The Pennsylvania State University

The Graduate School

PHENOTYPIC AND GENETIC ANALYSES OF THE LETTUCE BACTERIAL LEAF SPOT PATHOGEN, *XANTHOMONAS HORTORUM* PV. *VITIANS*, AND THE DEVELOPMENT OF A RAPID DETECTION TECHNIQUE

A Dissertation in

Plant Pathology

by

Emma R. Martinez

Submitted in Partial Fulfillment
Of the Requirements
For the Degree of

Doctor of Philosophy

December 2023

The dissertation of Emma Rosenthal was reviewed and approved by the following:

Carolee Bull
Dissertation Advisor
Chair of Committee

Beth Gugino
Professor of Plant Pathology and Environmental Microbiology
Director of Graduate Studies

Istvan Albert
Research Professor of Bioinformatics

Germán Sandoya
Assistant Professor of Horticultural Sciences at the University of Florida

Emily Weinert
Associate Professor of Biochemistry and Molecular Biology

María del Mar Jiménez Gasco
Interim Department Head of Plant Pathology and Environmental Microbiology

ABSTRACT

Bacterial leaf spot (BLS) of lettuce, caused by *Xanthomonas hortorum* pv. *vitians* (*Xhv*), was first described in 1918 and it remains a significant threat to lettuce cultivation. Symptoms of disease include water-soaking and chlorosis in the leaves and small necrotic spots that merge to form larger lesions. Outbreaks can be devastating, with up to 100% crop loss, and sources of inoculum have been found in crop debris and weeds. The pathogen is also thought to be seedborne, though this has not yet been demonstrated. Integrated management strategies include the removal of crop debris and weeds, the use of clean seeds, and the use of resistant cultivars. Chemical application may be effective in preventing outbreaks when applied at the earliest detection of the pathogen, but they work best when applied prophylactically, which can be a large sunken cost in years that the pathogen does not appear. Further, bactericides vary in their effectiveness. Multiple research groups have tried to make PCR-based detection methods for the pathogen, but none yet specifically detect *Xhv*. Detection is complicated by the existence of three hypothetical races of the pathogen, defined by their host reactions on various lettuce cultivars and accessions. While *Xhv* populations are predominantly clonal, any observed phenotypic variation is likely the result of genetic recombination over mutation.

The major knowledge gaps in this field of study include a full characterization of the host reactions of the three *Xhv* races, an understanding of the genotypes underlying *Xhv* race-specificity, and a set of detection methods for *Xhv* strains and to distinguish the races. Chapter 1 summarizes the literature that is foundational to the study of *Xhv*. Chapter 2 describes our characterizations of the *Xhv* race 1, 2, and 3 reactions on differential lettuce cultivars. Strains representing each hypothesized race were inoculated by syringe into all three of the previously

determined differential cultivars. We confirmed previous findings that *Xhv* race 1 strains triggered the resistant hypersensitive response (HR) on *L. sativa* cv. Little Gem, *Xhv* race 2 strains triggered HR on *L. serriola* PI491114, and *Xhv* race 3 strains triggered HR on *L. serriola* ARM-09-161-10-1. HR is a plant's resistance strategy in which it kills off its own tissue and limits the spread of bacteria; differential host reactions like those in *Xhv* have been used to define races in other bacterial species. We also demonstrated that *Xhv* race 1 strains triggered HR in *L. serriola* ARM-09-161-10-1, and that two sets of strains collected in Florida and France did include *Xhv* strains. A multi-locus sequence analysis scheme that has been used previously to predict *Xhv* race predicted that nearly all these strains belonged to *Xhv* race 1. Finally, we found that three plant introductions deposited at the USDA-ARS Germplasm Resources Information Network (GRIN) of the National Plant Germplasm System may also have race-specific resistance.

In Chapter 3, we sought to identify race-specific DNA sequences that might underlie the race-specific host reactions. Whole genome sequences of eighteen *Xhv* strains representing the three hypothesized races, along with eight related *Xanthomonas* strains, were included in the analysis. A maximum likelihood phylogeny based on concatenated whole genome single nucleotide polymorphisms confirmed previous results describing two major lineages of *Xhv* strains. Gene clusters encoding secretion systems, secondary metabolites, and bacteriocins were assessed to identify putative virulence factors that distinguish the *Xhv* races. Genome sequences were mined for effector genes, which have been shown to be involved in race-specificity in other systems. Two effectors identified in this study, *xopAQ* and the novel variant *xopAF2,* were revealed as possible mediators of a gene-for-gene interaction between *Xhv* race 1 and 3 strains and wild lettuce *L. serriola* ARM-09-161-10-1. Transposase sequence identified downstream of

xopAF2 and prophage sequence found within *Xhv* race 1 and 3 insertion sequences suggest that this gene may have been acquired through phage-mediated gene transfer. No other factors were identified from these analyses that distinguish the *Xhv* races.

For Chapter 4, whole genome sequence alignments allowed us to identify gene clusters that were specific to *Xhv* or to each of its races that could be used as targets for pathovar- or race-specific PCR-based detection methods. Four gene clusters were identified for possible *Xhv*, *Xhv* race 1, *Xhv* race 2, or *Xhv* race 3 detection, and many primers were designed and evaluated by touchdown PCR. We found that our pathovar specific primer set successfully detected 97% of the *Xhv* strains tested and did not detect any of the other six *X. hortorum* pathovars. However, it was also able to detect a close relative, *X. hydrangea*, from symptomatic lettuce tissue. While this means our method is also not *Xhv*-specific, we recommend the use of our protocol in tandem with a published LAMP assay developed for detection of *Xhv*, *X. hortorum* pv. *gardneri*, and *X. hortorum* pv. *cynarae*. The use of these two methods together should reveal whether an unknown sample is *Xhv*. Further work is necessary to develop a single *Xhv*-specific method, as well as the race-specific detection methods as each of these three methods resulted in false positives or false negatives.

In Chapter 5, we share our progress in evaluating the roles of *Xhv* race 1 and 3 specific effector genes, *xopAQ* and *xopAF2*, in HR induction in *L. serriola* ARM-09-161-10-1. Effector genes function in promoting bacterial colonization, but their detection by plant host proteins have been shown to trigger HR. We aimed to delete each effector gene from the *Xhv* race 1 BS0347 genome and tested two methods: scarless deletion and marker exchange deletion with a gentamycin resistance cassette. The deletion constructs were designed from upstream and

downstream regions flanking each effector, which were then synthesized into the nonreplicable plasmid pk18mobsacB. Movement of the constructs into a rifampicin resistant mutant of *Xhv* race 1 BS0347 and the occurrence of homologous recombination was expected to result in gene deletion from the chromosomal DNA and its transfer onto the plasmid, which would then be lost in counterselection on sucrose and rifampicin amended media. With the marker exchange method, the gentamycin resistance cassette would replace the effector gene at its site in the bacterial chromosome, and so double recombinants could be also selected for with additional gentamycin in the selective media. We have evidence of a successful *xopAQ* gene deletion. Expression constructs were also designed for *xopAQ* and *xopAF2* expression in a rifampicin resistant *Xhv* race 2 strain, but their transformation into the *Xhv* race 2 strain has not yet been successful. We outline additional strategies for expressing the effector genes in *Xhv* race 2. Once produced, these altered *Xhv* strains, the wild type strains, and a set of empty vectors will be inoculated into *L. serriola* ARM-09-161-10-1 to determine whether the genetic alterations impact the strains' ability to induce HR. Continuing this work is expected to reveal the genes involved in HR induction.

The final chapter of this dissertation summarizes our findings, highlights the importance of the work, and suggests avenues for future study. Overall, this dissertation advances the field of plant pathology by defining the *Xhv* race reactions on lettuce cultivars, predicting two genes that may be responsible for race-specific HR induction in a lettuce host, demonstrating methods that could be used for pathovar-specific PCR-based detection and showing progress made toward similar methods for race-specific detection, and finally showing progress toward demonstrating in planta the potential role of two race-specific genes in HR induction.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Dr. Carolee Bull, for her constant support, expert guidance, and impeccable mentorship. She provided me with countless opportunities for professional growth, and I would not be the scientist I am today without her influence. I thank my committee members, Dr. Beth Gugino, Dr. Istvan Albert, Dr. Emily Weinert, and Dr. Germán Sandoya, for their insights and advice throughout my graduate career. I thank Dr. Cristina Rosa for her excellent mentorship while I served as her teaching assistant, Dr. Neha Potnis for sharing her knowledge and advice with me, and Maria Gorgo for teaching me how best to serve Pennsylvania Latinx communities. I am grateful to all members, past and present, of the Bull Lab, for their wisdom and friendship, both of which contributed greatly to my success in this program. I thank the College of Agricultural Sciences faculty, extension educators, staff, and students for their support, especially those of the Department of Plant Pathology and the Latinx Agricultural Network.

I offer personal thanks to my husband for his never-ending kindness, encouragement, and love. I am tremendously grateful to my parents, who raised me to believe in my potential and to work hard to achieve even the most challenging goals. I am thankful for my brother, for always inspiring me to seek the best version of myself. I am forever grateful to my Colombian family for their love and spiritual guidance.

Funding for this work was provided by the Bunton Waller Graduate Assistantship at Penn State University; the 2014 Specialty Crops Block Grant Farm Bill App. CDFA #10.170, Project 44; and the NSF GRFP Grant No. 2018265841. The findings and conclusions presented here do not necessarily reflect the view of these funding agencies.

Chapter 1

Literature Review and Thesis Objectives

1.1 The Lettuce Industry and Bacterial Leaf Spot Management

Lettuce is ubiquitous in American cuisine; it is a staple in salads and sandwiches, it is favored for being low in calories while being a great source of vitamins and minerals, and it grows even more popular as American diets shift toward healthier eating. It is no surprise then that lettuce (*L. sativa*) is the highest valued vegetable crop in the United States at $3.8 billion in utilized production value in 2020, which was a 21% increase over the year prior (USDA-NASS, 2022). The main categories of cultivated lettuce are head, leaf, and romaine and while most are field grown, the hydroponics sector of the lettuce industry has seen growth due to consumer preference for fresh, locally grown produce and a loss of agricultural land to urbanization. China is the largest lettuce producer by annual gross yield, followed by the United States, India, Spain, Italy, Japan, and Iran (Shatilov et al., 2019).

Bacterial leaf spot (BLS), a disease caused by *Xanthomonas hortorum* pv. *vitians* (*Xhv*) Morinière *et al*. 2020 and formerly known as *Xanthomonas campestris* pv. *vitians*, is a significant threat to lettuce industries worldwide. Most commercial lettuce cultivars are susceptible to BLS infection or only have partial disease resistance (Carisse et al., 2000; Sahin and Miller et al., 1997). The first documented outbreak of lettuce bacterial leaf spot occurred on several South Carolina lettuce farms (Brown 1918). Disease prevalence and severity varied widely between inspected farms; while some reported little to no crop loss, others reported nearly total crop loss. Lettuce that was infected while very young did not produce heads, and while older lettuce heads sometimes completely succumbed to rot, in other cases only the outer leaves developed necrotic spots. Even lower levels of disease severity caused significant losses

to the sellable yield, however, because the necrotic outer leaves required removal prior to sale. Brown reported that greater disease prevalence was associated with the sudden drop in temperature to 22 degrees Fahrenheit, a lack of protection from wind, and the use of incompletely degraded compost material that could harbor the pathogen.

Unusually cold and humid conditions, the presence of the pathogen, and the use of susceptible lettuce cultivars also led to the first reported BLS outbreak in California in 1963 (Schroth et al., 1964). *Xhv* was identified morphologically, and the researchers posited that nearby weeds or contaminated seeds were possible inoculum sources. To prevent further BLS outbreaks, they suggested eliminating inoculum sources and limiting moisture conditions that could spread the pathogen, especially by forgoing the use of overhead irrigation. They investigated chemical application for the prevention of pathogen spread. Copper application (0.2% concentration) was effective but caused plant injury. Streptomycin application (50 or 100ppm) was also effective and did not result in plant injury, but it did not become a common practice due to a lack of registered status for use on lettuce crops.

Additional epidemiological studies began following a third period of severe outbreaks in Ohio, Florida, and Quebec during the 1990s (Sahin and Miller, 1997; Pernezny et al., 1995; Toussaint et al., 1999). These outbreaks spanned multiple growing seasons and caused significant losses to sellable yield. At the time, Ohio had the third largest leaf lettuce production, and the significance of these outbreaks prompted further research into preventing future sources of inoculum. Isolates from the Ohio outbreaks were identified as *Xhv* through morphological and biochemical observations, and their pathogenicity was confirmed by performing Koch's postulates (Sahin and Miller, 1997). The researchers observed a significant correlation between the disease severity of inoculated lettuce cultivars and the population level of *Xhv* recovered

from inoculated tissue. They also reported that the susceptible cultivars inoculated with *Xhv* produced a significantly greater proportion of infested seed compared to the inoculated resistant cultivars. Lastly, they showed that treating the seeds harvested from infected plants with 0.52% sodium hypochlorite, a surface sterilizing agent, resulted in significant reduction in the level of seed contamination. These results indicated that the pathogen might be transmitted externally on seed.

However, studies of contemporaneous BLS outbreaks on the western coast of the United States yielded different results. In California*, Xhv* was not isolated from seed lots prior to planting, including those that later developed BLS (Umesh et al., 1996). It was hypothesized that another source of inoculum caused the outbreaks (Barak et al., 2001, 2002). These researchers found that the pathogen could survive on nearby weed species and on crop debris between growing seasons and thus serve as an inoculum source in the following season. Another major output of this study is the development of an *Xhv*-specific, PCR-based detection method; however, our attempts to validate this detection method show that it also detects other *X. hortorum* species and so is not truly *Xhv*-specific. A LAMP-based method to distinguish members of the *X. hortorum-X. hydrangeae* species complex was later developed, however this likewise does not distinguish *Xhv* from its close relatives *X. hortorum* pv. *gardneri* and *X. hortorum* pv. *cynarae* (Dia et al., 2022a).

The results of these studies have led to the current strategies for managing bacterial leaf spot, including the use of certified clean seed, the elimination of nearby weeds and crop debris, and the practices of rotation with slow-maturing crops and long fallow periods between plantings to reduce *Xhv* survival in the environment into the next growing season. Copper-based bactericides are commonly used to reduce disease impact, but they can cause injury to plant

tissues and environmental damage; the evolution of copper-tolerant strains may further limit their effectiveness in the future (Abbasi et al., 2015; Khanal et al., 2020). Another effective strategy is the use lettuce cultivars that are resistant to BLS, and there are several breeding programs around the United States currently working on the identification of resistant germplasm and the transfer of genetic sources of resistance into lettuce cultivars (Bull et al., 2007).

Breeding lettuce for favorable traits like broad leaves, improved shelf life, and pleasant taste reduces the selective pressure to maintain disease resistance genes, and so breeding for resistance usually involves crossing desirable lettuce cultivars with wild lettuce accessions that retain genetic sources of resistance (Brown and Rant, 2013; Damerum et al., 2020). Offspring are selected that retain both desirable traits and disease resistance. Common techniques to breed new lettuce cultivars include the backcrossing, single seed descent (SSD), or pedigree methods. For backcrossing, a desirable line (with desirable horticultural traits) and a resistant accession (commonly a primitive accession) are crossed to produce the F_1 generation, and then a heterozygous F_1 individual is backcrossed with one of the parents (recurrent parent). Another desirable individual is selected among those progenies for its desired traits and then backcrossed again with the recurrent parent. This process may be repeated many times until the desired traits are fixed, or all progenies are homozygous for the resistant and other desired traits. For SSD (Brim, 1966) the F_1 population is created in the same way as above, but then the F_1 individual is self-pollinated to generate the F_2 generation. The F_2s are planted and allowed to self-pollinate to produce F_3 families. The F_3 families are planted to produce the F_4 seed, and single seed is collected from each plant to produce the subsequent generations. This process is repeated through the F_7 generation, at minimum, until the desired traits are fixed. The SSD method is mostly used to produce mapping populations for genetic studies. Lastly, the pedigree method

(Lowe, 1927) resembles SSD, but progenies with desirable traits are selected in the F_2 generation for their horticultural characteristics and allowed to self-pollinate to produce F_3 families. Selection occurs on each generation and not just after many generations. The goal of each method is to achieve a breeding line that is homozygous and fixed for the desired traits; for example, an ideal candidate might have the broad leaves of the parent *L. sativa* cultivar and the disease resistance of the parent *L. serriola* accession. Genetic sources of resistance to *Xhv* are discussed in section 1.4 of this dissertation.

1.2 Lettuce Bacterial Leaf Spot Pathogen: Description and Taxonomic History

A little over a century ago, United States Department of Agriculture (USDA) researcher Nellie Brown identified the causal pathogen of lettuce bacterial leaf spot (Brown 1918). Under the microscope, she observed bacterial streaming from stem-cross sections cut from infected lettuce. She described the bacterial cells as motile, rod-shaped, aerobic, and gram-negative with bi-polar flagella. In cultures, she observed the bacterial growth was cream-yellow, with round or elliptical colonies. She used her isolate to inoculate healthy lettuce plants, and later she observed the same BLS symptoms and re-isolated bacteria with the same description. Her completion of Koch's postulates established this isolate as the pathogen causing BLS.

Brown named the newly identified lettuce pathogen *Bacterium vitians* Brown 1918 using the classification scheme proposed by Erwin Smith in 1905. The name of the organism would change several times over the 20th century, as new technologies were developed for describing bacteria and philosophies shifted toward the use objective and multi-trait, rather than subjective and single-trait, methods of classification. The original three genera of plant pathogens were defined by a single trait, motility: *Bacterium* (non-motile), *Pseudomonas* (motile with polar

flagella), and *Bacillus* (motile with peritrichous flagella) (Migula 1900). The genus *Xanthomonas* (from the Greek xanthos, "yellow," and monas, "unit") was later proposed to include bacteria that were rod-shaped, gram negative, motile by polar flagella or nonmotile, mucoid, and yellow when grown on nutrient or potato agar (Dowson 1939). Following this proposal, the BLS pathogen was reclassified as *Xanthomonas vitians* Dowson 1943.

In 1980, the International Code of Nomenclature of Prokaryotes set the criteria for all correct names: to be valid, bacteria needed to be published, to have an authentic type strain, and to have a modern description to distinguish it from other species in its genus, one which is based on multiple traits rather than a single environmental character like host range or pathogenicity (Lapage et al., 1992). Many existing species that did not meet these criteria were re-classified under a comprehensive species name that did meet the criteria, including *Xanthomonas vitians* which was reclassified under the correct name *Xanthomonas campestris* (Pammel 1895) Dowson 1943. Additionally, The International Standards for Naming Pathovars of Phytopathogenic Bacteria provided criteria for naming pathovars, a term which refers to a group of strains pathogenic on the same host. Every pathovar must be validly published with a defined pathotype strain (Young et al., 1978). Following this convention, the BLS pathogen was again reclassified to *Xanthomonas campestris* pv. *vitians* (Brown 1918) Dye 1978, referring to the group of *X. campestris* strains pathogenic on lettuce. Plant pathologists saw the pathovar designation as a temporary solution and intended to re-establish many of the plant pathogenic species as sufficient data became available (Dye et al., 1980).

In 1995, Vauterin et al. used several molecular and biochemical methods to assess the relatedness of the *X. campestris* pathovars, including DNA-DNA hybridization, restriction patterns of rRNAs, protein expression via SDS-PAGE, and fatty acid methyl ester (FAME)

profiles. The synthesis of these data was used to re-classify more than 15 species of *Xanthomonas*, and many pathovars of *Xanthomonas campestris*, each with a new description and designated type strain. This included their proposed name for the BLS pathogen: *X. hortorum* pv. *vitians* Vauterin et al., 1995. However, because they did not propose new pathotype strains with descriptions, their pathovar designations were illegitimate. In 2020 the BLS pathogen was reclassified as *X. hortorum* pv. *vitians* Morinière *et al*. 2020 based whole genome phylogenetic analysis, and the publication included the required type and pathotype strain descriptions (Morinière et al., 2020).

Maintaining a clear and accurate record of bacterial taxonomy translates to the practical application of technology for managing plant diseases. For example, pathogen detection following an outbreak in an agricultural setting would allow diagnosticians to provide precise recommendations to prevent further damage or future outbreaks. They might recommend specific ways to control inoculum levels or cultivars to grow in the next planting that have genetic resistance to the pathogen. Furthermore, use of detection methods can provide insight into the distribution and movement of a pathogen, and with further study we can pursue an understanding of pathogen diversity and evolution, as well as plant-pathogen interaction. This, in turn, facilitates the development of new cultivars with genetic resistance to the plant pathogen. Therefore, studies of bacterial taxonomy provide a strong foundation for plant disease management and the development of new management strategies.

