Factor	Gene	Organism	Accession/ Locus tag
Coronatine-coronafacic acid	cfl	Pseudomonas syringae pv. tomato DC3000	PSPTO_4680
Coronatine-coronafacic acid	cfa-1	Pseudomonas syringae pv. tomato DC3000	PSPTO_4681
Coronatine-coronafacic acid	cfa-2	Pseudomonas syringae pv. tomato DC3000	PSPTO_4682
Coronatine-coronafacic acid	cfa-3	Pseudomonas syringae pv. tomato DC3000	PSPTO_4683
Coronatine-coronafacic acid	cfa-4	Pseudomonas syringae pv. tomato DC3000	PSPTO_4684
Coronatine-coronafacic acid	cfa-5	Pseudomonas syringae pv. tomato DC3000	PSPTO_4685
Coronatine-coronafacic acid	cfa-6	Pseudomonas syringae pv. tomato DC3000	PSPTO_4686
Coronatine-coronafacic acid	cfa-7	Pseudomonas syringae pv. tomato DC3000	PSPTO_4687
Coronatine-coronafacic acid	PSPTO_4688	Pseudomonas syringae pv. tomato DC3000	PSPTO_4688
Coronatine-coronafacic acid	cfa-8	Pseudomonas syringae pv. tomato DC3000	PSPTO_4689
Coronatine-coronafacic acid	cfa-9	Pseudomonas syringae pv. tomato DC3000	PSPTO_4690
Coronatine-coronamic acid	cmaA	Pseudomonas syringae pv. tomato DC3000	PSPTO_4709
Coronatine-coronamic acid	cmaB	Pseudomonas syringae pv. tomato DC3000	PSPTO_4710
Coronatine-coronamic acid	cmaC	Pseudomonas syringae pv. tomato DC3000	PSPTO_4711
Coronatine-coronamic acid	cmaD	Pseudomonas syringae pv. tomato DC3000	PSPTO_4707
Coronatine-coronamic acid	cmaE	Pseudomonas syringae pv. tomato DC3000	PSPTO_4708
Coronatine-coronamic acid	cmaT	Pseudomonas syringae pv. tomato DC3000	PSPTO_4712
Coronatine-coronamic acid	cmaU	Pseudomonas syringae pv. tomato DC3000	PSPTO_4714
Coronatine-coronamic acid	cmaX	Pseudomonas syringae pv. tomato DC3000	PSPTO_2295
Coronatine-coronamic acid	alanyl tRNA synthetase	Pseudomonas syringae pv. tomato DC3000	PSPTO_4713
Mangotoxin	mboA	Pseudomonas syringae pv. syringae CFBP3388	JX878402
Mangotoxin	mboB	Pseudomonas syringae pv. syringae CFBP3388	JX878402
Mangotoxin	mboC	Pseudomonas syringae pv. syringae CFBP3388	JX878402
Mangotoxin	mboD	Pseudomonas syringae pv. syringae CFBP3388	JX878402
Mangotoxin	mboE	Pseudomonas syringae pv. syringae CFBP3388	JX878402
Mangotoxin	mboF	Pseudomonas syringae pv. syringae CFBP3388	JX878402
Phaseolotoxin	sir2	Pseudomonas syringae pv. phaseolicola NPS3121	DQ141263
Phaseolotoxin	tp	Pseudomonas syringae pv. phaseolicola NPS3121	DQ141263

Phaseolotoxin	argK	Pseudomonas syringae pv. phaseolicola NPS3121	DQ141263
Phaseolotoxin	ptx2	Pseudomonas syringae pv. phaseolicola NPS3121	DQ141263
Phaseolotoxin	ptx3	Pseudomonas syringae pv. phaseolicola NPS3121	DQ141263
Phaseolotoxin	ptx4	Pseudomonas syringae pv. phaseolicola NPS3121	DQ141263
Phaseolotoxin	ptx5	Pseudomonas syringae pv. phaseolicola NPS3121	DQ141263
Phaseolotoxin	ptx6	Pseudomonas syringae pv. phaseolicola NPS3121	DQ141263
Phaseolotoxin	ptx7	Pseudomonas syringae pv. phaseolicola NPS3121	DQ141263
Phaseolotoxin	ptx8	Pseudomonas syringae pv. phaseolicola NPS3121	DQ141263
Phaseolotoxin	ptx9	Pseudomonas syringae pv. phaseolicola NPS3121	DQ141263
Phaseolotoxin	ptx10-desA	Pseudomonas syringae pv. phaseolicola NPS3121	DQ141263
Phaseolotoxin	ptx11	Pseudomonas syringae pv. phaseolicola NPS3121	DQ141263
Phaseolotoxin	ptx12	Pseudomonas syringae pv. phaseolicola NPS3121	DQ141263
Phaseolotoxin	ptx13	Pseudomonas syringae pv. phaseolicola NPS3121	DQ141263
Phaseolotoxin	ptx14	Pseudomonas syringae pv. phaseolicola NPS3121	DQ141263
Phaseolotoxin	ptx15	Pseudomonas syringae pv. phaseolicola NPS3121	DQ141263
Phaseolotoxin	ptx16	Pseudomonas syringae pv. phaseolicola NPS3121	DQ141263
Phaseolotoxin	amtA	Pseudomonas syringae pv. phaseolicola NPS3121	DQ141263
Phaseolotoxin	ptx18	Pseudomonas syringae pv. phaseolicola NPS3121	DQ141263
Phaseolotoxin	ptx19	Pseudomonas syringae pv. phaseolicola NPS3121	DQ141263
Phaseolotoxin	ptx20	Pseudomonas syringae pv. phaseolicola NPS3121	DQ141263
Phaseolotoxin	ptx21	Pseudomonas syringae pv. phaseolicola NPS3121	DQ141263
Phaseolotoxin	ptx22	Pseudomonas syringae pv. phaseolicola NPS3121	DQ141263
Phaseolotoxin	is-sir2	Pseudomonas syringae pv. phaseolicola NPS3121	DQ141263
Phaseolotoxin	tp	Pseudomonas syringae pv. phaseolicola NPS3121	DQ141263
Syringofactin	syfA	Pseudomonas syringae pv. syringae B64	CM001763
Syringofactin	syfB	Pseudomonas syringae pv. syringae B64	CM001763
Syringofactin	syfR	Pseudomonas syringae pv. syringae B64	CM001763
Syringolin	salA	Pseudomonas syringae pv syringae B301 D-R	AJ548826
Syringolin	sylE	Pseudomonas syringae pv syringae B301 D-R	AJ548826
Syringolin	sylD	Pseudomonas syringae pv syringae B301 D-R	AJ548826

Syringolin	sylC	Pseudomonas syringae pv syringae B301 D-R	AJ548826
Syringolin	sylB	Pseudomonas syringae pv syringae B301 D-R	AJ548826
Syringolin	sylA	Pseudomonas syringae pv syringae B301 D-R	AJ548826
Syringomycin	pseC	Pseudomonas syringae pv. syringae B64	PSSB64_3170
Syringomycin	pseB	Pseudomonas syringae pv. syringae B64	PSSB64_3171
Syringomycin	pseA	Pseudomonas syringae pv. syringae B64	PSSB64_3172
Syringomycin	syrD	Pseudomonas syringae pv. syringae B64	PSSB64_3180
Syringomycin	syrP	Pseudomonas syringae pv. syringae B64	PSSB64_3181
Syringomycin	syrB1	Pseudomonas syringae pv. syringae B64	PSSB64_3182
Syringomycin	syrB2	Pseudomonas syringae pv. syringae B64	PSSB64_3183
Syringomycin	syrC	Pseudomonas syringae pv. syringae B64	PSSB64_3184
Syringomycin	syrE-1	Pseudomonas syringae pv. syringae B64	PSSB64_3185
Syringomycin	syrE-2	Pseudomonas syringae pv. syringae B64	PSSB64_3186
Syringomycin	syrF	Pseudomonas syringae pv. syringae B64	PSSB64_3187
Syringomycin	orf-1	Pseudomonas syringae pv. syringae B64	PSSB64_3190
Syringomycin	orf-2	Pseudomonas syringae pv. syringae B64	PSSB64_3191
Syringomycin	orf-3	Pseudomonas syringae pv. syringae B64	PSSB64_3192
Syringomycin	syrG	Pseudomonas syringae pv. syringae B64	PSSB64_3193
Syringopeptin	sypA	Pseudomonas syringae pv. syringae B64	PSSB64_3179
Syringopeptin	sypB	Pseudomonas syringae pv. syringae B64	PSSB64_3178
Syringopeptin	sypC-1	Pseudomonas syringae pv. syringae B64	PSSB64_3176
Syringopeptin	sypC-2	Pseudomonas syringae pv. syringae B64	PSSB64_3177
Tabtoxin	tabA	Pseudomonas syringae BR2	DQ187985
Tabtoxin	tabB	Pseudomonas syringae BR2	DQ187985
Tabtoxin	tabC	Pseudomonas syringae BR2	DQ187985
Tabtoxin	tabD	Pseudomonas syringae BR2	DQ187985
Tabtoxin	tabP	Pseudomonas syringae BR2	DQ187985
Tabtoxin	tblA	Pseudomonas syringae BR2	DQ187985
Tabtoxin	tblC	Pseudomonas syringae BR2	DQ187985
Tabtoxin	tblD	Pseudomonas syringae BR2	DQ187985

