GENOMIC CHARACTERIZATIONS OF *XANTHOMONAS CUCURBITAE* AND USING COMPARATIVE GENOMICS TO PREDICT NOVEL MICROBE-ASSOCIATED MOLECULAR PATTERNS IN *XANTHOMONAS*

BY

JULIUS PASION

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Bioinformatics
with a concentration in Crop Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2020

Urbana, Illinois

Master's Committee:

Assistant Professor Sarah Hind, Chair
Assistant Professor Julian Catchen
Professor Stephen Moose

ABSTRACT

Bacterial spot is a major plant disease caused by many plant-pathogenic members of the genus *Xanthomonas*. While each species is narrow in host range, bacterial spot *Xanthomonads* infect a large variety of plant hosts, leading to large economic losses for farmers around the world. Although *Xanthomonas* utilizes a wide array of virulence and pathogenicity factors to infect their hosts, plants have a range of methods to recognize invaders and prevent infection. Understanding the genomic and molecular interactions between *Xanthomonas* and their hosts are an important part of developing effective crop protection strategies and breeding plants for resistance.

While *X. cucurbitae* has been identified as the causal agent of bacterial spot on cucurbits, no genomic-level analyses have been carried out regarding the pathogen. Using the first reference quality *X. cucurbitae* genome assembly, an RNA-seq analysis was carried out to assess virulence characteristics of the pathogen. By analyzing the *X. cucurbitae* transcriptome, we observed behavioral changes between nutrient-sufficient and host-mimicking conditions, as well as the upregulation of genes related to virulence and pathogenicity. We also identified virulence genes likely to be essential in successful bacterial spot infection. In addition, a RAD-seq analysis was performed to characterize populations clusters of *X. cucurbitae* isolated throughout the Midwestern United States. We revealed multiple populations of *X. cucurbitae* present throughout the region and demonstrated clear genetic differences between these populations using population genetics analyses. These studies demonstrate clear value in future genomic studies regarding *X. cucurbitae.*

X. euvesicatoria and *X. perforans* are two bacterial spot *Xanthomonads* affecting tomatoes and peppers. We conducted a comparative genomics study in *X. euvesicatoria* and *X. perforans* populations to identify genes under selection pressure, and to characterize potential genes involved in plant-pathogen interactions. By calculating the test statistic Tajima's D, we found evidence of purifying selection throughout the genomes of both bacterial spot Xanthomonads. In addition, Tajima's D was successfully able to detect known microbe-associated molecular patterns (MAMPs), and we were able to characterize the recognition of these MAMPs between species in luminol-based reactive oxygen species (ROS) assays. While this study was not yet able to identify novel MAMPs, we show that Tajima's D is a powerful tool in detecting genes that are important to plant-pathogen interactions.

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Sarah Hind for giving me the opportunity to further my education. Her excellent counseling and encouragement have pushed me to work hard and take advantage of the many opportunities a graduate level education can provide. I also would like to thank my other committee members Dr. Julian Catchen and Dr. Stephen Moose for their guidance and constructive criticisms throughout my research.

I would like to express gratitude towards members of the Hind Lab, especially Maria and Rikky, both for sharing their knowledge and for fostering an inviting lab atmosphere. I am grateful to members of the Catchen Lab, especially Angel, for teaching me how to analyze RAD-seq data and for giving great suggestions for my data analyses. I am also grateful to Dr. Dave Zhao for his guidance in my RNA-seq research.

Finally, I would like to thank my parents, my sister, and my partner Alice for their unwavering love and support throughout my master's education.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Pumpkin

The terms “pumpkin” and “winter squash” refers to multiple members of the *Cucurbita* family. In general, the most commonly used scientific name for pumpkin is *C. pepo*, but some cultivars of *C. moschata* and *C. maxima* are also referred to as pumpkins; the term winter squash has been used interchangeably to refer to some cultivars of pumpkins as well (Babadoost and Zitter, 2009). Initial domestication of *C. pepo* occurred between 8,000 and 10,000 years ago in Mexico, pre-dating the domestication of other crops such as maize and beans (Smith, 1997). Today, *C. pepo* is grown for a variety of uses, including human consumption, agricultural products, and ornamental purposes. Pumpkin cultivars can vary in many aspects, including size, shape, rind thickness, and color.

Depending on the cultivar, fruits and/or seeds of *C. pepo* can be consumed at an immature stage (i.e., thin-skinned summer types) or at a mature stage (i.e., hard-skinned winter types) (Paris, 1989). Pumpkin seeds are low in fat and rich in proteins, while the fleshy part of the fruits is high in β-carotene, carbohydrates, and other nutrients. Consuming pumpkins is also reported to bestow many health benefits, such as potentially acting in anti-carcinogenic or anti-diabetic capacities (Yadav *et al.*, 2010). Beyond fresh consumption, mature pumpkin fruits can be processed into products such as canned pumpkin puree. Pumpkin fruits and seeds are also processed for medicinal and cosmetic purposes, such as dietary supplements and skin scrubbers.

Agronomic Production of Pumpkin

From 2017 to 2018, over 27 million tons of pumpkins, squash, and gourds were produced worldwide, indicating an important crop for farmers around the world (FAOSTAT, 1997). Countries that are top producers of these crops include China, India, and Ukraine, with the United States (US) ranking within the top ten countries. Five states within the US, including Illinois, Indiana, Pennsylvania, Texas, and California, produce roughly 40% of the pumpkins nationally (USDA, 2019). Illinois is the largest pumpkin producer in the US; in 2018, Illinois produced about 500 million pounds of pumpkins, greater than the next four states combined. Nearly 80% of Illinois-grown pumpkins are used for processing, while 90% of the US processing pumpkins are grown in Illinois (USDA, 2019). In 2018, pumpkin production for fresh pumpkins and processed pumpkins combined totaled nearly $200 Million dollars (USDA, 2019).

Tomato

Tomato (*Solanum lycopersicum*) is a member of the *Solanaceae* family. This plant family contains many species that are staple foods in human diets worldwide such as tomatoes, potatoes, and peppers (Bergougnoux, 2014). The ancestor of modern tomato plants originated in the Andes Mountain Range in South America. Tomato plants were domesticated by the Incans and Aztecs, and subsequently brought to Europe in the 16th century. While initially avoided because of similarities to the related toxic nightshade plants, tomatoes have become the most widely consumed, non-starchy vegetable, and are a staple ingredient in food cultures worldwide (Burton-Freeman *et al.*, 2011).

Tomato fruits come in a variety of shapes and sizes and can differ in many taste-related qualities such as texture, color, and flavor. Fresh-market varieties are commonly eaten raw, while canning varieties are usually processed and consumed as a sauce or paste. Tomato fruits are nutritionally dense and contain key/important/essential nutrients such as potassium and vitamin C. Additionally, tomatoes are one of the main human dietary sources of lycopene, a carotenoid that is responsible for the wide range of colors exhibited by the fruits. Recent data from epidemiological studies suggest that lycopene may confer a wide range of health benefits, such as lowering cancer and cardiovascular disease risk (Story *et al.*, 2010). While these claims need to be substantiated, it is clear that tomato fruits are an integral component of human diets around the world.

Agronomic Production of Tomato

The top tomato producing countries are China, India, and the US. Within the US, Florida and California are consistently the top tomato-producing states, although other states also contribute to the industry. Florida and California account for about two-thirds of fresh tomato production in the US, with California producing 90% of the tomatoes used for processing (Guan *et al.*, 2016). While production in the US has slowed down due to the rise of tomato imports from Mexico, the demand for tomatoes continues to grow (Baskins *et al.*, 2019). In 2015, the value of the US fresh market and processing industries combined was more than $2.6 Billion dollars (Guan *et al.*, 2016). In addition to conventional field cropping systems, greenhouse tomato production has grown in popularity throughout the US, enabling year-round tomato production (Cook and Calvin, 2015).

Bacterial Spot Disease

Bacterial spot is a major plant disease caused by many species in the bacterial genus *Xanthomonas*. For tomato and pepper, the causal agents of bacterial spot include four *Xanthomonas* species: *X. euvesicatoria, X. perforans*, *X. vesicatoria*, and *X. gardneri*. Symptoms of this disease include dark lesions over the flesh of the fruit and large necrotic areas on leaves and stems. This disease is responsible for large percentages in tomato yield losses every year, and in extreme cases can even lead to complete crop failure. In 2010, the Midwest tomato processing industry estimated that nearly $8 million dollars of potential revenue was lost due to bacterial spot disease (NIFA-USDA, 2016). Bacterial spot disease is also a worldwide problem, hindering tomato and pepper production in countries including Brazil and Korea (Kyeon *et al.*, 2016).

In pumpkin, the causal agent of bacterial spot is *X. cucurbitae.* Symptoms include dark angular lesions on cucurbit leaves, and circular, sunken lesions on the rinds of fruit that enlarge as the fruit matures. *X. cucurbitae* has been found to cause yield losses of more than 50% in Illinois commercial pumpkin fields and has been isolated worldwide on various cultivars of *X. pepo* such as squash and jack-o'-lantern pumpkins (Babadoost and Zitter, 2009).

Transmission of Bacterial Spot and Management Practices to Mitigate Disease Spread

Bacterial spot thrives in warm and humid environments and can spread through seed and plant debris as well as by water splashes (Jones *et al.*, 2000). It is more prevalent during seasons with high humidity, while occurrences decrease in drier or cooler conditions. Wind-dispersed water droplets can rapidly disseminate the pathogen throughout a field, and the bacteria can also be

spread via contaminated farm tools and transplant materials. In addition, it can infect seeds, as well as survive for weeks or months on infected plant debris.

Growers take many measures to reduce bacterial spot disease occurrences in their fields. The primary mode of action is to prevent accidental introduction of the bacteria to uncontaminated fields and greenhouses, as the disease is difficult to manage once established. Crop rotations, seed treatments, and equipment decontaminations are common procedures used to prevent the spread of the bacterial (Ritchie, 2000). Additionally, copper-based chemical sprays have been widely used to control the disease; however, they are only effective prior to establishment of the disease. Unfortunately, the rise of copper-resistant and copper-tolerant xanthomonads have dulled the effectiveness of copper sprays, prompting research into other avenues of disease control (Behlau *et al.*, 2011). Biological controls such as bacteriophages have had some success in controlling bacterial spot infection (Flaherty *et al.*, 2000). Also, tomato breeding programs develop plants that resistant to bacterial spot, either using conventional breeding approaches with introgressions from wild relatives, or by developing transgenic plants. Together, cultural practices and resistant cultivars are integral components of successful disease management practices.

Bacterial Spot Xanthomonads

The genus *Xanthomonas* is comprised of gram-negative straight-rod shaped bacteria with a single polar flagellum for motility. *Xanthomonas* form yellow-pigmented colonies when cultured on growth media such as peptone-sucrose, and they produce the exopolysaccharide xantham gum, which is an important ingredient in the food production industry. Many species of

Xanthomonas are plant-pathogenic and infect staple food crops worldwide. While the host range for the genus is large, host specificity for each *Xanthomonas* species is generally narrow (Leyns *et al.*, 1984) with only rare crossover in host plants between different *Xanthomonas* species. Common hosts of *Xanthomonas* bacteria include tomato, rice, and citrus.

Phylogeny of Bacterial Spot Xanthomonads on Pepper and Tomato

The phylogeny of tomato- and pepper-specific bacterial spot Xanthomonads has changed greatly since their discovery in 1921 (Doidge, 1921). While initially thought to be comprised of a single species, studies in the 1990s identified two distinct groups (Stall *et al.*, 1994; Vauterin *et al.*,1995), while further phenotypic and genotypic studies elucidated three evolutionary lineages of *Xanthomonas* that caused disease on tomato and pepper (Jones *et al.*, 2000). In 2004, Jones *et al*. used DNA-DNA relatedness studies to reclassify these *Xanthomonas* into the four distinct lineages that are used today: *X. euvesicatoria, X. perforans, X. gardneri,* and *X. vesicatoria* (Jones *et al.*, 2004).

While all four *Xanthomonas* species share a common host range, there are varying degrees of genetic relatedness between them. Although a small core set of secreted proteins called type III effectors (T3Es) is shared amongst the four species, comparative genome studies have confirmed that these species are genetically distinct (Potnis *et al.*, 2011). Each species of *Xanthomonas* that cause bacterial spot disease uses distinct groups of cell-wall degrading enzymes during infection, but not all enzymes are shared between the four species, and the genomic arrangement of these genes also varies (Potnis *et al.,* 2011). If multiple *Xanthomonas* species are found in the same field, they can act as competitors against each other, since they share the same host. For example,

strains of *X. perforans* have been shown to outcompete and inhibit growth of *X. euvesicatoria* using *X. perforans*-derived bacteriocins (Hert *et al.*, 2005; Hert *et al.*, 2009).

Recent studies have re-examined the distinctions between these four species. 16s rRNA analysis reveals that *X. euvesicatoria* and *X. perforans* are more related to each other than to *X. vesicatoria* and *X. gardneri*, but still indicated that *X. euvesicatoria* and *X. perforans* are separate species (Jones *et al.*, 2004; Potnis *et al.*, 2011). However, later taxonomic analyses of these species indicate that these two Xanthomonads could be considered a single species. Multi-locus sequence analysis, DNA-DNA hybridizations, and average nucleotide identity studies conducted by Constantin *et al.* suggested that *X. perforans* should be merged with *X. euvesicatoria* (Constantin *et al.*, 2016). To add further complication, some bacterial spot strains show evidence of horizontal gene transfer and recombination-mediated evolution (Jibrin *et al.*, 2018). Further research is necessary to resolve the phylogeny between *X. euvesicatoria* and *X. perforans.*

Pattern Triggered Immunity

The first line of defense in the plant immune system is known as Pattern Triggered Immunity (PTI). Microbe-Associated Molecular Patterns (MAMPs) are molecules that often are essential for growth and survival of the microbe, and thus can be highly conserved. Plant cells use cell surface pattern recognition receptors (PRRs) to recognize MAMPs (Zipfel, 2014). Upon PRR recognition of a cognate MAMP, the PRR is activated and triggers intracellular signals that includes kinase phosphorylation cascades, production and release of reactive oxygen species (ROS), and calcium bursts, ultimately leading to the expression of primary immune response

genes (Li *et al.*, 2016). Through MAMP recognition, PTI protects plants from a broad range of possible pathogens.

The best characterized example is the MAMP flagellin, which is the major protein component of bacterial flagella. The specific peptide region designated flg22 is the specific MAMP ligand that is recognized by the broadly-conserved PRR called FLS2 (Chinchilla *et al.*, 2006). Another flagellin peptide region, designated flgII-28, is a MAMP recognized by FLS3, a PRR found in many solanaceous plants such as potatoes, tomatoes, and peppers (Hind *et al.*, 2016).

PRRs and MAMP Interactions

The interplay between microbial MAMPs and plant PRRs constitute a significant part of selective evolutionary pressures in both pathogens and their host plants. Because PRRs recognize specific sequences or structures in MAMPs, mutations in those MAMP regions can prevent the plant immune system from detecting the pathogen. For pathogens, selection pressure will favor mutations that prevent MAMP recognition but do not impede pathogenicity. For plants, PRRs that have evolved flexibility in the MAMP recognition sites may be able to recognize variations of target MAMPs. Because avoiding recognition is so crucial to successful bacterial infections, the DNA sequences encoding some MAMPs can be highly variable.

Effector Triggered Immunity

Plants have additional defenses to protect against disease-causing microbes. Many bacteria, including bacterial spot Xanthomonads, have a type III secretion system, where a syringe-like protrusion from the bacteria can penetrate through the plant cell and ‘inject’ type III effector

proteins (T3Es) directly into the cytosol. For many *Xanthomonas* species, one important class of effector proteins are Transcription Activator-Like (TAL) effectors, which function as transcription factors inside the host cell. Once translocated into the cytosol, these TAL effectors target promoter regions of genes inside the host nucleus, thus affecting genes involved in immune response as well as general cellular housekeeping functions, and ultimately increasing host susceptibility to disease (White, 2016). If the plant recognizes these effector proteins inside the cytosol using resistance genes (*R* genes), Effector Triggered Immunity (ETI) occurs; this mode of plant resistance is more robust and occurs much faster than the immunity conferred by PTI. One component of ETI is the hypersensitive response (HR), which leads to programmed cell death of the infected plant cells and restriction of pathogen growth at the site of infection.

Research Objectives

The overall objective of this research was to study genes and gene products involved in plant-pathogen interactions for bacterial species in the genus *Xanthomonas.* An RNA-seq analysis was carried on *X. cucurbitae* to identify genes important to infection on cucurbit plants. In addition, a RAD-seq analysis was used to characterize genetic clusters in *X. cucurbitae* isolates collected throughout the Midwestern United States. Finally, a comparative genomics approach was used to screen for novel MAMPs in *X. euvesicatoria* and *X. perforans,* the causal agent of bacterial spot disease on tomato and pepper plants.

LITERATURE CITED

Babadoost, M. and T. A. Zitter (2009). "Fruit Rots of Pumpkin: A Serious Threat to the Pumpkin Industry." Plant Disease **93**(8): 772-782.

Baskins, S., J. Bond and T. Minor (2019). "Unpacking the Growth in Per Capita Availability of Fresh Market Tomatoes." Retrieved 2019 May 20 from https://www.ers.usda.gov/webdocs/outlooks/92442/vgs-19c-01.pdf?v=6728.3

Behlau, F., B. I. Canteros, G. V. Minsavage, J. B. Jones and J. H. Graham (2011). "Molecular Characterization of Copper Resistance Genes from *Xanthomonas citri* subsp. *citri* and *Xanthomonas alfalfae* subsp. *citrumelonis*." Applied and Environmental Microbiology **77**(12): 4089.

Bergougnoux, V. (2014). "The history of tomato: From domestication to biopharming." Biotechnology Advances **32**(1): 170-189.

Burton-Freeman, B., & Reimers, K. (2011). "Tomato Consumption and Health: Emerging Benefits." American Journal of Lifestyle Medicine *5*(2): 182–191.

Chinchilla, D., Z. Bauer, M. Regenass, T. Boller and G. Felix (2006). "The *Arabidopsis* Receptor Kinase FLS2 Binds flg22 and Determines the Specificity of Flagellin Perception." The Plant Cell **18**(2): 465.

Constantin, E. C., I. Cleenwerck, M. Maes, S. Baeyen, C. Van Malderghem, P. De Vos and B. Cottyn (2016). "Genetic characterization of strains named as *Xanthomonas axonopodis* pv. *dieffenbachiae* leads to a taxonomic revision of the *X. axonopodis* species complex." Plant Pathology **65**(5): 792-806.