1.3 Pathogen Diversity

In addition to epidemiological research, the 1990s kickstarted studies to understand *Xhv* diversity. A 2003 study described two distinct groups of strains: Group A, which included 28

isolates from an Ohio outbreak that caused both localized and systemic symptoms in lettuce, and Group B, which included 10 strains from Ohio and California outbreaks that only caused localized symptoms (Sahin et al., 2003). These groups could also be distinguished by their genomic profiles, determined using data sets derived from enterobacterial repetitive intergenic consensus (ERIC), BOX-, and repetitive extragenic palindromic (REP) PCR results.

In the 2000s, the increasing efficiency and lowering cost of DNA sequencing platforms enabled a movement toward sequence analysis for understanding microbial diversity. Young et al.(2008), described a multi-locus sequence typing (Maiden et al. 1998) scheme for *Xanthomonas* strains using the four housekeeping gene sequences of *rpoD*, *dnaK*, *fyuA*, and *gyrB*. These sequences were concatenated and aligned, and the alignments were used to generate a phylogenetic tree based on nucleotide differences and using a maximum likelihood statistical model to estimate genetic relatedness. This procedure is termed multi-locus sequencing analysis (MLSA), and when applied to a set of 83 *Xhv* strains collected from California, Ohio, and Florida outbreaks, researchers found that the strains clustered into three distinct groups (Fayette et al., 2016). These results supported Sahin et al.'s (2003) claim that genetically distinct groups exist among *Xhv* strains, and the same study also revealed that while there is genetic evidence of some homologous recombination between *Xhv* strains, the population studied were predominantly clonal.

An adaptation of Young's MLSA procedure for *Xanthomonas* strains was applied to another set of 120 *Xhv* isolates collected from BLS outbreaks around the world (Bull et al., 2016). In one iteration, the four housekeeping genes were cut to form a shorter alignment, still capable of distinguishing five *Xhv* clades upon building the maximum likelihood phylogeny (A, B, C, D, and E). A second iteration was completed that added another housekeeping gene, *gap1*,

to the analysis and revealed eight distinct *Xhv* clades (A, B1, B2, C, D, E, F, and G). Importantly, all *Xhv* strains also belonged to a distinct clade from the other *Xanthomonas hortorum* pathotype strains included in the analysis.

1.4 Host-specificity and disease resistance

Host plants have strategies to defend themselves against pathogens, first with constitutive physical and chemical components to prevent colonization and establishment and then with inducible defenses, such as the triggering of salicylic acid and jasmonic acid production against biotrophic and necrotrophic pathogens, respectively. *Xanthomona*s species are hemibiotrophic; they can survive as epiphytes on plant surfaces without causing symptoms, but will later enter plant tissues through wounds, stomata, and hydathodes and their establishment in the tissues leads to necrosis (Dia et al., 2022b). Pathogen recognition at the plant cell surface triggers the basal immune response: an oxidative burst that causes damage to bacterial cells and amplifies salicylic acid signaling, which in turn mediates a variety of responses to strengthen plant cell defenses (War et al., 2011). *Xanthomonas* strains secrete polysaccharides to hinder this recognition. This back and forth between plant defense and pathogenicity is often described as an arms race; as the host evolves pathogen recognition and defense strategies, the pathogen evolves new ways to evade detection and colonize tissue.

A critical step in this interaction is the delivery via the type III secretion system (T3SS) of bacterial effector proteins, which serve to interfere with the host's basal immune response and facilitate bacterial establishment. The T3SS includes a needle-like protein structure to deliver effectors directly from the bacterial to the plant cell cytoplasm (Hueck, 1998). Effectors can preferentially bind to and inhibit plant signaling molecules (Macho et al, 2014; Zhou et al, 2014;

Xiang et al, 2008), tag resistance-associated plant proteins with ubiquitin, a protein signal that triggers degradation by the host proteasome (Rosebrock et al, 2007), degrade the proteins that suppress JA production, which triggers lower SA levels and thereby reduces the effectiveness of SAR (Gimenez-Ibanez et al., 2014). It is hypothesized that there are many more mechanisms by which effectors overcome plant defenses that are yet to be determined (Furutani et al, 2008).

Xanthomonas, and some *Ralstonia*, species have an additional class of effectors known as transcription activator-like effectors (TALEs) (Boch et al., 2009). They are composed of a nuclear localization signal (NLS), a central repeat region with repeat variable di-residues (RVDs) at the 12-13 position of each larger repeat, and a transcription activation domain. The NLS directs the TALE to enter the plant nucleus. The RVDs correspond to effector binding elements (EBEs) in the plant genome and allow for specific binding of the TALE to that sequence. The transcriptional activation domain triggers the plant to transcribe its own susceptibility (S) genes. A common class of susceptibility genes are called Sugars Will Eventually be Exported Transport (SWEET) genes that, when activated, upregulate sugar transport and thereby provide increased nutritional stock for bacterial populations to grow (Chen et al., 2010; Verdier et al., 2012).

Effector genes are often gained and lost through horizontal gene transfer, evidenced by their usual closeness to mobile genetic elements. The local environment then provides the selective pressure for their maintenance or loss in the genome (McCann and Guttman, 2008; Merda et al., 2017; Sarkar et al., 2006; Hajri et al., 2009). Host gene expression is one those selective pressures; one crucial mechanism is through plant resistance (R) genes that recognize effectors and trigger a secondary immune response (Scofield et al., 1996). This hypersensitive response (HR) is a programmed cell death that limits the spread of bacteria through the tissues.

When effectors fail to promote pathogenicity and instead are recognized by the host and trigger HR, they are termed as 'avirulence genes'.

We hypothesize that this type of gene-for-gene interaction is at play for *Xhv* and its lettuce host, at a sub-pathovar (race) level. *Xhv* strains can be divided into three races based on lettuce cultivar interactions: *Xhv* race 1 strains trigger HR in *L. sativa* cultivars Little Gem, Pavane, La Brillante, and PI358001-1 and *L. serriola* accession ARM-09-161-10-1; *Xhv* race 2 strains trigger HR in *L. serriola* PI491114; and *Xhv* race 3 strains trigger HR in *L. serriola* ARM-09-161-10-1 (Bull et al., 2016; Sandoya et al., in preparation). It is possible that differences effector repertoires between the *Xhv* races are responsible for this cultivar-specificity, and the likely presence of R-genes in the lettuce genome supports this hypothesis. For La Brillante and PI358001-1, these loci are termed *Xanthomonas resistance 1* (*Xar1*) and *Xanthomonas campestris vitians resistance* (*Xcvr*), respectively (Hayes et al, 2014; Wang et al., 2016). These loci may contain R genes that interact with a *Xhv* race 1 effector to trigger resistance in the form of the hypersensitive response. Further, similar interactions may be at play for *Xhv* race 2 and 3 determinant cultivars PI491114 and ARM-09-161-10-1, respectively.

The preliminary *Xhv* diversity study by Bull et al., 2016 with *Xhv* provided additional evidence to suggest that the HR phenotypes are dictated by the genotypes of the *Xhv* races. In addition to using MLSA to define *Xhv* sequetypes, they found that those sequetypes corresponded to the race-specific HR phenotypes. MLSA group B1, B2, D, E, F, and G strains belonged to *Xhv* race 1, MLSA group A strains belonged to *Xhv* race 2, and MLSA group C strains belonged to *Xhv* race 3. Host-specificity due to effector repertoire variation was demonstrated previously (Hajri et al., 2009; Sarkar et al., 2006), including at the level of cultivar for *Xanthomonas euvesicatoria* pv. *perforans* (Astua-Monge et al, 2000). Further research was

necessary to determine whether differences in *Xhv* effector repertoires were responsible for their cultivar-specificity in HR induction.

1.5 Thesis Objectives and Justification

I. **To confirm the *Xhv* race reactions on lettuce cultivars and introduce new *Xhv* strains and lettuce cultivars into the analyses.** The three hypothesized races of *Xhv* were previously defined by their different host reactions: hypothetical *Xhv* race 1 triggered HR on *L. sativa* cv. Little Gem; hypothetical *Xhv* race 2 triggered HR on *L. serriola* PI491114; hypothetical *Xhv* race 3 triggered HR on *L. serriola* ARM-09-161-10-1. Several of the other reactions, such as that of hypothesized *Xhv* race 1 on *L. serriola* ARM-09-161-10-1, had not yet been evaluated. These reactions and the reactions of new *Xhv* strains on novelplant introduction lines were evaluated.

II. **To characterize *Xhv* effector repertoires and identify candidate avirulence genes that may be responsible for cultivar-specificity.** We completed a comparative genomic analysis of *Xhv* strains and related type and pathotype strains and identified two effector genes, *xopAQ* and *xopAF2*, that are only present in *Xhv* race 1 and 3 strains, which trigger HR in *L. serriola* accession ARM-09-161-10-1. This study helped to generate the hypothesis that one or both genes serve as avirulence genes are recognized by an R-gene in ARM-09-161-10-1 to trigger HR.

III. **To develop a PCR-based detection method for *Xhv* that is both pathovar-specific and race-specific.** Comparative genomic analysis of *Xhv* strains and related type and

pathotype strains revealed genetic regions that were specific to *Xhv*, and did not include other *X. hortorum* strains, as well as regions that were specific to one of the three known races of *Xhv*. These regions were used as sequence targets to develop pathovar- and race-specific detection methods. Our touchdown PCR methods will be a crucial tool for tracking the movement of the pathogen and its races in global outbreaks, allowing clinicians to provide farmers with precise cultivar recommendations to plant crops resistant to *Xhv* detection in prior plantings, and rapidly detecting *Xhv* in seed lots for clean seed certification programs.

IV. **To investigate the role of putative avirulence genes in cultivar-specific HR elicitation.** The two putative avirulence genes that were identified in only *Xhv* race 1 and 3 strains were expressed in *Xhv* race 2 strains to determine whether those genes are sufficient for triggering HR in ARM-09-161-10-1, which typically experiences HR in response to race 1 and 3 strains and disease in response to race 2 strains. Identifying the avirulence genes functioning in *Xhv* may help to identify interacting R-genes in ARM-09-161-10-1, which would later allow plant breeders to rapidly screen germplasm for that R gene sequence. More generally, this study could advance our understanding of the evolution of pathogenicity in phytobacterial pathogens.

Chapter 2

Resistance phenotype defines the races of *Xanthomonas hortorum* pv. *vitians* strains

2.1 Introduction

Outbreaks of bacterial leaf spot on lettuce, although sporadic, can devastate the lettuce industry. The causal bacterium, *Xanthomonas hortorum* pv. *vitians* (*Xhv*; previously *Xanthomonas campestris* pv. *vitians*), was first described in 1918 after its isolation from symptomatic lettuce tissue from an outbreak in South Carolina (Brown 1918). The symptoms were described as chlorosis, water-soaking, and small necrotic spots that later merged into large black lesions. The completion of Koch's postulates confirmed *Xhv's* pathogenicity, and subsequent outbreaks throughout the 20th and into the 21st century demonstrated that crop loss due could be as high as 100% (Lu and Raid, 2013). *Xhv* strains were isolated from around the world for epidemiological study, and researchers found both phenotypic and genotypic variation.

Plant reactions to *Xhv* strains differentiate them from other pathogens and they also differentiate the groups within *Xhv*. The pathovar designation is used to group bacterial strains that are pathogenic on the same host. All strains classified as *Xhv* are pathogenic on lettuce (Morinière et al., 2020; Young et al., 1978). Sahin et al. (2003) saw differences in whether *Xhv* strains produced both local and systemic infections, or only local infections, and called pathogenic groups A and B, respectively. Biochemical analyses, including fatty acid profiling, serotyping, carbon substrate analysis, and protein fingerprints provided further evidence to support these groupings. Another study, Bull et al. (2016), revealed that upon inoculation of many lettuce cultivars and accessions that some groups of *Xhv* strains trigger plant resistance in

the form of a hypersensitive response (HR). The HR is a form of resistance against biotrophic pathogens in which the plant detects the pathogen and kills off its own infected tissue to limit the spread of the pathogen to healthy tissues (Morel and Dangl, 1997). For *Xhv*, the HR phenotype has been used to describe three hypothetical races: strains that cause HR in *L. sativa* cultivar Little Gem belong to hypothetical *Xhv* race 1, those that cause HR in *L. serriola* PI491114 belong to hypothetical *Xhv* race 2, and those that cause HR in *L. serriola* ARM-09-161-10-1 only belong to hypothetical *Xhv* race 3.

Xhv strain diversity has also been demonstrated through genotyping, which provided information on genetic and evolutionary relationships, as well as evidence to support the infraspecies classification schemes. Morinière et al. 2020 presented a maximum likelihood phylogeny built from the proteomes of *Xhv* and related strains, and they showed that *X. hortorum* groupings based on host-specificity correlated to groupings based on genetic relatedness. This is likely due to the clonality of *Xhv* strains (Fayette et al., 2016), and the phylogeny showed a correlation between the genetic relatedness of the strains and the pathovars designated within *X. hortorum*. Bull et al. (2016), showed through multi-locus sequence analysis that *Xhv* sequetype correlated to the compatible or incompatible response observed upon inoculation into several lettuce cultivars. Rosenthal et al. (2022) corroborated the results of Morinière and Bull using whole genome sequences of various *Xhv* strains and related *Xanthomonas* strains; *Xhv* strains clustered together and apart from other *X. hortorum* strains, and they also clustered in groups based on the cultivar upon which they induce HR.

Infraspecies classifications are useful frameworks for studying genetic sources of pathogenicity and for developing effective detection methods. Defining the set of *X. hortorum* strains that are pathogenic on lettuce, *Xhv*, established an in-group that could be compared to

out-group strains, the other *X. hortorum* strains that are pathogenic on other hosts. Comparative genomic analysis would later reveal *Xhv*-specific sequences that may be responsible for host-specificity or that can be utilized for pathovar-specific detection (Rosenthal et al., 2022). Similarly, groups of strains within the same pathovar can be classified into races, defined by their similar responses on differential hosts. This additional method of characterizing strain diversity facilitates study on cultivar-specificity and the genetic origin of pathogenicity, and race-level detection allows for the deployment of specific plant hosts to combat the pathogenic races present.

The race descriptions for *Xhv* were based on preliminary data from Bull et al. (2016), and so this study sought to confirm the *Xhv* race reactions on lettuce cultivars and introduce new *Xhv* strains and lettuce cultivars into the analyses. The *Xhv* strains tested here included representatives of the three hypothesized races, as well as six suspect *Xhv* strains isolated from Florida outbreaks. Multi-locus sequence analysis was completed to predict which strains would elicit HR in which cultivar based on the results of Bull et al. 2016. Each strain was then syringe-inoculated into the three previously described race-determinant lettuce cultivars and three new plant introduction lines: PI342498, PI667690, PI667709. An additional set of suspect *Xhv* strains were received from a French outbreak and these strains were also evaluated using MLSA. Our results confirm the *Xhv* race reactions, and we reveal that the new strains isolated in Florida likely belong to *Xhv* race 1 due to their HR elicitation on *L. sativa* cv. Little Gem. The Florida PI lines appear to have HR-based resistance to either *Xhv* race 1 or race 2 strains. The French strains we received cluster with strains known to belong to *Xhv* races 1 and 2 in MLSA, so we hypothesize that these strains would elicit HR in *L. sativa* cv. Little Gem and *L. serriola* PI491114, respectively, and future HR screening will be done to evaluate this hypothesis.

2.2 Materials and Methods

2.2.1 Bacterial strains and growth conditions

The bacterial strains included in this study are listed in Table 2-1. We included a representative strain for each of the three hypothesized *Xhv* races (Bull et al., 2016), as well as six strains isolated during an outbreak in a Florida lettuce field in 2017 and one strain isolated from an outbreak in a Florida field outbreak in 1995 (Pernezy et al., 1995). Eight strains were also received from France and added to our sequence-based analysis. All *Xhv* strains were routinely cultured on nutrient agar (NA) plates and incubated at 28°C. Colony suspensions to serve as PCR templates were created by streaking strains for colonies in pure culture and then mixing a single colony into each 30 µL aliquot of sterile water using a vortex.

2.2.2 BoxA1 rep-PCR of Florida isolates and re-isolates

To determine whether the isolates and re-isolates we received from the Florida 2017 BLS outbreak were identical, rep-PCR was completed using colony suspensions of these strains. The reaction was run with a primer targeting the repetitive palindromic BOX1A sequences known to be dispersed throughout bacterial genomes and the intervening sequences were amplified (Versalovic et al., 1994). The 25 µL reaction mixes contained 1X reaction buffer, 0.16 mg/mL BSA, 10% DMSO, 1.25 mM dNTPs, 1.76 mM BoxA1R primer, 0.08 U/µL Taq polymerase, and 1 µL of colony suspension. The cycling steps included two minutes at 95 °C, three seconds at 94 °C, thirty seconds at 92 °C, one minute at 40 °C, eight minutes at 65 °C, cycling twenty-nine times to step two, and eight minutes at 65 °C. The resulting fingerprints, or series of bands, were visualized using gel electrophoresis (60 V for 17 hours in 1% agarose gel) and the results for

each isolate and reisolate were compared to a control strain that was run alongside: *Xhv* B0347, of hypothesized race 1. PCR run with sterile water was included as a negative control. A 1 KB DNA ladder from New England Biolabs (Ipswich, Massachusetts, USA) was run alongside in this gel and all subsequent gels in this study.

Table 2-1. Bacterial strains included in this study. Multi-locus sequetypes are included for the strains that were analyzed in Bull et al., 2016 and the strains determined by this study. The asterisk applies to strain identities determined through 16S rRNA subunit sequence analysis.

Organism	**Strain**	**MLSA Sequetype**	**Origin**	**Collector or Citation**
X. hortorum pv. *vitians*	BS0341	B1	Salinas, California, USA	S. Koike
	BS3270	B1	Florida, USA	Pernezy et al., 1995
	BS0336	D	Salinas, California, USA	J. Barak
	BS0339	B2	Salinas, California, USA	J. Barak
	BS0340	B2	Salinas, California, USA	J. Barak
	BS0344	E	Salinas, California, USA	J. Barak
	BS0347	B1	Salinas, California, USA	J. Barak
	BS0348	E	Salinas, California, USA	J. Barak
	BS0541	F	Colorado, USA	S. Koike
	BS2869	G	Santa Maria, California, USA	S. Koike
	BS3046^{PT}	B1	Zimbabwe	R. Bailey

	BS3129	B2	Canada	V. Toussaint
	BS3298	D	CA, USA	C. T. Bull
	BS2995	A	Ithaca, NY, USA	W.H. Burkholder
	BS3043	A	USA	Sahin et al., 2003
	BS3127	A	Canada	V. Toussaint
	BS2861	C	King City, CA, USA	S. Koike, Rianda
	BS2862	C	King City, CA, USA	S. Koike, Rianda
X. hortorum pv. *hederae*, pathogen of English ivy	BS3107	N/A	USA	(Arnaud 1920) Dye 1978. Vauterin et al., 1995
Unknown, suspect *Xhv* strains isolated from infected lettuce	BP5150	*Pantoea deleyi**	Florida, USA	W. Wadlington and G. Sandoya
	BP5151	B2	Florida, USA	W. Wadlington and G. Sandoya
	BP5152	*Pantoea anthophila**	Florida, USA	W. Wadlington and G. Sandoya

	BP5153	B2	Florida, USA	W. Wadlington and G. Sandoya
	BP5154	B2	Florida, USA	W. Wadlington and G. Sandoya
	BP5164	B2	Florida, USA	W. Wadlington and G. Sandoya
	BP5168	B1	Florida, USA	Pernezy et al., 1995
	BP5156	B2	France	M.P. Starr, 1949
	BP5157	B1	France	C. Audusseau, 1994
	BP5158	A	France	L. Morinière, 2016
	BP5159	B2	France	L. Morinière, 2016
	BP5160	B2	France	L. Morinière, 2016
	BP5161	B1	France	L. Morinière, 2016
	BP5162	B1	France	D. Chapulliot, 2016
	BP5163	B2	France	D. Chapulliot, 2016

2.2.3 Multi-locus sequence analysis

PCR amplification of the four housekeeping genes *rpoD* (693 bp), *fyuA* (762 bp), *gyrB* (522 bp), and *gap1* (774 bp) was completed in 25 µL reactions consisting of 1X ImmoMix (Meridian Bioscience, Cincinnati, OH), 0.5 uM of each forward and reverse primer, and 1 µL of colony suspensions. The cycling steps included ten minutes at 95 °C, thirty seconds at 95 °C, one minute at 58 °C, one minute at 72 °C, cycling thirty-four times to step two, and six minutes at 72 °C. Following PCR amplification, gel electrophoresis was completed to confirm the fragment sizes. PCR products were then purified using the EXOSAP-IT (Thermo Fisher Scientific, Waltham, MA, USA). Cleaned products were submitted for Sanger sequencing at the Pennsylvania State University (PSU) Genomics Core Facility. Forward and reverse sequencing reads were aligned to the corresponding gene sequence for a reference strain, the *X. hortorum* pv. *vitians* type strain BS3046 to obtain a consensus sequence for each strain. Sequence analysis was completed using Qiagen CLC Genomics Workbench version 21 (Valencia, CA, USA). The consensus sequences for each strain were joined in the order of *rpoD*, *fyuA*, *gyrB*, and *gap1* to form a 2,751 bp long sequence (Zacaroni et al., 2012). These merged sequences were aligned and trimmed, and model testing was completed to determine the appropriate nucleotide substitution model for this sequence data. A maximum likelihood phylogeny was constructed using the neighbor-joining algorithm and the general time reversible substitution model with rate variation (4 categories) and topology estimation (Tavaré, 1986). Bootstraps were set to 1000 and a threshold of 60% was applied to the resulting phylogenetic tree. Branch lengths corresponded to the expected number of nucleic acid substitutions per site.