Tabtoxin	tblE	Pseudomonas syringae BR2	DQ187985
Tabtoxin	tblF	Pseudomonas syringae BR2	DQ187985
Tabtoxin	tblR	Pseudomonas syringae BR2	DQ187985
Tabtoxin	tblS	Pseudomonas syringae BR2	DQ187985
Ethylene Production	efe	Pseudomonas syringae pv. cannabina	AF101059
Auxin Biosynthesis	iaaM	Pseudomonas syringae pv. syringae Y30	U04358
Auxin Biosynthesis	iaaH	Pseudomonas syringae pv. syringae Y30	U04358
Auxin Inactivation	iaaL	Pseudomonas syringae pv. tomato DC3000	PSPTO_0371
Ice Nucleation	inaZ	Pseudomonas syringae S203	X03035
Syringolide	avrD	Pseudomonas syringae pv. tomato PT23	MSDS01000059
Type Iva Pilus	ponA	Pseudomonas aeruginosa PA96	PA96_5309
Type Iva Pilus	pilM	Pseudomonas aeruginosa PA96	PA96_5308
Type Iva Pilus	pilN	Pseudomonas aeruginosa PA96	PA96_5307
Type Iva Pilus	pilO	Pseudomonas aeruginosa PA96	PA96_5306
Type Iva Pilus	pilP	Pseudomonas aeruginosa PA96	PA96_5305
Type Iva Pilus	pilQ	Pseudomonas aeruginosa PA96	PA96_5304
Type Iva Pilus	aroK	Pseudomonas aeruginosa PA96	PA96_5303
Type Iva Pilus	aroB	Pseudomonas aeruginosa PA96	PA96_5302
Type Iva Pilus	pilB	Pseudomonas aeruginosa PA96	PA96_4749
Type Iva Pilus	pilC	Pseudomonas aeruginosa PA96	PA96_4750
Type Iva Pilus	pilD	Pseudomonas aeruginosa PA96	PA96_4751
Type Iva Pilus	coaE	Pseudomonas aeruginosa PA96	PA96_4752
Type Iva Pilus	yacG	Pseudomonas aeruginosa PA96	PA96_4753
Type Iva Pilus	pilU	Pseudomonas aeruginosa PA96	PA96_0401
Type Iva Pilus	pilT	Pseudomonas aeruginosa PA96	PA96_0400
Type Iva Pilus	yggS	Pseudomonas aeruginosa PA96	PA96_0399
Type Iva Pilus	ispG	Pseudomonas aeruginosa PA96	PA96_1132
Type Iva Pilus	yfgA	Pseudomonas aeruginosa PA96	PA96_1131
Type Iva Pilus	pilF	Pseudomonas aeruginosa PA96	PA96_1130
Type Iva Pilus	yfgB	Pseudomonas aeruginosa PA96	PA96_1129

Type Iva Pilus	pilE	Pseudomonas aeruginosa PA96	PA96_4780
Type Iva Pilus	pilY2	Pseudomonas aeruginosa PA96	PA96_4779
Type Iva Pilus	pilY1	Pseudomonas aeruginosa PA96	PA96_4778
Type Iva Pilus	pilV	Pseudomonas aeruginosa PA96	PA96_4775
Type Iva Pilus	fimU	Pseudomonas aeruginosa PA96	PA96_4774
Type Iva Pilus	pilA	Pseudomonas aeruginosa PA96	PA96_4748

Table AII.3(Sup. Table 3): Top BLAST hits for iaaM and iaaH genes in *P. syringae* cotton pathogen genomes.

Gene	Top BLAST Hit Name (Based on e-value)	% Query Coverage	E-value	% Pairwise Identity
iaaM	tryptophan 2-monooxygenase	94.08	$1.34x10^{-55}$	30.3
iaaH	GatA aspartyl/glutamyl-tRNA amidotransferase subunit A	98.21	$6.13x10^{-31}$	29.7

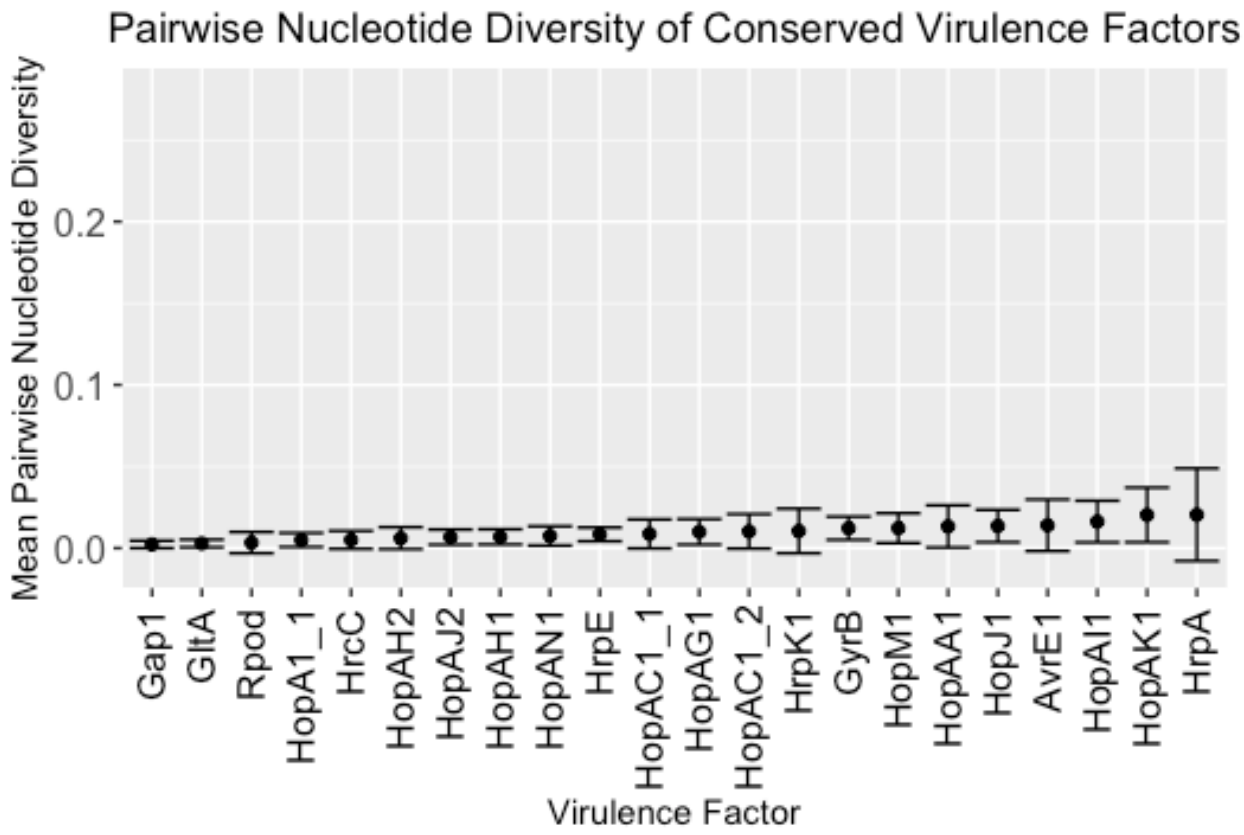


Figure AII.5(Sup. Fig 5): Pairwise nucleotide diversity of *P. syringae* cotton pathogen housekeeping genes (*gap1*, *gltA*, *rpoD*, and *gyrB*) and conserved type III effectors.