Cook, R., U. E. R. Service and L. Calvin (2015). Greenhouse Tomatoes Change the Dynamics of the North American Fresh Tomato Industry - Scholar's Choice Edition, Creative Media Partners, LLC.

Doidge, E. M. (1921). "A TOMATO CANKER." Annals of Applied Biology **7**(4): 407-430.

Flaherty, J. E., J. Jones, B. K. Harbaugh, G. C. Somodi and L. E. Jackson (2000). "Control of Bacterial Spot on Tomato in the Greenhouse and Field with H-mutant Bacteriophages." HortScience **35**: 882-884.

Food and Agriculture Organization of the United, N. (1997). FAOSTAT statistical database, [Rome] : FAO, c1997-2020 Retrieved 2020 May 5 from http://www.fao.org/faostat/en/#data/QC/visualize.

Guan, Z., T. Biswas and F. Wu. (2016). "The U.S. Tomato Industry: An Overview of Production and Trade." Retrieved 2019 May 20 from https://edis.ifas.ufl.edu/fe1027.

Hert, A. P., M. Marutani, M. T. Momol, P. D. Roberts, S. M. Olson and J. B. Jones (2009). "Suppression of the bacterial spot pathogen *Xanthomonas euvesicatoria* on tomato leaves by an attenuated mutant of *Xanthomonas perforans*." Applied and Environmental Microbiology **75**(10): 3323-3330.

Hert, A. P., P. D. Roberts, M. T. Momol, G. V. Minsavage, S. M. Tudor-Nelson and J. B. Jones (2005). "Relative Importance of Bacteriocin-Like Genes in Antagonism of *Xanthomonas perforans* Tomato Race 3 to *Xanthomonas euvesicatoria* Tomato Race 1 Strains." Applied and Environmental Microbiology **71**(7): 3581.

Hind, S. R., S. R. Strickler, P. C. Boyle, D. M. Dunham, Z. Bao, I. M. O'Doherty, J. A. Baccile, J. S. Hoki, E. G. Viox, C. R. Clarke, B. A. Vinatzer, F. C. Schroeder and G. B. Martin (2016). "Tomato receptor FLAGELLIN-SENSING 3 binds flgII-28 and activates the plant immune system." Nature Plants **2**(9): 16128.

Jibrin, M. O., N. Potnis, S. Timilsina, G. V. Minsavage, G. E. Vallad, P. D. Roberts, J. B. Jones and E. M. Goss (2018). "Genomic Inference of Recombination-Mediated Evolution in *Xanthomonas euvesicatoria* and *X. perforans*." Applied and Environmental Microbiology **84**(13): e00136-00118.

Jones, J. B., H. Bouzar, R. E. Stall, E. C. Almira, P. D. Roberts, B. W. Bowen, J. Sudberry, P. M. Strickler and J. Chun (2000). "Systematic analysis of xanthomonads (*Xanthomonas* spp.) associated with pepper and tomato lesions." International Journal of Systematic and Evolutionary Microbiology **50**(3): 1211-1219.

Jones, J. B., G. H. Lacy, H. Bouzar, R. E. Stall and N. W. Schaad (2004). "Reclassification of the Xanthomonads Associated with Bacterial Spot Disease of Tomato and Pepper." Systematic and Applied Microbiology **27**(6): 755-762.

Kyeon, M.-S., S.-H. Son, Y.-H. Noh, Y.-E. Kim, H.-I. Lee and J.-S. Cha (2016). "*Xanthomonas euvesicatoria* Causes Bacterial Spot Disease on Pepper Plant in Korea." The Plant Pathology Journal **32**(5): 431-440.

Leyns, F., M. Cleene, J. Swings and J. Ley (1984). "The host range of the genus *Xanthomona*." The Botanical Review **50**: 308-356.

Li, B., X. Meng, L. Shan and P. He (2016). "Transcriptional Regulation of Pattern-Triggered Immunity in Plants." Cell Host & Microbe **19**(5): 641-650.

Paris, H. S. (1989). "Historical records, origins, and development of the edible cultivar groups of *Cucurbita pepo* (*Cucurbitaceae*)." Economic Botany **43**(4): 423-443.

Potnis, N., K. Krasileva, V. Chow, N. F. Almeida, P. B. Patil, R. P. Ryan, M. Sharlach, F. Behlau, J. M. Dow, M. T. Momol, F. F. White, J. F. Preston, B. A. Vinatzer, R. Koebnik, J. C. Setubal, D. J. Norman, B. J. Staskawicz and J. B. Jones (2011). "Comparative genomics reveals diversity among xanthomonads infecting tomato and pepper." BMC Genomics **12**(1): 146.

Ritchie, D. F. (2000). "Bacterial spot of pepper and tomato." Retrieved 2019 May 21 from https://www.apsnet.org/edcenter/disandpath/prokaryote/pdlessons/Pages/Bacterialspot.aspx.

Smith, B. D. (1997). "The Initial Domestication of *Cucurbita pepo* in the Americas 10,000 Years Ago." Science **276**(5314): 932.

Stall, R. E., C. Beaulieu, D. Egel, N. C. Hodge, R. P. Leite, G. V. Minsavage, H. Bouzar, J. B. Jones, A. M. Alvarez and A. A. Benedict (1994). "Two Genetically Diverse Groups of Strains Are Included in *Xanthomonas campestris* pv. *vesicatoria*." International Journal of Systematic and Evolutionary Microbiology **44**(1): 47.

Story, E. N., R. E. Kopec, S. J. Schwartz and G. K. Harris (2010). "An Update on the Health Effects of Tomato Lycopene." Annual Review of Food Science and Technology **1**(1): 189-210.

USDA-ERS. (2019). "Pumpkins: Background & Statistics." Retrieved 2020 May 5 from https://www.ers.usda.gov/newsroom/trending-topics/pumpkins-background-statistics/.

USDA, N. I. o. F. a. (2016). "Fact Sheet: Bacterial Spot of Tomato." Retrieved 2018 December 2 from https://nifa.usda.gov/sites/default/files/resource/Tomato_Spot_Fact_Sheet.pdf.

USDA, NASS. (2019). "Quick Stats." Retrieved 2020 June 20 from https://quickstats.nass.usda.gov/results/60BA40C8-4A48-3CDF-A4E2-9A328035037E.

Vauterin, L., B. Hoste, K. Kersters and J. Swings (1995). "Reclassification of *Xanthomonas*." International Journal of Systematic and Evolutionary Microbiology **45**(3): 472.

White, F. (2016). "*Xanthomonas* and the TAL Effectors: Nature's Molecular Biologist." Methods in molecular biology (Clifton, N.J.) **1338**: 1-8.

Yadav, M., S. Jain, R. Tomar, G. B. K. S. Prasad and H. Yadav (2010). "Medicinal and biological potential of pumpkin: an updated review." Nutrition Research Reviews **23**(2): 184-190.

Zipfel, C. (2014). "Plant pattern-recognition receptors." Trends in Immunology **35**(7): 345-351.

CHAPTER 2:

VIRULENCE CHARACTERIZATION AND POPULATION GENETICS ANALYSES OF *XANTHOMONAS CUCURBITAE*

ABSTRACT

RNA-seq and RAD-seq analyses were carried out to assess genetic and phylogenetic characteristics of *X. cucurbitae*, a non-vascular plant pathogen commonly found throughout US Midwestern commercial pumpkin fields. Analysis of the *X. cucurbitae* transcriptome in growth-promoting versus host-mimicking conditions revealed many differences in expression of genes involved in cellular processes related to infection and metabolism. Under host-mimicking conditions, we observed increased expression in genes involved in virulence responses, such as type II cell wall-degrading enzymes and type III effector proteins and system genes. We also noted downregulation of genes involved in key homeostatic processes, indicative of alterations that are often observed when pathogens switch from growth to virulence phases. Analysis of our RAD-seq data revealed clear stratifications of *X. cucurbitae* populations throughout the Midwestern US. These population clusters were separated geographically, indicating little genetic admixture throughout our sampled population of *X. cucurbitae* isolates.

INTRODUCTION

Xanthomonas cucurbitae is a plant pathogen that infect members of the cucurbit plant family worldwide. *X. cucurbitae* was first isolated in 1926 in New York (Bryan, 1926), and since then has been identified all over the world, infecting various economically important cucurbit plants such as pumpkin, squash, and watermelon, among others (Liu *et al.*, 2016). *X. cucurbitae* infection is characterized by symptoms on the leaves and fruits of its cucurbit host plants.

Cucurbit leaves infected with bacterial spot exhibit small lesions beige in color that are surrounded by yellow halos; these lesions eventually cover large areas of infected leaves as the infection progresses (Ravanlou and Babadoost, 2015). Symptoms on the fruit of infected pumpkin include lesions and necrotic areas across the surface (Ravanlou and Babadoost, 2015). In addition, these lesions and necrotic spots allow entryways for secondary infections by outside bacteria and fungi, leading to further decay of infected fruit (Liu *et al.*, 2016).

In recent years, *X. cucurbitae* has become a large threat to pumpkin production in the United States (US) (Zhang *et al.*, 2018). Illinois ranks first among US states in terms of annual pumpkin production, accounting for approximately 90% of processing pumpkin production nationwide (USDA, 2019). In commercial fields throughout Illinois, bacterial spot infection frequently occurs and has been estimated to cause anywhere between 3% and 90% yield losses (Babadoost and Ravanlou, 2012). *X. cucurbitae* has been found throughout the Midwestern US as well as globally in other cucurbit production areas, indicating a clear need to study the pathogen and develop stronger crop management practices against this disease. While copper-based chemical sprays have been widely used to combat *X. cucurbitae* infection, their efficacy is limited and they present a risk of fomenting the development of copper-resistant strains due to selection pressures. Hence, more information is needed to develop more effective management strategies against bacterial spot of cucurbits.

Plant pathogens utilize an array of strategies to successfully infect and colonize their hosts. In gram-negative species such as those found in the *Xanthomonas* genus, bacteria have evolved specialized machinery such as the type II and type III secretion systems that increase their

pathogenicity or virulence (Costa *et al.*, 2015). The type II secretion system secretes molecules and proteins from the periplasm of the bacteria towards the extracellular space; these molecules include toxins, adhesins, and tissue-degrading enzymes like cellulase and protease (Ciancotto *et al.*, 2017). The type III secretion system injects effector proteins from the bacteria directly into the host cell, where they can affect host processes and increase the virulence of pathogen infection. Furthermore, some *Xanthomonas* species contain Transcription Activator-Like (TAL) effector proteins that can enter the cell nucleus and manipulate gene expression. The array of type III effectors that a plant pathogen possesses has a large effect on pathogen phenotypes such as host range and recognition by host plants (Escalon *et al.*, 2013). Therefore, knowledge of pathogenicity factors is useful for developing durable disease resistance in major crop plants (Potnis *et al.*, 2015, Zhang and Coaker, 2017).

While bacterial leaf spot is a widespread plant disease that affects cucurbits worldwide, no studies have been done previously to characterize the underlying genetic structure or molecular interactions of the pathogen *X. cucurbitae.* In this study we utilize two techniques, RNA sequencing (RNA-seq) and restriction site-associated DNA sequencing (RAD-seq), to understand more about *X. cucurbitae* on the molecular and genetic level.

RNA-seq is a powerful technique for transcriptomic profiling based on short next-generation sequencing (NGS) reads. It has a large variety of demonstrated applications for biological studies, including showing differences in overall gene expression between genotypes and determining differentially expressed genes over time (Kogenaru *et al.*, 2012, Socquet-Juglard *et al.*, 2013). In plant pathogen studies, RNA-seq is a useful tool for determining pathogenic gene

expression, leading to a deeper understanding of pathogen behavior at the transcriptomic level (Kim *et al.*, 2016). In this study, we grew *X. cucurbitae* cultures *in vitro* using host-mimicking media and evaluated different gene expression using RNA-seq to identify genes important for pathogenicity and infection.

Population genetics analyses have been used for many years to determine genetic variation and evolutionary forces acting upon a pathogen (Linde, 2010). These studies can uncover many aspects of the evolutionary history of the pathogen, such as the spread of important pathogenic genes as well as the lineages that develop due to spatial and temporal isolation (Chien *et al.*, 2019, Restrepo *et al.*, 2004). This information is important for the development of crop protection strategies and breeding programs (Trujillo *et al.*, 2014). In this study, we generated double digested RAD-seq libraries of 81 isolates of *X. cucurbitae* collected throughout the Midwestern US and performed population genetics analyses to identify and characterize genetic clustering of the isolates.

MATERIALS AND METHODS

RNA extraction and preparation

RNA was extracted using E.Z.N.A. total RNA kit (Omega Bio-tek, Norcross, GA, USA) from bacteria grown on nutrient-sufficient peptone-sucrose (PS) or Hrp-inducing (XVM2) (Wengelnik, Marie *et al.* 1996) media for 24 hrs prior to harvesting. Ribosomal RNA was removed with the Ribo-Zero rRNA Removal Kit for Gram-Negative Bacteria and libraries were prepared using the TruSeq Stranded mRNA Library Prep (Illumina). Paired end 250 bp reads were sequenced using the MiSeq Nano V2 platform and yielded 985,288 reads.

Aligning RNA-seq reads

Read quality was checked using FastQC and adapters were trimmed using Cutadapt (version 2.0) (Martin 2011). We used an unpublished reference genome of *X. cucurbitae* strain ATCC 23378 for RNA-seq read alignments. A reference genome index of *X. cucurbitae* was generated using the alignment tool STAR (version 2.7) (Dobin, Davis *et al.* 2013) with the parameter of genomeSAindexNbases = 8. Reads were aligned to the genome index using options alignIntronMax 1 and outSAMtype BAM. featureCounts (version 1.6.4) (Liao, Smyth *et al.* 2013) was used to count mapped reads for genomic features on the *X. cucurbitae* reference genome.

Differential expression analysis

The R package EdgeR (Robinson, McCarthy *et al.* 2010) was used to perform differential expression analysis. Genes were considered differentially expressed if the absolute value of log-fold change was greater than 1.5, and the Benjamini-Hochberg adjusted *P*-value was less than 0.05 after EdgeR analysis. Functional and pathway analysis was carried out using the DAVID Functional Annotation platform (Huang, Sherman *et al.*, 2009, Huang, Sherman *et al.*, 2009). Figures were made using R packages ggplot2 (version 3.1.1) (Wickham 2009) and pheatmap (version 1.0.12).

***X. cucurbitae* RAD library preparation**

Based on the information provided by our collaborators (Liu, 2015), 81 bacterial isolates used in this analysis were collected from commercial pumpkin fields in the North Central Region (NCR) during 2012-2013. In addition, the reference strain *X. cucurbitae* ATCC 23378 was included in

this study. Bacterial isolates were cultured on Laurie Bertani agar (LB) in Petri plates and plates were incubated in dark at 24 ± 1°C for 48 hrs. One loop of bacterial cells from a colony on LB plate was transferred into a centrifuge tube and suspended in 1 ml of sterile distilled water (SDW). Bacterial DNA was extracted using FastDNA™ Spin Kits (MP Biomedicals, Solon, OH) according to the manufacturer's instruction for a double-digest RAD-seq analysis (Peterson *et al*., 2012). Restriction endonucleases MspI (New England BioLabs, #R0106T) and SacI (New England BioLabs Inc., Ipswich, MA) were used to digest DNA fragments. Digested DNA fragments for each isolate were ligated to adapter 1 and adapter 2 each with indexed barcodes (Table 2.1).

Barcoded DNA fragments for each individual isolate were PCR amplified, and size selection using gel electrophoresis was used to confirm that digestion and ligation were performed correctly. DNA fragments with lengths between 200 and 500 bp were excised from the agarose gel, and DNA was recovered using Wizard SV gel and PCR CleanUp system (Promega Corporation, Madison, WI) according to the manufacturer's instruction. The DNA libraries for the 82 *X. cucurbitae* isolates were pooled and submitted to the High-Throughput Sequencing and Genotyping Unit of the Biotechnology Center at the University of Illinois, Urbana-Champaign. Sequencing runs were performed on one flow cell lane for Illumina HiSeq2000 using 100 bp single-end reads.

Generating RAD loci

RAD-seq analysis was carried out according to the protocol established by Rochette and Catchen, 2017. Initial quality of the sequencing runs was checked using FASTQC (version 0.11.2) software. Sample reads were demultiplexed using the *process_radtags* tool in STACKS

(version 2.5) (Catchen *et al.*, 2013) using default parameters. Using *process_radtags*, low quality reads and reads missing either barcodes or RAD cut-sites were filtered out during demultiplexing. Filtered reads were then aligned to *X. cucurbitae* strain ATCC 23378 using default parameters in BWA-MEM (version 0.7.17) (Li *et al.*, 2009). Alignment files were assessed using the *flagstats* function as part of the SAMtools suite (version 1.1) (Li *et al.*, 2009). isolates Iowa 384 and Iowa 385 were removed from the study due to poor alignments with the reference genome; The 79 remaining *X. cucurbitae* strains collected throughout the Midwestern US, plus strain ATCC 23378, aligned more than 80% of reads to the reference genome. The remaining filtered reads of these 80 *X. cucurbitae* strains were assembled into loci and genotyped by the reference-based Stacks pipeline, *ref_map.pl.* 79913 SNP loci were seen from Stacks using the *populations* program. *Populations* was then run with the following parameters: --smooth, -R .5, --genepop, --phylip, --structure, --write-single-snp, --ordered-export, resulting in 3,548 SNP loci kept.

Population genetics analyses

To analyze phylogenetic and genetic structure of the *X. cucurbitae* populations, we used principal component analysis, maximum likelihood phylogenetic inference, and Bayesian clustering. Genetic clusters and cluster membership probabilities were computed using STRUCTURE (version 2.3.4) (Pritchard *et al.*, 2000). To estimate K, we searched from K=2 to K=15 with 10 replicates; 10,000 burnin reps and 5,000 after burnin reps was used for each replicate, with admixture allowed. An optimum K was computed using Structure Harvester (Earl and vonHoldt, 2012), and a Structure plot for K=3 were generated using Structure Plot (Ramamasy *et al.*, 2014). For Illinois substructure analysis, we searched from K=2 to K=7.

Cluster membership probability analysis was carried out identically to the whole population analysis. A maximum likelihood phylogenetic inference tree was built using concatenated SNP data derived from *populations.* IQTree (version 1.6.12) (Nguyen *et al.*, 2015) was used to infer an ML-tree using the parameters *auto* and *ultrafast bootstrap inferences* = 10000. Using *auto*, K2P+ASC was determined to be the best-fit model. The resulting consensus tree was visualized in Figtree (version 1.4.4). PCA was performed using the dudi.pca function within the R package *adegenet* (version 2.1.2) (Jombart, 2008) and visualized using R package *factoextra* (version 1.0.7).