2.2.4 Analysis of 16S sequences

Two of the suspect *Xhv* strains isolated from the 2017 Florida outbreak did not produce fragments following PCR for the four housekeeping genes used for MLSA, and so these were selected for analysis of the 16S sequences to determine the identity of these strains. The 16S rRNA subunit was amplified via colony PCR as described previously (Lane et al., 1985), the fragment sizes were verified using gel electrophoresis, and after cleaning the PCR product with EXOSAP-IT (Thermo Fisher Scientific, Waltham, MA, USA), and the cleaned PCR products were submitted for sequencing at the PSU Genomics Core Facility. 16S rRNA sequences were trimmed and then submitted as queries into the Ribosomal Database Project search function with the option to compare to type strains only selected. The best matches reported were those with the highest percent identity.

2.2.5 Resistance screening in lettuce

Xhv strains were spread on NA and incubated for three to four days at 28 °C to achieve single colonies, and then individual colonies were sub-cultured on NA for two days to produce lawns of bacteria. Pellets of bacterial lawn were suspended in 10 mL of sterile 0.1 M, pH 7.0 phosphate buffer and adjusted to approximately 10^8 CFU/mL, which corresponded to OD600 = 0.60 +/- 0.05 using a spectrophotometer. The suspended bacteria served as inoculum for the HR screen with four lettuce lines: *Lactuca sativa* cvs. Little Gem and cv. Vista Verde, *L. serriola* PI491114 and ARM-09-161-10-1. Three novel lettuce lines that were identified as resistant by a Florida breeding program were also included: PI342498, PI667690, PI667709. Lettuce plants were germinated in soil within a growth chamber set to 60% humidity, 10 ℃, and in total darkness for two days. They were then moved to a growth room kept at 25 ℃ and under a 12/12

light and dark cycle until mature. Bacteria were inoculated into leaf tissue via needleless syringe on the underside of the leaves of lettuce plants grown for 4-6 weeks. Two leaves were inoculated per plant, three plants (replicates) were inoculated per treatment, and the experiment was completed twice. Plants were then incubated for a week, and the resulting responses were rated at 36 hours post-inoculation (hpi), 48 hpi, 72 hpi, and one-week post-inoculation (wpi). The rating scale was as follows for 36, 48, and 72 hpi: 0 for no reaction, + for slight discoloration, ++ for discoloration and cell collapse, and +++ for discoloration, cell collapse, and drying. Ratings for wpi were defined as 'C' for compatible (disease) interaction and 'HR' for hypersensitive response or incompatible interaction.

2.3 Results and Discussion

2.3.1 Rep-PCR showed that 2017 Florida *Xhv* isolates and re-isolates had identical banding patterns.

Our lab received both isolates and re-isolates of the strains from the Florida 2017 outbreak, following a demonstration of Koch's postulates completed by our University of Florida colleagues. We intended to demonstrate that the isolates and re-isolates were the same using a molecular method to further prove their pathogenicity. Comparing the DNA fingerprints of each 2017 Florida *Xhv* isolate to that of its re-isolate revealed them to be identical (Figure 2-1). BP5150 and BP5151 (lanes 1-2) match their reisolates (lanes 11-12). BP5152 (lanes 3-4) matches its reisolate (lane 13). BP5153 (lanes 5-6) matches two of the other Florida strains, BP5154 (lanes 7-8) and BP5164 (lanes 9-10), and all three match their re-isolates (lanes 14, 15, and 16, respectively). The hypothesized *Xhv* race 1 strain run in lane 17, BS0341 is nearly

identical to the BP5153, BP5154, and BP5164 strains but for a missing band near 1.1 KB, indicated by an arrow in Figure 1. No bands appeared for the sterile water control.

These results support the experiments that were performed at the University of Florida that confirmed the pathogenicity of these strains and provided the final demonstration of Koch's postulates. They also suggest that the strains BP5153, BP5154, and BP5156 are more genetically similar to each other and to BS0341 than they are to the other strains included in this analysis. Further analysis was necessary to understand the relatedness of these strains and to determine their identities.

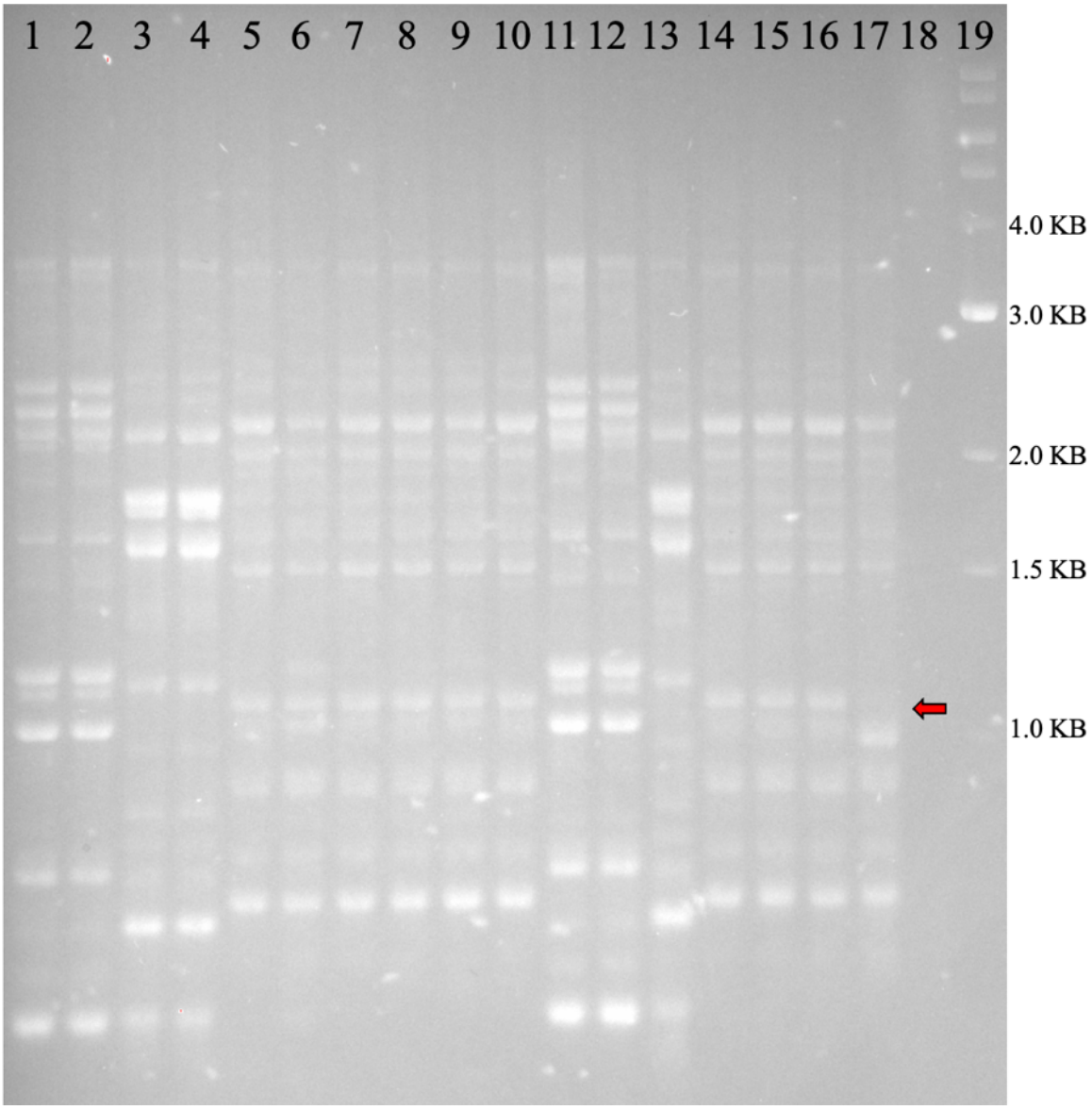


Figure 2-1. Rep-PCR with BoxA1 primer for 2017 FL isolates and re-isolates[1]. 1.5% agarose was gel run for 17 hours at 60V. Lane 1 was BP5150, lane 2 was BP5151, lanes 3-4 were BP5152, lanes 5-6 were BP5153, lanes 7-8 were BP5154, lanes 9-10 were BP5164, lane 11 was the re-isolate of BP5150, lane 12 was the re-isolate of BP5151, lane 13 was the re-isolate of

[1] These results were published in Sandoya, G.V., **Rosenthal, E.R.**, Rodrigues-Porto, L., Wadlington, W.H., Bull, C.T., Simko, I., Carroll, A. (2022). "Germplasm Resistant to Race 1 of *Xanthomonas hortorum pv. vitians* (Brown, 1918) Moriniére et al. 2020 Causing Bacterial Leaf Spot of Lettuce (*Lactuca sativa L.*)." **Journal of Plant Pathology**. https://doi.org/10.1007/s42161-022-01123-0.

BP5152, lane 14 was the reisolate of BP5153, lane 15 was the reisolate of BP5154, lane 16 was the reisolate of BP5155, lane 17 was the hypothesized *Xhv* race 1 strain BS0341, lane 18 was a sterile water control, and lane 19 was a 1 kb DNA ladder from NEB (Ipswich, Massachusetts, USA). The red arrow points to a missing fragment in the BS0341 strain that is present in the most similar FL strains BP5153, BP5154, and BP5164.

2.3.2 Suspect *Xhv* strains from Florida outbreak clustered with known *Xhv* MLSA-types.

The PCRs with the *rpoD*, *fyuA*, *gyrB*, and *gap1* genes all produced DNA fragments at the expected sizes. Multilocus sequence analysis with known *Xhv* strains revealed clusters that corresponded to those previously reported by Bull et al. 2016. Their 5-gene MLSA further broke down the sequetypes, adding sequetypes B1, B2, F, and G to their original 4-gene MLSA clusters A, B, C, D, and E. Here we report a 4-gene MLSA as this is sufficient to separate the clusters of strains that are hypothesized to correspond to HR phenotypes on various lettuce cultivars (Figure 2-2).

Regarding the known *Xhv* strains, BS3127, BS3043, and BS2995 cluster together and BS2860 and BS2861 cluster together, as expected, and these groups correspond to Bull's definitions sequetypes A and C, respectively (Figure 2). BS3046 (Bull sequetype: B1), BS0347 (B1), and BS2869 (G) cluster together, BS0340 (B2), BS3129 (B2), BS0339 (B2), and BS0541 (F) cluster together, and BS3298 (D) and BS0336 (D) cluster together. The *X. hortorum* pv. *hederae* type strain, BS3107, formed the outgroup clade separated from all known *Xhv* strains.

The strains from the Florida lettuce outbreaks cluster together. Strains from the 2017 outbreak, BP5151, BP5153, BP5154, and BP5164 cluster together, though the first appears to have some sequence variation that differentiates it from the other three and the other members of

sequetype B2. The 1995 Florida strains, BS3270 and BP5168, form a clade with sequetype D strains, and they also appear to have sequence variation that differentiates it from the other strains in the same cluster. Because the sequetypes B2 and D are made up of members of *Xhv* race 1, these results led us to hypothesize that the FL strains also belong to *Xhv* race 1.

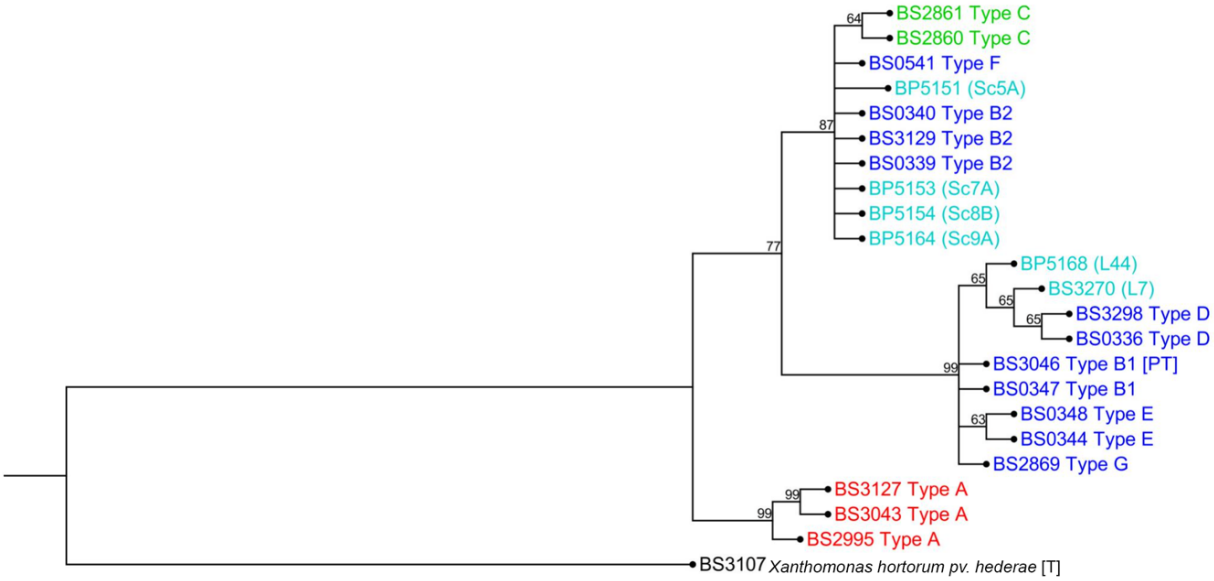


Figure 2-2. MLSA of suspect *Xhv* strains isolated in Florida (teal); hypothesized *Xhv* races 1 (blue), 2 (red), and 3 strains (green); and a *X. hortorum* pv. *hederae* (T) strain (black)[2].

Maximum likelihood phylogeny generated using the neighbor-joining algorithm and the general time reversible nucleotide substitution model (including rate variation and estimated topology). Above the nodes the bootstrap values greater than 60% are shown. The branch lengths represent the expected number of nucleotide substitutions per site.

[2] These results were published in Sandoya, G.V., **Rosenthal, E.R.**, Rodrigues-Porto, L., Wadlington, W.H., Bull, C.T., Simko, I., Carroll, A. (2022). "Germplasm Resistant to Race 1 of *Xanthomonas hortorum pv. vitians* (Brown, 1918) Moriniére et al. 2020 Causing Bacterial Leaf Spot of Lettuce (*Lactuca sativa L.*)." **Journal of Plant Pathology**. https://doi.org/10.1007/s42161-022-01123-0.

2.3.3 Non-*Xhv* Florida isolates belong to *Pantoea* genus and may be opportunistic pathogen.

Two of the six suspect *Xhv* strains that were submitted to the lab did not produce fragments when running PCR for the four-gene MLSA scheme designed for *Xanthomonas* strains (Young et al., 2008; Zacaroni et al., 2012). These strains were BP5150 and BP5152. Amplification of their 16S ribosomal subunit genes was successful, and subsequent sequence analysis revealed their best matches in the ribosomal database project to be *Pantoea deleyi* at 98.9% identity and *Pantoea anthophila* at 99.3% identity, respectively (Table 2-2). A proven cutoff for species classification is 99% identity (Johnson et al., 2019). Additional analysis, such as a measure of average nucleotide identity higher than 95%, could clarify whether BP5150 is *P. deleyi*. BP5151 can be classified as *P. anthophil*a because it surpassed this cutoff. *P. deleyi* was first isolated from eucalyptus experiencing bacterial blight and dieback in Uganda and *P. anthophila* was first isolated from *Impatiens balsamina* and other flowering shrubs in India (Brady et al., 2009). Other closely related *Pantoea* species, such as *P. stewartii* and *P. anantis*, are known to have either necrotrophic phases in their life cycle (Gentzel et al., 2022), or be fully necrotrophic (Stice et al., 2020), respectively. Due to these results, and the confirmation of pathogenicity by Koch's postulates, we hypothesize that some of the necrotic spot symptoms seen on the Florida lettuce samples were caused by *P. deleyi* and *P. anthophila* strains.

Table 2-2. Best matches of 16S rRNA sequencing and alignment using RDP tools. For each strain, there were many high-scoring matches of the same genus. Shown here are the matches with the highest percent identities.

BP Name	Percent Identity	Type Strain	Host and Location of Isolation
BP5150	98.9	*Pantoea deleyi*	Eucalyptus in Uganda (Brady et al., 2009).
BP5152	99.3	*Pantoea anthophila*	*Impatiens balsamina* and other flowering shrubs in India (Brady et al., 2009).

2.3.4 HR screening confirmed three *Xhv* races.

The hypothesized *Xhv* races from Bull et al., 2016, included race 1 strains that induced HR in *L. sativa* cv. Little Gem, race 2 strains that induced HR in *L. serriola* PI491114, and race 3 strains that induced HR in *L. serriola* ARM-09-161-10-1. The strains that represented each race were not fully tested, however, on the other two differential cultivars; for example, HR screens for *Xhv* race 1 strains on *L. serriola* ARM-09-161-10-1 were not previously reported.

Our results revealed a scheme for characterizing the *Xhv* races (Table 2-3). BS3270 induced HR in *L. sativa* cv. Little Gem; BS3127 induced HR in *L. serriola* PI491114; and BS2861 induced HR in *L. serriola* ARM-09-161-10-1. These reactions define the races of *Xhv*, as races 1, 2, and 3 respectively. Additionally, we found that BS3270 triggered HR in *L. serriola* ARM-09-161-10-1 and disease in *L. serriola* PI491114; BS3127 triggered disease in *L. sativa* cv. Little Gem and *L. serriola* ARM-09-161-10-1; and BS2861 triggered disease in *L. sativa* cv. Little Gem *and L. serriola* PI491114. All three of these strains triggered disease in an additional

L. sativa cultivar, Vista Verde, which was previously described to be susceptible to all three hypothetical races. No reactions were visible on plants inoculated with sterile phosphate buffer.

Table 2-3. Summary of HR Screen Results. Symptoms were rated at 36 hours, 48 hours, and 1-week post-infection. Typical HR can be observed as dry, tan tissue with cell collapse, often appearing by 48 hours in the infiltration zone. Disease symptoms include water-soaking and blackening of the infiltration zone. This experiment was repeated once with the same results.