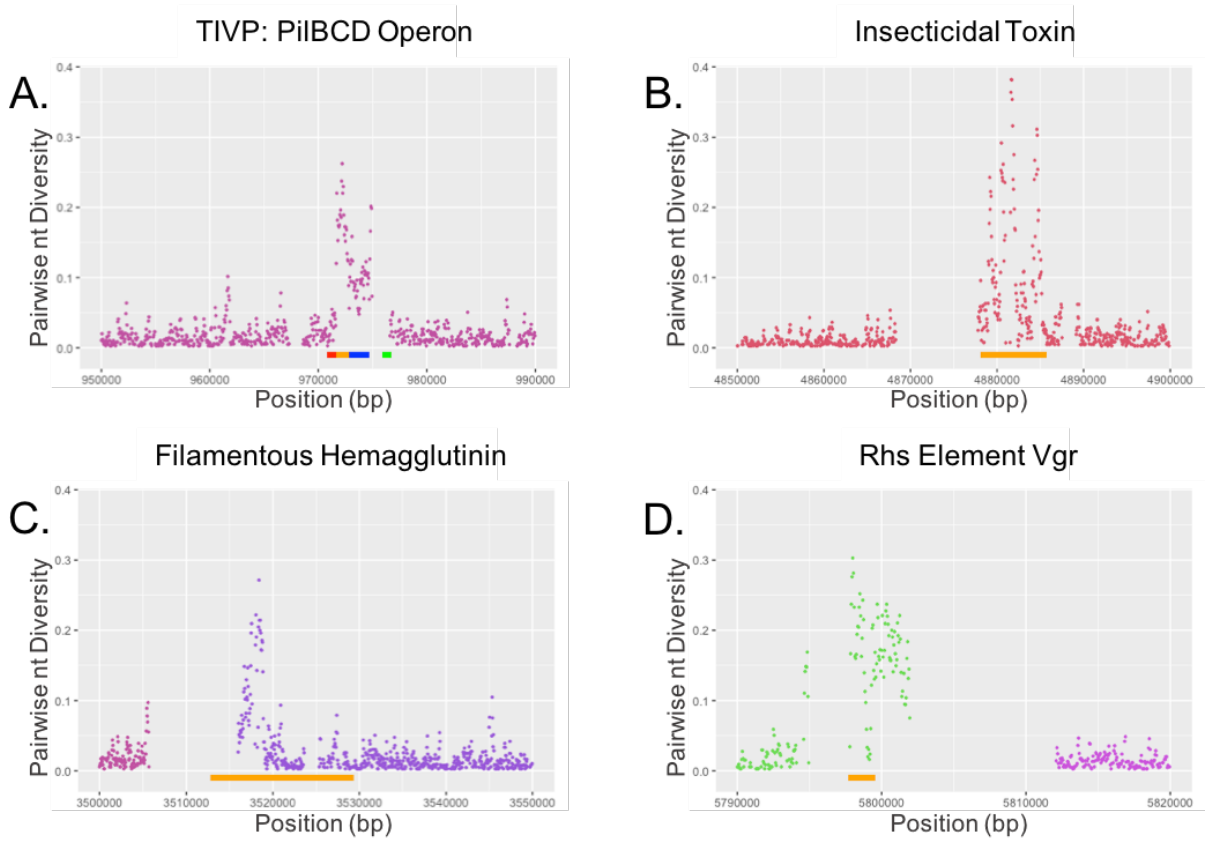


Figure AII.6(Sup. Fig. 6): Regions of high pairwise nucleotide diversity among *P. syringae* cotton pathogens include possible PTI targets and Polymorphic Toxins (From Figure 4b). Lines under X-axis represent the position of each gene in the Ps480 genome. A) Type IV pilus operon with *pilB* (blue), *pliC* (orange), *pilD* (red), and a conserved hypothetical protein (green) B) Insecticidal toxin C) Filamentous hemagglutinin, and D) Rhs Element Vgr.

```
Ps183_pilB      MTDVVLTGLAKQLVLAELLTETAAQQAYQQARREKVSLVSYLVQNKLVRSLTLAEIASDQ 60
Ps248_pilB      MTDAVLTGLAKQLVLAELLTETVAQQAYQHARRDKVSLVSYLVQSKLVKSLTLAEMASEQ 60
Ps203_pilB      MTEAVLTGLAKQLVLAELLTETAAQQAYIQARRDKISLVSYLVQNKLVKSLTVAEMASDQ 60
Ps234_pilB      MTEAVLTGLAKQLVLAELLTETAAQQAYIQARRDKISLVSYLVQNKLVKSLTVAEMASDQ 60
Ps236_pilB      MTDAVLTGLAKQLVLAELLTEPVAQQAYQQARRDKISLVSYLVQHKLVKSLTLAELASDQ 60
Ps238_pilB      MTDAVLTGLAKQLVLAELLTEPVAQQAYQQARRDKISLVSYLVQHKLVKSLTLAELASDQ 60
Ps480_pilB      MTDAVLTGLAKQLVLAELLTEPVAQQAYQQARRDKISLVSYLVQHKLVKSLTLAELASDQ 60
                **:.***************** .***** :***:*:******** ***:***:**:**:*

Ps183_pilB      FGVPFVDLSSLDKESQPKGLVSEKLVRQHNALPLWRRGNKLFIGISDPTNHQAITDIQFS 120
Ps248_pilB      FGVPFMDLASLDRESQPKGLVSEKLVRQHHALPLWRRGNKLFIGISDPTNHQAVTDIQFS 120
Ps203_pilB      FGVPFMDLTSLDRESQPKGLVSEKLVRQHHALPLWRRGNKLFIGISDPTNHQAVTDIQFS 120
Ps234_pilB      FGVPFMDLTSLDRESQPKGLVSEKLVRQHHALPLWRRGNKLFIGISDPTNHQAVTDIQFS 120
Ps236_pilB      FGVPFMDLASLDKESQPKGLVSEKLVRQHHALPLWRRGNKLFIAISDPTNHQAVTDIQFS 120
Ps238_pilB      FGVPFMDLASLDKESQPKGLVSEKLVRQHHALPLWRRGNKLFIAISDPTNHQAVTDIQFS 120
Ps480_pilB      FGVPFMDLASLDKESQPKGLVSEKLVRQHHALPLWRRGNKLFIAISDPTNHQAVTDIQFS 120
                *****:**:***:****************:*************.*********:******

Ps183_pilB      TGLNTEAILVEDDKLTVAIDRFFDTESGLGELADIDLGLEIEPASDKETSLATQSDADDA 180
Ps248_pilB      TGLNTEAILVEDDKLTTAIDRFFDSDSGLGNLEDVDLGLDIEPADGKETSLSTQNDADDA 180
Ps203_pilB      TGLNTEAILVEDDKLTVAIDRFFDSDSGLGNLEDVDLGLDIEPADGKETSIATQNDADDA 180
Ps234_pilB      TGLNTEAILVEDDKLTVAIDRFFDSDSGLGNLEDVDLGLDIEPADGKETSIATQNDADDA 180
Ps236_pilB      TGLNTEAILVEDDKLTIAIDRFFDSDSGLGNLEDVDLGLDIEPADGKETTLATQNDADDA 180
Ps238_pilB      TGLNTEAILVEDDKLTIAIDRFFDSDSGLGNLEDVDLGLDIEPADGKETTLATQNDADDA 180
Ps480_pilB      TGLNTEAILVEDDKLTIAIDRFFDSDSGLGNLEDVDLGLDIEPADGKETTLATQNDADDA 180
                **************** *******::****:* *:****:****..***:::**.*****

Ps183_pilB      PVVRFVNKMLMDAIRLGSSDLHFEPYEKIFRVRLRTDGILHEVAKPPIHLANRIAARLKV 240
Ps248_pilB      PVVRFVNKMLMDAIRLGSSDLHFEPYEKIFRVRLRTDGILHEVARPPIHLANRIAARLKV 240
Ps203_pilB      PVVRFVNKMLMDAIRLGSSDLHFEPYEKIFRVRLRTDGILHEVARPPIHLANRIAARLKV 240
Ps234_pilB      PVVRFVNKMLMDAIRLGSSDLHFEPYEKIFRVRLRTDGILHEVARPPIHLANRIAARLKV 240
Ps236_pilB      PVVRFVNKMLMDAIRLGSSDLHFEPYEKIFRVRLRTDGILHEVARPPIHLANRIAARLKV 240
Ps238_pilB      PVVRFVNKMLMDAIRLGSSDLHFEPYEKIFRVRLRTDGILHEVARPPIHLANRIAARLKV 240
Ps480_pilB      PVVRFVNKMLMDAIRLGSSDLHFEPYEKIFRVRLRTDGILHEVARPPIHLANRIAARLKV 240
                ****************************************:***************

Ps183_pilB      MASLDISERRKPQDGRVKLRVSKNKAIDFRMNTLPTLWGEKIVMRILDPTSAQMGIDALG 300
Ps248_pilB      MASLDISERRKPQDGRVKLRVSKTKAIDFRMNTLPTLWGEKIVMRILDPTSAQMGIDALG 300
Ps203_pilB      MASLDISERRKPQDGRVKLRVSKTKAIDFRMNTLPTLWGEKIVMRILDPTSAQMGIDALG 300
Ps234_pilB      MASLDISERRKPQDGRVKLRVSKTKAIDFRMNTLPTLWGEKIVMRILDPTSAQMGIDALG 300
Ps236_pilB      MASLDISERRKPQDGRVKLRVSKTKAIDFRMNTLPTLWGEKIVMRILDPTSAQMGIDALG 300
Ps238_pilB      MASLDISERRKPQDGRVKLRVSKTKAIDFRMNTLPTLWGEKIVMRILDPTSAQMGIDALG 300
Ps480_pilB      MASLDISERRKPQDGRVKLRVSKTKAIDFRMNTLPTLWGEKIVMRILDPTSAQMGIDALG 300
                **********************.************************************

Ps183_pilB      YEPDQKALYMEALKQPQGMILVTGPTGSGKTVSLYTGLNILNTVDINISTAEDPVEINLE 360
Ps248_pilB      YEPEQKALYLEALKQPQGMILVTGPTGSGKTVSLYTGLNILNTVDINISTAEDPVEINLE 360
Ps203_pilB      YEPEQKALYLEALKQPQGMILVTGPTGSGKTVSLYTGLNILNTVDINISTAEDPVEINLE 360
Ps234_pilB      YEPEQKALYLEALKQPQGMILVTGPTGSGKTVSLYTGLNILNTVDINISTAEDPVEINLE 360
Ps236_pilB      YEPEQKALYLEALKQPQGMILVTGPTGSGKTVSLYTGLNILNTVDINISTAEDPVEINLE 360
Ps238_pilB      YEPEQKALYLEALKQPQGMILVTGPTGSGKTVSLYTGLNILNTVDINISTAEDPVEINLE 360
Ps480_pilB      YEPEQKALYLEALKQPQGMILVTGPTGSGKTVSLYTGLNILNTVDINISTAEDPVEINLE 360
                ***:*****:**************************************************

Ps183_pilB      GINQVNVNPRQGLDFSQALRAFLRQDPDVIMVGEIRDLETAEIAIKASQTGHMVLSTLHT 420
Ps248_pilB      GINQVNVNPRQGLDFSQALRAFLRQDPDVIMVGEIRDLETAEIAIKASQTGHMVLSTLHT 420
Ps203_pilB      GINQVNVNPRQGLDFSQALRAFLRQDPDVIMVGEIRDLETAEIAIKASQTGHMVLSTLHT 420
Ps234_pilB      GINQVNVNPRQGLDFSQALRAFLRQDPDVIMVGEIRDLETAEIAIKASQTGHMVLSTLHT 420
Ps236_pilB      GINQVNVNPRQGLDFSQALRAFLRQDPDVIMVGEIRDLETAEIAIKASQTGHMVLSTLHT 420
Ps238_pilB      GINQVNVNPRQGLDFSQALRAFLRQDPDVIMVGEIRDLETAEIAIKASQTGHMVLSTLHT 420
Ps480_pilB      GINQVNVNPRQGLDFSQALRAFLRQDPDVIMVGEIRDLETAEIAIKASQTGHMVLSTLHT 420
                ************************************************************
```

```
Ps183_pilB      NSAAETLTRLHHMGVAAFNIATAINLIIAQRLARKLCSHCKKELDIPRETLIQEGFPEEK 480
Ps248_pilB      NSAAETLTRLHHMGVAAFNIATAINLIIAQRLARKLCSHCKKELDIPRETLIKEGFPEAE 480
Ps203_pilB      NSAAETLTRLHHMGVAAFNIATAINLIIAQRLARKLCSHCKKELDIPRETLIQEGFPEAK 480
Ps234_pilB      NSAAETLTRLHHMGVAAFNIATAINLIIAQRLARKLCSHCKKELDIPRETLIQEGFPEAK 480
Ps236_pilB      NSAAETLTRLHHMGVAAFNIATAINLIIAQRLARKLCSHCKKELDIPRETLIKEGFPEAK 480
Ps238_pilB      NSAAETLTRLHHMGVAAFNIATAINLIIAQRLARKLCSHCKKELDIPRETLIKEGFPEAK 480
Ps480_pilB      NSAAETLTRLHHMGVAAFNIATAINLIIAQRLARKLCSHCKKELDIPRETLIKEGFPEAK 480
                **************************************************:***** :

Ps183_pilB      VGTFKIYGPVGCEHCNGGYRGRVGIYEVVKKTPELERIIMEEGNSLEISRQMRKDGFNDL 540
Ps248_pilB      IGKFKVYGPVGCEHCNGGYRGRVGIYEVVKKTSELERIIMEEGNSLEISRQMRKDGFSDL 540
Ps203_pilB      IGTFKIYGPVGCEHCNGGYRGRVGIYEVVKKTPELERIIMEEGNSLEISRQMRKDGFNDL 540
Ps234_pilB      IGTFKIYGPVGCEHCNGGYRGRVGIYEVVKKTPELERIIMEEGNSLEISRQMRKDGFNDL 540
Ps236_pilB      IGTFKIYGPMGCEHCNGGYRGRVGIYEVVKKTAELERIIMEEGNSLEISRQMRKDGFNDL 540
Ps238_pilB      IGTFKIYGPMGCEHCNGGYRGRVGIYEVVKKTAELERIIMEEGNSLEISRQMRKDGFNDL 540
Ps480_pilB      IGTFKIYGPMGCEHCNGGYRGRVGIYEVVKKTAELERIIMEEGNSLEISRQMRKDGFNDL 540
                :*.**:***:********************** ***********************.**

Ps183_pilB      RTSGLSKAMQGITSLEEVNRVTKD  564
Ps248_pilB      RTSGLSKAMQGITSLEEVNRVTKD  564
Ps203_pilB      RTSGLSKAMQGITSLEEVNRVTKD  564
Ps234_pilB      RTSGLSKAMQGITSLEEVNRVTKD  564
Ps236_pilB      RTSGLSKAMQGITSLEEVNRVTKD  564
Ps238_pilB      RTSGLSKAMQGITSLEEVNRVTKD  564
Ps480_pilB      RTSGLSKAMQGITSLEEVNRVTKD  564
                ************************
```