RESULTS

RNA-seq analysis reveals genes important for infection

Differential gene expression was analyzed between cultures of the *X. cucurbitae* strain ATCC 23378 grown on nutrient-sufficient PS media or host-mimicking XVM media. We observed 441 genes that were differentially expressed between the two treatments; using the PS media treatment as a control, 296 genes were downregulated in XVM media, while 145 genes were upregulated (Figure 2.1A-B). We observed a variety of genes that were differentially expressed under host-mimicking conditions. Many genes found to be upregulated in host-mimicking media included virulence and pathogenicity factors (Figure 2.1C). We observed an increase in gene expression for type II secreted proteins such as extracellular proteases and cellulases as well as an exoglucanase (1,4-beta-cellobiosidase) encoded by the gene *cbhA* (Table 2.2). Type III secretion system genes and type III effector proteins were also more highly expressed in cultures grown in host-mimicking media (Table 2.3). In addition, we observed upregulation in genes for motility, chemotaxis, and some cell membrane transporters, all of which are cellular processes

important for pathogen virulence (Figure 2.1C). Genes related to sulfur metabolism were upregulated as well, likely due to the high sulfur concentration in XVM media (Table 2.4). In contrast, genes related to general stress response, as well as other cellular processes and metabolism genes, were downregulated under host-mimicking conditions.

The log-fold gene expression patterns observed in one replicate for the nutrient-sufficient control treatment varied considerably from the other two replicate samples, as well as compared to the host-mimicking samples (Figure 2.2). While FASTQC verified all sequencing data to be high quality, it is possible that there were other mitigating factors that impacted sample collection or processing, or the alignment and quantification of reads. Upon further investigation, we found an overrepresentation of ribosomal RNA reads in our abnormal control replicate compared to other samples in the study, indicating failure of complete rRNA depletion during library preparation. Nevertheless, results for differential gene expression were consistent between the other replicate samples for each treatment, and analysis using all samples resulted in identification of several hundred genes that were significantly altered in their gene expression between treatments.

DAVID analysis yielded gene enrichment groups similar to observed in Figure 2.1. This study uses the first reference quality *X. cucurbitae* genome, ATCC 23378, annotated using the NCBI Prokaryote Genome Annotation Pipeline. Since there was no published reference genome for *X. cucurbitae* at the time of analysis, the DAVID gene database was not able to map all differentially expressed genes towards enrichment groups. Regardless, the DAVID analysis was able to find significant enrichment groups for sulfur metabolism, sulfur assimilation, amino acid

synthesis, glycoside hydrolases and enzymes, and outer membrane receptor and transporters (Table 2.4).

Population structures and phylogenetic analysis of *X. cucurbitae*

Isolates of *X. cucurbitae* were taken from many commercial pumpkin and gourd farms throughout the Midwestern United States (Figure 2.3). For our study, an optimum K of 3 was determined for *X. cucurbitae* based on delta K and mean log probabilities (Table 2.5). Michigan, Wisconsin, and Iowa strains built a single population cluster, along with the reference strain from New York. Ohio and Indiana formed a separate cluster, and Illinois formed a population cluster with Kansas. Interestingly, we observed possible substructures as K increased: at K greater than or equal to 5, Kansas began to form its own distinct population cluster, while Illinois separated into multiple cluster groupings (Figure 2.4). This is likely due to the diversity of sampling locations within the state of Illinois (Figure 2.3). To follow up on this observation, further STRUCTURE analysis was performed on Illinois strains, separating individuals by sampling location. An optimum K of 2 was observed, with select isolates from Keenes, Illinois, clustered separately from the rest of Illinois strains (Table 2.6).

Maximum likelihood trees were built using RAD loci generated from *stacks.* An unrooted tree was generated to visualize the relationships between *X. cucurbitae* strains isolated from different sampling locations (Figure 2.5). Multiple soft polytomies were observed across the tree; more genetic data may be required to resolve these polytomies at a higher resolution. South Charleston and Elmwood were found to have identical alignments; therefore, we computed our tree with those populations as a single group. Regardless, two distinct clades were observed in our

phylogenetic tree, separating Iowa, Wisconsin, and Michigan strains from Illinois, Kansas, Ohio, and Indiana strains (Figure 2.5).

PCA clustering revealed similar trends as our other analyses based on RAD loci. *X. cucurbitae* strains generally clustered near individuals taken from the same sampling location (Figure 2.6). Illinois strains again were the most diverse grouping of *X. cucurbitae* strains, separating from other populations on both PC1 and PC2 axes. The separation of Iowa, Wisconsin, and Michigan strains from the rest of our individuals mirror results seen in our phylogenetic tree, showing clear genetic differences between *X. cucurbitae* strains.

DISCUSSION

Plant pathogens undergo many transcriptional changes during periods of invasion and colonization of host tissues. Behavioral changes in organisms can be understood at the molecular level by analyzing the transcriptome in response to new environmental factors; studying the transcriptome under different conditions can elucidate which genes are important to the infection process. In this study, we compared the transcriptome of *X. cucurbitae* under nutrient-sufficient and host-mimicking conditions to observe cellular changes during infection. We observed significant upregulation in predicted cell-wall degrading enzymes such as glucanases, cellulases, and proteases. These upregulated proteins show that, like other species of plant pathogenic *Xanthomonas*, *X. cucurbitae* uses a variety of enzymes secreted by the type II secretion system during infection (Ciancotto *et al.*, 2017). Whether or not these enzymes are necessary for pathogenicity or virulence remains to be tested, and future studies into these questions will allow

a greater understanding of the enzyme arsenal *X. cucurbitae* deploys during infection (Tayi *et al.*, 2016).

cbhA is one enzyme found to be significantly upregulated in *X. cucurbitae* when grown under host-mimicking conditions. *cbhA* encodes a 1,4-beta-cellobiosidase enzyme that is regularly observed in vascular species of *Xanthomonas* such as *X. oryzae* pv. *oryzae* and *X. campestris* pv. *campestris,* where it serves an important role for invading vascular tissues enriched in cellulose like xylem cells (An *et al., 2020)*. In addition, genome sequences from non-vascular species of *Xanthomonas* such as *X. oryzae* pv. *oryzicola* and *X. sacchari* have been observed to not encode *cbhA.* Hence, it is unusual for the nonvascular pathogen *X. cucurbitae* to both possess a *cbhA* gene and for us to observe upregulation of these gene during conditions mimicking host infection. Since other *Xanthomonas* species have been observed to break the boundaries between vascular and non-vascular methods of infection (Mensi *et al.*, 2014), it is possible that *X. cucurbitae* acquired this enzyme as an adaptation to infecting cucurbit hosts. Further studies are needed to ascertain the role of *cbhA* in *X. cucurbitae* pathogenicity on pumpkin plants, as well as tracing its evolutionary origin.

Many predicted type III effector proteins were found to be upregulated under host-mimicking conditions. These predicted effectors include *AvrBs2*, *HpaA,* and several *Xop* and *Hop* family effectors, which are well-characterized genes involved in plant pathogen virulence (Kearney and Staskawicz, 1990, Li *et al.*, 2015, Lorenz *et al.*, 2008. White *et al.*, 2009). Interestingly, the one TAL effector predicted in the *X. cucurbitae* genome, *tal1*, was not differentially expressed, with a log-fold change of -0.276. The type III effector repertoire present within species of

Xanthomonas has been observed to vary between species, pathovars. and even different strains, likely due to phenomena such as horizontal gene transfer and homologous recombination (Jibrin *et al.*, 2018). Since effector repertoires can widely vary between strains of *Xanthomonas*, studying type III effectors in genetically distinct isolates of *X. cucurbitae* could potentially reveal key virulence determinants among isolates and track the spread of important or novel virulence factors.

Evolutionary forces are important factors impacting the relationships between the host and the pathogen. With the rise of high-throughput sequencing, we can better understand the effects of these evolutionary forces and use *in silico*-derived insights to develop stronger crop protection strategies. In our analysis, we obtained 3,548 RAD loci using a stringent cutoff that required a locus to be found in at least 50% of individual isolates to be considered. Our data shows clear differences in genome content between geographically isolated strains of *X. cucurbitae,* with little genetic admixture overall. STRUCTURE plots reveal little overlap between individuals across different regions, and phylogenetic tree construction shows large genetic distances between two regional clusters. Gene flow in *X. cucurbitae* may be at a low rate due to its main vector of transmission, as rain splashes and wind dispersal may not be able to effectively disseminate *X. cucurbitae* far enough to mingle with geographically distant populations (McDonald and Linde, 2002). This may explain why there are clearly delineated populations of *X. cucurbitae* throughout the Midwestern US region. To further support that theory, geographically close isolates tended to cluster together. While pumpkin production is highest in Illinois, farms in other states are smaller scale, and mostly grown for ornamental, recreational, or cooking purposes. We observed clustering in both PCA space and the phylogenetic tree among

Iowa, Michigan, and Wisconsin isolates, all states that are in close geographic contact with each other. Similarities between Indiana and Ohio isolates were also observed. Kansas and Illinois isolates tended to cluster together, but we found that as predicted K increased, Kansas isolates trended to become their own population, while Illinois isolates broke down into two sub-populations. Overall, there seems to be little gene flow between strains at both the state level as well as within individual states. While our study was informative in assessing the diversity of *X. cucurbitae,* whole-genome sequencing may further resolve polytomies present in our dataset as well as provide a higher resolution to perform population genetic analyses.

CONCLUSIONS

Genomic analyses such as RNA-seq and RAD-seq are powerful tools in studying plant pathogens. Here, we uncovered transcriptional differences between *X. cucurbitae* cultures grown under nutrient-sufficient and host-mimicking conditions. We observed significant differences between these transcriptomic profiles, mainly involving genes related to bacterial virulence and pathogenicity. Also, we used genomic data from 80 total bacterial strains to generate population genetics analyses for the endemic cucurbit pathogen *X. cucurbitae.* We observed differences in genetic structure between different sampling locations and revealed the diversity of *X. cucurbitae* strains throughout the Midwestern US. Due to the decreasing costs of whole-genome sequencing, future studies could utilize entire genome sequences from representative isolates to resolve the phylogenetic relationships between *X. cucurbitae* isolates and to understand the complexity of the Midwestern US *X. cucurbitae* population at a deeper resolution. Overall, further genomic studies of *X. cucurbitae* strains would be helpful in determining the breadth of diversity within this pathogen and inform us regarding future crop protection strategies to combat this disease.

TABLES AND FIGURES

Table 2.1. *X. cucurbitae* isolates and corresponding barcode adaptors used for restriction associated DNA sequence analysis (RAD-seq). Isolates were collected from Jack O'Lantern (JOL), processing pumpkins (PP), or winter squash (S) cultivars. The isolates indicated in red (IA_3_384 and IA_4_385) were removed from the final analysis due to poor alignment of sequences to the reference genome.

Isolate	State	Field Location (City, County)	Year Collected	Plant	Adaptor 1 Sequence	Adaptor 2 Sequence
IA_1_380	IA	Wever, Lee	2013	JOL	GCGT	ACTTGA
IA_2_382	IA	Wever, Lee	2013	JOL	TGCGA	ATCACG
IA_3_384	IA	Wever, Lee	2013	JOL	TGCGA	CGATGT
IA_4_385	IA	Wever, Lee	2013	JOL	TGCGA	TTAGGC
IA_5_386	IA	Wever, Lee	2013	JOL	TGCGA	TGACCA
IA_6_3881	IA	Wever, Lee	2013	JOL	TGCGA	GCCAAT
IA_7_389	IA	Wever, Lee	2013	JOL	TGCGA	CAGATC
IA_8_390	IA	Wever, Lee	2013	JOL	TGCGA	ACTTGA
IL_21_3241	IL	Belleville, St. Clair	2013	JOL	AGTA	ACTTGA
IL_22_3251	IL	Belleville, St. Clair	2013	JOL	CAGA	ATCACG
IL_23_3751	IL	Elmwood, Peoria	2013	S	CAGA	CGATGT
IL_5_185	IL	Huntley, McHenry	2013	JOL	CTCC	GCCAAT
IL_6_222	IL	Keenes, Wayne	2013	JOL	CTCC	CAGATC
IL_7_225	IL	Keenes, Wayne	2013	JOL	CTCC	ACTTGA
IL_8_226	IL	Keenes, Wayne	2013	JOL	TGCC	ATCACG
IL_9_228	IL	Keenes, Wayne	2013	JOL	TGCC	CGATGT
IL_10_2301	IL	Keenes, Wayne	2013	JOL	TGCC	TTAGGC
IL_11_231	IL	Keenes, Wayne	2013	JOL	TGCC	TGACCA
IL_12_232	IL	Keenes, Wayne	2013	JOL	TGCC	GCCAAT
IL_13_2341	IL	Keenes, Wayne	2013	JOL	TGCC	CAGATC
IL_14_236	IL	Keenes, Wayne	2013	JOL	TGCC	ACTTGA
IL_15_2381	IL	Keenes, Wayne	2013	JOL	AGTA	ATCACG
IL_3_138	IL	Malta, DeKalb	2013	JOL	CTCC	TTAGGC
IL_4_139	IL	Malta, DeKalb	2013	JOL	CTCC	TGACCA
IL_16_307	IL	Pontoon Beach, Madison	2013	JOL	AGTA	CGATGT
IL_17_3081	IL	Pontoon Beach, Madison	2013	JOL	AGTA	TTAGGC

Table 2.1. (cont.)

Isolate	State	Field Location (City, County)	Year Collected	Plant	Adaptor 1 Sequence	Adaptor 2 Sequence
IL_18_3101	IL	Pontoon Beach, Madison	2013	JOL	AGTA	TGACCA
IL_19_318	IL	Pontoon Beach, Madison	2013	JOL	AGTA	GCCAAT
IL_20_3212	IL	Pontoon Beach, Madison	2013	JOL	AGTA	CAGATC
IL_1_1221	IL	Princeville, Peoria	2013	PP	CTCC	ATCACG
IL_2_123	IL	Princeville, Peoria	2013	PP	CTCC	CGATGT
IN_1_332	IN	Brookston, White	2013	JOL	CGCTT	ATCACG
IN_2_333	IN	Brookston, White	2013	JOL	CGCTT	CGATGT
IN_3_334	IN	Brookston, White	2013	JOL	CGCTT	TTAGGC
IN_4_336	IN	Brookston, White	2013	JOL	CGCTT	TGACCA
IN_5_337	IN	Brookston, White	2013	JOL	CGCTT	GCCAAT
IN_6_338	IN	Brookston, White	2013	JOL	CGCTT	CAGATC
IN_7_339	IN	Brookston, White	2013	JOL	CGCTT	ACTTGA
IN_8_340	IN	Brookston, White	2013	JOL	TCACC	ATCACG
IN_9_341	IN	Brookston, White	2013	JOL	TCACC	CGATGT
IN_10_3441	IN	Brookston, White	2013	JOL	TCACC	TTAGGC
IN_11_345	IN	Brookston, White	2013	JOL	TCACC	TGACCA
IN_12_346	IN	Brookston, White	2013	JOL	TCACC	GCCAAT
IN_13_347	IN	Brookston, White	2013	JOL	TCACC	CAGATC
IN_14_348	IN	Brookston, White	2013	JOL	TCACC	ACTTGA
IN_15_349	IN	Brookston, White	2013	JOL	CTAGC	ATCACG
KS_1_430	KS	Courtland, Republic	2013	S	CAGA	ACTTGA
KS_2_433	KS	Courtland, Republic	2013	S	AACT	ATCACG
KS_3_4551	KS	Norway, Republic	2013	JOL	AACT	CGATGT
KS_4_456	KS	Norway, Republic	2013	JOL	AACT	TTAGGC
KS_5_4581	KS	Norway, Republic	2013	JOL	AACT	TGACCA
KS_6_4591	KS	Norway, Republic	2013	JOL	AACT	GCCAAT
KS_7_460	KS	Norway, Republic	2013	JOL	AACT	CAGATC
KS_8_461	KS	Norway, Republic	2013	JOL	AACT	ACTTGA
KS_9_462	KS	Norway, Republic	2013	JOL	GCGT	ATCACG
KS_10_4631	KS	Norway, Republic	2013	JOL	GCGT	CGATGT
KS_11_466	KS	Norway, Republic	2013	JOL	GCGT	TTAGGC
KS_12_467	KS	Norway, Republic	2013	JOL	GCGT	TGACCA
KS_13_4681	KS	Norway, Republic	2013	JOL	GCGT	GCCAAT
KS_14_470	KS	Norway, Republic	2013	JOL	GCGT	CAGATC
MI_1_358	MI	Benton Harbor, Berrien	2013	JOL	CTAGC	CGATGT

Table 2.1. (cont.)

Isolate	State	Field Location (City, County)	Year Collected	Plant	Adaptor 1 Sequence	Adaptor 2 Sequence
MI_2_359	MI	Benton Harbor, Berrien	2013	JOL	CTAGC	TTAGGC
ATCC 23378	NY	Ithaca, Thompkins (Bryan 1930)	1926	S	CGAT	CAGATC
OH_1_2561	OH	South Charleston, Clark	2013	JOL	CAGA	TTAGGC
OH_2_261	OH	South Charleston, Clark	2013	JOL	CAGA	TGACCA
OH_3_2831	OH	South Charleston, Clark	2013	JOL	CAGA	GCCAAT
OH_4_2871	OH	South Charleston, Clark	2013	JOL	CAGA	CAGATC
WI_1_2001	WI	Janesville, Rock	2013	JOL	CTAGC	TGACCA
WI_2_201	WI	Janesville, Rock	2013	JOL	CTAGC	GCCAAT
WI_3_292	WI	Janesville, Rock	2013	JOL	CTAGC	CAGATC
WI_4_2031	WI	Janesville, Rock	2013	JOL	CTAGC	ACTTGA
WI_5_2051	WI	Janesville, Rock	2013	JOL	ACAAG	ATCACG
WI_6_206	WI	Janesville, Rock	2013	JOL	ACAAG	CGATGT
WI_7_2071	WI	Janesville, Rock	2013	JOL	ACAAG	TTAGGC
WI_8_2081	WI	Janesville, Rock	2013	JOL	ACAAG	TGACCA
WI_9_2091	WI	Janesville, Rock	2013	JOL	ACAAG	GCCAAT
WI_10_2101	WI	Janesville, Rock	2013	JOL	ACAAG	CAGATC
WI_11_2111	WI	Janesville, Rock	2013	JOL	ACAAG	ACTTGA
WI_12_2121	WI	Janesville, Rock	2013	JOL	CGAT	ATCACG
WI_13_2131	WI	Janesville, Rock	2013	JOL	CGAT	CGATGT
WI_14_2141	WI	Janesville, Rock	2013	JOL	CGAT	TTAGGC
WI_15_219	WI	Janesville, Rock	2013	JOL	CGAT	TGACCA

Table 2.2. Differentially expressed predicted type II cell wall-degrading enzymes (CWDE). Protein IDs and predicted CWDEs were generated by the NCBI Prokaryote Genome Annotation Pipeline. LogFC and *P*-values were generated by the differential gene expression R package EdgeR.