	Treatment:					
	BS3270	BP5168	BS3530	BS2861	BP5154	Phosphate Buffer
Little Gem	**HR**	**HR**	DS	DS	HR*	NR
PI358001-1	**HR**	**HR**	DS	DS	HR	NR
PI491114	DS	DS	**HR**	DS	DS	NR
ARM-09-161-10-1	HR	HR	DS	**HR**	HR*	NR
Vista Verde	DS	DS	DS	DS	ID	NR
PI342498	DS	HR*	HR	DS	DS	NR
PI667690	HR	HR	DS	DS	HR*	NR
PI667709	HR	HR	DS	DS	HR*	NR

DS = Typical disease symptoms appearing at or after 48 hours in more than half of plants rated

HR = Hypersensitive response

HR* = Atypical HR

NR = No reaction

ID = Insufficient data; no plant survived past 36 hours

2.3.5 Florida *Xhv* isolate is likely race 1.

Screening with the Florida isolate BP5154 showed that it triggered an atypical HR in *L. sativa* cv. Little Gem and *L. serriola* ARM-09-161-10-1. The result is atypical because the tissue at the site of inoculation turned black rather than the tan color that is typically observed for HR; in fact, black tissue is more commonly associated with disease. However, we observed clear tissue collapse that occurred rapidly post-inoculation and these are both typically observed for HR, indicating quick pathogen recognition and programmed cell death to limit bacterial spread. BP5154 resulted in typical disease symptoms for *L. serriola* PI491114, and the results were inconclusive for *L. sativa* cv. Vista Verde because the plants died by 36 hpi and before sufficient ratings could be conducted. Based on the HR elicitation in *L. sativa* cv. Little Gem, however atypical, it is likely that BP5154 belongs to *Xhv* race 1.

Further research is necessary to understand why HR presents atypically for this strain, to clarify the strain's race designation, and to evaluate the race for the other Florida *Xhv* strains BP5151, BP5153, and BP5164. We hypothesize due to the genetic similarity of the latter two strains to BP5154, evidenced by the rep-PCR and MLSA results, that these strains would present the same HR results; however, it is unclear how the genetic variation observed for strain BP5151 would contribute to its pathogenicity on lettuce.

2.3.6 *Xhv* strains trigger HR in novel FL PI lines.

The novel line *L. sativa* PI342498 was susceptible to infection with *Xhv* race 1 strain BS3270, *Xhv* race 3 strain BS2861, and Florida isolate BP5154. It appeared to have an atypical HR upon inoculation with *Xhv* race 1 strain BP5168, and a clear HR upon inoculation with *Xhv* race 2 strain BS3530. Another two novel lines, *L. sativa* PI667690 and PI667709 were

susceptible to BS3530 and BS2861, had HRs to BS3270 and BP5168, and had an atypical HR to BP5154. These novel lettuce lines may be useful as germplasm for the development of new cultivars resistant to race 1 or 2 strains.

2.3.7 MLSA shows French strains cluster together and are distinct from known *Xhv* MLSA-types.

The PCRs with the suspect *Xhv* strains from France produced fragments of the expected sizes for *rpoD*, *fyuA*, *gyrB*, and *gap1*. Multi-locus sequence analysis with these strains, a set of *Xhv* strains of known race, and the *X. hortorum* pv. *hederae* (*Xhh*) type strain BS3107 revealed that these French strains cluster with the known *Xhv* strains and apart from the *Xhh* outgroup (Figure 2-3). Strains BP5156, BP5159, BP5160, and BP5163 all clustered with known *Xhv* strains of sequetype B2, while BP5157, BP5161, and BP5162 all clustered with known *Xhv* strains of sequetypes B1, D, E, and G. All of these sequetypes represent strains that belong to *Xhv* race 1, and so these results led to the hypothesis that these seven French strains may also belong to *Xhv* race 1. The strain BP5158 clusters with the known *Xhv* strains of sequetype A, which represents *Xhv* race 2 strains, and so we hypothesized that this strain may belong to *Xhv* race 2. Based on our hypotheses, we might anticipate that the group of seven strains would elicit HR on *L. sativa* cv. Little Gem and *L. serriola* ARM-09-161-10-1, while BP5158 would elicit HR on *L. serriola* PI491114. HR screening with these race-determinant lettuce cultivars would be necessary to know the race of these strains. Such screening has not yet been completed due to the limited availability of the *Xhv* race 2 and 3 differential lettuce lines, *L. serriola* PI491114 and ARM-09-161-10-1.

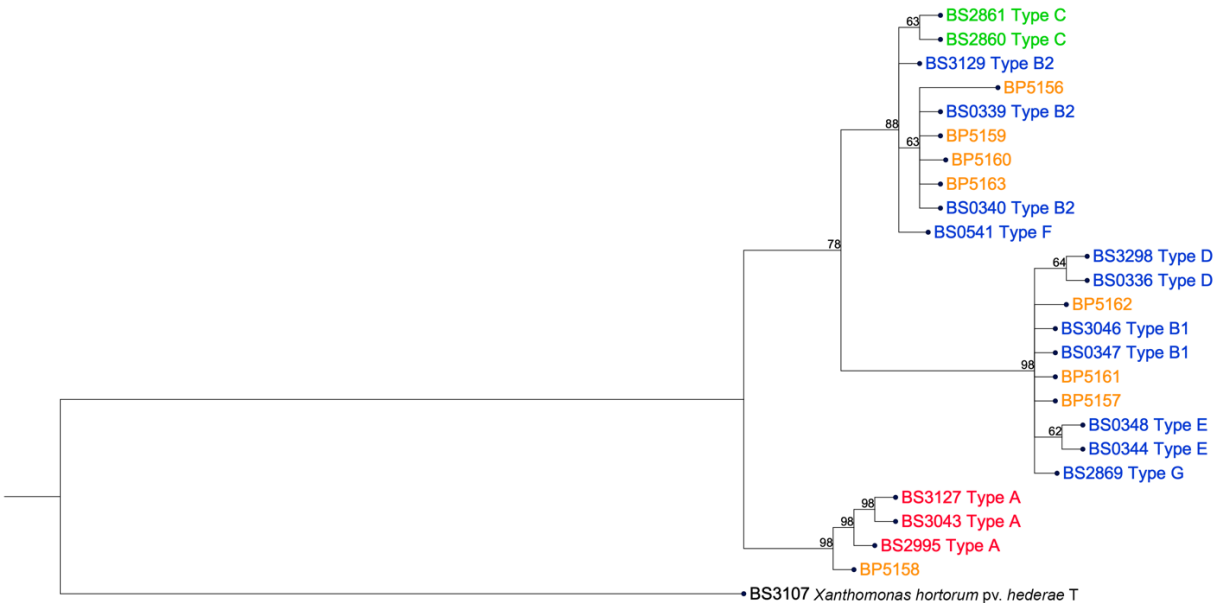


Figure 2-3. MLSA of suspect *Xhv* strains isolated in France (orange); *Xhv* race 1 (blue), 2 (red), and 3 strains (green); and a *X. hortorum* pv. *hederae* (T) strain (black). Maximum likelihood phylogeny generated using the neighbor-joining algorithm and the general time reversible nucleotide substitution model (including rate variation and estimated topology). Above the nodes the bootstrap values greater than 60% are shown. The branch lengths represent the expected number of nucleotide substitutions per site.

2.4 Conclusions

A previous study completed by Bull et al. (2016) had hypothesized three *Xhv* races based on their host reactions in several lettuce lines: the hypothesized race 1 strains triggered HR in *L. sativa* cv. Little Gem, the hypothesized race 2 strains triggered HR in *L. serriola* PI491114, and the hypothesized race 3 strains triggered HR in *L. serriola* ARM-09-161-10-1. They also demonstrated that the MLSA sequetype of the *Xhv* strain corresponded to the hypothesized *Xhv* race: sequetypes B, D, E, F, and G strains belonged to hypothesized race 1, sequetype A strains belonged to hypothesized race 2, and C strains belonged to hypothesized race 3. This study fills in a particular knowledge gap, namely the cross reactivity of each *Xhv* race on the cultivars that were used to hypothesize the races. This is important information as it substantiated the hypothesized race structure for *Xhv* in preparation for publication in a peer-reviewed journal.

Our screening for HR response in various lettuce cultivars confirmed that the *Xhv* sequetype previously identified by MLSA in Bull et al., 2016, correlates to the lettuce cultivar on which HR elicitation occurs. We demonstrated that two sequetype B strains, BS3270 and BP5168 representing the clusters of sequetypes B, D, E, F, and G strains, triggered HR in *L. sativa* cv. Little Gem. The sequetype A strain BS3530 triggered HR in *L. serriola* PI491114 and the sequetype C strain BS2861 triggered HR in *L. serriola* ARM-09-161-10-1. These reactions defined the *Xhv* races 1, 2, and 3, respectively. We then determined that the *Xhv* race 1 strains triggered HR in *L. serriola* ARM-09-161-10-1, while causing disease in *L. serriola* PI491114. The *Xhv* race 2 strain caused disease in both *L. sativa* cv. Little Gem and *L. serriola* ARM-09-161-10-1. The *Xhv* race 3 strain caused disease in *L. sativa* cv. Little Gem and *L. serriola* PI491114. All of these *Xhv* strains caused disease on *L. sativa* cv. Vista Verde.

HR screening of the entire collection of over 100 *Xhv* strains would be necessary to determine whether this correlation between sequetype and race remains consistent. This, as well as screening with additional lettuce cultivars, could also reveal whether there are three *Xhv* races in total or if there are additional races that have not yet been described. In this study, screening a new Florida *Xhv* isolate with the race-determinant lettuce lines and three novel parental inbred lines did not reveal additional *Xhv* races; MLSA led us to predict the Florida isolate would belong to *Xhv* race 1 based on its clustering with sequetype B2 and F strains, and it did produce a pattern of reactions on lettuce cultivars that matches *Xhv* race 1. However, one unexpected finding was the presentation of the hypersensitive response to this strain: normally we observe collapsed, tan tissue but in this case, it was collapsed but black. Even the known *Xhv* race 2 strain appeared to produce this atypical HR upon inoculation with one of the novel Florida lettuce lines. It is possible that the inoculum preparation and culturing or the age of the plant at inoculation changed the appearance of the HR we observed. Johansson et al., 2015 found that the formulation of their culture media influenced their measures of electrolyte leakage, a measure taken to quantify HR, upon inoculation of Arabidopsis plants with *Pseudomonas syringae* pv. *tomato* DC3000. It is also possible that the mechanism by which HR was triggered in these interactions was somehow weakened or delayed, resulting in the death of some tissue due to disease while or before the cell collapse due to HR took effect.

Comparative genomic analysis may be useful for understanding this atypical HR, as well as the differences in HR elicitation of the *Xhv* races. The hypersensitive response is a secondary plant immune response that is known to be elicited by bacterial proteins called effectors. If the primary response occurs, or recognition of small molecules termed pathogen associated molecular patterns (PAMPs) and subsequent physiological changes to prevent disease, effectors

may be delivered by the bacterium to thwart this response (Elstner, 1991; Macho et al, 2014; Zhou et al, 2014; Xiang et al, 2008). In turn, the plant can gain the ability to recognize these effectors, and this is what can elicit the hypersensitive response to limit the spread of bacteria through the plant tissues (Boch and Bonas, 2010). Comparisons of the whole genomes of these *Xhv* strains may reveal differences in their encoded effectors that are race-specific, or specific to the strains that produce atypical HR. By analyzing these differences, one might determine the genetic basis for the race-specificity we observed in HR elicitation and for the atypical HR presentation.

Another interesting finding was that two strains isolated from lettuce tissue with the appearance of bacterial leaf spot, were not *Xhv* but instead were *Pantoea* species. This genus includes several plant pathogens that are either necrotic or hemibiotrophic, and so they have necrotic stages built into their life cycles (Gentzel et al., 2022; Stice et al., 2020). *Pantoea* species have previously been shown to be common endophytes in lettuce tissue (Hou et al., 2013). However, Koch's postulates confirmed that these strains were pathogenic on lettuce. The Koch's postulates will be repeated to better characterize the appearance of the necrotic lesions, and if the pathogenicity is again demonstrated, it will be the first report of *Pantoea* species causing disease on lettuce.

Lastly, the eight French strains were analyzed using MLSA to determine whether they were *Xhv* and if so, identify their sequetypes and form hypotheses regarding their race. Seven of these strains clustered with known *Xhv* race 1 strains, and another strain clustered with *Xhv* race 2 strains. While the genetic analysis offered by MLSA is useful for generating hypotheses about race, in this case we might suspect that the seven strains would elicit HR on *L. sativa* cv. Little Gem and the eighth would elicit HR on *L. serriola* PI491114, the completion of HR screening is

a necessary step to confirm the races of these strains. This work was postponed due to the unavailability of seeds belonging to the *Xhv* race 2 and 3 differential lettuce lines, *L. serriola* PI491114 and ARM-09-161-10-1, but it is expected to be completed soon.

As researchers continue to embrace sequence analysis as a way to delineate taxa, this case provides an example of how classifications based on phenotype, such as pathovar or race, can offer important information regarding the biology of the group of strains. For this reason, we highlight that while MLSA and other sequence-based analyses may serve as powerful predictive tools and allow researchers to make inferences based on the genetic relatedness of *Xhv* strains, HR screening with the race-determining cultivars continues to be essential to accurately determine the races of those strains.

Chapter 3

Comparative genomic analysis of the lettuce bacterial leaf spot pathogen, *Xanthomonas hortorum* pv. *vitians,* to investigate race specificity[3]

3.1 Introduction

Bacterial leaf spot (BLS) of lettuce is caused by *Xanthomonas hortorum* pv. *vitians* (*Xhv*) Morinière *et al*. 2020 (formally *X. campestris* pv. *vitians*), which was first described in 1918 following an outbreak in South Carolina lettuce fields (Brown, 1918). BLS is a sporadic, yet significant threat to lettuce production worldwide, and several subsequent outbreaks in the last century have prompted research to better understand this plant-pathogen interaction. Investigation into *Xhv* strain diversity has been vital to the development and application of effective strategies for the management of BLS. In 2016, Bull et al. demonstrated that some, but not all, of the 120 *Xhv* strains evaluated resulted in a hypersensitive response (HR) to limit the spread of infection upon inoculation into *Lactuca sativa* cvs. Little Gem, Pavane, and La Brillante, and that these strains did not cause HR in the other lettuce cultivars that were tested. This subset of *Xhv* strains was designated race 1 to correspond to the HR reaction elicited in Little Gem, Pavane, and La Brillante and the hypothesized interaction with the resistance gene (R-gene) *Xar1* identified in these strains (Hayes et al., 2014). Another identified source of resistance to *Xhv* race 1 strains isolated from Florida lettuce fields was the *Xcvr* R-gene from *L. sativa* PI358001-1 (Wang et al., 2016). The presence of these R-genes led to the hypothesis that *Xhv* race 1 strains engage in a gene-for-gene interaction with these lettuce cultivars. Additional races of the pathogen were designated among those *Xhv* strains that did not result in HR in the

[3] This chapter was published as **Rosenthal, E. R.**, Potnis, N., & Bull, C. T. (2022). Comparative Genomic Analysis of the Lettuce Bacterial Leaf Spot Pathogen, *Xanthomonas hortorum* pv. *vitians*, to Investigate Race Specificity. **Frontiers in Microbiology**, *13*, 840311. https://doi.org/10.3389/fmicb.2022.840311.

Xar1- or *Xcvr*-encoding cultivars, but instead triggered HR in either *L. serriola* PI491114 (designated race 2) or *L. serriola* ARM-09-161-10-1 (designated race 3). Recent HR screening also suggests that *Xhv* race 1 strains might also induce HR in ARM-09-161-10-1 (Rosenthal et al., unpublished).

Bull et al. 2016, found that the diversity they observed in the resistance phenotypes of the lettuce cultivars corresponded to the genetic diversity of the pathogen. They demonstrated the presence of five distinct sequetypes among 120 *Xhv* strains using a multi-locus sequence analysis (MLSA) scheme originally designed for distinguishing *Xanthomonas* spp. (Young et al., 2008; Maiden et al., 1998). The results of this study corroborated previous work on *Xhv* diversity (Sahin et al., 2003; Fayette et al., 2016), and they found that the strains that induce HR upon injection into the *Xar1*- or *Xcvr*-encoding cultivars all belonged to sequetypes B, D, or E. The strains which triggered HR in PI491114 all belonged to sequetype A, and those that triggered HR in ARM-09-161-10-1 belonged to sequetype C. These results supported the hypothesis that genotypic differences underly the differences in resistance phenotype. The relationship between sequetype and race also supported the finding that *Xhv* populations are predominantly clonal, though with some sequence variation due to recombination (Fayette et al., 2016).

Studies of plant-pathogen interactions have suggested a constant cycle in which plants and their pathogens evolve to outperform each other, in defense strategy and pathogenicity respectively (Zipfel and Felix, 2005; Dangl and Jones, 2001). Pathogen presence can trigger basal plant defenses, such as the production of reactive oxygen species (ROS), to prevent further colonization. This response can then be interrupted by the delivery of bacterial effector proteins through type three-secretion system. In turn, effector activity can be thwarted by plant host recognition of those effectors and activation of resistance (R-)genes, such as those that activate

HR to kill off invaded tissues and limit the spread of infection. The presence of the R-genes *Xar1* and *Xcvr* in lettuce cultivars capable of HR exclusively to *Xhv* race 1 (Hayes et al, 2014; Wang et al., 2016) suggested the possibility that *Xhv* race 1 might produce an effector that is recognized by those R-genes. Other gene-for-gene interactions might also be at play for *Xhv* races 2 and 3 and their respective HR-inducible lettuce cultivars, though R-genes have not yet been demonstrated for PI491114 and ARM-09-161-10-1.

In other bacterial plant pathogens, the proximity of many type III effectors (T3Es) to mobile genetic elements suggests that they were acquired or lost through horizontal gene transfer with other bacteria in their local microbiome (McCann & Guttman, 2008; Merda et al., 2017). Following effector gene acquisition or loss, the selective environment created by the host's defenses may have caused the new genotype to be maintained and to be host-specific (Hajri et al., 2009; Sarkar et al., 2006) or even race-specific. Such a relationship between race-specific genotype and tomato cultivar has already been demonstrated for *X. euvesicatoria* pv. *perforans* (Astua-Monge et al, 2000).

Other gene products, such as secondary metabolites and bacteriocins, have been shown to play a role in virulence. The *Xanthomonas* pigment Xanthomonadin and siderophore Xanthoferrin contribute to bacterial fitness by protecting bacterial cells from UV damage and sequestering iron in low-iron conditions, respectively; functional knockouts of these gene clusters resulted in reduced virulence according to Rajagopal et al., 1997, and Pandey et al., 2017. A variety of bacteriocins identified in *Xanthomonas* spp. enhance bacterial competitiveness through antimicrobial activity (Holtsmark et al., 2008). Although a role in host-specific pathogenicity for these components has not yet been demonstrated, it is possible that race-specific inclusion of these genes may be used to further distinguish the *Xhv* races.

This study was designed to investigate variation in effector repertoires between the *Xhv* races as a possible genetic source for race-specific HR elicitation. Several important virulence factors were also studied to look for genotypic differences that could be used to distinguish the *Xhv* races. Eighteen geographically diverse isolates of *Xhv* and seven related plant pathogenic *Xanthomonas* strains were sequenced and assembled. A genome assembly for *X. hortorum* pv. *carotae* M081 was also retrieved and included for comparison. Genome statistics and predicted features, including plasmids and secretions systems, were determined for each strain. A maximum likelihood phylogeny based on SNP data corroborated the taxonomic relationships predicted by MLSA (Bull et al., 2016 and Fayette et al., 2016). Mining the genomes for effector sequences and other potential virulence factors revealed insights into the race-specificity of the BLS pathogen. A preliminary report of this data has been given (Rosenthal et al., 2018).

3.2 Materials and Methods

3.2.1 Bacterial strains and culturing methods

All strains used in this study and their sources are shown in Table 3-1. Eighteen strains of *Xhv* were selected as representatives of the three pathogenic races designated in previous studies: race 1 (12 strains), race 2 (4 strains), and race 3 (2 strains). The type and pathotype strains of *X. hortorum* pv. *hederae, X. hortorum* pv. *taraxaci, X. hortorum* pv. *pelargonii, X. hortorum* pv. *gardneri, and X. hortorum* pv. *cynarae* were included for comparison to the *Xhv* strains, as well as one *Xanthomonas hortorum* strain from radicchio. *X. campestris* pv. *coriandri* was included as an outgroup. Bacteria were routinely cultured on nutrient agar (NA) and in nutrient broth (NB), both at room temperature (20-28 C).

Table 3-1. *Xanthomonas* strains included in this study. Type and pathotype strains are marked T and PT, respectively, in the strain designation column.