Figure AII.7(Sup. Fig 7): Amino acid clustal alignment of *P. syringae* cotton pathogen PilB sequences.

```
Ps236_pilC      MASKAAKVIVYTWEGVDKKGTKTSGELSGHNLALVKAQLRKQGINPTKVRKKSASIFGKG 60
Ps238_pilC      MASKAAKVIVYTWEGVDKKGTKTSGELSGHNLALVKAQLRKQGINPTKVRKKSASIFGKG 60
Ps480_pilC      MASKAAKVIVYTWEGVDKKGTKTSGELSGHNLALVKAQLRKQGINPTKVRKKSASIFGKG 60
Ps203_pilC      MASKAVKVTVYTWEGVDKKGTKTSGELSGHNLALVKAQLRKQGINPTKVRKKSASLFGKG 60
Ps234_pilC      MASKAVKVTVYTWEGVDKKGTKTSGELSGHNLALVKAQLRKQGINPTKVRKKSASLFGKG 60
Ps183_pilC      MASKSVKVSVYTWEGVDKKGGKLSGEVNGHNLALVKAQLRKQGINVTKVRKKPVSIFGKG 60
Ps248_pilC      MASKAIKVIVYTWEGVDKKGAKTSGELSGHNLALVKAQLRKQGINPTKVRKKSVSIFGKG 60
                ****: ** ********** * ***:.***************** ****** .*:****

Ps236_pilC      KKIKPLDIAFFSRQMATMMKAGVPLLQSFDIISEGAENPNMRTLVSSLKQEVSAGNSFAT 120
Ps238_pilC      KKIKPLDIAFFSRQMATMMKAGVPLLQSFDIISEGAENPNMRTLVSSLKQEVSAGNSFAT 120
Ps480_pilC      KKIKPLDIAFFSRQMATMMKAGVPLLQSFDIISEGAENPNMRTLVSSLKQEVSAGNSFAT 120
Ps203_pilC      KKIKPLDIAFFSRQMATMMKAGVPLLQSFDIISEGAENPNMRALVNSLKQEVSAGNSFAT 120
Ps234_pilC      KKIKPLDIAFFSRQMATMMKAGVPLLQSFDIISEGAENPNMRALVNSLKQEVSAGNSFAT 120
Ps183_pilC      KKIKPLDIAFFSRQMATMMKAGVPLLQSFDIIAEGAENPNMRALVDSLKQEVSAGNSFAT 120
Ps248_pilC      KKIKPLDIAFFARQMATMMKAGVPLLQSFDIIGEGAENPNMRALVSSLKQEVSAGNSFAM 120
                ***********:********************.*********:**.*************

Ps236_pilC      ALRQKPEYFDDLFCNLVDAGEQAGALESLLDRVASYKEKTEKLKAKIKKAMTYPIAVLIV 180
Ps238_pilC      ALRQKPEYFDDLFCNLVDAGEQAGALESLLDRVASYKEKTEKLKAKIKKAMTYPIAVLIV 180
Ps480_pilC      ALRQKPEYFDDLFCNLVDAGEQAGALESLLDRVASYKEKTEKLKAKIKKAMTYPIAVLIV 180
Ps203_pilC      ALRQKPEYFDELFCNLVDAGEQAGALESLLDRVASYKEKTEKLKAKIKKAMTYPAAVVVV 180
Ps234_pilC      ALRQKPEYFDELFCNLVDAGEQAGALESLLDRVASYKEKTEKLKAKIKKAMTYPAAVVVV 180
Ps183_pilC      ALRQKPEYFDDLFCNLVDAGEQAGALESLLDRVASYKEKTEKLKAKIKKAMTYPAAVLIV 180
Ps248_pilC      ALRQKPEYFDDLFCNLVDAGEQAGALESLLDRVASYKEKTEKLKAKIKKAMTYPAAVLIV 180
                **********:****************************************** **::*

Ps236_pilC      AIIVSGILLIKVVPQFQSVFAGFGADLPAFTLMVIGLSNIVQEWWLIIVGLLFAGFFLFK 240
Ps238_pilC      AIIVSGILLIKVVPQFQSVFAGFGADLPAFTLMVIGLSNIVQEWWLIIVGLLFAGFFLFK 240
Ps480_pilC      AIIVSGILLIKVVPQFQSVFAGFGADLPAFTLMVIGLSNIVQEWWLIIVGLLFAGFFLFK 240
Ps203_pilC      AIIVSGILLIKVVPQFQVVFAGFGAELPGFTLMVIGLSEIVQKWWLAISLAFFAGAFFLK 240
Ps234_pilC      AIIVSGILLIKVVPQFQVVFAGFGAELPGFTLMVIGLSEIVQKWWLAISLAFFAGAFFLK 240
Ps183_pilC      AVIVSGILLIKVVPQFQSVFTGFGAELPAFTLMVIGLSNIVQEWWLGIVGLFFFGFFVFK 240
Ps248_pilC      AVIVSGILLIKVVPQFQSVFAGFGAELPTFTLMVIGLSEVVQKWWLAIVGLFFASVFIFK 240
                *:*************** **:****:** *********::**:*** *   :* . *.:*

Ps236_pilC      RAYKKSQKFRDGLDRLLLKAPLIGPLIFKSSVARYARTLATTFAAGVPLVEALDSVAGAT 300
Ps238_pilC      RAYKKSQKFRDGLDRLLLKAPLIGPLIFKSSVARYARTLATTFAAGVPLVEALDSVAGAT 300
Ps480_pilC      RAYKKSQKFRDGLDRLLLKAPLIGPLIFKSSVARYARTLATTFAAGVPLVEALDSVAGAT 300
Ps203_pilC      RAYKQSQKFRDSLDRFLLKVPIIGPLIFKSSVARYARTLATTFAAGVPLVEALDSVAGAT 300
Ps234_pilC      RAYKQSQKFRDSLDRFLLKVPIIGPLIFKSSVARYARTLATTFAAGVPLVEALDSVAGAT 300
Ps183_pilC      KSYKQSQKFRDSIDRFLLKVPVIGPLIFKSSVARYARTLATTFAAGVPLVEALDSVAGAT 300
Ps248_pilC      RAYKRSQKFRDSLDRFLLKVPVIGPLIFKSSVARYARTLATTFAAGVPLVEALDSVAGAT 300
                ::**:******.:**:***.*:************************************

Ps236_pilC      GNVVFKNAVIKVKQDVSTGMQLNFSMRSTGVFPSLAIQMTAIGEESGALDTMLDKVATYY 360
Ps238_pilC      GNVVFKNAVIKVKQDVSTGMQLNFSMRSTGVFPSLAIQMTAIGEESGALDTMLDKVATYY 360
Ps480_pilC      GNVVFKNAVIKVKQDVSTGMQLNFSMRSTGVFPSLAIQMTAIGEESGALDTMLDKVATYY 360
Ps203_pilC      GNVVFKNAVNKVKQDVSTGMQLNFSMRSTGVFPSLAIQMTAIGEESGALDNMLDKVATYY 360
Ps234_pilC      GNVVFKNAVNKVKQDVSTGMQLNFSMRSTGVFPSLAIQMTAIGEESGALDNMLDKVATYY 360
Ps183_pilC      GNVVFRNAVNQVKQDVSTGMQLNFSMRSTGVFPSLAIQMTAIGEESGALDNMLDKVATYY 360
Ps248_pilC      GNVVFRNAVNKVKQDVSTGMQLNFSMRSTGVFPSLAIQMTAIGEESGALDNMLDKVATYY 360
                *****:*** :*************************************.*********
```

```
Ps236_pilC      EDEVDNMVDNLTSLMEPMIMAFLGVIVGGLVIAMYLPIFKLGSIV      405
Ps238_pilC      EDEVDNMVDNLTSLMEPMIMAFLGVIVGGLVIAMYLPIFKLGSIV      405
Ps480_pilC      EDEVDNMVDNLTSLMEPMIMAFLGVIVGGLVIAMYLPIFKLGSIV      405
Ps203_pilC      EDEVDNMVDSLTSLMEPVIMGVLGVIVGGLVIAMYLPIFKLGGAV      405
Ps234_pilC      EDEVDNMVDSLTSLMEPVIMGVLGVIVGGLVIAMYLPIFKLGGAV      405
Ps183_pilC      EEEVDNMVDSLTSLMEPMIMAVLGVIVGGLVIAMYLPIFKLGNVV      405
Ps248_pilC      EDEVDNMVDSLTSLMEPMIMAVLGVVVGGLVIAMYLPIFKLGDVV      405
                *:*******.*******:**..***:****************. *
```