Protein ID	Predicted CWDE	logFC	*P*-Value
EBN15_16330	endo-1,4, β-D-glucanase	2.122322871	0.000084
EBN15_16380	endo-1,4, β-D-glucanase	3.460835471	0.000000656
EBN15_14660	1,4-beta-cellobiosidase	2.369366016	0.000744918
EBN15_10500	polygalacturonase	1.62114129	0.005630933
EBN15_11860	extracellular proteases	2.359483444	0.0000166

Table 2.3. Differentially expressed predicted type III effector proteins. Protein IDs and predicted effector proteins were generated by the NCBI Prokaryote Genome Annotation Pipeline. LogFC and *P*-values were generated by the differential gene expression R package EdgeR.

Protein ID	Predicted Effector	logFC	*P*-Value
EBN15_00540	AvrBs2	2.37024304	0.00029867
EBN15_16965	HpaA	2.556234889	0.00108774
EBN15_11365	XopE2	1.703475429	0.00200625
EBN15_13550	XopL	1.872431048	0.00035353
EBN15_17030	XopAD	2.143826874	0.0000808
EBN15_17060	HopAE	2.26402211	0.0000640

Table 2.4. Summary of DAVID Analysis for Differentially Expressed Genes (DEGs). Gene functional annotation terms are indicated as "Key Terms/Categories" and the number of *X. cucurbitae* ATCC 23378 genes that grouped into different clusters based on these terms are indicated. An enrichment score ≥ 1.3 indicates significantly overrepresented terms as compared to the whole-genome proportion of genes within each category.

Cluster	Key Terms/Categories	Number of Genes	Enrichment Score
1	Sulfur metabolism; sulfate assimilation	8	3.21
2	Amino acid biosynthesis	15	1.86
3	Glycoside hydrolase; starch and sugar metabolism	12	1.80
4	Membrane; receptor; transport	19	1.63

Table 2.5. Bayesian clustering results for *X. cucurbitae* isolates collected throughout the Midwestern US. The calculated optimal K is indicated in bold.

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	ΔK
2	10	-17099.35	511.1908	NA	NA	NA
3	**10**	**-11713.75**	**8.2243**	**5385.6**	**3585.47**	**435.963**
4	10	-9913.62	488.8794	1800.13	600	1.227297
5	10	-8713.49	555.7292	1200.13	1864.04	3.354223
6	10	-9377.4	1517.902	-663.91	19605.48	12.91617
7	10	-29646.79	64847.33	-20269.39	25450.56	0.392469
8	10	-24465.62	50998.11	5181.17	10608.94	0.208026
9	10	-8675.51	229.0476	15790.11	16028.3	69.97803
10	10	-8913.7	458.8687	-238.19	443.87	0.967314
11	10	-9595.76	1514.06	-682.06	553.42	0.365521
12	10	-9724.4	2088.304	-128.64	11184.96	5.356001
13	10	-21038	39540.1	-11313.6	4972.21	0.125751
14	10	-27379.39	59253.28	-6341.39	13418.11	0.226453
15	10	-47138.89	81645.24	-19759.5	NA	NA

Table 2.6. Bayesian clustering results for *X. cucurbitae* isolates sampled within Illinois. Calculated optimal K is in bold.

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln"(K)|	ΔK
1	10	-7634.63	4.6819	NA	NA	NA
2	**10**	**-5911.26**	**2.806**	**1723.37**	**602.82**	**214.8306**
3	10	-4790.71	4.094	1120.55	535.64	130.8347
4	10	-4205.8	7.9846	584.91	332.81	41.68165
5	10	-3953.7	71.0455	252.1	2175.87	30.62645
6	10	-5877.47	3498.045	-1923.77	3322.3	0.949759
7	10	-4478.94	1663.46	1398.53	NA	NA

Figure 2.1. Transcriptional profile of *X. cucurbitae* 23378 identifies genes differentially expressed (DEGs) under host-mimicking conditions compared to nutrient-sufficient conditions. **(A)** Volcano plot showing gene expression vs. significance on the x and y axes, respectively. Vertical dashed lines indicate a log-fold change ≥ 1.5 or ≤ -1.5, and above the horizontal dashed line indicate a *P*-value < 0.05. Pink and turquoise represent genes with upregulated and downregulated expression, respectively. **(B)** Total numbers of upregulated and downregulated differentially expressed genes (DEGs). **(C)** DEGs categorized by predicted function. The number of genes grouped within each category is indicated.

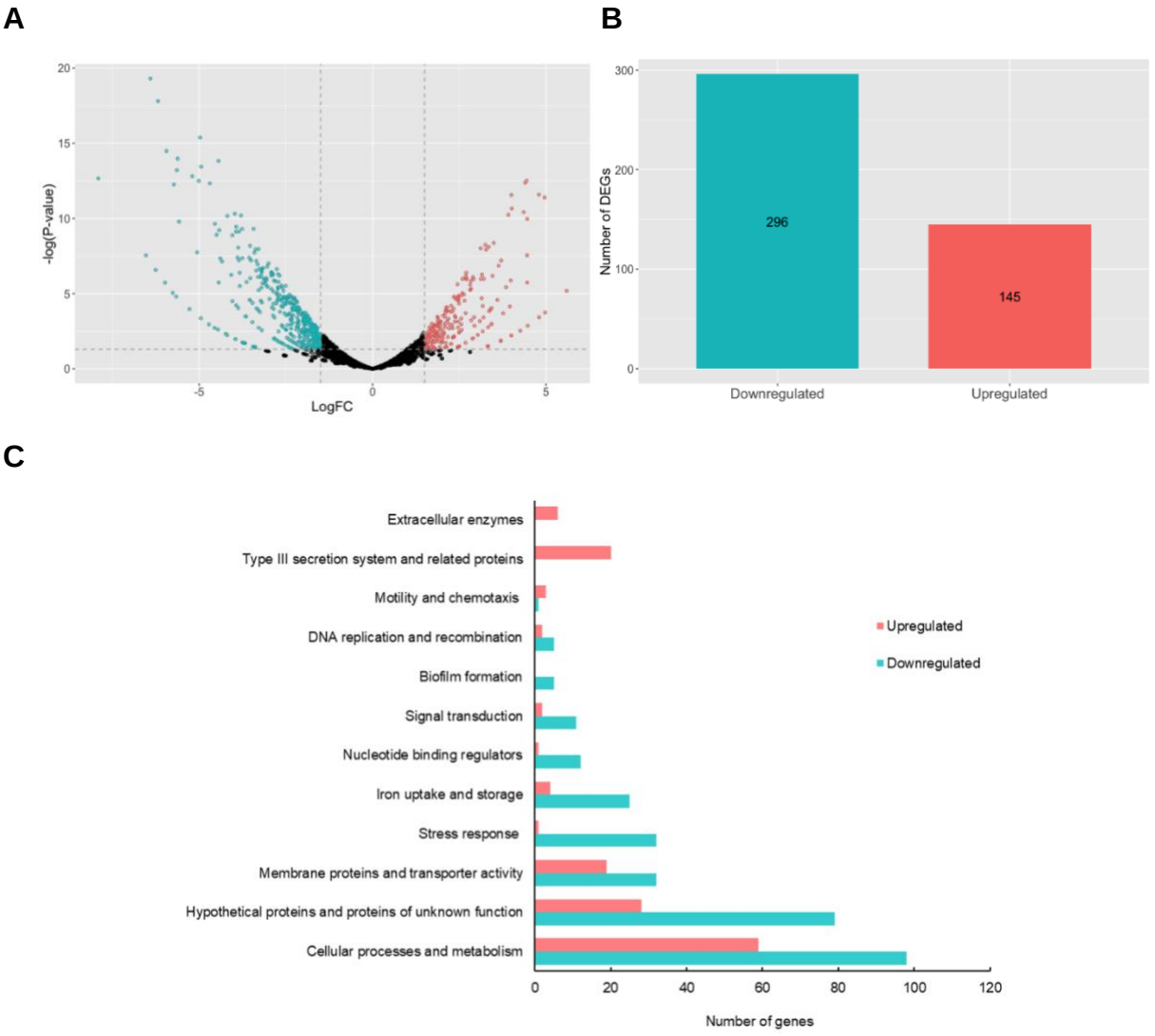


Figure 2.2. Heatmap showing differences in gene expression between *X. cucurbitae* cultures grown in nutrient-sufficient (PS) or host-mimicking (XVM) conditions. Scaled counts of reads to genomic features derived from FeatureCounts are shown. Euclidian distance clusters on rows and columns were generated by pheatmap. Legend indicates whether each bacterial replicate was grown on nutrient-sufficient (green) or host-mimicking (pink) media.

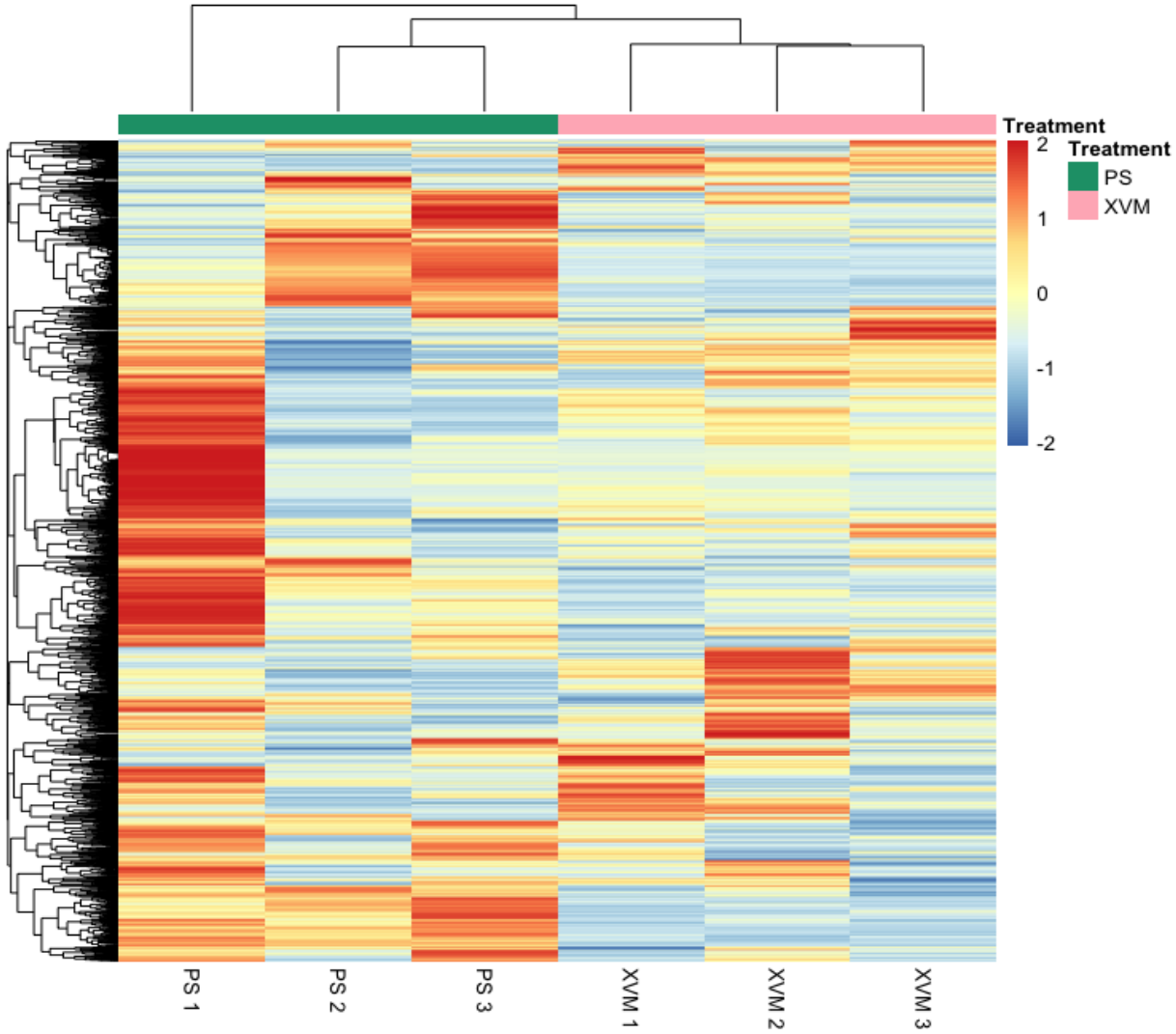


Figure 2.3. Map of sampling sites where *X. cucurbitae* samples were isolated. Data provided by Google Maps (n.d.).

Figure 2.4. STRUCTURE clustering at **(A)** K = 3, **(B)** K= 4, and **(C)** K = 5 for *X. cucurbitae* isolates collected from the Midwestern US. Labeled boxes underneath each plot signifies grouping by the state in which each isolate was sampled.

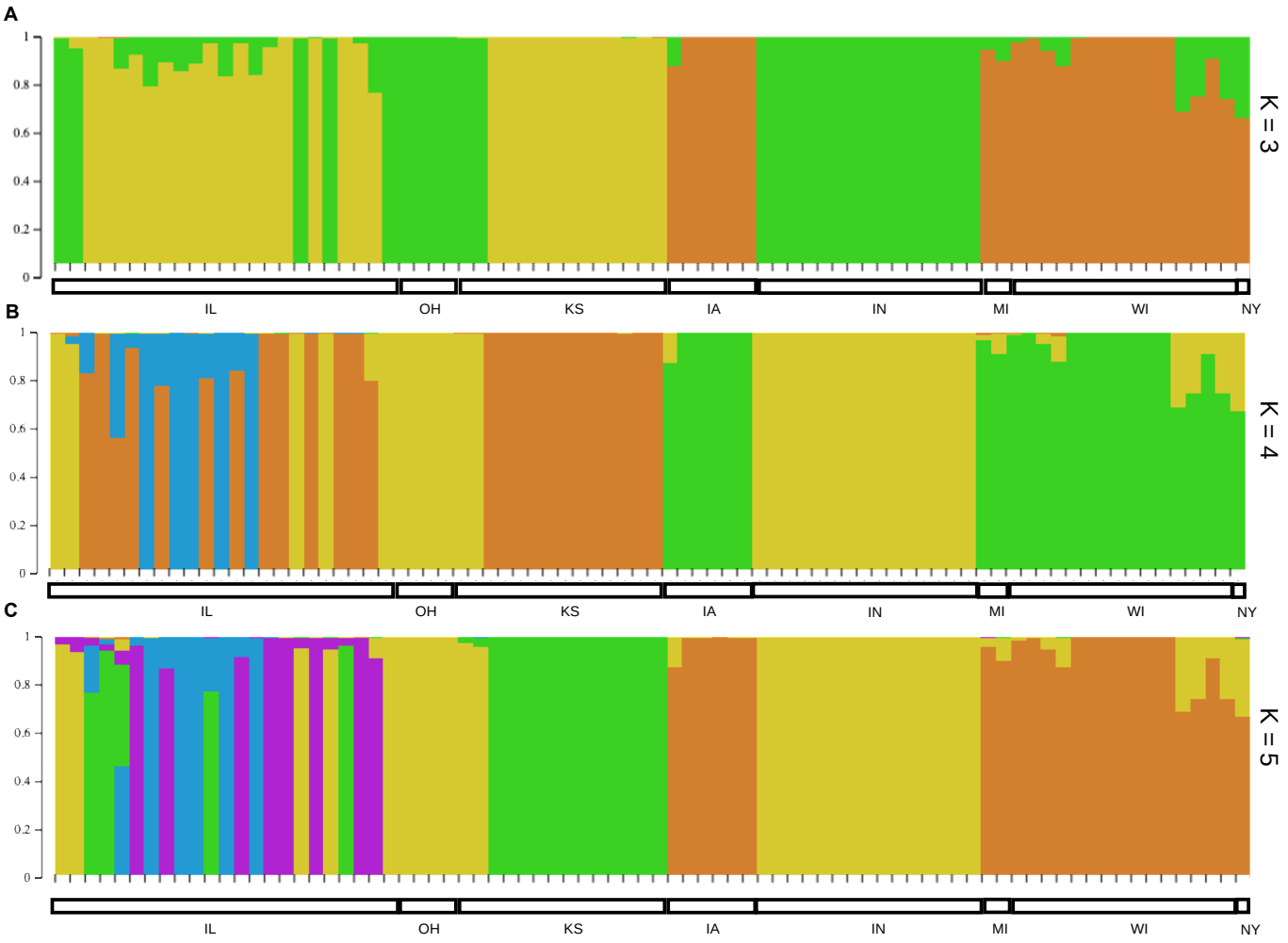


Figure 2.5. Unrooted maximum likelihood tree for *X. cucurbitae* isolates from the Midwest. Branch labels show bootstrap support for 10,000 ultrafast bootstrap replicates from IQTREE; only bootstrap values ≥ 60 are shown. Each tip label signifies sampling site location and the number of isolates taken from each sampling site. Colors represent different states while the scale bar indicates maximum phylogeny pairwise distances. South Charleston, Ohio, and Elmwood, Illinois, were found to have identical concatenated SNP sequences