Organism	Strain	Other Strain IDs	Geographic Origin	Host of isolation (Common Name)	Race*	Source or Citation
X. hortorum* pv. *vitians	BP5172	*Xav* 98-37 2/01	Salinas, CA, U.S.A.	*L. sativa* (Lettuce)	1	J Barak
	BS0339	Salinas 2/01	Salinas, CA, U.S.A.	*L. sativa* (Lettuce)	1	J Barak
	BS0340	*Xav* 98-23 2/01	Salinas, CA, U.S.A.	*L. sativa* (Lettuce)	1	J Barak
	BS0347	*Xcv* 5/01	Salinas, CA, U.S.A.	*L. sativa* (Lettuce)	1	J Barak
	BP5176	*Xcv* 5/01	Salinas, CA, U.S.A.	*L. sativa* (Lettuce)	1	J Barak
	BP5177	"Edge A"	Colorado, U.S.A.	*L. sativa* (Lettuce)	1	S. Koike
	BP5179	"Daniel Rom"	Salinas, CA, U.S.A.	*L. sativa* (Lettuce)	1	S. Koike
	BP5182	"Moreno Let"	Santa Maria, CA, U.S.A.	*L. sativa* (Lettuce)	1	S. Koike
	NCPPB 4058	N/A	United Kingdom	*L. sativa* (Lettuce)	1	H. Stanford
	CFBP 8686^{PT}	LMG 938^{PT}, NCPPB 2248^{PT}, MR20213^{PT}	Zimbabwe	*L. sativa* (Lettuce)	1	R. Bailey
	BP5191	VT111	Canada	*L. sativa* (Lettuce)	1	V. Toussaint
	BP5192	*Xcv*-2	CA, U.S.A.	*L. sativa* (Lettuce)	1	C.T. Bull
	ICMP 1408	PDDCC 1408	Ithaca, NY, U.S.A.	*L. sativa* (Lettuce)	2	W.H. Burkholder
	ICMP 4165	LMG 7508, PDDCC 4165	New Zealand	*L. sativa* (Lettuce)	2	H.J. Boesewinkel
	BS3127	VT106	Canada	*L. sativa* (Lettuce)	2	V. Toussaint
	BP5194	917	Ohio	*L. sativa* (Lettuce)	2	Sahin et al., 2003
	BS2861	"Christy BuLet 2"	King City, CA, U.S.A.	*L. sativa* (Lettuce)	3	S. Koike, Rianda

	BP5181	"Christy BuLet 3"	King City, CA, U.S.A.	*L. sativa* (Lettuce)	3	S. Koike, Rianda
X. hortorum* from *radicchio	BP5178	N/A	Salinas, CA, U.S.A.	*Cichorium intybus* (Radicchio)		Zacaroni et al., 2012
X. hortorum* pv. *hederae	CFBP 4925^{T}	ICMP 453^{T}, NCPPB 939^{T}, LMG 733^{T}	U.S.A.	*Hedera helix* (English Ivy)		(Arnaud 1920) Dye 1978. Vauterin et al., 1995
X. hortorum* pv. *taraxaci	CFBP 410^{PT}	ATCC 19318^{PT}, NCPPB 940^{PT}, LMG 870^{PT}	Ithaca, NY, U.S.A.	*Taraxacum kok-saghyz* (Russian dandelion)		(Niederhauser 1943) Dye 1978. Vauterin et al., 1995.
X. hortorum* pv. *pelargonii	CFBP 2533^{PT}	ICMP 4321^{PT}, LMG 7314^{PT}, NCPPB 2985^{PT}	Auckland, New Zealand	*Pelargonium peltatum L. L'Hér.* (Pelargonium)		(Brown 1923) Dye 1978
X. hortorum* pv. *gardneri	CFBP 8163^{PT}	LMG 962^{PT}, ATCC19865^{PT}, NCPPB 881^{PT}, PDCC 1620^{PT}	Yugoslavia	*Lycopersicon esculentum* Mill. (Tomato)		(ex Šutić 1957) Jones et al., 2006.
X. hortorum* pv. *cynarae	CFBP 4188^{PT}	ICMP 16775^{PT}	France	*Cynara scolymus* (Artichoke)		Trébaol et al., 2000.
X. hortorum* pv. *carotae	CFBP 7900	M081	Hungary	*Daucus carota* L. var. sativus (Carrot)		(Kendrick 1934) Dye 1978. Vauterin et al., 1995.
X. campestris* pv. *coriandri	CFBP 8452^{PT}	LMG 687^{PT}, ATCC 17996^{PT}, ICMP 5725^{PT}, NCPPB 1758^{PT}, PDDCC 5725^{PT}	India	*Coriandrum sativum* (Coriander)		(Srinivasan et al. 1961) Dye. Vauterin et al., 1995.

*There are currently ten *Xhv* race 2 strains and four *Xhv* race 3 strains described.

3.2.2 DNA extraction and whole genome sequencing

DNA extraction was performed using the Qiagen Dneasy UltraClean Microbial Kit (Valencia, CA, USA) according to the manufacturer's instructions, with the following change: 5 replicates were completed for each strain, and after eluting each first replicate, that elute was used to elute the next replicate, and so on. This resulted in >2.0 ug of total DNA per strain, as measured using the ThermoFischer Qubit Fluorometer 3.0 and the Invitrogen dsDNA Broad Range Assay Kit (cat nos. Q33216 and Q32850; Waltham, MA). Library preparation and 250 x 250 paired-end whole genome sequencing was performed at the Pennsylvania State University Genomics Facility using the TruSeq DNA PCR Free Library Construction and the Illumina MiSeq System (cat. nos. FC-121-3001 and SY-410-1003; San Diego, CA).

3.2.3 Quality control and whole genome and plasmid assembly

Trimmomatic PE v.0.39 (Bolger et al., 2014) was used to remove Illumina adapters and poor-quality sequence, defined as four consecutive bases with phred scores lower than 20 and DNA fragments shorter than 50 bases. Sequence quality reports were generated using FastQC v0.11.9 (Andrews, 2010) before and after using Trimmomatic PE to verify an improvement in sequence quality which would improve the accuracy of the genome assembly. This procedure was completed in a loop using trimming.sh, spades.sh, and spades_loop.txt (Appendices A3-1 to A3-3).

Trimmed reads were assembled de novo using SPAdes v3.12.0 (Bankevich et al. 2012) and Quast v5.0.0 (Gurevich et al., 2013) was used to provide descriptive statistics about the assemblies (Table 2-2). Plasmid counts were predicted using the Recycler pipeline (Rosov, et al.,

2017) and plasmid sequences were predicted using plasmidSPAdes (Antipov et al., 2016); using both programs for plasmid analysis has been recommended by the creators of plasmidSPAdes due to the differing predictive strengths of each tool. Raw reads and whole genome sequence assemblies are available at BioProject PRJNA790934. Completeness of the whole genome assemblies was assessed using BUSCO v.5.2.2 (lineage=xanthomonadales_odb10 and mode=genome; Manni et al., 2021) to compare these assemblies against several publicly available genome assemblies for related *Xanthomonas* strains, obtained from NCBI's genome database, including *X. campestris* pv. *campestris* (ATCC 33913^{T}), *X. populi* (CFBP 1817^{T}), *X. arboricola* pv. *juglandis* (CFBP 2528^{T}), *X. hortorum* pv. *pelargonii* (CFBP 2533^{PT}), *X. hortorum* pv. *taraxaci* (CFBP 410^{PT}), *X. hortorum* pv. *cynarae* (CFBP 4188^{PT}), *X. hortorum* pv. *carotae* (CFBP 7900), *X. hortorum* pv. *gardneri* (CFBP 8163^{PT}), *X. hortorum* pv. *vitians* (CFBP 8686^{PT}), *X. campestris* pv. *coriandri* (ICMP 5725^{PT}), *X. fragariae* (PD 885^{T}), and *X. hortorum* pv. *hederae* (WHRI 7744^{T}).

Table 3-2. Genome characteristics for *Xanthomonas hortorum* pv. *vitians* strains and related type and pathotype strains.

	Strain	*Genome size (Mb)*	*N50*	*GC %*	*Contig No.*	*CDSs No.*	*CDSs Length (Mb)*	*% Genome CDS*	*Plasmid No.*	*Iss compared to BS3127*
X. hortorum pv. vitians race 1	BP5172	5.1	90636	63.7	128	4590	4.6	91.1	2	97
	BS0339	5.2	45509	63.6	223	4747	4.7	90.8	1	93
	BS0340	5.4	82765	63.4	324	4653	4.7	86.7	0	105
	BS0347	5.1	114952	63.7	114	4646	4.7	91.0	0	91
	BP5176	5.1	40310	63.7	264	4644	4.7	91.4	3	72
	BP5177	5.2	89202	63.7	138	4674	4.7	90.8	1	104
	BP5179	5.2	26996	63.7	364	4702	4.7	91.0	4	87
	BP5182	5.2	55827	63.7	179	4712	4.7	90.9	2	94
	NCPPB 4058	5.3	80295	63.6	130	4793	4.8	90.6	0	95
	CFBP 8686^{PT}	5.1	53428	63.8	183	4591	4.6	91.4	2	88
	BP5191	5.3	49126	63.5	220	4848	4.8	90.8	1	64
	BP5192	5.1	106975	63.7	130	4665	4.7	91.0	0	100
	ICMP 1408	5.2	107536	63.7	109	4720	4.7	91.3	0	0

X. hortorum pv. vitians race 2	ICMP 4165	5.3	129151	63.6	107	4865	4.9	91.2	1	0
	BS3127	5.1	106407	63.7	98	4647	4.7	91.6	0	0
	BP5194	5.2	33155	63.6	271	4725	4.7	91.5	0	0
X. hortorum pv. vitians race 3	BS2861	5.2	118085	63.7	120	4690	4.7	90.9	0	no data
	BP5181	5.2	98381	63.7	128	4684	4.7	91.0	1	111
X. hortorum from radicchio	BP5178	5.2	58212	63.7	199	4655	4.7	91.1	1	no data
X. hortorum pv. hederae	CFBP 4925^{T}	5.4	43552	63.8	270	4916	5.0	91.7	1	no data
X. hortorum pv. taraxaci	CFBP 410^{PT}	5.0	53166	63.9	173	4535	4.6	92.5	1	no data
X. hortorum pv. pelargonii	CFBP 2533^{PT}	5.2	128949	63.8	83	4788	4.9	92.9	1	no data
X. hortorum pv. gardneri	CFBP 8163^{PT}	5.2	112448	63.7	95	4707	4.7	91.5	1	no data
X. hortorum pv. cynarae	CFBP 4188^{PT}	5.4	129044	63.4	335	4633	4.7	87.3	1	no data
X. campestris pv. coriandri	CFBP 8452^{PT}	5.1	422303	64.9	38	4294	4.4	87.1	0	no data

3.2.4 Variant-based phylogeny using whole genome sequences.

To investigate the genetic relatedness of the *Xhv* strains and related strains used in this study, a phylogenetic tree was constructed based on a core alignment of single nucleotide polymorphisms (SNPs). The whole genome assembly of *Xanthomonas hortorum* pv. *carotae* M081 was downloaded from NCBI (GenBank assembly accession GCA_000505565.1https://www.ncbi.nlm.nih.gov/bioproject/PRJNA50443/) to be used for comparison. The snippy-multi pipeline v4.3.6 (Seeman, 2015) was used to map trimmed Illumina sequencing reads to the *Xhc* M081 reference and to identify and align a set of core SNPs. In the resulting core alignment, all non-[AGTCN-] characters were replaced using snippy-clean_full_aln and then loci with high levels of base substitution, indicating possible recombination, were removed using run_gubbins.py v2.4.1 (Croucher et al., 2015). SNPs were extracted from the filtered multi-FASTA alignment using SNP-sites with the option '-c' to only output AGTC into a final alignment file. This file was imported into CLC Genomics Workbench and General Time Reversible with rate variation (4 categories) and estimated topology was selected as the appropriate substitution model following model testing with the neighbor-joining clustering method. This model was used to produce a maximum-likelihood phylogeny with 1000 bootstraps and branch lengths representing the expected number of nucleotide substitutions per sequence site.

3.2.5 Effector gene mining

BLAST databases were constructed from each whole genome assembly scaffold file or plasmid scaffold file. A FASTA file containing known bacterial effector protein sequences, including those previously published for plant pathogenic *Xanthomonas* and *Pseudomonas*

species, was provided (White et al. 2009). Translated nucleotide BLAST was used to align each entry of this effector sequence catalogue to the whole genome sequence databases. Matches were filtered to include those with greater than 60% identity, greater than 40% query coverage per high scoring sequence pair, and less than 0.0001 expect value. Other potential matches were found using a filter of greater than 45% identity, greater than 40% qcovhsp, and less than 0.00001 e-value; this allowed for inclusion of some other effectors which may be present but with moderate sequence variation compared to those in our database. The same searches were performed using the predicted plasmid sequences to determine which effectors may be carried on plasmids. The filtered matches constitute the hypothesized effector repertoires of that strain. Mining for these effector repertoires was completed using the script effectorgene_mining.sh (Appendix A3-4).

3.2.6 Gene alignments and trees

Nucleotide sequences for two genes of interest, *xopAQ* and *xopAF*, were extracted from our assembly files using the program samtools v1.9 faidx and the gene positions found from the effector gene mining procedure (Li et al., 2009). These sequences were used as input for the ORFfinder program available online at https://www.ncbi.nlm.nih.gov/orffinder/, and the longest ORFs were selected for protein alignment using MEGA11: Molecular Evolutionary Genetics Analysis (Tamura, Stecher, and Kumar, et al., 2021), along with the original protein sequences published for the homologs XopAQ, HopAQ1, Rip6, and Rip11 (Potnis et al., 2011; Guttman et al., 2002; Mukaihara et al., 2010); or XopAF (AvrXv3) and HopAF1 (Astua-Monge et al., 2000; Petnicki-Ocwieja et al., 2002). Overhanging sequences were trimmed. Maximum-likelihood phylogenies were generated from these alignments to show the relatedness of the extracted

genes. The Jones-Taylor-Thornton (JTT) amino acid substitution model, an assumption of uniform substitution rates among sites, and the nearest-neighbor-interchange (NNI) method were used to generate the phylogenies. The phylogenies were tested using 1000 bootstraps and branch lengths represented the number of amino acid substitutions per site.

3.2.7 Mobile genetic elements

Mobile genetic elements were assessed for their effector content, especially for *xopAQ* and *xopAF* presence, to find evidence of recent acquisition by a common ancestor of *Xhv* race 1 and 3 strains. Large insertion sequences in *Xhv* race 1 and 3 strains were identified by aligning their whole genome sequencing reads to the assembled sequence for *Xhv* race 2 strain BS3127, identifying the sequence regions present in their genomes that were not present in BS3127, and aligning the regions flanking the putative insertion to their whole genome assemblies to find the complete insertion sequences. These steps were completed following the MGEfinder pipeline (Durrant et al., 2020). *Xhv* race 3 strain BS2861 was excluded from this analysis because its inclusion caused the program to fail. The identified insertion sequences were then mined for effector genes using the method described above, as well as transposase sequences included in the ISEScan package (Xie and Tang, 2017) using the same method and phage elements using PHASTER's online platform (Arndt, et al., 2016) and the default settings.

3.2.8 *xopAF* and *xopAQ* gene regions in *Xhv* race 2 strains

Whether there were other known genes present at the sites in *Xhv* race 2 strains where *xopAF* and *xopAQ* were absent was analyzed. Sequences flanking the *xopAF* and *xopAQ* genes from BP5172 (*Xhv* race 1) were extracted using samtools faidx, including 1,000 nucleotides on

either side. *Xhv* race 2 strains were then parsed for these sequences using BLAST to identify the same regions of their genome sequence. Alignments of these regions were completed for a subset of *Xhv* race 1 and 2 strains, and both the *Xhv* race 3 strains. These alignments revealed the sequence within the *Xhv* race 2 genomes that were present instead of the *xopAF* and *xopAQ* genes. These sequences were copied into NCBI's ORFfinder to look for coding sequence and NCBI's Conserved Domain Search to determine the possible function of that coding sequence. One protein-coding gene of interest which was present in all *Xhv* race 2 strains instead of *xopAF* was used as a query in a BLAST search of *Xhv* race 1 and 3 genomes to determine whether this gene was specific to *Xhv* race 2 strains or present in all *Xhv* strains tested.

3.2.9 Secretion system gene mining

Genes associated with or comprising secretion systems were compiled into a database by downloading their sequences from representative species from NCBI (Pena et al., in preparation). Nucleotide BLAST using the script Ssgene_mining.sh (Appendix A3-5) revealed genes from our database which aligned to the whole genome sequence assemblies. Plasmid sequences were also queried for secretion system genes. Matches were filtered to retain only those that had 65% identity and 65% query coverage per high scoring sequence pair or higher and an e-value less than 0.00001. To confirm these results, additional global alignments of extracted secretion system gene clusters were completed in CLC Genomics Workbench, and aligned genes were again filtered for 65% identity. Reported results represent the sum of these two methods.

3.2.10 Secondary metabolite/bacteriocin gene mining

Secondary metabolite gene clusters were predicted for each strain using the antiSMASH v5.0 online server using the 'relaxed strictness' parameter and the 'KnownClusterBlast' feature (Blin et al., 2019). Individual genes within these clusters were manually filtered to include only hits with greater than 60% identity and 40% query coverage.

To verify the antiSMASH results, each whole genome assembly was also uploaded to the BAGEL4 webserver and queried against a database of known bacteriocin genes and motifs using default settings (de Jong et al., 2016). Areas of interest that were scored as significant by the program (where the sum of weight factors is greater than 175) are reported here along with their identified class.

3.3 Results and Discussion

3.3.1 Comparative Genomic Analysis

Eighteen assembled genomes representing the three pathogenic races of *Xhv*, as well as those of seven related *Xanthomonas* strains, are available at BioProject PRJNA790934. The bioinformatics pipeline used in this study is shown in Figure 3-1. Genome-associated statistics are reported in Table 3-2. The average *Xhv* genome size was 5.2 Mbp and in this study the number of contigs ranged from 98 for the *Xhv* race 2 strain BS3127 to 364 for the *Xhv* race 1 strain BP5179, with an average of 180 contigs per strain. The N50 for *Xhv* strains varied from 27,000 bp to 129,000 bp. Whole genome annotation showed that each *Xhv* strain contained about

4,700 coding sequences (CDS), with an average total length of CDS per genome being 4.7 Mb and accounting for 91% of the total genome size. The number of plasmids predicted for each *Xhv* strain varied from zero to four, with no pattern by race observed.

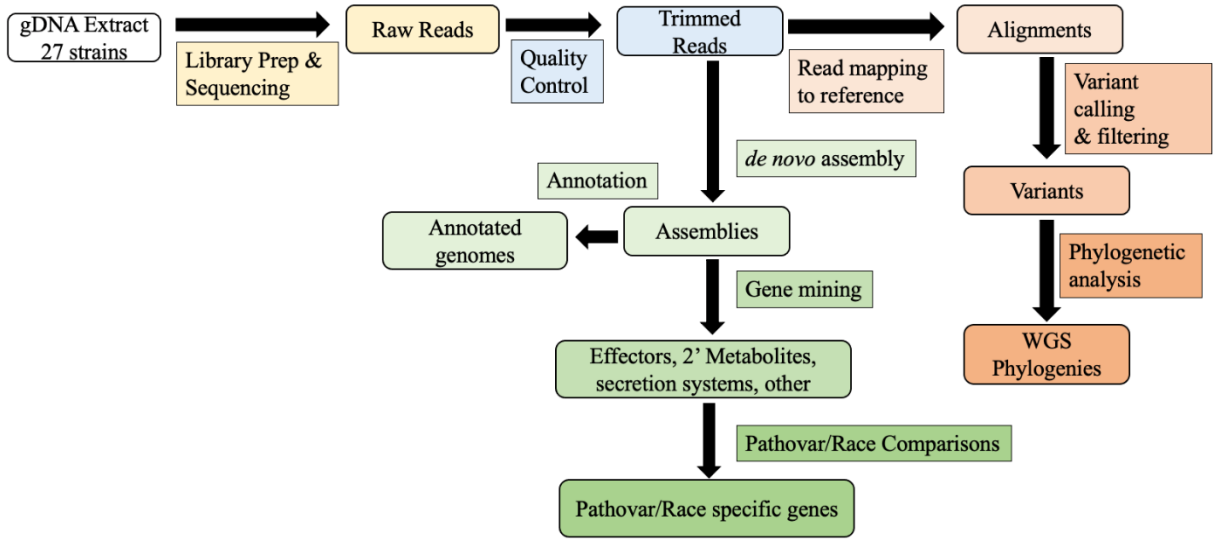


Figure 3-1. Bioinformatics workflow for comparing whole genome sequences.