Figure AII.8(Sup. Fig 8): Amino acid clustal alignment of *P. syringae* cotton pathogen PilC sequences.

```
Ps248_pilD      MPLLDLLASSPLAFVTTCCILGLIIGSFLNVVVYRLPIMMERDWKAQSRELLGLPAEPDQ 60
Ps183_pilD      MPLLDLLASSPLAFVTTCCILGLIIGSFLNVVVYRLPIMMERDWKAQSRELLGLPAEPDQ 60
Ps203_pilD      MPLLDLLASSPLAFVTTCCILGLILGSFLNVVVYRLPIMMERDWKAQSRELLGLPAEPDQ 60
Ps234_pilD      MPLLDLLASSPLAFVTTCCILGLILGSFLNVVVYRLPIMMERDWKAQSRELLGLPAEPDQ 60
Ps236_pilD      MPLLDLLASSPLAFVTTCCILGLIIGSFLNVVVYRLPIMMERDWKAQSREMLGLPAEPDQ 60
Ps238_pilD      MPLLDLLASSPLAFVTTCCILGLIIGSFLNVVVYRLPIMMERDWKAQSREMLGLPAEPDQ 60
Ps480_pilD      MPLLDLLASSPLAFVTTCCILGLIIGSFLNVVVYRLPIMMERDWKAQSREMLGLPAEPDQ 60
                ************************:*************************:*********

Ps248_pilD      PVFNLNRPRSSCPHCAHKIRPLENLPVISYLLLRGKCSQCKAPISKRYPLVELTCAVLSA 120
Ps183_pilD      PVLNLNRPRSSCPHCAHKIRPWENLPVISYLLLRGKCSQCKAPISKRYPLVELTCAVLSA 120
Ps203_pilD      PVFNLNRPRSSCPHCAHKIRPWENLPVISYLLLRGKCSQCKAPISKRYPLVELTCAVLST 120
Ps234_pilD      PVFNLNRPRSSCPHCAHKIRPWENLPVISYLLLRGKCSQCKAPISKRYPLVELTCAVLST 120
Ps236_pilD      PVFNLNRPRSSCPHCAHKIRPWENLPVISYLLLRGKCSQCKAPISKRYPLVELTCAVLSA 120
Ps238_pilD      PVFNLNRPRSSCPHCAHKIRPWENLPVISYLLLRGKCSQCKAPISKRYPLVELTCAVLSA 120
Ps480_pilD      PVFNLNRPRSSCPHCAHKIRPWENLPVISYLLLRGKCSQCKAPISKRYPLVELTCAVLSA 120
                **:****************** ************************************:

Ps248_pilD      YVAWHFGFGWQAAAMLVLGWGLLAMSLIDADHQLLPDSLVLPLLWLGLIVNAFGLFTSLN 180
Ps183_pilD      YVAWHFGFGWQAAAMLVLVWGLLAMSLIDADHQLLPDSLVLPLLWLGLIVNAFGLFTSLN 180
Ps203_pilD      YVAWHFGFGWQAAAMLVLSWGLLAMSLIDADHQLLPDSLVLPLLWLGLIVNAFGLFTSLN 180
Ps234_pilD      YVAWHFGFGWQAAAMLVLSWGLLAMSLIDADHQLLPDSLVLPLLWLGLIVNAFGLFTSLN 180
Ps236_pilD      YVAWHFGFGWQAAAMLVLSWGLLAMSLIDADHQLLPDSLVLPLLWLGLIVNAFGLFTSLN 180
Ps238_pilD      YVAWHFGFGWQAAAMLVLSWGLLAMSLIDADHQLLPDSLVLPLLWLGLIVNAFGLFTSLN 180
Ps480_pilD      YVAWHFGFGWQAAAMLVLSWGLLAMSLIDADHQLLPDSLVLPLLWLGLIVNAFGLFTSLN 180
                ****************** ****************************************

Ps248_pilD      DALWGAVAGYLALWSVFWLFKLVTGKEGMGYGDFKLLAMLGAWGGWQILPLTILLSSLVG 240
Ps183_pilD      DALWGAVAGYLALWSVFWLFKLVTGKEGMGYGDFKLLAMLGAWGGWQILPLTILLSSLVG 240
Ps203_pilD      DALWGAVAGYLALWSVFWLFKLVTGKEGMGYGDFKLLAMLGAWGGWQILPLTILLSSLVG 240
Ps234_pilD      DALWGAVAGYLALWSVFWLFKLVTGKEGMGYGDFKLLAMLGAWGGWQILPLTILLSSLVG 240
Ps236_pilD      DALWGAVAGYLALWSVFWLFKLVTGKEGMGYGDFKLLAMLGAWGGWQILPLTILLSSLVG 240
Ps238_pilD      DALWGAVAGYLALWSVFWLFKLVTGKEGMGYGDFKLLAMLGAWGGWQILPLTILLSSLVG 240
Ps480_pilD      DALWGAVAGYLALWSVFWLFKLVTGKEGMGYGDFKLLAMLGAWGGWQILPLTILLSSLVG 240
                ************************************************************

Ps248_pilD      AVLGVIMMRVRRVESGTPIPFGPYLAIAGWIALLWGGQITDSYMQFAGFR  290
Ps183_pilD      AVLGVIMMRVRRVESGTPIPFGPYLAIAGWIALLWGGQITDSYMQFAGFR  290
Ps203_pilD      AVLGVIMMRVRRVESGTPIPFGPYLAIAGWIALLWGGQITDSYMQFAGFR  290
Ps234_pilD      AVLGVIMMRVRRVESGTPIPFGPYLAIAGWIALLWGGQITDSYMQFAGFR  290
Ps236_pilD      AVLGVIMMRVRRVESGTPIPFGPYLAIAGWIALLWGGQITDSYMQFAGFR  290
Ps238_pilD      AVLGVIMMRVRRVESGTPIPFGPYLAIAGWIALLWGGQITDSYMQFAGFR  290
Ps480_pilD      AVLGVIMMRVRRVESGTPIPFGPYLAIAGWIALLWGGQITDSYMQFAGFR  290
                ************************************************
```

Figure AII.9(Sup. Fig 9): Amino acid clustal alignment of *P. syringae* cotton pathogen PilD sequences.

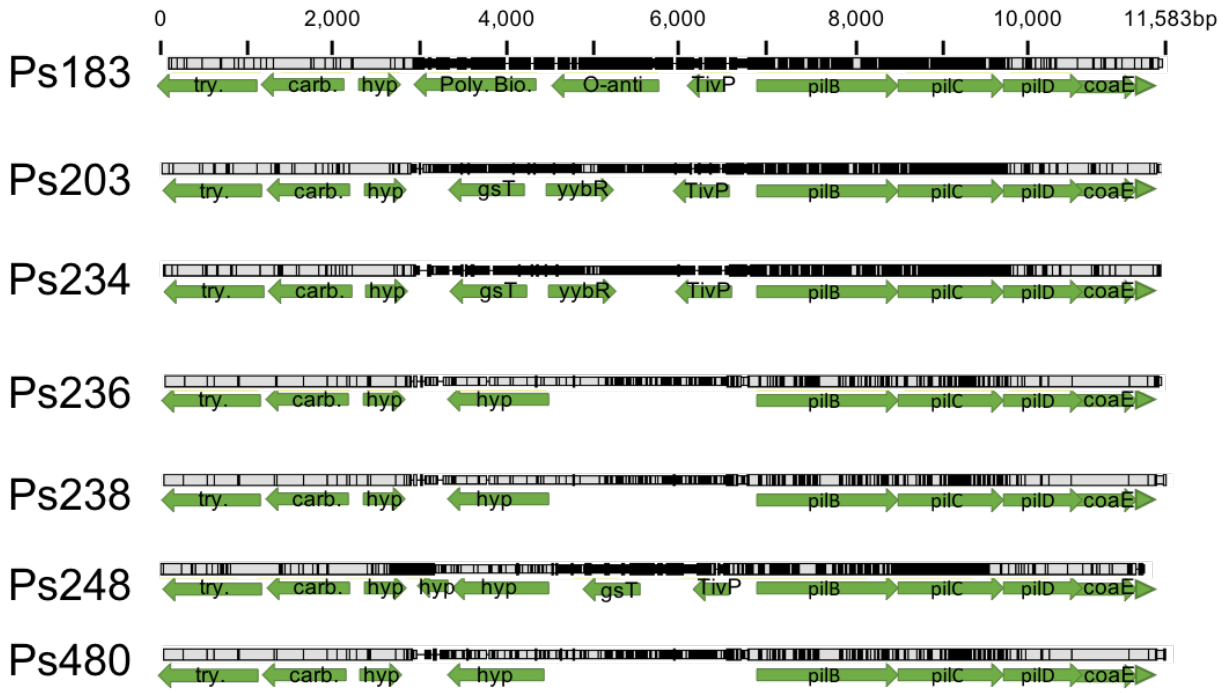


Figure AII.10(Sup. Fig. 10): Nucleotide alignment of the PilBCD operon and 5' region.