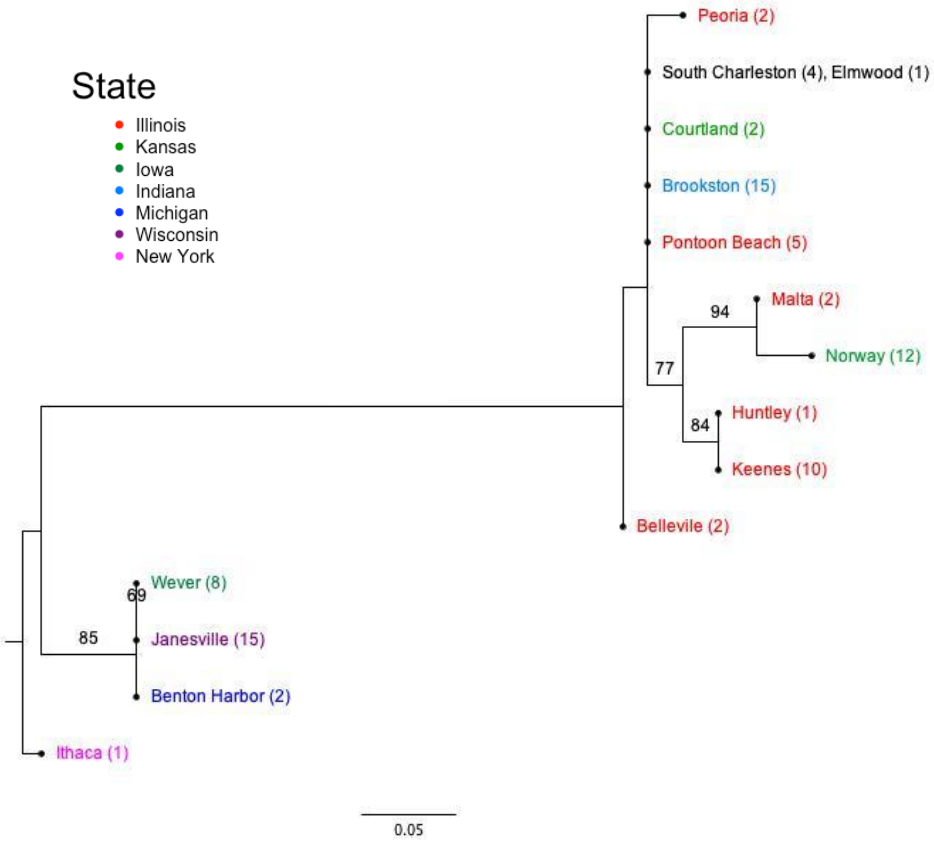
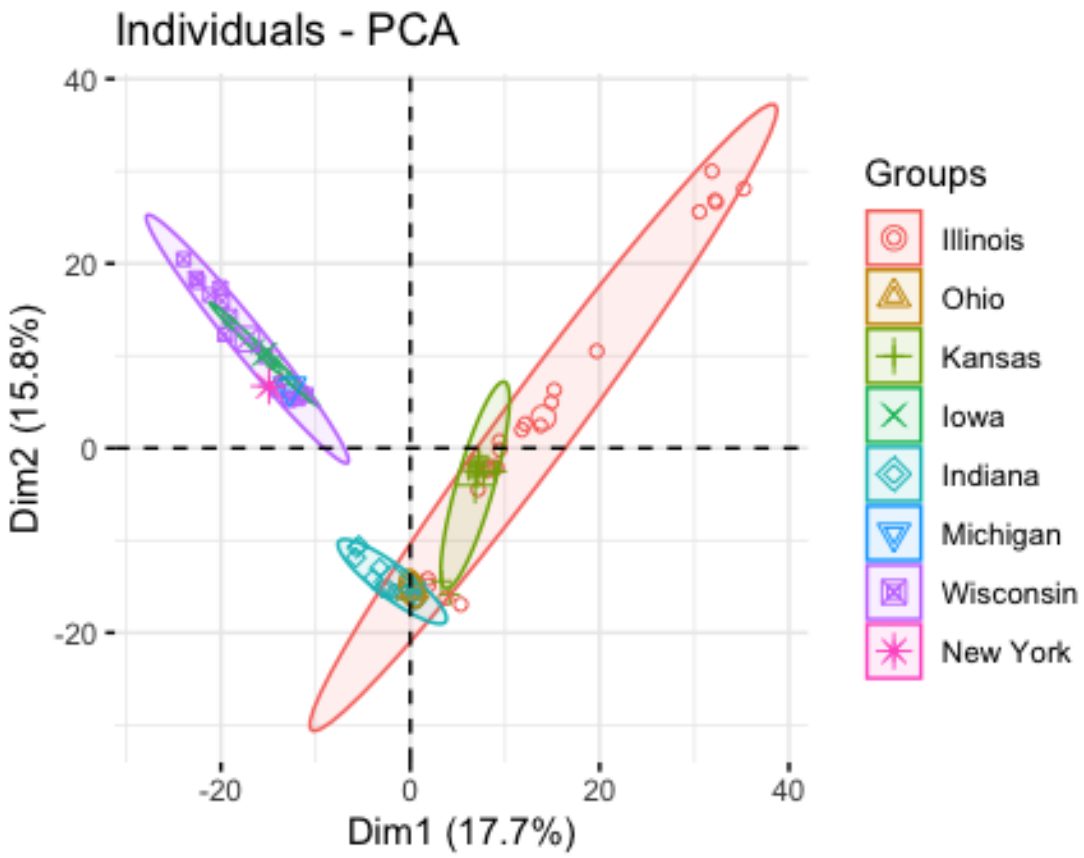


Figure 2.6. Principal components analysis (PCA) visualization of *X. cucurbitae* strains isolated throughout the Midwestern US. Each color represents a different state. Results of the first two PCs are shown.



LITERATURE CITED

Abbasali, R. and B. Mohammad (2015). "Development of Bacterial Spot, Incited by *Xanthomonas cucurbitae*, in Pumpkin Fields." HortScience **50**(5): 714-720.

An, S.-Q., N. Potnis, M. Dow, F.-J. Vorhölter, Y.-Q. He, A. Becker, D. Teper, Y. Li, N. Wang, L. Bleris and J.-L. Tang (2019). "Mechanistic insights into host adaptation, virulence and epidemiology of the phytopathogen *Xanthomonas*." FEMS Microbiology Reviews **44**(1): 1-32.

Babadoost, M. and A. Ravanlou (2012). "Outbreak of Bacterial Spot (*Xanthomonas cucurbitae*) in Pumpkin Fields in Illinois." Plant Disease **96**(8): 1222-1222.

Bryan, M. K. (1926). "Bacterial Leafspot on Hubbard Squash." Science **63**(1623): 165.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores and W. A. Cresko (2013). "Stacks: an analysis tool set for population genomics." Molecular Ecology **22**(11): 3124-3140.

Chien, C.-C., M.-Y. Chou, C.-Y. Chen and M.-C. Shih (2019). "Analysis of genetic diversity of *Xanthomonas oryzae* pv. *oryzae* populations in Taiwan." Scientific Reports **9**(1): 316.

Cianciotto, N. P. and R. C. White (2017). "Expanding Role of Type II Secretion in Bacterial Pathogenesis and Beyond." Infection and Immunity **85**(5): e00014-00017.

Costa, T. R. D., C. Felisberto-Rodrigues, A. Meir, M. S. Prevost, A. Redzej, M. Trokter and G. Waksman (2015). "Secretion systems in Gram-negative bacteria: structural and mechanistic insights." Nature Reviews Microbiology **13**(6): 343-359.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson and T. R. Gingeras (2013). "STAR: ultrafast universal RNA-seq aligner." Bioinformatics (Oxford, England) **29**(1): 15-21.

Earl, D. A. and B. M. vonHoldt (2012). "STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method." Conservation Genetics Resources **4**(2): 359-361.

Escalon, A., S. Javegny, C. Vernière, L. D. Noël, K. Vital, S. Poussier, A. Hajri, T. Boureau, O. Pruvost, M. Arlat and L. Gagnevin (2013). "Variations in type III effector repertoires, pathological phenotypes and host range of *Xanthomonas citri* pv. *citri* pathotypes." Molecular Plant Pathology **14**(5): 483-496.

Google (n.d). “Sampling sites for *X. Cucurbitae* in the Midwestern US.” Retrieved 2020 June 30 from https://www.google.com/maps/d/edit?mid=1AM9z3kiQY4AzNYhEaU5VYJLxKD7m5vDV&usp=sharing

Huang, D. W., B. T. Sherman and R. A. Lempicki (2008). "Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists." Nucleic Acids Research **37**(1): 1-13.

Huang, D. W., B. T. Sherman and R. A. Lempicki (2009). "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources." Nature Protocols **4**(1): 44-57.

Jibrin, M. O., N. Potnis, S. Timilsina, G. V. Minsavage, G. E. Vallad, P. D. Roberts, J. B. Jones and E. M. Goss (2018). "Genomic Inference of Recombination-Mediated Evolution in *Xanthomonas euvesicatoria* and *X. perforans*." Applied and Environmental Microbiology **84**(13): e00136-00118.

Jombart, T. (2008). "adegenet: a R package for the multivariate analysis of genetic markers." Bioinformatics **24**(11): 1403-1405.

Kearney, B. and B. J. Staskawicz (1990). "Widespread distribution and fitness contribution of *Xanthomonas campestris* avirulence gene *avrBs2*." Nature **346**(6282): 385-386.

Kim, S., Y.-J. Cho, E.-S. Song, S. H. Lee, J.-G. Kim and L.-W. Kang (2016). "Time-resolved pathogenic gene expression analysis of the plant pathogen *Xanthomonas oryzae* pv. *oryzae*." BMC Genomics **17**(1): 345.

Kogenaru, S., Q. Yan, Y. Guo and N. Wang (2012). "RNA-seq and microarray complement each other in transcriptome profiling." BMC Genomics **13**(1): 629.

Li, H. and R. Durbin (2009). "Fast and accurate short read alignment with Burrows-Wheeler transform." Bioinformatics **25**(14): 1754-1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin and S. Genome Project Data Processing (2009). "The Sequence Alignment/Map format and SAMtools." Bioinformatics **25**(16): 2078-2079.

Li, S., Y. Wang, S. Wang, A. Fang, J. Wang, L. Liu, K. Zhang, Y. Mao and W. Sun (2015). "The Type III Effector AvrBs2 in *Xanthomonas oryzae* pv. *oryzicola* Suppresses Rice Immunity and Promotes Disease Development." Molecular Plant-Microbe Interactions **28**(8): 869-880.

Liao, Y., G. K. Smyth and W. Shi (2014). "featureCounts: an efficient general purpose program for assigning sequence reads to genomic features." Bioinformatics (Oxford, England) **30**(7): 923-930.

Linde, C. C. (2010). "Population genetic analyses of plant pathogens: new challenges and opportunities." Australasian Plant Pathology **39**(1): 23-28.

Liu, Q., A. Ravanlou and M. Babadoost (2016). "Occurrence of Bacterial Spot on Pumpkin and Squash Fruit in the North Central Region of the United States and Bacteria Associated with the Spots." Plant Disease **100**(12): 2377-2382.
Liu, Q. (2015). "Determining role of bacteria associated with bacterial spot on pumpkin and squash fruit in the North Central Region of the United States." Retrieved 2020 May 5 from http://hdl.handle.net/2142/78674

Lorenz, C., O. Kirchner, M. Egler, J. Stuttmann, U. Bonas and D. Büttner (2008). "HpaA from *Xanthomonas* is a regulator of type III secretion." Molecular Microbiology **69**(2): 344-360.

Martin, M. (2011). "Cutadapt removes adapter sequences from high-throughput sequencing reads." EMBnet.journal; Vol 17, No 1: Next Generation Sequencing Data Analysis.

McDonald, B. A. and C. Linde (2002). "Pathogen Population Genetics, Evolutionary Potential, and Durable Resistance." Annual Review of Phytopathology **40**(1): 349-379.

Mensi, I., M.-S. Vernerey, D. Gargani, M. Nicole and P. Rott (2014). "Breaking dogmas: the plant vascular pathogen *Xanthomonas albilineans* is able to invade non-vascular tissues despite its reduced genome." Open Biology **4**(2): 130116.

Nguyen, L.-T., H. A. Schmidt, A. von Haeseler and B. Q. Minh (2015). "IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies." Molecular Biology and Evolution **32**(1): 268-274.

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher and H. E. Hoekstra (2012). "Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species." PLOS ONE **7**(5): e37135.

Potnis, N., S. Timilsina, A. Strayer, D. Shantharaj, J. D. Barak, M. L. Paret, G. E. Vallad and J. B. Jones (2015). "Bacterial spot of tomato and pepper: diverse *Xanthomonas* species with a wide variety of virulence factors posing a worldwide challenge." Molecular Plant Pathology **16**(9): 907-920.

Pritchard, J. K., M. Stephens and P. Donnelly (2000). "Inference of population structure using multilocus genotype data." Genetics **155**(2): 945-959.

Ramasamy, R. K., S. Ramasamy, B. B. Bindroo and V. G. Naik (2014). "STRUCTURE PLOT: a program for drawing elegant STRUCTURE bar plots in user friendly interface." SpringerPlus **3**: 431-431.

Restrepo, S., C. M. Velez, M. C. Duque and V. Verdier (2004). "Genetic Structure and Population Dynamics of *Xanthomonas axonopodis* pv. *manihotis* in Colombia from 1995 to 1999." Applied and Environmental Microbiology **70**(1): 255.

Robinson, M. D., D. J. McCarthy and G. K. Smyth (2010). "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data." Bioinformatics (Oxford, England) **26**(1): 139-140.

Rochette, N. C. and J. M. Catchen (2017). "Deriving genotypes from RAD-seq short-read data using Stacks." Nature Protocols **12**(12): 2640-2659.

Socquet-Juglard, D., T. Kamber, J. F. Pothier, D. Christen, C. Gessler, B. Duffy and A. Patocchi (2013). "Comparative RNA-Seq Analysis of Early-Infected Peach Leaves by the Invasive Phytopathogen *Xanthomonas arboricola* pv. *pruni*." PLOS ONE **8**(1): e54196.

Tayi, L., R. V. Maku, H. K. Patel and R. V. Sonti (2016). "Identification of Pectin Degrading Enzymes Secreted by *Xanthomonas oryzae* pv. *oryzae* and Determination of Their Role in Virulence on Rice." PLOS ONE **11**(12): e0166396.

Trujillo, C. A., J. C. Ochoa, M. F. Mideros, S. Restrepo, C. López and A. Bernal (2014). "A Complex Population Structure of the Cassava Pathogen *Xanthomonas axonopodis* pv. *manihotis* in Recent Years in the Caribbean Region of Colombia." Microbial Ecology **68**(1): 155-167.

USDA-ERS (2019). "Pumpkins: Background & Statistics." Retrieved 2020 May 5 from https://www.ers.usda.gov/newsroom/trending-topics/pumpkins-background-statistics/.

Wengelnik, K., C. Marie, M. Russel and U. Bonas (1996). "Expression and localization of HrpA1, a protein of *Xanthomonas campestris* pv. *vesicatoria* essential for pathogenicity and induction of the hypersensitive reaction." Journal of Bacteriology **178**(4): 1061-1069.

White, F. F., N. Potnis, J. B. Jones and R. Koebnik (2009). "The type III effectors of *Xanthomonas*." Molecular Plant Pathology **10**(6): 749-766.

Zhang, M. and G. Coaker (2017). "Harnessing Effector-Triggered Immunity for Durable Disease Resistance." Phytopathology **107**(8): 912-919.

Zhang, X. and M. Babadoost (2018). "Characteristics of *Xanthomonas cucurbitae* Isolates from Pumpkins and Survival of the Bacterium in Pumpkin Seeds." Plant Disease **102**(9): 1779-1784.

CHAPTER 3

UTILIZING TAJIMA'S D TO DISCOVER NEW MICROBE-ASSOCIATED MOLECULAR PATTERNS IN *XANTHOMONAS*

ABSTRACT

The coevolution between plants and pathogens constitute a significant proportion of evolutionary history for both host and parasite. The decreasing costs of whole-genome sequencing and the proliferation of publicly available sequencing data has led to the creation of large collections of genome sequences, and computational methods can be employed to look for genetic signatures of evolution. In this study, we used the population genetics test statistic Tajima's D to identify potential genes involved in plant-pathogen interactions between tomato and *Xanthomonas euvesicatoria* and *X. perforans*, the causal agents of bacterial spot disease of tomato. We assessed the usefulness of this method and experimentally confirmed alleles of known microbe-associated molecular patterns (MAMPs). While we were not yet able to identify any new MAMPs, we demonstrated the utility of population genetic statistics such as Tajima's D for the identification of genes that may be important for pathogen recognition by the host immune machinery.

INTRODUCTION

Bacterial spot of tomato (*Solanum lycopersicum*) and pepper (*Capsicum annum*) is a major disease caused by four species in the bacterial genus *Xanthomonas*, including *X. euvesicatoria*, *X. perforans*, *X. vesicatoria*, and *X. gardneri* (Jones *et al.*, 2004). Bacterial spot can lead to major crop losses in warm, humid regions and is characterized by dark lesions on the fruit as well as necrotic areas on stems and leaves (Potnis *et al.*, 2015). It can spread across a field through wind

and water droplets and often can be transferred via contaminated farm equipment (McInnes *et al.*, 1988). Pathogenic Xanthomonads can also survive on crop debris and residues for long periods of time, making bacterial spot difficult to fully control (Jones *et al.*, 1986). Farmers utilize common cultural practices, including crop rotations, seed treatments, and equipment decontaminations, to limit the spread of the disease (Ritchie *et al.*, 2000). Additionally, copper-based chemical sprays have been widely used to control bacterial spot (Behlau *et al.*, 2011); however, they are most effective prior to establishment of the disease. Unfortunately, the rise of copper-resistant and copper-tolerant Xanthomonads have reduced the effectiveness of copper sprays, prompting research into other avenues of disease control (Behlau *et al.*, 2011).

Pattern-triggered immunity (PTI) is the first set of immune defense response in plants and can stave off infection from many plant pathogens. Cell-surface pattern recognition receptors (PRRs) recognize conserved microbial features known as microbe-associated molecular patterns (MAMPs) (Zipfel, 2014). When these PRRs recognize a cognate MAMP, the signal transduction pathway for pattern-triggered immunity is activated, with the recognition of MAMPs triggering intracellular signaling that includes calcium bursts, kinase phosphorylation cascades, and production and release of reactive oxygen species (ROS), ultimately leading to the expression of primary immune response genes (Yu *et al*, 2017). While PTI responses can protect plants from a broad range of diseases, many factors influence their effectiveness. Variations in either MAMPs or PRRs can reduce or prevent the recognition of plant pathogens (Bhattarai *et al.*, 2016, Vetter *et al.*, 2016), and some pathogens attempt to bypass the PTI system by deploying effector proteins that suppress PTI response (Spoel *et al.*, 2012).