Compared to the *Xhv* strains, the related *Xanthomonas* strains had approximately the same genome sizes, fewer gaps in the sequence with an average of 170 contigs per strain, and similar variation in N50. Annotations for these reference strains revealed similar statistics for coding sequences, with 4,650 CDSs at an average length of 4.7 Mb and representing 91% of the total genome size. A single plasmid is predicted for each of the reference strains except for the pathotype of *X. campestris* pv. *coriandri*, which did not have any predicted plasmid sequence. BUSCO analysis for measuring the completeness of all these sequence assemblies showed that they are complete and similar to other published *Xanthomonas* genomes (Figure 3-2).

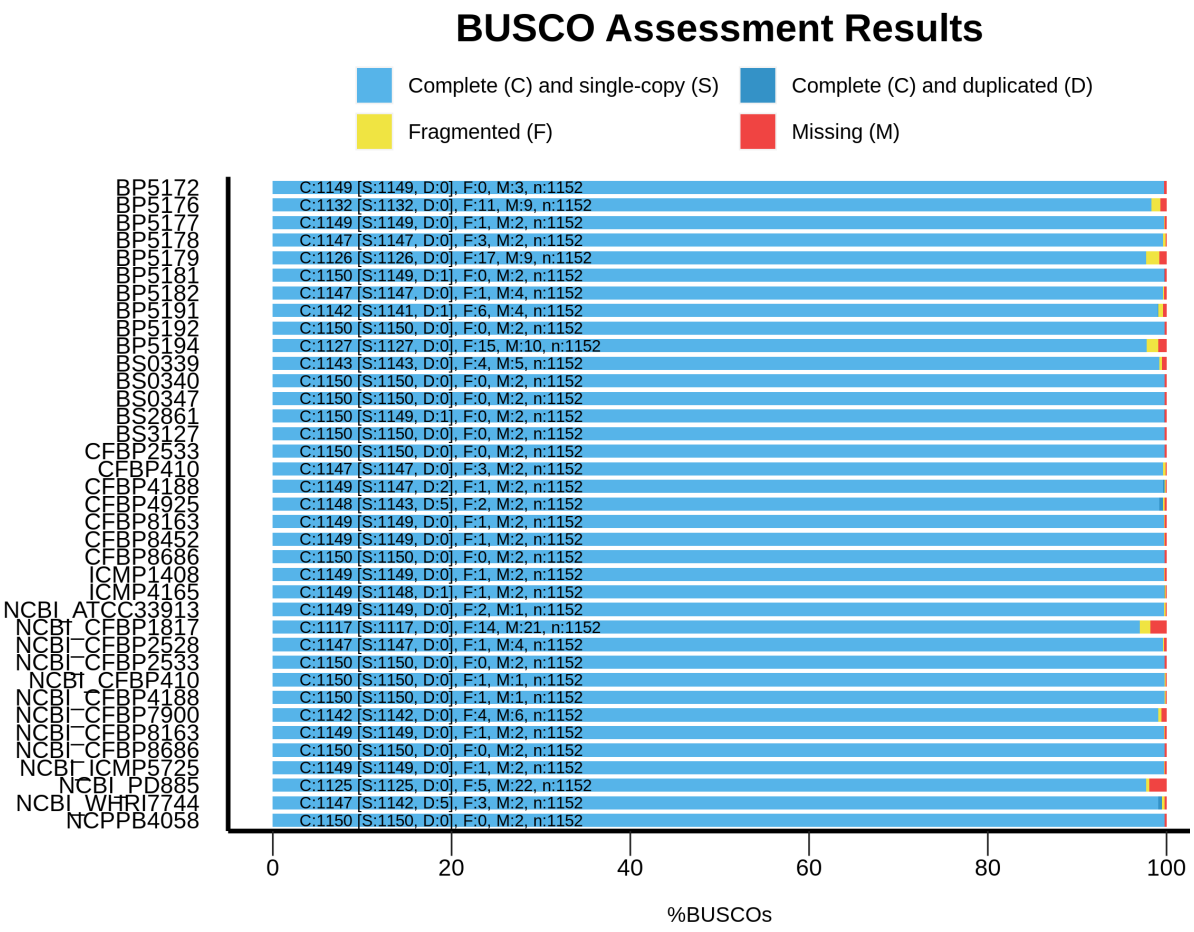


Figure 3-2. Genome assembly completeness assessment for *Xhv* and related strains. The completeness of the genome assemblies produced in this study was evaluated using the benchmarking universal single copy orthologs method (BUSCO; Manni et al., 2021). The proportions of complete and single-copy orthologs in these assemblies were compared to those of previously published *Xanthomonas* assemblies retrieved from NCBI's genome database, including *X. campestris* pv. *campestris* (ATCC 33913^{T}; GenBank assembly accession GCA_000007145.1), *X. populi* (CFBP 1817^{T}; GCA_002940065.1), *X. arboricola* pv. *juglandis* (CFBP 2528^{T}; GCA_001013475.1), *X. hortorum* pv. *pelargonii* (CFBP 2533^{PT}; GCA_012922215.1), *X. hortorum* pv. *taraxaci* (CFBP 410^{PT}; GCA_012922225.1), *X. hortorum* pv. *cynarae* (CFBP 4188^{PT}; GCA_002939985.1), *X. hortorum* pv. *carotae* (CFBP 7900;

GCA_000505565.1), *X. hortorum* pv. *gardneri* (CFBP 8163^{PT}; GCA_012922265.1), *X. hortorum* pv. *vitians* (CFBP 8686^{PT}; GCA_012922135.1), *X. campestris* pv. *coriandri* (ICMP 5725^{PT}; GCA_019201305.1), *X. fragariae* (PD 885^{T}; GCA_900183975.1), and *X. hortorum* pv. *hederae* (WHRI 7744^{T}; GCA_003064105.1).

3.3.2 Phylogenetic Analysis

The maximum likelihood phylogeny using whole genome SNP data showed a distinct clade to which all the *Xhv* strains belong (Figure 3-3). Within this larger clade there were three groups that corresponded to the three *Xhv* races, originally distinguished by their differing host reactions on various lettuce plant introduction lines (Bull et al., 2016). *Xhv* race 1 and 2 appear to have diverged from a common ancestor that itself diverged from strains of *X. hortorum* pv. *gardneri* and *X. hortorum* pv. *cynarae*, which are pathogenic on tomato and artichoke, respectively. The *Xhv* race 3 strains appeared to be more recently evolved, stemming from *Xhv* race 1 strains. This corroborated the finding of Fayette et al. (2016) in which *Xhv* race 2 strains made up a single MLST. The relationship between the *Xhv* strains and the other strains of *Xanthomonas hortorum* corroborated results presented in Morinière et al. (2020), which were based on a predicted core proteome rather than the whole genome SNP data presented here.

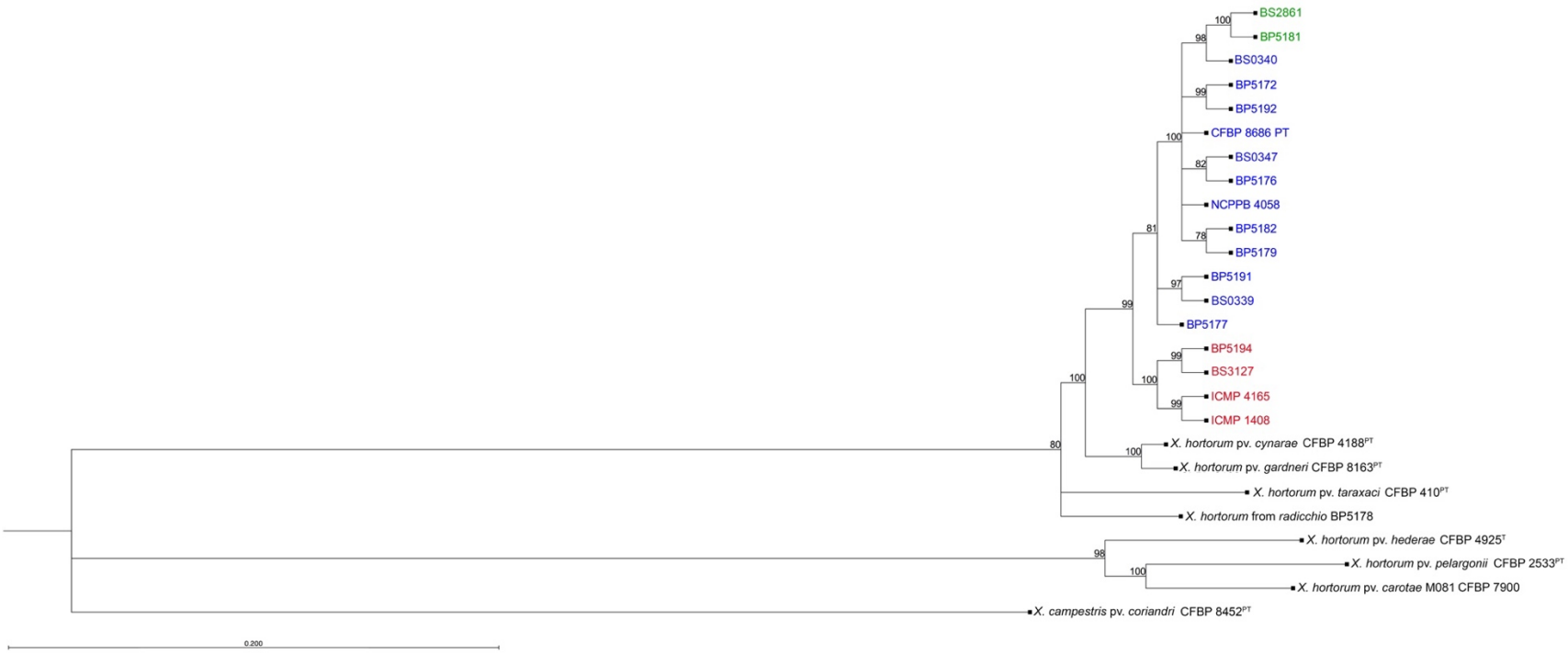


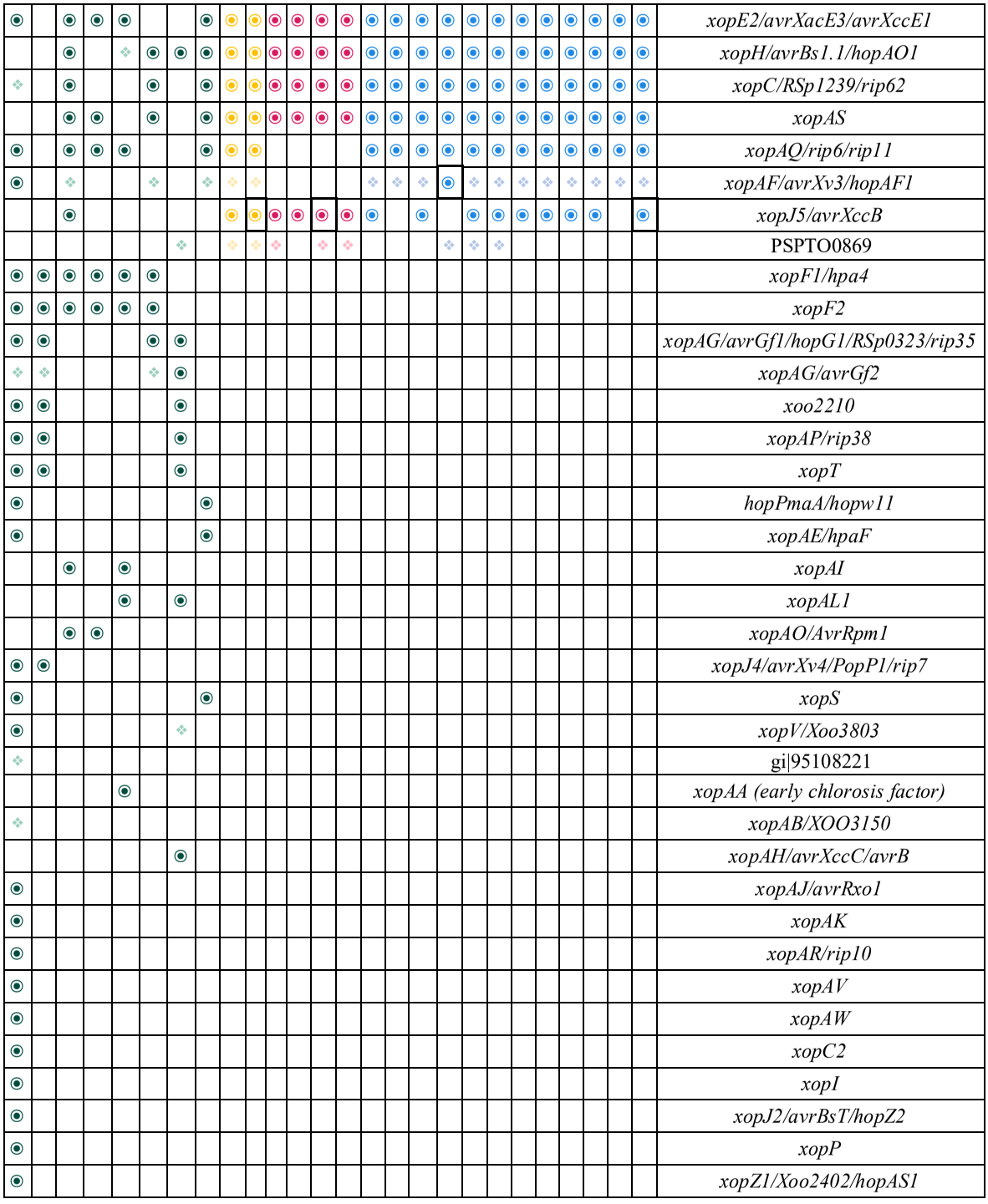
Figure 3-3. Phylogenetic tree constructed from whole genome SNP data. Maximum-likelihood phylogeny of *Xanthomonas hortorum* pv. *vitians* strains of race 1 (blue), race 2 (red), and race 3 (green) and related *Xanthomonas* type and pathotype strains (black) used in this study. SNP data was extracted from trimmed Illumina sequence reads and core aligned using the snippy-multi pipeline (Seemann, 2015) and the tree was built from sequence alignment using CLC Genomics Workbench. This phylogeny was constructed using the neighbor-joining clustering method and the general time-reversible substitution method with rate variation (4 categories) and estimated topology. Bootstrap values are shown above branches and branch length represents the expected number of nucleotide substitutions per site. Branches shorter than 0.0100 are shown as having a length of 0.0100

3.3.3 Effector Repertoire Variations

The focus of this study was to identify race-specific avirulence genes among the *Xhv* strains that could explain the race-specific HR phenotypes observed upon inoculation into various lettuce-cultivars. Of the total sixty-seven effector homologs predicted, fifteen were identified in all *X. hortorum* strain sequences tested here and constitute the core effector repertoire of the species (Table 3-3). Regarding the *Xhv* strains, thirty-four of the predicted effector homologs were present in all *Xhv* strains tested, representing the core repertoire of the pathovar. Homologs for the transcription-activator like effector (TALE) *avrHah1* were identified in all *Xhv* strains tested here, as well as in the *X. hortorum* pv. *hederae*, *X. hortorum* pv. *taraxaci*, *X. hortorum* pv. *gardneri*, *X. hortorum* pv. *cynarae*, *and X. campestris* pv. *coriandri* sequences evaluated. This class of effectors is significant for its ability to enter the host cell nucleus and alter transcription in such a way that makes the host susceptible to disease (Bogdanov et al., 2010). Although race-specific variation in *avrHah1* sequence was not observed in our strains, this could be a limitation in the ability of our short-read sequencing method to resolve the repeat sequences characteristic of TALEs. Further research is needed using a long-read sequencing procedure that can demonstrate whether there are race-specific differences as well as pathovar-specific differences in the *avrHah1* gene.

Table 3-3. Filtered matches for effector sequences in *Xhv* and related *Xanthomonas s*trains. Bolded effectors are those predicted for all *X. hortorum* strains. The effector with an asterisk indicates that it is a TALE. Circles indicate a match of >60% identity, >40% qcovhsp, and >0.00001 e-value; diamonds indicate a lower percent identity threshold of >45%; and empty boxes correspond to absence of that effector gene. Blue color corresponds to *Xhv* race 1 hits, magenta to *Xhv r*ace 2 hits, yellow to *Xhv* race 3 hits, and dark green to related *Xanthomonas* reference strain hits. Thick box borders indicate gene presence verified through targeted gene amplification and visualization by gel electrophoresis.

	Race	Strains	***avrBs2***	***hrpA***	***xoo1914/hopPmaI***	***xopAD/skwp protein***	***xopE3/xoo0261/hopAJ***	***xopN/hopAU1***	***xopQ/hopQ1***	***xopX/holPsyAE/Xoo3022***	***xopAY***	***xopZ2***	*xopK/XOO1669*	*xopL/xcc4186/xoo1662*	*xopR*	*xopAU*	*xopB/hopD1/hopPtoD1*	*xopD/PsvA*	*xopG/hopH/hopAP/rip36*	*xopJ3/avrRxv/avrA*	*avrBs1*	*avrHah1/pth/avXa7/avrB6/avrBs3/avrX*	*xopAM/hopR1*	*xopE1/avrXacE1*
X. hortorum pv. *vitians*	1	BP5172	⦿	❖	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	❖	⦿	⦿	⦿	❖	⦿	⦿	⦿	⦿
		BS0339	⦿	❖	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	❖	⦿	⦿	⦿	❖	⦿	⦿	⦿	⦿
		BS0340	⦿	❖	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	❖	⦿	⦿	⦿	❖	⦿	⦿	⦿	⦿
		BS0347	⦿	❖	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	❖	⦿	⦿	⦿	❖	⦿	⦿	⦿	⦿
		BP5176	⦿	❖	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	❖	⦿	⦿	⦿	❖	⦿	⦿	⦿	⦿
		BP5177	⦿	❖	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	❖	⦿	⦿	⦿	❖	⦿	⦿	⦿	⦿
		BP5179	⦿	❖	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	❖	⦿	⦿	⦿	❖	⦿	⦿	⦿	⦿
		BP5182	⦿	❖	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	❖	⦿	⦿	⦿	❖	⦿	⦿	⦿	⦿
		NCPPB 4058	⦿	❖	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	❖	⦿	⦿	⦿	❖	⦿	⦿	⦿	⦿
		CFBP 8686^{PT}	⦿	❖	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	❖	⦿	⦿	⦿	❖	⦿	⦿	⦿	⦿
		BP5191	⦿	❖	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	❖	⦿	⦿	⦿	❖	⦿	⦿	⦿	⦿
		BP5192	⦿	❖	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	❖	⦿	⦿	⦿	❖	⦿	⦿	⦿	⦿
	2	ICMP 1408	⦿	❖	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	❖	⦿	⦿	⦿	❖	⦿	⦿	⦿	⦿
		ICMP 4165	⦿	❖	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	❖	⦿	⦿	⦿	❖	⦿	⦿	⦿	⦿
		BS3127	⦿	❖	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	❖	⦿	⦿	⦿	❖	⦿	⦿	⦿	⦿
		BP5194	⦿	❖	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	❖	⦿	⦿	⦿	❖	⦿	⦿	⦿	⦿
	3	BS2861	⦿	❖	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	❖	⦿	⦿	⦿	❖	⦿	⦿	⦿	⦿
		BP5181	⦿	❖	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	❖	⦿	⦿	⦿	❖	⦿	⦿	⦿	⦿
X. hortorum from radicchio		BP5178	⦿	❖	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	❖	⦿	⦿	⦿	⦿	⦿		⦿	⦿
X. hortorum pv. *hederae*		CFBP 4925^{T}	⦿	❖	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	❖	❖	❖	⦿	⦿		❖	⦿	⦿		
X. hortorum pv. *taraxaci*		CFBP 410^{PT}	⦿	❖	⦿	⦿	⦿	⦿	⦿	⦿	❖	⦿	⦿	⦿	⦿	❖	⦿	⦿	⦿	❖	⦿	⦿	⦿	
X. hortorum pv. *pelargonii*		CFBP 2533^{PT}	⦿	❖	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿		❖		⦿	⦿		⦿		⦿	⦿	⦿
X. hortorum pv. *gardneri*		CFBP 8163^{PT}	⦿	❖	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	❖		❖	⦿	⦿	⦿	⦿	❖	⦿	⦿	⦿
X. hortorum pv. *cynarae*		CFBP 4188^{PT}	⦿	❖	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿	❖	⦿	⦿	⦿	❖	⦿	⦿	⦿	⦿
X. hortorum pv. *carotae* M081		CFBP 7900	⦿	❖	⦿	⦿	⦿	⦿	⦿	⦿	⦿	⦿		⦿						❖				
X. campestris pv. *coriandri*		CFBP 8452^{PT}	⦿	❖	⦿	⦿	⦿	⦿	⦿	⦿			⦿	⦿	⦿	⦿			❖			⦿		⦿


xopE2/avrXacE3/avrXccE1
xopH/avrBs1.1/hopAO1
xopC/RSp1239/rip62
xopAS
xopAQ/rip6/rip11
xopAF/avrXv3/hopAF1
xopJ5/avrXccB
PSPTO0869
xopF1/hpa4
xopF2
xopAG/avrGf1/hopG1/RSp0323/rip35
xopAG/avrGf2
xoo2210
xopAP/rip38
xopT
hopPmaA/hopw11
xopAE/hpaF
xopAI
xopAL1
xopAO/AvrRpm1
xopJ4/avrXv4/PopP1/rip7
xopS
xopV/Xoo3803
gi|95108221
xopAA (early chlorosis factor)
xopAB/XOO3150
xopAH/avrXccC/avrB
xopAJ/avrRxo1
xopAK
xopAR/rip10
xopAV
xopAW
xopC2
xopI
xopJ2/avrBsT/hopZ2
xopP
xopZ1/Xoo2402/hopAS1

Two additional effectors, *xopAF* and *xopAQ*, were predicted in the genomes of all *Xhv* strains belonging to races 1 and 3 but were absent from those of *Xhv* race 2. The presence of *xopAF* was also predicted in the genomes of *X. hortorum* pv. *taraxaci*, *X. hortorum* pv. *cynarae*, and *X. campestris* pv. *coriandri* but not in the genomes of *X. hortorum* pv. *pelargonii*, *X. hortorum* pv. *gardneri*, *X. hortorum* pv. *carotae* M081, and the *X. hortorum* strain from radicchio. Further, the presence of *xopAQ* was predicted in the genomes of *X. hortorum* from radicchio, *X. hortorum* pv. *pelargonii*, *X. hortorum* pv. *gardneri*, *X. hortorum* pv. *cynarae*, and *X. campestris* pv. *coriandri* but not of *X. hortorum* pv. *hederae*, *X. hortorum* pv. *taraxaci,* and *X. hortorum* pv. *carotae* M081.