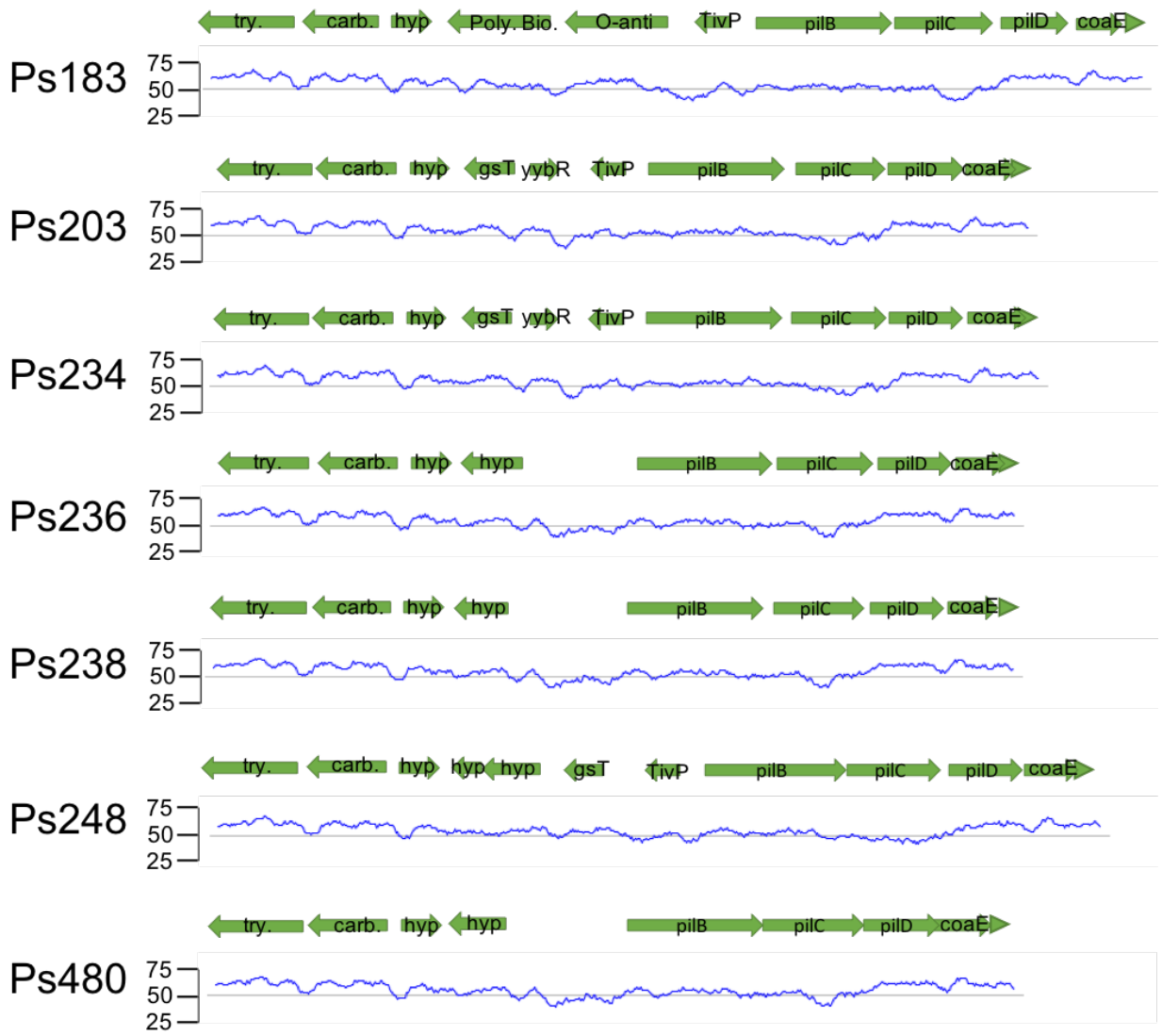


Figure AII.11(Sup. Fig. 11): GC content of PilBCD operon and 5' region based on 200bp windows.

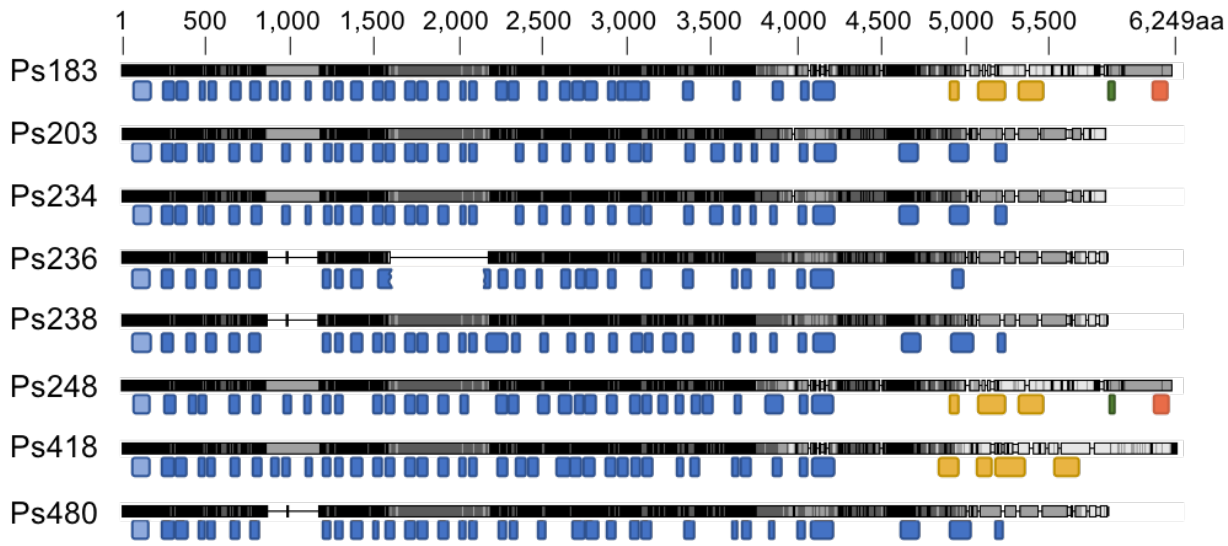


Figure AII.12(Sup. Fig. 12): Amino acid alignment and predicted domains of Filamentous Hemagglutinin: Hemagglutinin Activity Domains (light blue), Hemagglutinin Repeat Domains (blue and orange), Pre-toxin domain with VENN motif (green), and MafB19-like deaminase (red).

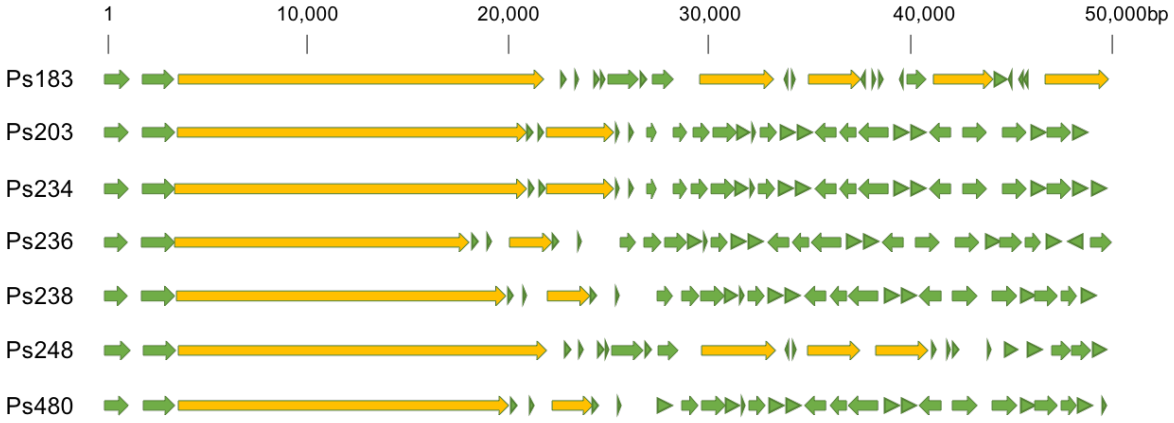


Figure AII.13(Sup. Fig. 13): Genomic context of Filamentous Hemagglutinin genes. Orange highlighted genes are annotated as being in the filamentous hemagglutinin family. The positions of other unrelated genes are shown in green.

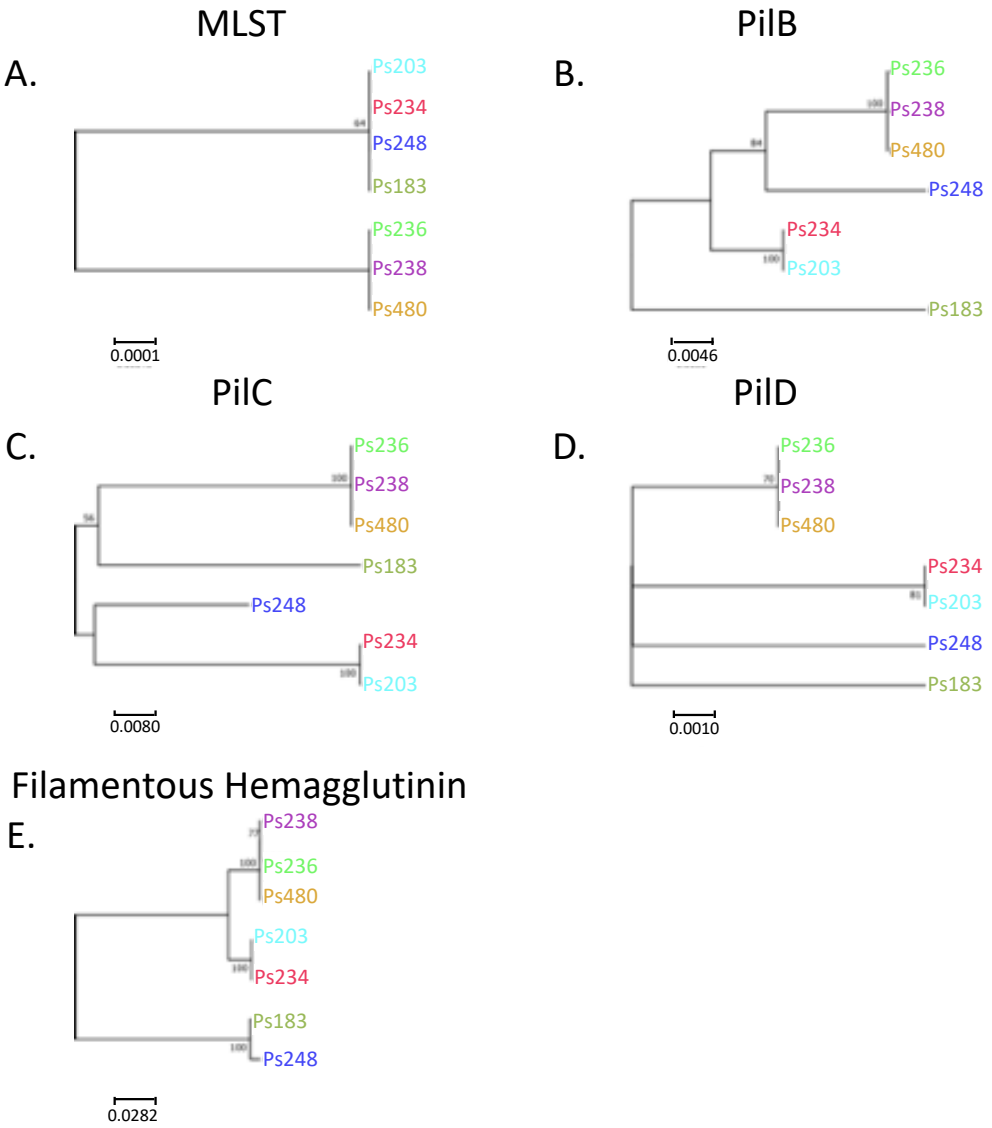


Figure AII-14 (Sup. Fig. 14): MLST and gene phylogenies of Type IV pilus genes and *filamentous hemagglutinin* within *P. syringae* cotton pathogen genomes. Maximum Likelihood phylogenetic tree based on amino acid sequences of A) Concatenated regions of house keeping genes: *gyrB*, *rpoD*, *gap1*, and *gltA*. B) *pilB,* C) *pilC,* D) *pilD,* and E) Filamentous hemagglutinin within *P. syringae* cotton pathogens. Branch labels are bootstrap percentage values based on 100 bootstraps.