Because PTI is an important mechanism for disease resistance, characterizing new PRR and MAMP relationships is a promising avenue for durable disease resistance. In controlled laboratory conditions, it has been observed that transferring the PRR EFR from *Arabidopsis thaliana* to susceptible *Nicotiana benthamiana* and *Solanum lycopersicum* can confer broad PTI resistance against certain pathogenic bacteria (Lacombe *et al.*, 2010). Similarly, immune system defenses can be primed when plants encounter disease resistance inducers (Thakur *et al.*, 2013). While the discovery rate of new MAMPs has increased in recent years, much of the research on PTI interactions focuses on well-characterized MAMPs and PRR receptors, such as flg22 and FLS2 (Felix *et al.*, 2004) and EF-Tu and EFR (Zipfel *et al.*, 2006). The flg22 region of flagellin is widely recognized in the plant kingdom by the cognate plant recognition receptor FLS2 (Gómez-Gómez and Boller, 2000, Zipfel *et al.*, 2004). In addition, the flgII-28 region of flagellin is recognized by the FLS3 receptor present in many solanaceous plants such as potato, tomato, and pepper (Hind *et al.*, 2016).

Screening for non-neutrally evolving genes products has been successfully used to identify new MAMPs in plant pathogens (Mott *et al.*, 2016, McCann *et al*, 2012). Because MAMPs are usually associated with the fitness of the pathogen, beneficial MAMP mutations are hypothesized to occur at low frequencies; these mutations allow the pathogen to evade PRR detection, while not impacting pathogen fitness. The statistic Tajima's D (Tajima, 1989) has been used to identify MAMPs in the plant pathogen *Ralstonia solanocearum* (Eckshtain-Levi *et al.*, 2018). In population genetics, purifying selection indicates the removal of deleterious alleles in a population, and subsequently the fixation of a single allele. Balancing selection indicates the maintenance of multiple alleles in a population. Since the Tajima's D statistic screens for both

purifying and balancing selection, it is a powerful tool in identifying potential MAMPs. We utilized a comparative genomics approach using Tajima's D to identify novel MAMPs in *X. euvesicatoria* and *X. perforans*, and tested peptide sequences related to those proteins using oxidative burst assays to confirm their ability to induce an immune response in tomato leaves.

MATERIALS AND METHODS

Core genome building

X. euvesicatoria and *X. perforans* were chosen for this study due to their close genetic relatedness. Genomic information for 41 *X. euvesicatoria* and 44 *X. perforans* strains was accessed and downloaded from NCBI (summarized in Table 3.1). All *Xanthomonas* genome assemblies present in the study were uploaded to the Bacterial Isolate Genome Sequence Database (BIGSdb; https://pubmlst.org/software/database/bigsdb/) (Jolley and Maiden, 2010), and each *Xanthomonas* species as separately analyzed by the GenomeComparator tool. For this analysis, the core genome was defined as genes present in 90% of strains within each species. *X. euvesicatoria* strain 85-10 and *X. perforans* strain 91-118 were used as reference genomes in GenomeComparator for *X. euvesicatoria* and *X. perforans* strains, respectively. Coding sequences from each assembly were extracted and used for comparison against sequences from other isolates. Pairwise all-vs-all blastn searches against other genomes was used to determine orthologs, which were defined as having a minimum 70% percent identity and 50% alignment coverage, and word-size of 20.

Tajima's D and GO term analysis

Core genome orthologs were aligned for each of the two *Xanthomonas* species using MAFFT. Resulting alignment files were imported into R and Tajima's D values were calculated using the R package 'pegas'. Benjamini-Hochberg adjusted *P*-values were calculated with a false discovery rate of 0.05 (Benjamini and Hochberg, 1995). The RefSeq Protein identifiers for each core genome sequence were extracted from GenomeComparator and translated into UniProt identifiers using the UniProt Retrive/ID mapping tool. Subsequent UniProt identifiers were submitted to the Web Gene Ontology Annotation Plot (WEGO; http://wego.genomics.org.cn/) (Ye *et al.*, 2018) tool to visualize GO term annotations of non-neutrally evolving genes.

Peptides

Predicted MAMP residues of OmpW were designed from the OmpW protein sequence of *X. perforans* 91-118 (Shwartz *et al., 2015*). Peptide sequences were synthesized (Genscript), with all peptides having greater than 80% purity. Peptide preparations for OmpW1-Allele 1, OmpW2-Allele 1, and OmpW2-Allele 2 were dissolved in water, while OmpW1-Allele 2 was initially solved in a small volume (i.e., 50 µl) of 1 M HCl prior to addition of water. Additional peptides included the consensus flg22 (Genscript), flg22 from *X. euvesicatoria* strain 85-10, flgII-28 from *Pseudomonas syringae* pv. *tomato* strain T1 (Cai *et al.*, 2011), consensus flgII-28 from *Xanthomonas* species, and flgII-28 from *X. euvesicatoria* strain 85-10.

Plant conditions and luminol based ROS assay

Six-week-old tomato plants (*Solanum lycopersicum* pv. Moneymaker) were grown in a greenhouse with daily cycles of 16 hours light/8 hours dark at 25°C day/21°C night

temperatures. Leaf discs were punched using a #2 cork borer and placed into wells of a 96-well luminometer plate containing 200 μL distilled water, then incubated at room temperature for 16 hours in the dark. After water removal, 100 μl of a solution containing 34 μg/ml luminol, 20 μg/ml horseradish peroxidase, and peptide solutions at the concentrations indicated, was placed into each well. The plate was placed into a Molecular Devices Filtermax F5, where the Relative Light Units (RLU) for each well was measured every three minutes for a total of 16 measurements over 45 minutes.

RESULTS

Publicly available *X. euvesicatoria* and *X. perforans* genomes with less than 200 scaffolds were chosen for core genome analysis. 41 *X. euvesicatoria* strains and 44 *X. perforans* strains were analyzed for each respective species (Table 3.1). Using the BIGSdb GenomeComparator tool, the core genome of *X. euvesicatoria* included 4,361 genes, while the *X. perforans* core genome contained 4,217 genes.

Orthologs identified using GenomeComparator were aligned and the gene alignments were extracted to calculate the Tajima's D values for each gene in the core genomes. Although many genes in *X. euvesicatoria* and *X. perforans* were found to be non-neutrally evolving, it is possible that many of these significant genes could be false positives. Therefore, Benjamini-Hochberg adjusted *P*-values were calculated to control for false discovery rate. *P*-values for Tajima's D were calculated according to a beta distribution, and a Benjamini-Hochberg correction was applied to control for false positives.

In general, a Tajima's D value less than -2 suggests significant purifying selection, and a value larger than 2 indicates significant balancing selection. The average Tajima's D value for *X. euvesicatoria* was -2.15 with a standard deviation of 0.61 (Figure 3.1A). Similarly, the average Tajima's D value for *X. perforans* was -1.58 with a standard deviation of 1.45 (Figure 3.1B). After adjusting the *P*-values, a single gene was found to have a significantly positive Tajima's D value in *X. euvesicatoria*, while 2,731 genes had significantly negative Tajima's D values. In *X. perforans*, 98 genes were found to have significantly positive Tajima's D values, while 521 genes had significantly negative Tajima's D values. Both core genomes for *X. euvesicatoria* and *X. perforans* were found to be trending towards purifying selection.

To observe trends in the types of genes that were identified to be evolving non-neutrally, we extracted GO Terms from the list of significant *X. perforans* genes and grouped them according to their Cellular Component and Molecular Function domains (Figure 3.2). Interestingly, most significant genes had Gene Ontology (GO) Terms related to the structure of the cell, the cell membrane, or cellular components. In addition, many genes with significant Tajima's D values are known to be involved in virulence and pathogenicity, such as secretion systems machinery, TonB-dependent receptors, and secreted proteases (Blanvillain *et al.*, 2007, Figaj *et al.*, 2019). Thus, use of Tajima's D statistic seems to be an effective tool in identifying genes involved in plant-pathogen interactions.

Bacterial flagellin is a highly abundant protein that functions as a MAMP in plants through perception of the peptides flg22 and flgII-28 (Gómez-Gómez and Boller, 2000, Cai *et al.*, 2011). In *X. euvesicatoria*, the Tajima's D value for flagellin was significantly negative at -3.47, while

X. perforans flagellin had a negative but non-significant Tajima's D value of -2.46. Looking at allele frequencies within our *Xanthomonas* populations, 39 out of 41 *X. evuvesicatoria* and all *X. perforans* strains shared identical flg22 peptide sequences (Table 3.2). However, the *X. euvesicatoria* flgII-28 allele shared by the majority of our *X. euvesicatoria* strains differed from the consensus allele present in our *X. perforans* population and generally present in many *Xanthomonas* species (designated 'flgII-28 *X*. spp'). 39 out of 41 of our *X. euvesicatoria* contained a lysine (K) at the 18th amino acid position, while 42 out of 44 of our *X. perforans* strains contained an asparagine (N) (Table 3.3).

To evaluate the accuracy of the Tajima's D statistic and to test for differences in MAMP perception between flgII-28 alleles, we used a luminol-based oxidative burst assay to measure rates of ROS production, which is a well-known component of the plant PTI response. Peptide sequences of flg22 and flgII-28 from *X. euvesicatoria* strain 85-10, the *Xanthomonas* consensus sequence of flgII-28, and control peptides for flg22 and flgII-28 were assayed for PTI elicitation (Figure 3.3). The flg22 allele present in *X. euvesicatoria* and *X. perforans* strains did not elicit a PTI response in *S. lycopersicum* cultivar Moneymaker leaves, compared to the flg22 consensus peptide (Figure 3.3A and C). The flgII-28 peptide from the *Xanthomonas* consensus sequence (flgII-28 *X*. spp) elicited an oxidative burst response in tomato plants similar to that elicited by the flgII-28 control peptide from *Pseudomonas syringae* pv. *tomato* T1; however, the flgII-28 peptide corresponding to the allele present in the majority of *X. euvesicatoria* strains did not cause an oxidative burst response (Figure 3.3B and C).

Non-neutrally evolving genes in *X. perforans* were analyzed to identify possible MAMP candidates, specifically those with significantly positive values, indicating that these genes may be undergoing balancing selection. The gene encoding the outer membrane protein OmpW was found to have a significantly positive Tajima's D value of 3.09. OmpW is a porin comprised of an 8-stranded beta-barrel, and its purported function is to transport hydrophilic molecules across the outer membranes of bacteria (Hong *et al.*, 2006). OmpW is widely conserved across gram-negative bacteria and may be involved in many processes related to homeostasis; it is also a largely abundant protein found in the cell membrane across *Xanthomonas* species (Watt *et al.*, 2005). While its specific biochemical function has yet to be determined, OmpW has been shown to be important for salinity tolerance (Xu *et al.*, 2005, Wu *et al.*, 2006, Fu *et al.*, 2018), iron transport (Catel-Ferriera *et al.*, 2015), and bacterial survival during environmental stress (Xiao *et al.*, 2016). Furthermore, it has been shown that the OmpW homolog in members of the *Burkolderia cepacia* complex cause immunogenic MAMP responses in mammalian cells (McClean *et al.*, 2016).

Alignments of the OmpW alleles within our *X. perforans* population showed the presence of two major alleles. Several of the allelic differences corresponded to amino acid residue changes in the surface-exposed regions of OmpW, specifically in the 2nd and 3rd extracellular loop structures (Figure 3.4). Because the side chain properties of these amino acids differed greatly (Table 3.4), these alleles could have structural variations, potentially leading to changes in PRR perception if these regions contained MAMP sequences. The location of these allelic differences, in tandem with a significant Tajima's D value, indicated to us that OmpW was a strong candidate for MAMP testing. Peptides containing the 2nd and 3rd extracellular regions of OmpW were

synthesized for both alleles and were tested in the luminol-based oxidative burst assay. At concentrations of 1 µM, none of the four OmpW peptides induced a ROS response in leaf disks from Moneymaker tomato plants in any of three independent experiments (Figure 3.5A and C). A fourth experiment was carried out using 10 µM of OmpW peptides, and similar results were obtained (Figure 3.5B).

DISCUSSION

Bacterial spot is a worldwide threat to tomato and pepper production, and growers have used a variety of methods to prevent or manage the disease (Potnis *et al.*, 2015). Due to the ubiquity of the disease, researchers have a significant interest in understanding more about the interactions between bacterial spot causing Xanthomonads and their hosts. With the rise of bioinformatics methods and high-throughput sequencing, large-scale comparative genomics analyses have led to the discovery of new MAMPs involved in plant-pathogen interactions (Mott *et al.*, 2016, McCann *et al.*, 2012, Eckshtain-Levi *et al.*, 2018). We utilized a similar approach to characterize the core genomes of *X. euvesicatoria* and *X. perforans*, identify non-neutrally evolving genes, and select promising MAMP candidates for testing using tomato plants. While our laboratory experiments were not able to confirm any novel MAMPs, our results allow for a deeper understanding of the genomes for these pathogens and the MAMP flgII-28 in *Xanthomonas*.

We observed a small reduction in core genome size compared to the average number of genes in a *Xanthomonas* genome. The average gene count for a *X. euvesicatoria* genome from the public NCBI genome database is 4,723 genes, while it is 4,689 genes for *X. perforans*. At a 90% core genome threshold, our core genome sizes are 92% and 90% of the average genome sizes for *X.*

euvesicatoria and *X. perforans*, respectively, indicating that most genes are conserved between strains for each species. The smaller percentage of core genome coverage in *X. perforans* is likely due to the greater genetic variability found in the species (Jibrin *et al*., 2018). In addition, both core genomes in our study are trending towards purifying selection, with more non-neutrally evolving genes exhibiting a significantly negative Tajima's D value than a positive value. While many of the genes identified by Tajima's D may not be directly involved in MAMP perception, they may instead be involved in other aspects of plant-pathogen interactions, be under other strong selection pressures, or be associated with genes that have strong selection pressures. The size of our *X. euvesicatoria* and *X. perforans* core genomes, as well as their measures of selection pressure, are indicative of highly specialized pathogens, which agrees with the small and specialized host range of these bacteria.

We set a cutoff for *Xanthomonas* genome assemblies consisting of no more than 200 scaffolds, in order to avoid the highly fragmented assemblies that likely would be missing several gene sequences within the gapped regions. The choice of strains in our studies directly impacts the characteristics of the core genome and adding different genome assemblies of these species would greatly impact the composition of the core genome (Roach *et al.*, 2019). Indeed, studies such as Roach *et al*. determine very different core and pan genomes when using different sets of *Xanthomonas* genome assemblies. Future comparative genomics studies of *Xanthomonas* would benefit greatly from using a diversity of high-quality genomes to capture the natural diversity of bacterial spot causing Xanthomonads.

We used a luminol-based oxidative burst assay to determine the recognition of different *Xanthomonas* flg22 and flgII-28 peptides in Moneymaker tomato plants. The most prevalent flg22 sequence found in our populations of *X. euvesicatoria* and *X. perforans* strains was not perceived by tomato leaves, indicating that this allele may provide a fitness advantage for the bacteria through evasion of FLS2-mediated PTI responses; similar findings were also observed for *Arabidopsis thaliana* plants treated with flg22 peptides from *Xanthomonas campestris* pv. *campestris* strains (Sun *et al*., 2006). While the majority of strains in both *Xanthomonas* populations had the same flg22 sequence, *X. euvesicatoria* and *X. perforans* strains had distinct flgII-28 sequences. The flgII-28 peptide present in most *X. perforans* strains is shared with many other *Xanthomonas* species and was found to cause PTI response in tomato. In contrast, a single amino acid difference present in the *X. euvesicatoria* flgII-28 sequence prevents induction of oxidative burst responses in tomato leaves treated with the peptide. This flgII-28 epitope likely evades recognition by its cognate PRR FLS3, and furthermore suggests that, unlike *X. perforans* strains, most *X. euvesicatoria* strains should be able to completely avoid flagellin-based recognition by tomato plants. The significant Tajima's D value of flagellin in *X. euvesicatoria* indicates evidence of purifying selection within that species. It is possible that this advantageous flgII-28 allele can be transferred from *X. euvesicatoria* into *X. perforans*, shown by evidence of homologous recombination between strains of these two species (Jibrin *et al.*, 2018).

OmpW was a strong MAMP candidate in our core genome dataset, as it had a significant Tajima's D value of 3.09 in *X. perforans*. Additionally, it is a highly abundant transmembrane protein present in the outer membranes of bacteria, and variable amino acid residues are present on extracellular domains of the OmpW proteins. However, OmpW peptide treatments failed to

induce an oxidative burst response in tomato leaves. This outcome mirrors that of OmpW in another *Xanthomonas* species, *Xanthomonas campestris* pv. *campestris* (Watt *et al.*, 2006). In that study, whole OmpW proteins were isolated from two-dimensional gels and used in an oxidative burst assay, but it failed to elicit responses in tobacco cell cultures. While OmpW was identified as a gene undergoing significant balancing selection, this selection pressure may not necessarily result from plant-pathogen interactions. Instead, OmpW may function in other essential processes that are under selection pressure, or alternatively OmpW may be genetically or functionally linked to other cellular factors that are under selection.

CONCLUSIONS

We established that while the widespread allele of flgII-28 present in *X. euvesicatoria* can evade the FLS3 receptor, the allele present in *X. perforans* and other *Xanthomonas* species is still perceived by the PRR. We also demonstrate that MAMPs can be identified using the statistic Tajima's D, although functional confirmation of putative MAMPs remains a time-consuming and difficult process. Utilizing population genomic screens to identify novel MAMPs has been shown for other plant pathogens such as *R. solanacearum* (Eckshtain-Levi *et al.*, 2018), but *R. solanacearum* a large and varied host range, while *Xanthomonas euvesicatoria* and *X. perforans* have a host range limited to a subset of solanaceous plants including tomato and pepper. While OmpW from these species may not elicit a PTI response in tomato plants, this does not preclude using signatures of balancing selection to successfully identify novel MAMPs s (Chen *et al.*, 2019). The quality of genomes assemblies used has a direct effect on core genome building and subsequent Tajima's D calculations. A diversity of high-quality genome assemblies, and subsequent gene alignments, can lead to a clearer understanding of selection pressures observed

in *Xanthomonas*. Further studies into the effectiveness of population genetics statistics like Tajima's D will help develop effective and more reliable pipelines for computationally identifying novel MAMPs.