The gene *xopAF* was first identified as *avrXv3* in *X. euvesicatoria* pv. *perforans* strains of tomato race 3 (T3) and was demonstrated to be responsible for race-specific HR elicitation in tomato cultivar Hawaii 7981 (Astua-Monge et al., 2000). Acting intracellularly, *xopAF* causes the upregulation of defense related genes (Balaji et al., 2007). Since its original identification, homologs have been identified in strains of *X. citri* pv. *citri* (*Xcc*) isolated only from Mexican lime and *X. vasicola* pv. *vasculorum* (*Xvv*) isolated from only from sugarcane. Experiments with *xopAF* mutants showed it was not responsible limiting the host range of *Xcc*, but did contribute to virulence (Jalan et al., 2013). Cultivar-specificity for *Xvv* strains does not appear to have been tested.

The other gene, *xopAQ*, was first identified as *rip6/rip11* in a screen for *Ralstonia solanacearum* proteins that are injected into host plant cells (Mukaihara et al., 2010). The *Xanthomonas* homolog *xopAQ* was first described in *X. hortorum* pv. *gardneri* and hypothesized to be part of a group of pathotype-specific genes responsible for its aggressive disease compatibility on tomato and pepper (Potnis et al., 2011). Homologs have since been found in *X.*

arboricola, *X. citri*, and *X. euvesicatoria* and have a proposed function in lipid modification that may be involved in host recognition and hypersensitive response (Barak et al., 2016; and Thieme et al., 2007).

Five of the effectors predicted for all *Xhv* strains were found within their plasmid sequences: *avrBs1*, *avrHah1*, *xopE1*, *xopE2*, and *xopH* (Appendix Table A3-6). Several other effector genes were predicted among various *Xhv* strains, including *avrBs2*, *xopL*, *xopJ5*, *xopAQ*, *xopB*, *xopC*, *xopD*, *xopG*, and *xopZ2*, but with no pattern that could underly the race-specific differences in HR elicitation in lettuce cultivars. Each of the reference strains had a different predicted repertoire of plasmid-borne effectors, and there were no plasmid-borne effectors that were predicted for all *X. hortorum* pathovars. The genes *avrBs1*, *avrHah1*, *xopH*, *xopAQ, avrBsT* (predicted only in *X. campestris* pv. *coriandri*), and *xopAO* (predicted only in *X. hortorum* pv. *gardneri* and *X. hortorum* pv. *cynarae*) are known to be plasmid-borne in other *Xanthomonas* species (Swanson et al., 1988; Schornack et al., 2008; Roach et al., 2019; Potnis et al., 2011; Minsavage et al., 1990; Newberry et al., 2019), and so were likely acquired through plasmid transfer in the strains tested here. All other predicted plasmid-borne effectors (see Appendix Table A3-6 for the complete list) are not yet demonstrated to be plasmid-borne in other species but may still have been acquired through plasmid transfer to a strain ancestral to *X. hortorum* pv. *vitians*. Further study with completed *Xhv* genomes produced using long-read sequencing may be able to address these questions regarding the acquisition of these effectors. All effectors not predicted on plasmid sequence here are assumed to be chromosomal genes.

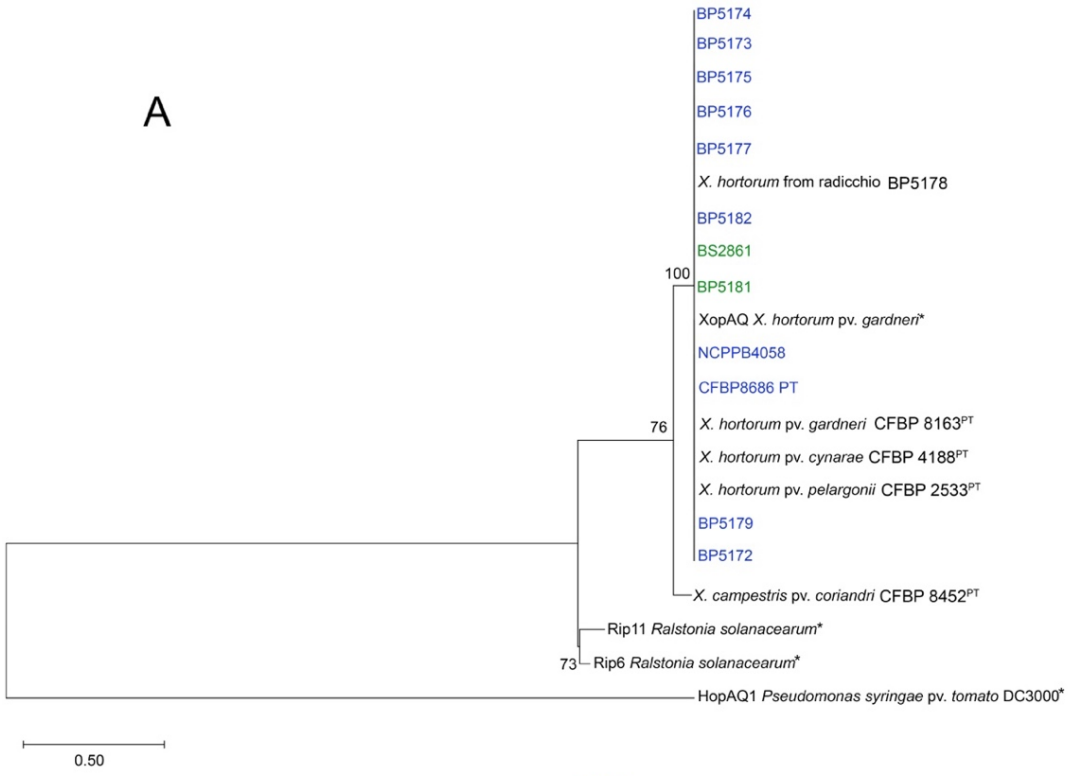
From these data we hypothesize that the differences in effector repertoire composition between *Xhv* races 1 and 3, and *Xhv* race 2, is responsible for the race-specific elicitation of HR in lettuce. It is possible that *xopAQ* and/or *xopAF* interact with an R-gene in ARM-09-161-10-1

to trigger resistance to *Xhv* race 1 and 3 strains, and that the absence of these genes in *Xhv* race 2 strains allows them to go undetected. Further work is needed to confirm the avirulence function of these *xopAQ* and *xopAF* homologs in ARM-09-161-10-1, and that the resistance in these cultivars is governed by a single dominant R gene.

3.3.4 XopAQ Homolog and New XopAF2 Variant

Alignment of the XopAQ protein sequences revealed that they were present among *X. hortorum* strains (Table 3-3), it was identical to XopAQ protein sequence from *X. hortorum* pv. *gardneri*. In the gene tree, all XopAQ protein sequences from *X. hortorum* strains clustered together and were separated from the HopAQ1 protein sequences from *Pst* DC3000 (Figure 3-4A).

The XopAF protein sequence alignment showed that all *Xhv* race 1 and 3 homologs were identical except for that of NCPPB 4058, which had significant variation. Phylogenetic analysis revealed that all XopAF protein sequences from *X. hortorum* strains clustered together and were more closely related to the HopAF1 protein sequences from *Pseudomonas syringae* pv. *tomato* (*Pst*) DC3000 than the XopAF/AvrXv3 from *Xanthomonas euvesicatoria* pv. *perforans* (Figure 3-4B). It is possible this effector gene was originally acquired by the ancestor of *X. hortorum* complex from a *Pseudomonas* species. Here, we designate this new effector variant XopAF2 for *X. hortorum* and propose to distinguish it from protein XopAF1 identified in *X. euvesicatoria*.


A
BP5174
BP5173
BP5175
BP5176
BP5177
X. hortorum from radicchio BP5178
BP5182
BS2861
100
BP5181
XopAQ X. hortorum pv. gardneri*
NCPPB4058
CFBP8686 PT
76
X. hortorum pv. gardneri CFBP 8163PT
X. hortorum pv. cynarae CFBP 4188PT
X. hortorum pv. pelargonii CFBP 2533PT
BP5179
BP5172
X. campestris pv. coriandri CFBP 8452PT
Rip11 Ralstonia solanacearum*
73
Rip6 Ralstonia solanacearum*
HopAQ1 Pseudomonas syringae pv. tomato DC3000*
0.50

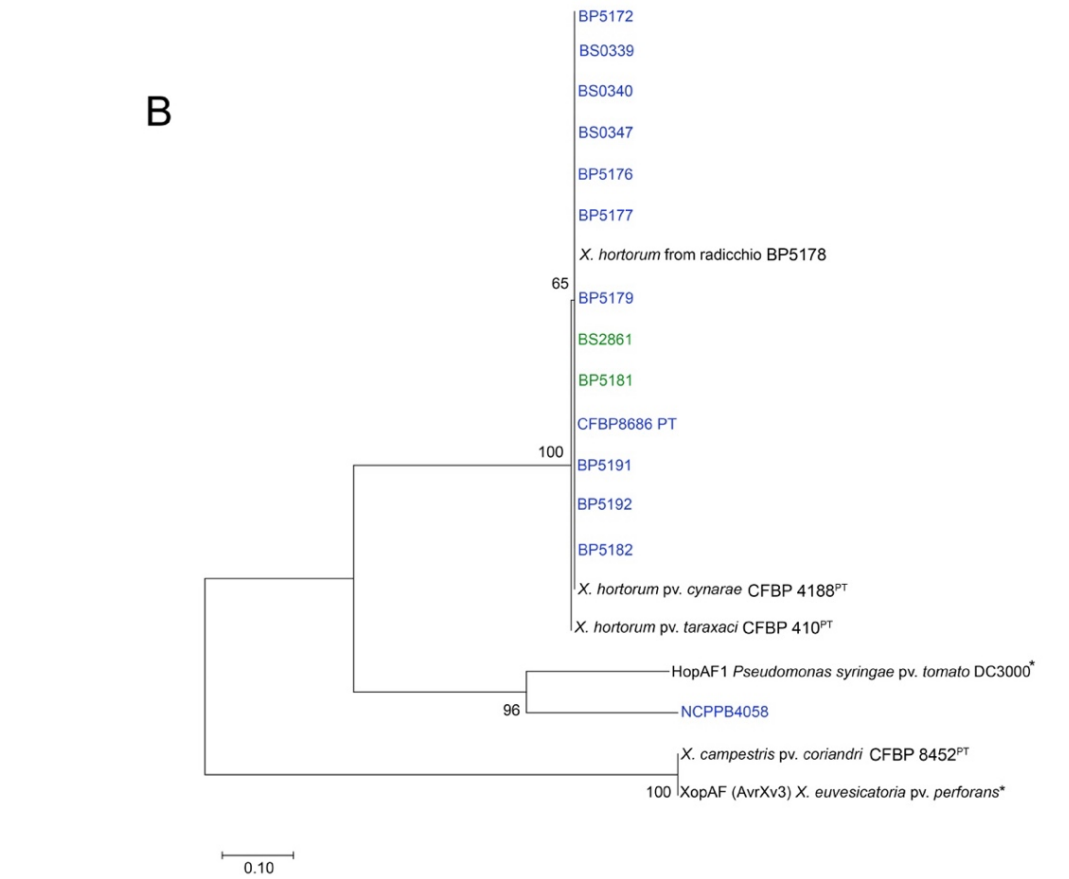

B
BP5172
BS0339
BS0340
BS0347
BP5176
BP5177
X. hortorum from radicchio BP5178
65
BP5179
BS2861
BP5181
CFBP8686 PT
100
BP5191
BP5192
BP5182
X. hortorum pv. cynarae CFBP 4188PT
X. hortorum pv. taraxaci CFBP 410PT
HopAF1 Pseudomonas syringae pv. tomato DC3000*
96
NCPPB4058
X. campestris pv. coriandri CFBP 8452PT
100
XopAF (AvrXv3) X. euvesicatoria pv. perforans*
0.10

Figure 3-4. Phylogenetic trees constructed from (A) XopAQ amino acid alignment and (B) XopAF amino acid alignment. Maximum-likelihood phylogenies were constructed for the two genes present in *Xhv* race 1 and race 3 strains (blue and green) but missing from *Xhv* race 2 strains. Nucleotide sequences for these genes were extracted from the whole genome sequence assemblies and converted to amino acid sequences using OrthoFinder. Alignments and trimming were completed using MEGA11, as well as the phylogeny building using the Jones-Taylor-Thornton (JTT) substitution model, uniform substitution rates among sites, and nearest-neighbor-interchange (NNI) method for tree inference. Bootstrap values are shown beside branches and branch length represents the expected number of amino acid substitutions per site. Asterisks indicate published sequences used as references for proteins HopAF1 and XopAF (Astua-Monge et al., 2000; Petnicki-Ocwieja et al., 2002) and XopAQ, HopAQ1, Rip6, and Rip11 (Potnis et al., 2011; Guttman et al., 2002; Mukaihara et al., 2010)

Determining the regions flanking *xopAQ* and *xopAF* in *Xhv* race 1 and 3 strains allowed for the identification of those same regions in *Xhv* race 2 strains, and alignments revealed the gene sequence present in place of those effector genes. Where *xopAQ* would have been present, the *Xhv* race 2 strains vary in the sequence at this position: conserved domain search revealed that ICMP 1408 encoded for a DNA breaking-rejoining enzyme, while ICMP 4165 encoded for an IS3 family transposase and BS3127 did not encode any conserved domain. In contrast, where *xopAF* would have been present, alignments showed that the *Xhv* race 2 strains all had the same sequence which was predicted to encode a glycosyl hydrolase family 3 C-terminal domain. Mining the *Xhv* race 1 and 3 genomes showed that this sequence could be found elsewhere in those genomes, and so it is not likely a distinguishing feature of a particular *Xhv* race.

In comparison to *Xhv* race 2 strain BS3127, the number of insertion sequences present in *Xhv* race 1 and 3 strains ranged from 64-111, with an average of 92 (Table 3-2). Searching these insertion sequences for effector genes yielded no matches using the same effector gene mining method used for the whole genome sequences. Searches for the transposase genes included in the ISEScan package reveals a negative sense transposase sequence on the same node as *xopAF* and 953 bp downstream. This transposase sequence belongs to insertion sequence family IS5, and the sequence itself is 371 bp in length. Further, searching the *Xhv* race 1 and 3 insertion sequences for phage elements using PHASTER reveals two intact prophage regions. These results indicate that the effector gene *xopAF* may have been acquired by a strain ancestral to *Xhv* race 1 and 3 strains through phage-mediated gene transfer. As expected, no insertion sequences were found in other *Xhv* race 2 strains when compared to BS3127, which supports the assertion that these strains are more related to one another than to the *Xhv* race 1 and 3 strains.

3.3.5 Secondary Metabolites

Though the main interest in this study was to identify effector repertoires as possible genetic sources of host-specificity, several other virulence factors were investigated for genetic differences among the strains tested that could be useful for characterization of the *Xhv* races. The secondary metabolites Xanthomonadin and Xanthoferrin are encoded by gene clusters that are conserved among *Xanthomonas* species. Xanthomonadin was first described in *X. oryzae* pv. *oryzae* (*Xoo*) and is a yellow pigment that protects bacteria from the increase in lipid-degrading free radicals caused by UV exposure (Rajagopal et al., 1997; Ayala et al., 2014). Xanthoferrin is a siderophore that enhances bacterial survival in low-iron conditions, such as the plant epidermis, due to its iron search and uptake function. Knock-out mutants for these metabolites resulted in

decreased virulence for *X. campestris* pv. *campestris*, the causal agent of black rot on cabbage. For Xanthomonadin, lipid peroxidation is higher in knock-out mutants than in WT strains (Rajagopal et al., 1997). For Xanthoferrin, the knock-out results in smaller leaf lesions and less migratory capacity through the tissue (Pandey et al, 2017).

Secondary metabolite gene cluster mining of the whole genome sequences using antiSMASH revealed the possible presence of two clusters (Table 3-4). Nine to eleven genes of the 14-gene cluster encoding Xanthomonadin I, as described for *Xathomonas oryzae* pv. *oryzae* (*Xoo*), were found only in the genome sequences of *Xhv* race 1 strains BP5172 and BS0340, and those of the reference strains of *X. hortorum* from radicchio and *X. hortorum* pv. *hederae*. If functional, the production of Xanthomonadin I would afford these strains a fitness advantage in their ability to resist UV damage. The entire 7-gene cluster for Xanthoferrin production, as described for *Xoo*, was found in all genome sequences tested except for those of *Xhv* race 1 strain BP5179, the *X. hortorum* pv. *hederae* pathotype, and the *X. hortorum* pv. *taraxaci* pathotype, which were missing one gene, *XOO1360*. The strains that do produce Xanthoferrin would be better equipped to survive in low-iron conditions and may be able to spread further through host tissues than strains that do not produce Xanthoferrin. Further research is necessary to demonstrate the production and activity of these metabolites empirically.

Table 3-4. Predicted secondary metabolite gene clusters. Solid color indicates predicted gene presence, no fill indicates predicted gene absence. Blue color corresponds to *Xhv* race 1 hits, red to *Xhv* race 2 hits, green to *Xhv* race 3 hits, and grey to related *Xanthomonas* reference strain hits. Fractions indicate the observed number of genes over the expected number of genes for that cluster.