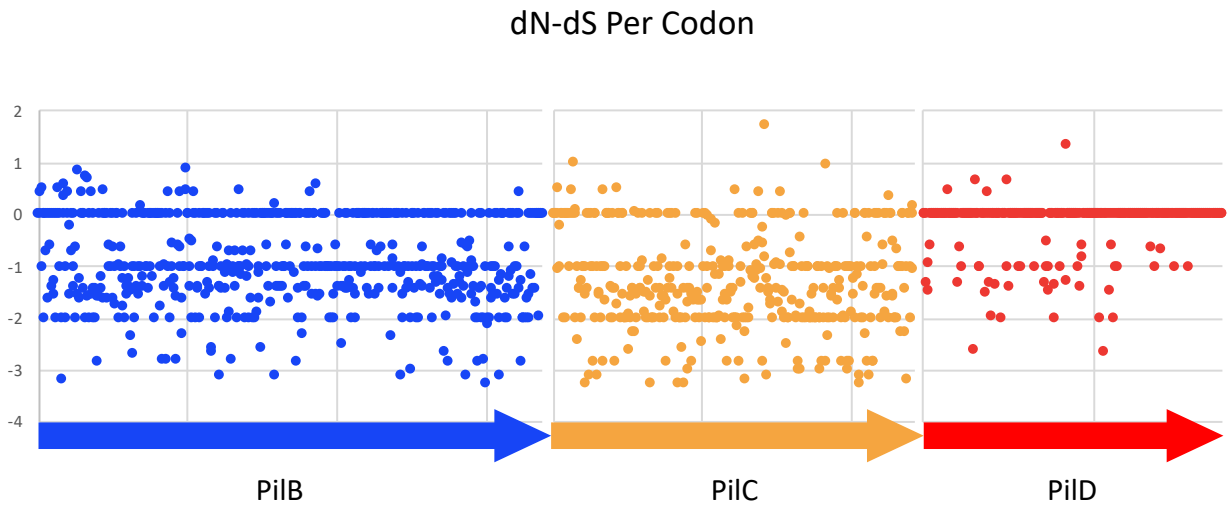


Figure AII-15 (Sup. Fig. 15): dN-dS of the operons encoding *pilB, pilC, and pilD* genes of the Type IV pilus. Nonsynonymous and synonymous substitutions were estimated based on maximum likelihood reconstructions of the ancestral states of these genes. No codons were found to be significantly under positive selection. Analyses were computed using Mega7 and included 7 nucleotide sequences encoded by *P. syringae* cotton pathogens.

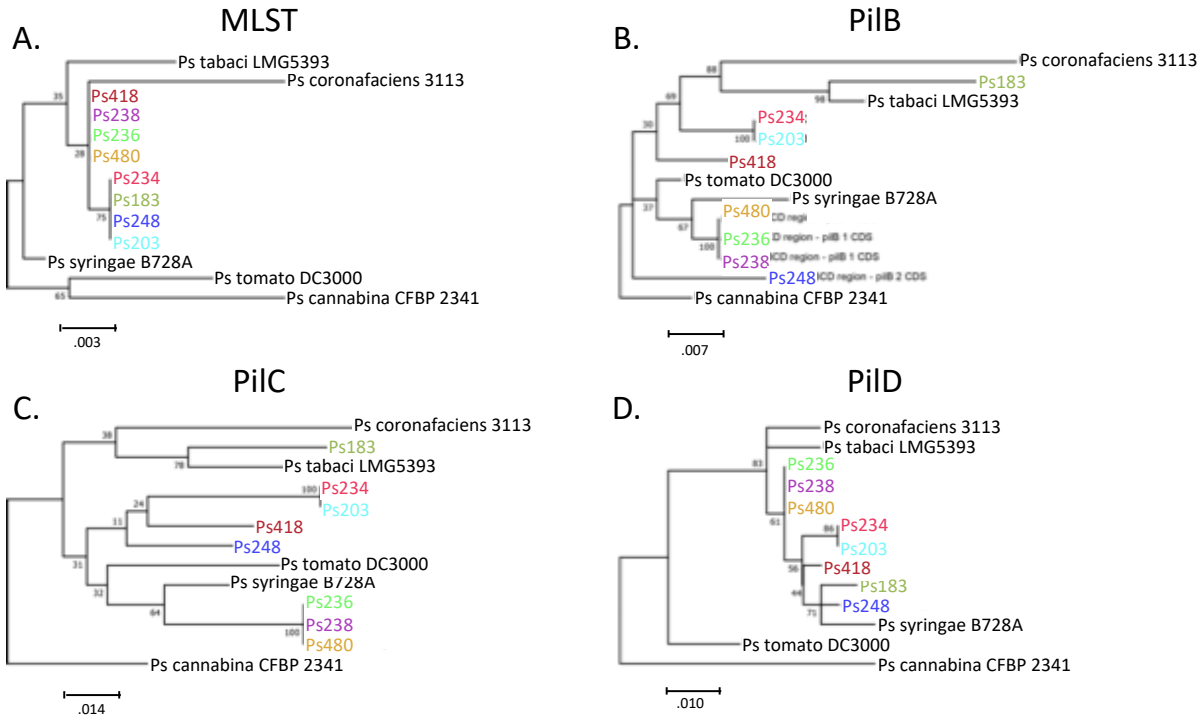


Figure AII-16 (Sup. Fig. 16): MLST and gene phylogenies of Type IV Pilus Genes spanning 5 *P. syringae* phylogroups. Maximum Likelihood phylogenetic tree based on amino acid sequences of A) concatenated regions of housekeeping genes: *gyrB*, *rpoD*, *gap1*, and *gltA*. B) *pilB* C) *pilC* D) *pilD* within *P. syringae* pathogens across 5 phylogroups. Branch labels are bootstrap percentage values based on 100 bootstraps.

Table AII.4(Sup. Table 4): Previously published genomes of *P. syringae* pathogens of tomato, kiwi, and cherry used in this study

Strain	pathovar	Isolation Source	Country, Year	Sequenced	GenBank Accession
NCPPB 1108	tomato	*Solanum lycopersicum*	UK 1961	Cai et al 2011	ADGA
T1	tomato	*Solanum lycopersicum*	Canada 1986	Almedia et al 2009	ABSM
K40	tomato	*Solanum lycopersicum*	USA 2005	Cai et al 2011	ADFY
LNPV 17.41	tomato	*Solanum lycopersicum*	France 1996	Cai et al 2011	ADFZ
A9	tomato	*Solanum lycopersicum*	USA 1996	Thapa et al 2016	LNKY
407	tomato	*Solanum lycopersicum*	USA 1997	Thapa et al 2016	LNKZ
PT23	tomato	*Solanum lycopersicum*	USA1990	Meaden et al 2017	MSDS
NYS-T1	tomato	*Solanum lycopersicum*	USA 2009	Jones et al 2015	JRRA
syr9097	syringae	*Prunus avium*	UK 2010	Hulin et al 2018	CP026568
syr5275	syringae	*Prunus avium*	UK 1990	Hulin et al 2018	NBAP
syr7924	syringae	*Prunus avium*	UK 2000	Nowell et al 2016	LIHR
syr9656	syringae	*Prunus avium*	UK 2012	Hulin et al 2018	MLEM
syr7928A	syringae	*Prunus avium*	UK 2000	Hulin et al 2018	NBAL
R1-5244	morsprunorum	*Prunus avium*	UK 1960	Hulin et al 2018	CP026557–CP026561
R2-leaf	morsprunorum	*Prunus avium*	UK 2014	Hulin et al 2018	CP026562–CP026567
avii3846	avii	*Prunus avium*	France 1991	Nowell et al 2016	LIIJ
PsaJ2	actinidiae	*Actinidia chinensis*	Japan 1988	Mazzaglia et al 2012	AGNQ
NZLV-14	actinidiae	*Actinidia deliciosa*	New Zealand 2010	McCann et al 2013	AOKG
NZLV-18	actinidiae	*Actinidia chinensis*	New Zealand 2010	McCann et al 2013	AOKE
NZLV-6	actinidiae	*Actinidia chinensis*	New Zealand 2010	McCann et al 2013	AOKJ
PsaK26	actinidiae	*Actinidia chinensis*	Korea 1997	McCann et al 2013	AOJW
NZV13	actinidiae	*Actinidia deliciosa*	New Zealand 2010	McCann et al 2013	AOKO
PsaJ35	actinidiae	*Actinidia deliciosa*	Japan 1984	McCann et al 2013	AOKP-3

Appendix III: Cotton Bacterial Blight Resistance Across the Parents of a Nested Association Mapping Population

Introduction

Canonical plant resistance genes are dominantly inherited genes that confer qualitative resistance to pathogens[1]. The most well studied family of resistance proteins are nucleotide binding leucine-rich repeat (NB-LRR) proteins. These proteins either directly or indirectly recognize the presence of pathogen effectors within the cell and trigger a rapid cell death at the site of the infection[2,3]. This reaction is microscopic under field conditions. However, when high concentrations of bacteria are injected into the host, a large necrotic lesion develops at the site of the injection approximately two days after inoculation. This phenotype allows for quick and accurate screening of plant germplasm.