TABLES AND FIGURES

Table 3.1. List of *Xanthomonas* genomes used in the study. Information in the table includes the species, strain identifier, level of genome assembly, number of contigs or scaffolds, and NCBI Assembly accession identifier. *X. euvesicatoria* is organized alphabetically by strain identifier, then *X. perforans.*

Species	**Strain identifier**	**Assembly Level**	**# Scaffolds**	**GenBank assembly accession**
X. euvesicatoria	181	Scaffold	107	GCA_001010095.1
X. euvesicatoria	199	Scaffold	110	GCA_001008975.1
X. euvesicatoria	206	Scaffold	133	GCA_001008815.1
X. euvesicatoria	376	Scaffold	128	GCA_001009045.1
X. euvesicatoria	455	Scaffold	137	GCA_001009055.1
X. euvesicatoria	490	Scaffold	132	GCA_001009075.1
X. euvesicatoria	515	Scaffold	118	GCA_001008825.1
X. euvesicatoria	526	Scaffold	178	GCA_001008835.1
X. euvesicatoria	586	Scaffold	120	GCA_001008885.1
X. euvesicatoria	679	Scaffold	109	GCA_001008895.1
X. euvesicatoria	681	Scaffold	104	GCA_001008905.1
X. euvesicatoria	683	Scaffold	108	GCA_001009095.1
X. euvesicatoria	684	Scaffold	157	GCA_001009125.1
X. euvesicatoria	685	Scaffold	124	GCA_001009135.1
X. euvesicatoria	689	Scaffold	143	GCA_001009205.1
X. euvesicatoria	695	Scaffold	110	GCA_001009215.1
X. euvesicatoria	BRIP38997	Contig	112	GCA_003993175.1
X. euvesicatoria	BRIP39016	Contig	96	GCA_003993195.1
X. euvesicatoria	BRIP62390	Contig	137	GCA_003993345.1
X. euvesicatoria	BRIP62391	Contig	131	GCA_003993335.1
X. euvesicatoria	BRIP62392	Contig	183	GCA_003993725.1

Table 3.1 (cont.)

Species	Strain identifier	Assembly Level	# Scaffolds	GenBank assembly accession
X. euvesicatoria	BRIP62395	Contig	168	GCA_003993315.1
X. euvesicatoria	BRIP62396	Contig	161	GCA_003993275.1
X. euvesicatoria	BRIP62400	Contig	131	GCA_003993685.1
X. euvesicatoria	BRIP62403	Contig	144	GCA_003993675.1
X. euvesicatoria	BRIP62425	Contig	163	GCA_003993265.1
X. euvesicatoria	BRIP62438	Contig	99	GCA_003993255.1
X. euvesicatoria	BRIP62441	Contig	135	GCA_003993655.1
X. euvesicatoria	BRIP62555	Contig	139	GCA_003993605.1
X. euvesicatoria	BRIP62757	Contig	149	GCA_003993615.1
X. euvesicatoria	BRIP62858	Contig	144	GCA_003993595.1
X. euvesicatoria	BRIP62959	Contig	129	GCA_003993225.1
X. euvesicatoria	BRIP63464	Contig	140	GCA_003993185.1
X. euvesicatoria	DAR26930	Contig	101	GCA_003993445.1
X. euvesicatoria	DAR34895	Contig	69	GCA_003992805.1
X. euvesicatoria	F4-2	Scaffold	123	GCA_001009165.1
X. euvesicatoria	G4-1	Scaffold	112	GCA_001009245.1
X. euvesicatoria	H3-2	Scaffold	108	GCA_001009175.1
X. euvesicatoria	L3-2	Scaffold	152	GCA_001009255.1
X. euvesicatoria	LMG12749	Contig	54	GCA_001401675.2
X. euvesicatoria	LMG930	Complete Genome	5	GCA_001908795.1
X. perforans	BRIP62383	Contig	61	GCA_003993135.1
X. perforans	BRIP62384	Contig	77	GCA_003993115.1
X. perforans	BRIP62386	Contig	80	GCA_003993575.1
X. perforans	BRIP62397	Contig	45	GCA_003993535.1
X. perforans	BRIP62398	Contig	57	GCA_003993095.1
X. perforans	BRIP623989	Contig	70	GCA_003993105.1

Table 3.1 (cont.)

Species	Strain identifier	Assembly Level	# Scaffolds	GenBank assembly accession
X. perforans	BRIP62404	Contig	109	GCA_003993055.1
X. perforans	BRIP62405	Contig	76	GCA_003993035.1
X. perforans	BRIP63262	Contig	98	GCA_003993015.1
X. perforans	BRIP63565	Contig	72	GCA_003993025.1
X. perforans	BRIP63666	Contig	59	GCA_003992975.1
X. perforans	CFBP 7293	Contig	32	GCA_001976075.1
X. perforans	GEV1001	Contig	78	GCA_001010025.1
X. perforans	GEV1026	Scaffold	70	GCA_001010035.1
X. perforans	GEV1044	Scaffold	82	GCA_001009935.1
X. perforans	GEV1054	Scaffold	85	GCA_001009925.1
X. perforans	GEV1063	Scaffold	93	GCA_001010085.1
X. perforans	GEV2120	Contig	85	GCA_004102205.1
X. perforans	GEV839	Contig	109	GCA_001009475.1
X. perforans	GEV872	Scaffold	80	GCA_001009485.1
X. perforans	GEV893	Scaffold	116	GCA_001009545.1
X. perforans	GEV904	Scaffold	134	GCA_001009795.1
X. perforans	GEV909	Contig	71	GCA_001009825.1
X. perforans	GEV915	Scaffold	70	GCA_001009855.1
X. perforans	GEV917	Scaffold	122	GCA_001009865.1
X. perforans	GEV936	Scaffold	104	GCA_001009845.1
X. perforans	GEV940	Contig	98	GCA_001009885.1
X. perforans	GEV968	Contig	96	GCA_001010015.1
X. perforans	GEV993	Scaffold	69	GCA_001010005.1
X. perforans	LH3	Complete Genome	5	GCA_001908855.1
X. perforans	NI1	Contig	64	GCA_003136155.1
X. perforans	TB6	Scaffold	148	GCA_001009945.1

Table 3.1 (cont.)

Species	**Strain identifier**	**Assembly Level**	**# Scaffolds**	**GenBank assembly accession**
X. perforans	Xp10-13	Contig	78	GCA_001009405.1
X. perforans	Xp11-2	Contig	50	GCA_001009445.1
X. perforans	Xp15-11	Contig	63	GCA_001009465.1
X. perforans	Xp17-12	Contig	84	GCA_001009745.1
X. perforans	Xp18-15	Contig	78	GCA_001009765.1
X. perforans	Xp2010	Scaffold	13	GCA_001009785.1
X. perforans	Xp3-15	Contig	93	GCA_001009675.1
X. perforans	Xp4-20	Contig	71	GCA_001009705.1
X. perforans	Xp5-6	Contig	70	GCA_001009365.1
X. perforans	Xp7-12	Contig	65	GCA_001009385.1
X. perforans	Xp8-16	Contig	62	GCA_001009685.1
X. perforans	Xp9-5	Contig	113	GCA_001009395.1

Table 3.2. Flg22 peptide sequences used in the study. Red amino acids indicate differences between peptides. * indicates alleles also shared by the majority of *X. perforans* strains.

Peptide Name	Sequence
flg22	QRLSTGSRINSAKDDAAGLQIA
flg22 *Xe* 85-10*	QQLSSGKRITSFAVDAAGGAIA

Table 3.3. FlgII-28 peptide sequences used in the study. Red amino acids indicate differences between peptides. * indicates alleles also shared by the majority of *X. perforans* strains.

Peptide Name	Sequence
flgII-28 *Pst* T1	ESTNILQRMRELAVQSRNDSNSSTDRDA
flgII-28 *X.* spp*	EIGNNLQRIRELSVQSANATNSSTDREA
flgII-28 *Xe* 85-10	EIGNNLQRIRELSVQSAKATNSSTDREA

Table 3.4. OmpW peptide sequences used in the study. Red amino acids indicate differences between peptides. OmpW peptide sequences were derived from *X. perforans* 91-118. OmpW-1 Allele 1 and 2 are found on the 2nd extracellular residue of OmpW. OmpW-2 Allele 1 and 2 are found on the 3rd extracellular residue of OmpW.

Peptide Name	Sequence
OmpW-1 Allele 1	DIALRGLGRVGST
OmpW-1 Allele 2	DIAIGGLGRVGST
OmpW-2 Allele 1	FDTDTGGSLAGSTLELD
OmpW-2 Allele 2	FDTDRGGSLAGSTKELD

Figure 3.1. Tajima's D distributions for the core genome of *X. euvesicatoria* (**A**) and *X. perforans* (**B**). X-axis indicates Tajima's D values found using the R package *pegas,* while Y-axis indicates counts of genes. Dashed line represents the mean Tajima's D value for each species, which were -2.15 for *X. euvesicatoria* (**A**) and -1.58 for *X. perforans* (**B**).

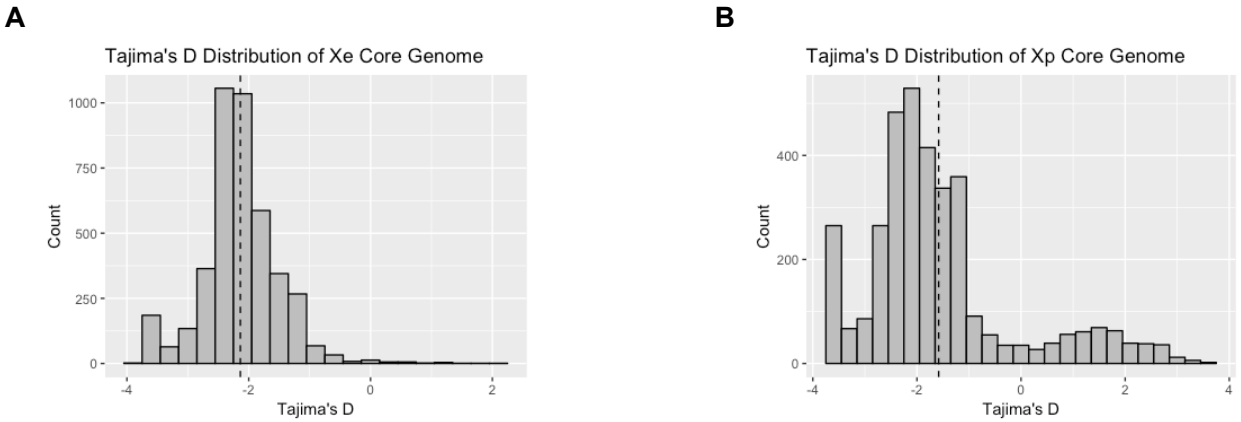


Figure 3.2. GO terms of genes with significantly positive Tajima's D values in *X. perforans*. GO terms were derived from Protein IDs and the UniProt Retrieve / ID mapping tool. Groupings of Go terms was performed using the Web Gene Ontology Annotation Plot (WEGO) web tool. X-axis indicates functional characteristics of genes as determined by Gene Ontology classification.

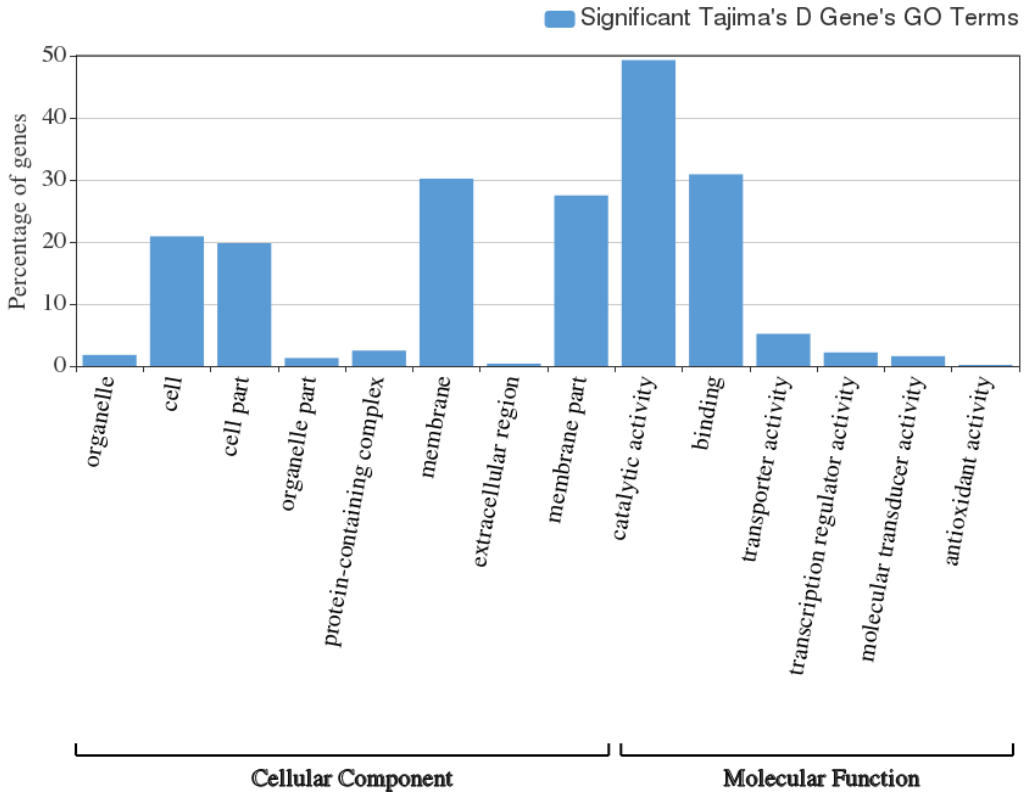


Figure 3.3. Oxidative burst produced by *S. lycopersicum* cv. 'Moneymaker' leaf disks treated with 100 nM flg22 (**A**) and flgII-28 (**B**) peptides listed in Tables 3.2 and 3.3, respectively. Each data point represents the mean of four biological replicates (n = 4 plants), with each plant represented by the average of four technical replicates. Error bars represent standard deviations. Similar results were obtained in three independent experiments, with one representative experiment shown here. (**C**) Total oxidative burst production for each peptide treatment. Each bar represents the average total relative light units (i.e., area under the curve) produced for each treatment for three independent experiments, with error bars representing the standard error. Groups were derived using the post-hoc analysis Tukey's Test.

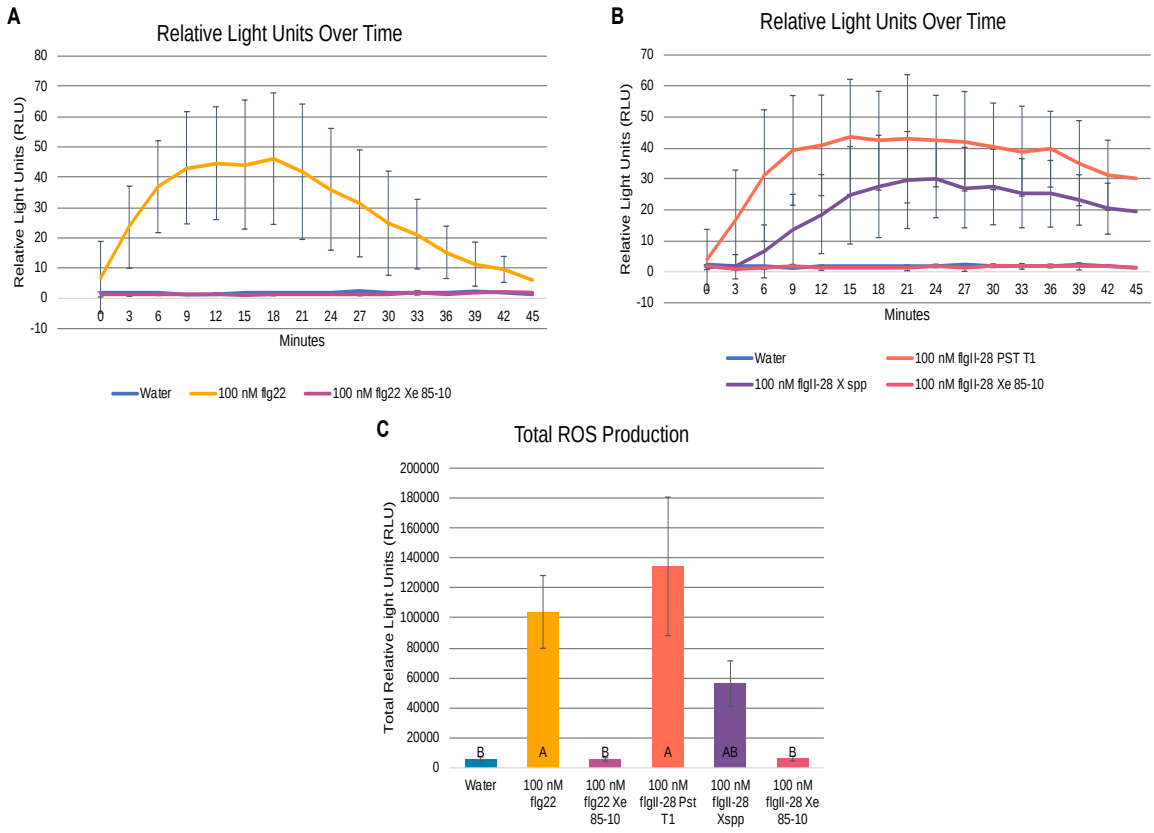


Figure 3.4. 3D transmembrane view of OmpW. Image from the RCSB PDB (rcsb.org) of PDB ID 2F1T (Hong, H., D. Patel, L. Tamm and B. van den Berg) (2006). "The Outer Membrane Protein OmpW Forms an Eight-Stranded beta-Barrel with a Hydrophobic Channel." The Journal of Biological Chemistry **281**(11): 7568-7577. Blue and red dots designate the cytoplasmic membrane and outer membrane respectively. Red colored protein region indicates transmembrane region of OmpW, while grey colored protein region indicates extracellular regions of OmpW.