Many *Xanthomonas citri* pv. *malvacearum* (*Xcm*) resistance genes have been roughly mapped in the *Gossypium hirsutum* genome[4,5]. When deployed in parallel, they confer durable resistance to this disease[6]. However, the genetic basis for this resistance is still unknown. A nested association mapping (NAM) population has recently been developed that will enable finer mapping of resistance genes and ultimately the identification of the resistance gene itself.

Tyagi et al. genotyped 381 diverse accessions of *G. hirsutum* and identified a core set of 53 accessions that span 96% of the diversity[7]. These 53 accessions have each been crossed with both (*Xcm* susceptible) parental lines of the population: accessions DIV 017 Acala Maxxa and DIV126 DES56. The progeny were allowed to self-pollinate for six generations, ensuring homozygosity of the resulting lines and maximum opportunities for cross-over to mix parental DNA. Screening the resulting lines will enable fine mapping of resistance genes after

preliminary experiments are completed. Here I describe a *Xcm* resistance screen for the parents of the NAM population to identify candidate accessions for mapping resistance genes.

Methods

Forty-seven of the 53 accessions were screened for resistance to ten strains of *Xcm* in order to identify candidates for mapping resistance genes. The remaining six accessions were not screened due to germination issues. Briefly, the screen was performed by inoculating bacterial cultures with an OD600 of 0.01 in 10 mM $MgCl_2$ into fully expanded leaves with a needleless syringe. Inoculated plants were kept in a room with 50% humidity. Symptoms were assessed two or more days later.

Results

Twenty-eight accessions displayed water soaking at the site of the inoculation, a typical susceptible response (Table AIII.1). Ten accessions displayed unclear responses due to plant damage and pest infestations that obscured results as well as incomplete resistance responses. The incomplete resistance may have been caused by minor resistance alleles or simply a response to abiotic stress. Nine accessions triggered a resistance response to each *Xcm* strain during 3 experimental replicates. Interestingly, these accessions are phylogenetically divergent, suggesting that multiple sources of resistance have been identified (Figure AIII.1).

In order to determine if the source of resistance within these genomes was a single dominantly inherited gene, the seven resistant accessions were crossed with DIV 017 Acala Maxxa, a susceptible accession. The progeny was self-fertilized to develop an F2 population. Three populations were screened for resistance. Individuals within two of these populations still showed unclear symptoms. However, the population created with DIV 042 Arkot 8102 showed clear symptoms including 26 resistant and eight susceptible individuals, demonstrating a 76%

resistant 24% susceptible population. This is close to the 3:1 ratio of resistant to susceptible lines that would be expected if one parent had a single, dominantly inherited resistance gene.

Table AIII.1: Resistance and susceptibility of *G. hirsutum* NAM parents to *Xcm*.

Resistant	Susceptible	Unclear
DIV 042 Arkot 8102	DIV 002 Acala 111Rogers	DIV 059 CA23
DIV 046 BJAGI NECT	DIV 012 Acala 5	DIV206-1 LBBCDBOAKH-1-90
DIV 063 CABD3CABCH-1-89	DIV 017 Acala Maxxa	DIV237
DIV 066 CAHUGIBBCS -1-88	DIV 034 Allen 33	DIV269
DIV 176-1 H1220	DIV 040 Arkansas 10	DIV237 Paymaster 101
DIV 206-1 LBBCDBOAKH-1-90	DIV079 Coker201	DIV245 Paymaster hs200
DIV 245 Paymaster HS200	DIV096 CS-8610	DIV256 PD2165
DIV 319 SPNXCHGIBH-1-94 -1	DIV106 Deltapine 14	DIV269 PD93009
DIV 347 Tamcot Luxor-1	DIV126 DES56	DIV272 PD93030
	DIV128 Dixie King	DIV 316 Southland M1
	DIV134 Earlistaple 7	
	DIV136 Empire	
	DIV141 Express 121	
	DIV144 FJA	
	DIV168 Gregg35	
	DIV 169 GSA74	
	DIV 178-1 HALF AND HALF	
	DIV 184-1 HOPI MOENCOPI	
	DIV 195-1 La.850082FN	
	DIV 209-1 LOCKETT 88	
	DIV 211-1 M.U.8B UA 7-44	
	DIV 230 NC88-95	
	DIV 232 New Boykin	
	DIV 246 Paymaster HS26	
	DIV 249 PD 0013	
	DIV 255 PD 2164	
	DIV 263 PD 781	
	DIV 264 PD 785	

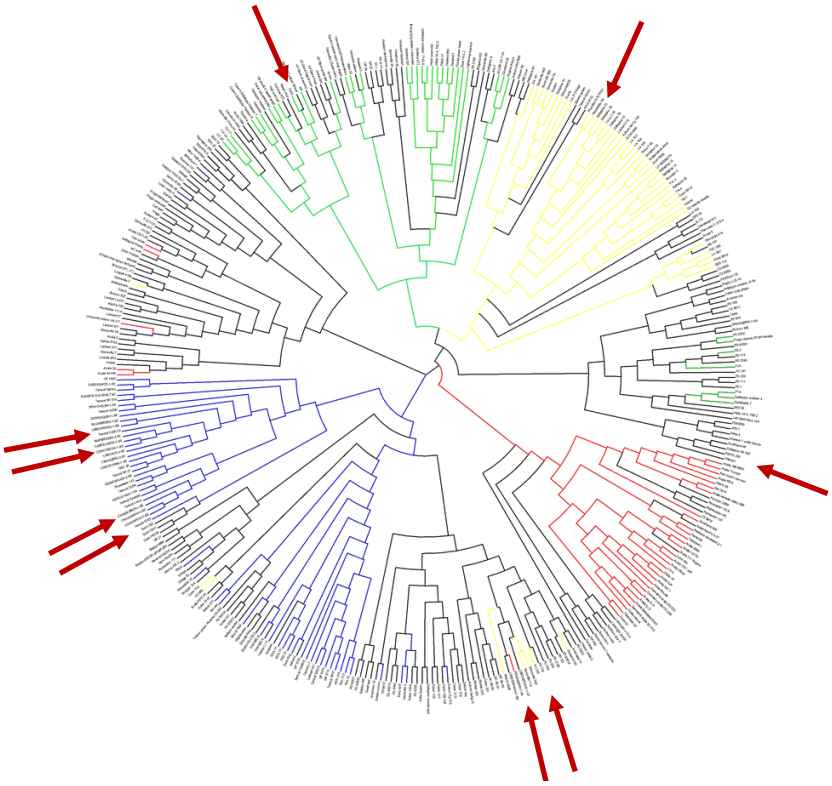


Figure AIII.1: Positions of nine resistant NAM parents on global dendrogram of 378 *G. hirsutum* accessions adapted from Tyagi et al. 2014. The dendrogram was created by Tyagi et al. using the neighbor joining method with a distance matrix generated with 120 SSR markers. Colors represent STRUCTURE analysis correlating to the four cotton growing areas: Western (red), Southwest (blue), Midsouth (yellow), and Eastern (green)[7].

Future Directions

Now that F6 populations have been generated, 100 progeny will be screened for resistance, starting with the populations created with DIV 042 Arkot 8102. These individuals will then be re-sequenced to identify DIV 042 Arkot 8102 genomic regions that correlate with resistance. In parallel, RenSeq could be performed to further narrow down the list of potential resistance genes. RenSeq (short for resistance gene enrichment and sequencing) uses a custom developed chip of putative NB-LRR genes to sequence segregating populations and identify resistance genes that are present in only resistant individuals[8]. BLAST searches of 560 annotated NB-LRR resistance genes from *Arabidopsis thaliana* and *Medicago truncatula* have already resulted in 673 putative NB-LRR genes that can be used to develop a chip for RenSeq. This work

will significantly contribute to the understanding of *Xcm* resistance in cotton and aid in breeding *Xcm* resistance into more farmer preferred varieties.

References

1. Jones JDG. Putting knowledge of plant disease resistance genes to work. *Curr Opin Plant Biol*. 2001;4(4):281-287. doi:10.1016/S1369-5266(00)00174-6

2. Caplan J, Padmanabhan M, Dinesh-Kumar SP. Plant NB-LRR Immune Receptors: From Recognition to Transcriptional Reprogramming. *Cell Host Microbe*. 2008;3(3):126-135. doi:10.1016/j.chom.2008.02.010

3. Heath MC. Hypersensitive response-related death. In: Lam E, Fukuda H, Greenberg J, eds. *Programmed Cell Death in Higher Plants*. Dordrecht: Springer Netherlands; 2000:77-90. doi:10.1007/978-94-010-0934-8_6

4. Delannoy E, Lyon BR, Marmey P, et al. Resistance of Cotton Towards Xanthomonas campestris pv. malvacearum. *Annu Rev Phytopathol*. 2005;43(1):63-82. doi:10.1146/annurev.phyto.43.040204.140251

5. Xiao J, Fang DD, Bhatti M, Hendrix B, Cantrell R. A SNP haplotype associated with a gene resistant to Xanthomonas axonopodis pv. malvacearum in upland cotton (Gossypium hirsutum L.). *Mol Breed*. 2010;25(4):593-602. doi:10.1007/s11032-009-9355-y

6. Essenberg M, Bayles MB, Pierce ML, Verhalen LM. Pyramiding B Genes in Cotton Achieves Broader But Not Always Higher Resistance to Bacterial Blight. *Phytopathology*. 2014;104(10):1088-1097. doi:10.1094/PHYTO-06-13-0167-R

7. Tyagi P, Gore MA, Bowman DT, Campbell BT, Udall JA, Kuraparthy V. Genetic diversity and population structure in the US Upland cotton (Gossypium hirsutum L.). *Theor Appl Genet*. 2014;127(2):283-295. doi:10.1007/s00122-013-2217-3

8. Jupe F, Witek K, Verweij W, et al. Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant J*. 2013;76(3):530-544. doi:10.1111/tpj.12307