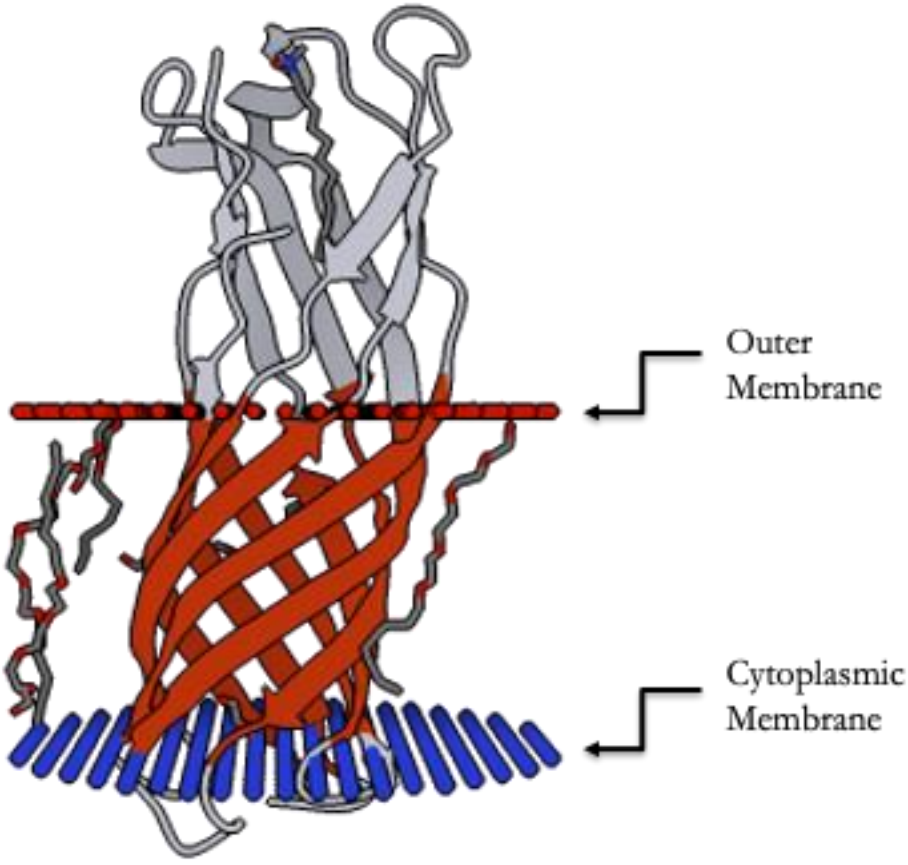
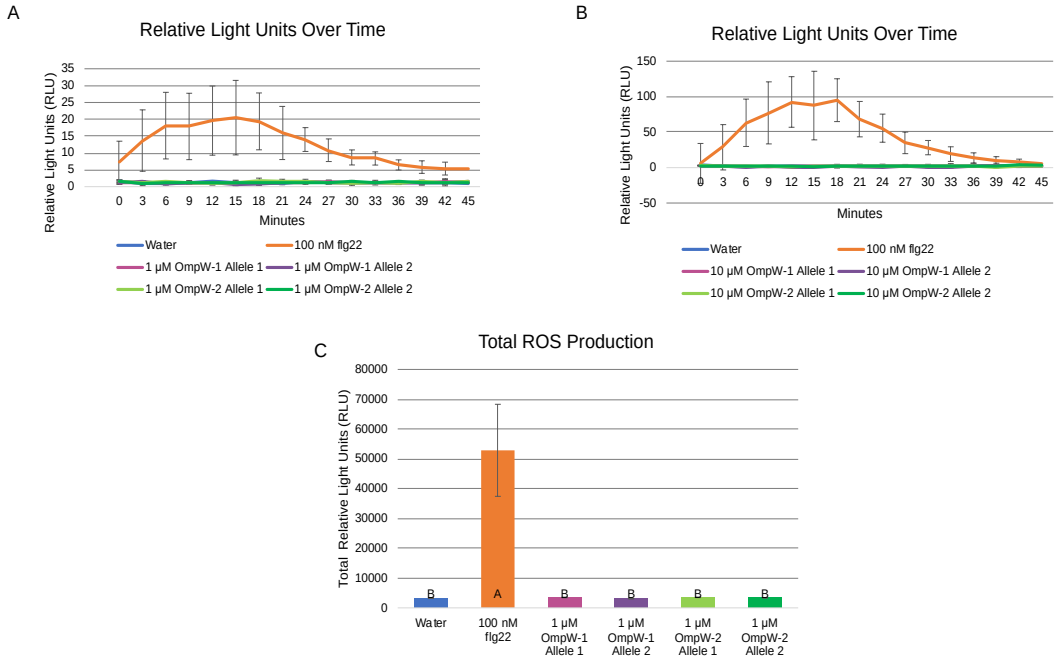


Figure 3.5. Oxidative burst produced by *S. lycopersicum* cv. 'Moneymaker' leaf disks treated with 1 µM **(A)** or 10 µM **(B)** OmpW peptides listed in Table 3.4. Each data point represents the mean of four biological replicates (n = 4 plants), with each plant represented by the average of four technical replicates. Error bars represent standard deviations. Similar results were obtained across three independent experiments in **(A)**, with one representative experiment shown here. A single experiment was performed using 10 µM OmpW peptides **(B)**, shown here. (**C**) Total oxidative burst production for each peptide treatment. Each bar represents the average total relative light units (i.e., area under the curve) produced for each treatment for three independent experiments, with error bars representing the standard error. Significant groups were derived using the post-hoc analysis Tukey's Test.



LITERATURE CITED

Behlau, F., B. I. Canteros, G. V. Minsavage, J. B. Jones and J. H. Graham (2011). "Molecular Characterization of Copper Resistance Genes from *Xanthomonas citri* subsp *citri* and *alfalfae* subsp. *citrumelonis*." Applied and Environmental Microbiology **77**(12): 4089.

Benjamini, Y. and Y. Hochberg (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." Journal of the Royal Statistical Society. Series B (Methodological) **57**(1): 289-300.

Bhattarai, K., F. J. Louws, J. D. Williamson and D. R. Panthee (2016). "Differential response of tomato genotypes to *Xanthomonas*-specific pathogen-associated molecular patterns and correlation with bacterial spot (*Xanthomonas perforans*) resistance." Horticulture Research **3**(1): 16035.

Blanvillain, S., D. Meyer, A. Boulanger, M. Lautier, C. Guynet, N. Denancé, J. Vasse, E. Lauber and M. Arlat (2007). "Plant Carbohydrate Scavenging through TonB-Dependent Receptors: A Feature Shared by Phytopathogenic and Aquatic Bacteria." PLOS ONE **2**(2): e224.

Cai, R., J. Lewis, S. Yan, H. Liu, C. R. Clarke, F. Campanile, N. F. Almeida, D. J. Studholme, M. Lindeberg, D. Schneider, M. Zaccardelli, J. C. Setubal, N. P. Morales-Lizcano, A. Bernal, G. Coaker, C. Baker, C. L. Bender, S. Leman and B. A. Vinatzer (2011). "The plant pathogen *Pseudomonas syringae* pv. *tomato* is genetically monomorphic and under strong selection to evade tomato immunity." PLOS Pathogens **7**(8): e1002130-e1002130.

Catel-Ferreira, M., S. Marti, L. Guillon, L. Jara, G. Coadou, V. Molle, E. Bouffartigues, G. Bou, I. Shalk, T. Jouenne, X. Vila-Farrés and E. Dé (2016). "The outer membrane porin OmpW of *Acinetobacter baumannii* is involved in iron uptake and colistin binding." FEBS Letters **590**(2): 224-231.

Chen, Y., C. Bendix and J. D. Lewis (2019). "Comparative Genomics Screen Identifies Microbe-Associated Molecular Patterns from '*Candidatus Liberibacter'* spp. That Elicit Immune Responses in Plants." Molecular Plant-Microbe Interactions **33**(3): 539-552.

Felix, G., J. D. Duran, S. Volko and T. Boller (1999). "Plants have a sensitive perception system for the most conserved domain of bacterial flagellin." The Plant Journal **18**(3): 265-276.

Figaj, D., P. Ambroziak, T. Przepiora and J. Skorko-Glonek (2019). "The Role of Proteases in the Virulence of Plant Pathogenic Bacteria." International Journal of Molecular Sciences **20**(3): 672.

Fu, X., J. Zhang, T. Li, M. Zhang, J. Li and B. Kan (2018). "The Outer Membrane Protein OmpW Enhanced *V. cholerae* Growth in Hypersaline Conditions by Transporting Carnitine." Frontiers in Microbiology **8**(2703).

Gómez-Gómez, L. and T. Boller (2000). "FLS2: An LRR Receptor-like Kinase Involved in the Perception of the Bacterial Elicitor Flagellin in *Arabidopsis*." Molecular Cell **5**(6): 1003-1011.

Hind, S. R., S. R. Strickler, P. C. Boyle, D. M. Dunham, Z. Bao, I. M. O'Doherty, J. A. Baccile, J. S. Hoki, E. G. Viox, C. R. Clarke, B. A. Vinatzer, F. C. Schroeder and G. B. Martin (2016). "Tomato receptor FLAGELLIN-SENSING 3 binds flgII-28 and activates the plant immune system." Nature Plants **2**(9): 16128.

Hong, H., D. Patel, L. Tamm and B. van den Berg (2006). "The Outer Membrane Protein OmpW Forms an Eight-Stranded beta-Barrel with a Hydrophobic Channel." The Journal of Biological Chemistry **281**(11): 7568-7577.

Jibrin, M. O., N. Potnis, S. Timilsina, G. V. Minsavage, G. E. Vallad, P. D. Roberts, J. B. Jones and E. M. Goss (2018). "Genomic Inference of Recombination-Mediated Evolution in *Xanthomonas euvesicatoria* and *X. perforans*." Applied and Environmental Microbiology **84**(13): e00136-00118.

Jolley, K. A. and M. C. J. Maiden (2010). "BIGSdb: Scalable analysis of bacterial genome variation at the population level." BMC Bioinformatics **11**(1): 595.

Jones, J., K. Pohronezny, R. Stall and J. Jones (1986). "Survival of *Xanthomonas campestris* pv. *vesicatoria* in Florida on tomato crop residue, weeds, seeds, and volunteer tomato plants." Phytopathology **76**(4): 430-434.

Jones, J. B., G. H. Lacy, H. Bouzar, R. E. Stall and N. W. Schaad (2004). "Reclassification of the Xanthomonads Associated with Bacterial Spot Disease of Tomato and Pepper." Systematic and Applied Microbiology **27**(6): 755-762.

Katoh, K. and D. M. Standley (2013). "MAFFT multiple sequence alignment software version 7: improvements in performance and usability." Molecular Biology and Evolution **30**(4): 772-780.

Lacombe, S., A. Rougon-Cardoso, E. Sherwood, N. Peeters, D. Dahlbeck, H. P. van Esse, M. Smoker, G. Rallapalli, B. P. H. J. Thomma, B. Staskawicz, J. D. G. Jones and C. Zipfel (2010). "Interfamily transfer of a plant pattern-recognition receptor confers broad-spectrum bacterial resistance." Nature Biotechnology **28**: 365.

McCann, H. C., H. Nahal, S. Thakur and D. S. Guttman (2012). "Identification of innate immunity elicitors using molecular signatures of natural selection." Proceedings of the National Academy of Sciences **109**(11): 4215.

McClean, S., M. E. Healy, C. Collins, S. Carberry, L. O'Shaughnessy, R. Dennehy, Á. Adams, H. Kennelly, J. M. Corbett, F. Carty, L. A. Cahill, M. Callaghan, K. English, B. P. Mahon, S. Doyle and M. Shinoy (2016). "Linocin and OmpW Are Involved in Attachment of the Cystic Fibrosis-Associated Pathogen *Burkholderia cepacia* Complex to Lung Epithelial Cells and Protect Mice against Infection." Infection and Immunity **84**(5): 1424-1437.

McInnes, T. (1988). "Airborne Dispersal of Bacteria in Tomato and Pepper Transplant Fields." Plant Disease **72**(7): 575-579.

Mott, G. A., S. Thakur, E. Smakowska, P. W. Wang, Y. Belkhadir, D. Desveaux and D. S. Guttman (2016). "Genomic screens identify a new phytobacterial microbe-associated molecular pattern and the cognate *Arabidopsis* receptor-like kinase that mediates its immune elicitation." Genome biology **17**: 98-98.

Paradis, E. (2010). "pegas: an R package for population genetics with an integrated-modular approach." Bioinformatics **26**(3): 419-420.

Potnis, N., S. Timilsina, A. Strayer, D. Shantharaj, J. D. Barak, M. L. Paret, G. E. Vallad and J. B. Jones (2015). "Bacterial spot of tomato and pepper: diverse *Xanthomonas* species with a wide variety of virulence factors posing a worldwide challenge." Molecular Plant Pathology **16**(9): 907-920.

Ritchie, D. F. 2000. Bacterial spot of pepper and tomato. The Plant Health Instructor. DOI: 10.1094/PHI-I-2000-1027-01 http://www.apsnet.org/education/LessonsPlantPath/BacterialSpot.

Roach, R., R. Mann, C. G. Gambley, T. Chapman, R. G. Shivas and B. Rodoni (2019). "Genomic sequence analysis reveals diversity of Australian *Xanthomonas* species associated with bacterial leaf spot of tomato, capsicum and chilli." BMC Genomics **20**(1): 310.

Schwartz, A. R., N. Potnis, S. Timilsina, M. Wilson, J. Patané, J. Martins, G. V. Minsavage, D. Dahlbeck, A. Akhunova, N. Almeida, G. E. Vallad, J. D. Barak, F. F. White, S. A. Miller, D. Ritchie, E. Goss, R. S. Bart, J. C. Setubal, J. B. Jones and B. J. Staskawicz (2015). "Phylogenomics of *Xanthomonas* field strains infecting pepper and tomato reveals diversity in effector repertoires and identifies determinants of host specificity." Frontiers in Microbiology **6**(535).

Spoel, S. H. and X. Dong (2012). "How do plants achieve immunity? Defence without specialized immune cells." Nature Reviews Immunology **12**(2): 89-100.

Sun, W., F. M. Dunning, C. Pfund, R. Weingarten and A. F. Bent (2006). "Within-Species Flagellin Polymorphism in *Xanthomonas campestris* pv *campestris* and Its Impact on Elicitation of *Arabidopsis FLAGELLIN SENSING2*-Dependent Defenses." The Plant Cell **18**(3): 764.

Thakur, M. and B. S. Sohal (2013). "Role of Elicitors in Inducing Resistance in Plants against Pathogen Infection: A Review." ISRN Biochemistry **2013**: 762412-762412.

Vetter, M., T. L. Karasov and J. Bergelson (2016). "Differentiation between MAMP Triggered Defenses in *Arabidopsis thaliana*." PLOS Genetics **12**(6): e1006068.

Watt, S. A., V. Tellström, T. Patschkowski and K. Niehaus (2006). "Identification of the bacterial superoxide dismutase (SodM) as plant-inducible elicitor of an oxidative burst reaction in tobacco cell suspension cultures." Journal of Biotechnology **126**(1): 78-86.

Watt, S. A., A. Wilke, T. Patschkowski and K. Niehaus (2005). "Comprehensive analysis of the extracellular proteins from *Xanthomonas campestris* pv. *campestris* B100." Proteomics **5**(1): 153-167.

Wu, L., X. Lin, F. Wang, D. Ye, X. Xiao, S. Wang and X. Peng (2006). "OmpW and OmpV are Required for NaCl Regulation in *Photobacterium damsela*." Journal of Proteome Research **5**(9): 2250-2257.

Xu, C., S. Wang, H. Ren, X. Lin, L. Wu and X. Peng (2005). "Proteomic analysis on the expression of outer membrane proteins of *Vibrio alginolyticus* at different sodium concentrations." Proteomics **5**(12): 3142-3152.

Ye, J., Y. Zhang, H. Cui, J. Liu, Y. Wu, Y. Cheng, H. Xu, X. Huang, S. Li, A. Zhou, X. Zhang, L. Bolund, Q. Chen, J. Wang, H. Yang, L. Fang and C. Shi (2018). "WEGO 2.0: a web tool for analyzing and plotting GO annotations, 2018 update." Nucleic Acids Research **46**(W1): W71-W75.

Yu, X., B. Feng, P. He and L. Shan (2017). "From Chaos to Harmony: Responses and Signaling upon Microbial Pattern Recognition." Annual Review of Phytopathology **55**: 109-137.

Zipfel, C. (2014). "Plant pattern-recognition receptors." Trends in Immunology **35**(7): 345-351.

Zipfel, C., G. Kunze, D. Chinchilla, A. Caniard, J. D. G. Jones, T. Boller and G. Felix (2006). "Perception of the Bacterial PAMP EF-Tu by the Receptor EFR Restricts *Agrobacterium*-Mediated Transformation." Cell **125**(4): 749-760.

Zipfel, C., S. Robatzek, L. Navarro, E. Oakeley, J. Jones, G. Felix and T. Boller (2004). "Bacterial disease resistance in Arabidopsis through flagellin perception." Nature **428**: 764-767.

CHAPTER 4

FUTURE DIRECTIONS

Throughout these studies, I have characterized genes important to pathogenicity and virulence in *X. cucurbitae*, the causal agent of bacterial spot disease in cucurbits such as pumpkin, squash, and cucumber. In addition, I have discovered genetic clusters between *X. cucurbitae* isolates collected throughout the Midwestern United States. This research represents the first steps in characterizing *X. cucurbitae* on a genomic level and provides a framework for further studies regarding the plant pathogen.

In the future, a dual RNA-seq approach can be used to further elucidate the relationship between *X. cucurbitae* and cucurbit host plants. Dual RNA-seq will allow us to study the host response to infection in cucurbit plants and determine genes important to defense against bacterial spot disease. Dual RNA-seq will also allow us to observe how *X. cucurbitae* behaves during infection *in vivo*, instead of simulating host cells on host-mimicking media. In addition, collecting dual RNA-seq data at different timepoints can be used to monitor genes in both organisms throughout the infection process, from initial stages of invasion to eventual host manipulation by the pathogen.

The gene *cbhA* was found to be upregulated in *X. cucurbitae* grown in host-mimicking conditions. Although *X. cucurbitae* is a non-vascular pathogen, *cbhA* is a gene largely found only in vascular plant pathogenic *Xanthomnas* species. It is currently unknown how our strain of *X. cucurbitae* acquired this gene, as well as how widespread this gene is throughout the Midwestern

US *X. cucurbitae* population. Future phenotypic and genomic studies regarding this gene should be carried out to understand its importance in pathogenicity and virulence.

Further genome wide studies of *X. cucurbitae* can be carried out to discover the evolutionary connections between isolates. While RAD-seq is a useful tool for characterizing regions of a genome, whole genome comparisons between isolates can be carried out due to the continually lowering costs of whole-genome sequencing. Type II enzymes and type III effectors have a direct effect on host specificity and pathogenicity; utilizing whole-genome sequencing to compare these repertoires between isolates can uncover the adaptability and pathogenicity of the Midwestern US *X. cucurbitae* population. Whole-genome sequencing can also be used to resolve phylogenetic relationships between isolates at a deeper resolution. In addition, using whole-genome sequencing for core genome and pan genome analyses can determine the diversity of *X. cucurbitae* isolates in the Midwestern US population.

While our studies were not able to discover novel MAMPs in *X. euvesicatoria* and *X. perforan*s, other studies have shown success in using a comparative genomics approach to identifying novel MAMPs. Future research using this approach will benefit from using higher quality, more diverse genome assemblies, as they will capture more diversity between bacterial populations. These higher quality genome assemblies will also increase the power of population genetic statistics such as Tajima's D. While Tajima's D was an effective statistic for identifying MAMPs in our *Xanthomonas* populations, other population genetics statistics such as Li's D or d_N/d_S may also provide further insight into non-neutrally evolving genes or gene products.