	Type	**Arylpolyene**	**Lassopeptide**	**Siderophore**
	Strains	**Xanthomonadin I**	**Xanthomonin I/II**	**Xanthoferrin**
***X. hortorum* pv. *vitians* race 1**	BP5172	11/14	4/4	7/7
	BS0339	0/4	4/4	7/7
	BS0340	10/14	4/4	7/7
	BS0347	0/4	4/4	7/7
	BP5176	0/4	4/4	7/7 (split contigs)
	BP5177	0/4	4/4	7/7
	BP5179	0/4	4/4	6/7
	BP5182	0/4	4/4	7/7
	NCPPB 4058	0/4	4/4	7/7
	CFBP 8686^{PT}	0/4	4/4	7/7
	BP5191	0/4	4/4	7/7
	BP5192	0/4	4/4	7/7
***X. hortorum* pv. *vitians* race 2**	ICMP 1408	0/4	4/4	7/7
	ICMP 4165	0/4	4/4	7/7
	BS3127	0/4	4/4	7/7
	BP5194	0/4	4/4	7/7
***X. hortorum* pv. *vitians* race 3**	BS2861	0/4	4/4	7/7
	BP5181	0/4	4/4	7/7
***X. hortorum* from radicchio**	BP5178	10/14	4/4	7/7
X. hortorum* pv. *hederae	CFBP 4925^{T}	10/14	2/4	6/7
X. hortorum* pv. *taraxaci	CFBP 410^{PT}	9/14	4/4	6/7

X. hortorum* pv. *pelargonii	CFBP 2533^{PT}	10/14	4/4	7/7
X. hortorum* pv. *gardneri	CFBP 8163^{PT}	11/14	4/4	7/7
X. hortorum* pv. *cynarae	CFBP 4188^{PT}	0/4	4/4	7/7
X. hortorum* pv. *carotae	CFBP 7900	0/4	4/4	7/7
X. campestris* pv. *coriandri	CFBP 8452^{PT}	10/14	0/4	7/7

2.3.6 Bacteriocins

Bacteriocins are another potential virulence factor. Three classes of bacteriocins have been proposed: class I which are heat-stable and ribosomally produced, post-translationally modified peptides (RiPPs); class II which are heat-stable and unmodified; and class III which are thermolabile and unmodified. Xanthomonin I/II, microcin, and rhodandoin are all class I bacteriocins called lassopeptides, so named for their 3-D looped structure (Hegemann et al., 2013). Sactipeptides and lanthipeptides also belong to class I, and all class I peptides have a variety of functions and may aid in bacterial fitness via antibacterial or antiviral activity (Maksimov et al, 2012). Zoocin A is a class III bacteriocin known as a bacteriolysin, which also serves as an antimicrobial peptide but differs from the others in its ability to degrade target bacterial cell walls (Simmonds et al. 1996).

Searches for bacteriocin gene clusters using antiSMASH revealed the presence of one bacteriocin among the strains tested (Table 3-4), and a parallel analysis using BAGEL4 produced more detailed as well as some contradictory results (Table 3-5). The entire 4-gene cluster encoding Xanthomonin I/II, as previously described for *Xanthomonas hortorum* pv. *gardneri* (ATCC 19865), was found in all strains except the *X. hortorum* pv. *hederae* pathotype which was missing half of the cluster, and *Xanthomonas campestris* pv. *coriandri*, which was missing the

entire cluster. This might suggest that all strains except *X. hortorum* pv. *hederae* have an advantage in bacterial fitness due to their production of Xanthominin I/II. However, searches with BAGEL4 revealed that five of the *Xhv* race 1, two of the *Xhv* race 2, and both *Xhv* race 3 strains encoded for the Xanthomonin I cluster, and the remaining seven *Xhv* race 1 and two of the *Xhv* race 2 strains encoded for the Xanthomonin II cluster. Among the non-*Xhv* strains, *X. hortorum* from radicchio, *X. hortorum* pv. *hederae*, *X. hortorum* pv. *taraxaci*, *X. hortorum* pv. *pelargonii*, and *X. hortorum* pv. *cynarae* all encoded for the Xanthomonin I cluster; *X. hortorum* pv. *gardneri* and *X. hortorum* pv. *carotae* M081 encoded for the Xanthominin II cluster. *X. campestris* pv. *coriandri* did not encode for either Xanthomonin cluster. These results suggest that all strains tested except *X. campestris* pv. *coriandri* might lack the fitness advantage afforded by Xanthomonin I/II production, however further study would be necessary to determine whether Xanthomonin I or II is expressed by these strains and confer a fitness advantage.

Several other bacteriocins were found to be encoded in strains tested. A cluster encoding for microcin production was found in five of the twelve *Xhv* race 1 strains, all the *Xhv* race 2 and 3 strains, and all the related *Xanthomonas* strains except for *X. hortorum* pv. *hederae* and *X. hortorum* pv. *taraxaci*. *X. hortorum* pv. *hederae*^{PT} and *X. hortorum* pv. *carotae* M081 were the only strains encoding for rhodandoin and lanthipeptide, respectively. Nine *Xhv* race 1 strains, all the *Xhv* race 2 and 3 strains, *X. hortorum* from radicchio, *X. hortorum* pv. *hederae*, and *X. hortorum* pv. *taraxaci* encoded for sactipeptide. For each strain in which they are produced, these class I bacteriocins likely enhance that strain's ability to compete for resources through antibacterial and antiviral activity. All strains tested encoded for zoocin A production, a type of class III bacteriocin known as a bacteriolysin. The presence of this cluster in all strains suggests

that they have a fitness advantage due to the ability to degrade the cell walls of competitor bacteria.

Table 3-5. Predicted bacteriocin genes. Solid color indicates predicted gene presence, no fill indicates predicted gene absence. Blue color corresponds to *Xhv* race 1 hits, red to *Xhv* race 2 hits, green to *Xhv* race 3 hits, and grey to related *Xanthomonas* reference strain hits.

	Class:	**I**	**I**	**I**	**I**	**I**	**I**	**II**	**III**
		Lasso peptide	**Lasso peptide**	**Lasso peptide**	**Lasso peptide**	**Sactipeptides**	**Lanthipeptide**	**Leaderless**	**Bacteriolysin**
		Xanthomonin I	**Xanthomonin II**	**Microcin**	**Rhodanodin**	**Sactipeptides**	**Lanthipeptide**	**Enterocin**	**Zoocin A**
	Strains								
***X. hortorum* pv. *vitians* race 1**	BP5172								
	BS0339								
	BS0340								
	BS0347								
	BP5176								
	BP5177								
	BP5179								
	BP5182								
	NCPPB 4058								
	CFBP 8686^{PT}								
	BP5191								
	BP5192								
X. *hortorum* pv. *vitians* race 2	ICMP 1408								
	ICMP 4165								
	BS3127								

	BP5194								
X. *hortorum* pv. *vitians* race 3	BS2861								
	BP5181								
***X. hortorum* from radicchio**	BP5178								
X. hortorum* pv. *hederae	CFBP 4925^{T}								
X. hortorum* pv. *taraxaci	CFBP 410^{PT}								
X. hortorum* pv. *pelargonii	CFBP 2533^{PT}								
X. hortorum* pv. *gardneri	CFBP 8163^{PT}								
X. hortorum* pv. *cynarae	CFBP 4188^{PT}								
X. hortorum* pv. *carotae	CFBP 7900								
X. campestris* pv. *coriandri	CFBP 8452^{PT}								

3.3.7 Type II Secretion System

The type II, type III, type IV, and type VI bacterial secretion systems are important delivery systems for pathogenicity and virulence factors, and so the whole genome sequence data was also used to assess the possible presence and composition of these secretions system gene clusters. Type II secretion systems (T2SSs) enhance the virulence of plant pathogens by releasing cell wall degrading enzymes and toxins (Jha et al., 2005). The secretion complex is composed of an inner membrane platform, a pseudopilin extending from that platform into the periplasm, and a secretin forming a pore in the outer membrane. Its substrates are transported into the periplasm by the Sec or Tat pathways and into the path of the type II pseudopilin, which then pushes the substrate out through the outer membrane pore. There are two variations of the type II secretion system, known as Xps and Xcs; although only the former contributes to virulence, the latter has some homology that allows for it to complement the former. Both variations can be present in the same strain, or the Xps system may be present alone, and they are both encoded in gene clusters (*xpsE-N*, *xpsD*; *xcsC-N*, respectively).

T2SS gene cluster mining of each of the 26 whole genome sequences included in this study revealed that all encoded for the complete Xcs-T2SS (Figure 3-5A). However, the *xcs* gene cluster has not been implicated in the virulence of a plant pathogen to date, except in its ability to complement the *xpsD-H* genes of the Xps system, and so it is unlikely to play a role in virulence for the strains in included in this study. Regarding the *xps* gene cluster, *X. hortorum* pv. *taraxaci* and *X. hortorum* pv. *gardneri* genomes encoded for its entirety, and all the other genomes encoded for the entire cluster minus the *xpsD* gene (Figure 3-5A). One exception was the *Xhv* race 1 strain BP5179 genome sequence, which lacked *xpsI* and parts of *xpsJ* and *xpsD* – all of

which may be explained by gaps between nodes at their expected positions in the sequence. The incomplete or absent *xpsD* genes may be complemented by the presence of *xcsD* due to significant homology. No T2SS genes were identified within predicted plasmid sequences, suggesting that these genes are chromosomal. Overall, these results suggest that all strains studied here contained a functional Xps-T2SS that may contribute to virulence through the delivery of toxins and cell-wall degrading enzymes, though empirical study is necessary to confirm this hypothesis.

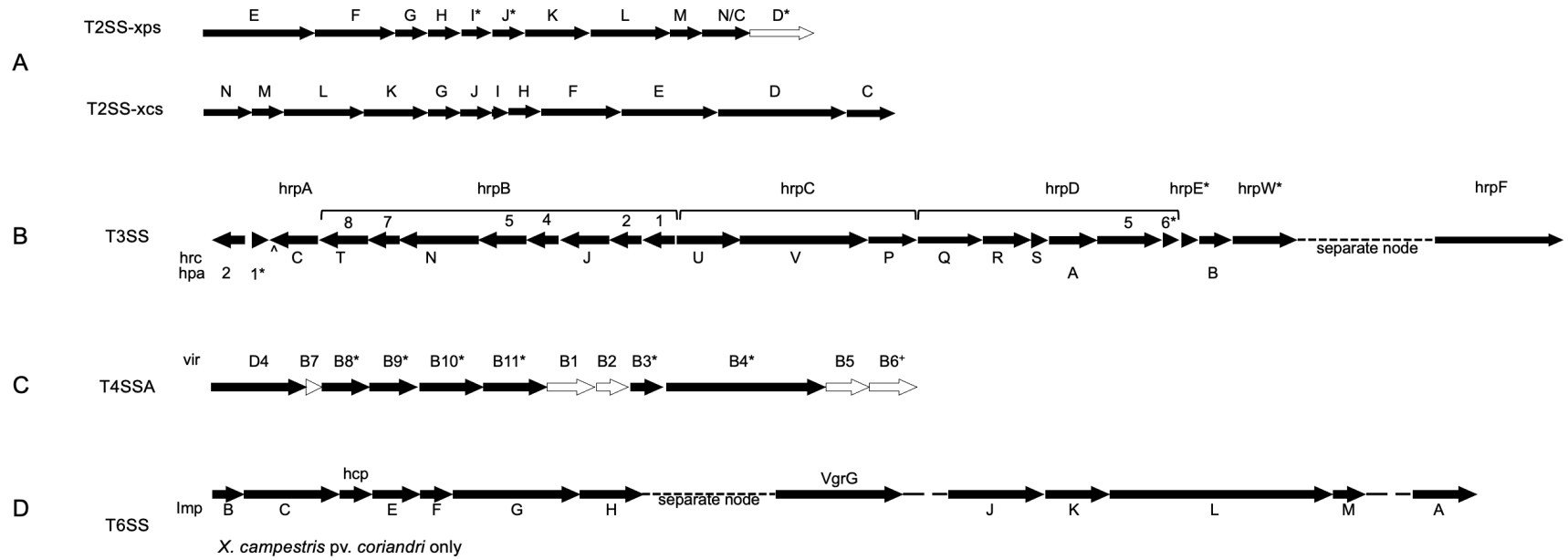


Figure 3-5. Predicted Secretion system gene clusters for *Xhv* and related *Xanthomonas* strains. Presence of each labeled gene in the majority of *Xhv* and related strains is indicated by a black arrow, and absence from all strains is indicated by a white arrow. Asterisks, carets, and crosses indicate exceptions listed here. A: T2SSs xps and xcs - exceptions include *X. hortorum* pv. *taraxaci* and *X. hortorum* pv. *gardneri*, for which the entire xps gene cluster is predicted, and *Xhv* race 1 strain BS2853, for which the genes *xpsI*, *xpsJ*, and *xpsD* were truncated; B, T3SS - exceptions include *X. hortorum* pv. *hederae*, which had a large insertion between *hpa1* and *hrcC* and lacked *hrpE*, *X. campestris* pv. *coriandri*, which lacked *hpa1* and *hrpW*, and *X. hortorum* pv. *pelargonii*, which lacked *hrpE* and *hrpD6*; C, T4SSA - exceptions include *Xhv* race 1 strains BS0541 and BS3044, which encoded the additional gene *virB6*, *X. hortorum* pv. *gardneri*, which encoded the additional *virB6* gene and lacked *virB11*, *X. hortorum* from radicchio, which only encoded

for *virD4*, and *X. campestris* pv. *coriandri*, which only encoded for *virB4*, *virB8*, *virB9*, and *virD4*; and D, T6SS - which was only predicted for the *X. campestris* pv. *coriandri* strain and not found in any other strain tested.

3.3.8 Type III Secretion System

Type III secretion systems (T3SSs) play a major role in the virulence and pathogenicity of bacterial plant pathogens (Büttner et al., 2016). The transmembrane complex includes a needle-like structure that extends into the extracellular space and can pierce the host plant cell wall to directly deliver effector proteins into the host cell cytoplasm. Once inside the host cell, the effectors can halt the basal immune response, tag R-proteins for degradation, disrupt protective phytohormones, or enter the nuclease and alter gene transcription – all of these are strategies to enhance virulence or enable pathogenicity. A cluster of approximately 20 genes encodes the T3SS (Rossier et al., 1999). The effector genes that they translocate are usually dispersed throughout the genome and near to mobile genetic elements, which facilitates their transfer and potential for gain or loss (Cesbron et al., 2015; Darrasse et al., 2013).

Twenty-four genes contributing to the T3SS structure and function were queried against the 25 *Xanthomonas* strains included in this study, and matches are demonstrated in Figure 3-5B. The genome sequences of the *X. hortorum* pv. *vitians* strains, *X. hortorum* from radicchio, *X. hortorum* pv. *taraxaci*, *X. hortorum* pv. *gardneri,* and *X. hortorum* pv. *cynarae* all encoded for the complete hrp gene cluster and likely have functional systems (Hueck 1998). The *X. campestris* pv. *coriandri* genome sequence tested here lacked *hpa1* and *hrpW.* The *hpa1* gene product transports effectors across the T3SS apparatus in *X. oryzae* pv. *oryzae* (Wang et al., 2018) and the *hrpW* gene is a type III secreted protein that resembles both a harpin and pectate lyase (Charkowski et al., 1998). Either of these two genes may contribute to virulence, but they are not required for a functional T3SS; therefore, this *X. campestris* pv. *coriandri* strain likely has a functional T3SS. On the other hand, the genome sequences of *X. hortorum* pv. *hederae* and

X. hortorum pv. *pelargonii* both lacked *hrpE*, and that of *X. hortorum* pv. *pelargonii* also lacked *hrpD6*. The *hrpE* gene encoded the type 3 pilus in *X. vesicatoria* (Weber et al., 2005) and the *hrpD6* gene was responsible for regulating the expression of other T3SS structural genes and type three secreted effectors in *X. oryzae* pv. *oryzicola* (Li et al., 2011); without these genes the T3SS is likely not functional in these strains. Further research would be necessary to demonstrate the translocation activity of these secretion systems empirically, especially for *X. hortorum* pv. *hederae* and *X. hortorum* pv. *pelargonii*, which both encode T3Es but seem to lack key T3SS structural gene sequences in their genomes. No T3SS genes were found to be encoded by predicted plasmid sequences, indicating that they are chromosomal.

3.3.9 Type IV Secretion System

There are two classes of type IV secretion systems (T4SSs); class A has 12 core genes *virB1-11* and *virD4* (Tzfira and Citovsky, 2006) and class B has 5 core genes (*Dot*/*IcmCDFG*[*H*/*K*]) and up to 27 total genes (Hilbi et al., 2017; Qiu and Luo, 2017). Class A substrates have antimicrobial activity to reduce competition for resources; Class B substrates interfere with plant host cell signaling. Class B has not yet been found in *Xanthomonas* spp. The database used here for type IV secretion system gene mining only includes the ten class A core genes for which DNA sequences are available.

All genome sequences tested here encoded for incomplete T4SSs of class A (Figure 3-5C). The *Xhv* genome sequences and those of *X. hortorum* pv. *hederae*, *X. hortorum* pv. *pelergoni*, and *X. hortorum* pv. *taraxaci*, all encoded for the genes *virB3*, *virB4*, *virB8*, *virB9*, *virB10*, *virB11*, and *virD4*, and those of two *Xhv* race 1 strains also encoded for *virB6* (BP5177 and NCPPB 4058). For three of the *Xhv* race 2 strains (ICMP 4165, BS3127, and BP5194),

virD4 was encoded for on predicted plasmid sequence and all other predicted T4SS genes were predicted on chromosomal sequence. In other *Xanthomonas* species, T4SS the genes *virB1*, *virB2*, *virB4*, *virB9*, *virB10*, *virD2*, and *virD4* have been shown to be plasmid-borne, and among those all but *virD2* can be found duplicated in a larger T4SS gene cluster located on the chromosome (Potnis et al., 2011). The genomic sequence of *X. hortorum* pv. *gardneri* encoded for nearly the same T4SS composition as the *Xhv* strains but was missing *virB11* and had the additional *virB6* gene. The genome sequence of *X. hortorum* isolated from radicchio encoded for only *virD4*. *X. campestris* pv. *coriandri* genome sequence encoded for only *virB4, virB8*, *virB9*, and *virD4*. All chromosomal and plasmid sequences assessed here lacked genes that encoded the core complex and pilus of this system (Sgro et al., 2019). Without core structural genes, it is likely that none of these strains have functional T4SSs. More research is necessary to confirm this hypothesis.

3.3.10 Type VI Secretion System

The type VI secretion system (T6SS) in plant pathogenic bacteria has antimicrobial ability; they directly inject toxins into competitor bacteria cells to reduce competition for resources (Bernal et al., 2017). The contribution of T6SS in plant host interactions, however, has not yet been investigated (Liyanapathiranage, Jones, and Potnis, et al., 2021). A cluster of thirteen type six secretion genes, *tssA*-*tssM*, encode the system and were formerly referred to as imp (impaired in nitrogen fixation) genes. The injection structure they form resembles a bacteriophage and toxin delivery relies on its direct contact with bacterial cell walls. The genes *tssD* and *tssI*, also known as *hcp* and *vgrG*, respectively, are thought to be both structural components and substrates of this system (Russell et al., 2014).

The only whole genome sequence included in this study that encoded for a T6SS was *X. campestris* pv. *coriandri* (Figure 3-5D). The gene cluster was split between two nodes; node 6 contained genes *tssB*, *tssC*, *tssD*, *tssE-H* and node 15 contained genes *tssI-M* and *tssA*. None of these genes were found among predicted plasmid sequences, suggesting that they are all encoded chromosomally. The presence of all core genes suggests that the T6SS is functional in this strain of *X. campestris* pv. *coriandri* and may contribute to its fitness against competitor bacteria.

Secretion system mining revealed that the *X. hortorum* strains, including all *Xhv* strains tested here, likely have functional type II and III secretion systems and do not have type IV or type VI secretion systems. *X. campestris* pv. *coriandri* also likely has functional type II, III, and VI secretion systems but likely lacks the type IV secretion system. The presence of type II and type III secretion system gene clusters suggest that these strains can deliver type II secreted toxins and enzymes and type III secreted effectors that may enhance virulence. In *X. campestris* pv. *coriandri*, the type VI secretion system may offer a fitness advantage in the ability to secrete type VI effectors that target competitor bacteria.

3.4 Conclusions

This study revealed insights into the race structure of *X. hortorum* pv. *vitians*. Phylogenetic analysis shows that strains of *Xhv* races 1 and 3 are more closely related to each other than to strains of *Xhv* race 2. HR induction in ARM-09-161-10-1 upon inoculation with *Xhv* race 1 and 3 strains may be due to a gene-for-gene interaction, in which a race-specific effector is recognized by a resistance protein expressed by this cultivar. Such a race-specific interaction has been demonstrated for tomato cultivars and three races of *X. euvesicatoria* pv. *perforans* (Astua-Monge et al, 2000). The one TALE identified among the genome sequences tested, *avrHah1*, does not show race-specific variation here, but long-read sequencing to resolve the repeat regions may yet reveal such variation. Two other putative effectors identified in this study are possible candidate genes responsible for race-specificity because they are present only in *Xhv* races 1 and 3 genome sequences and not in *Xhv* race 2 genome sequences. These genes are a *xopAQ* homolog and the novel effector variant *xopAF2.* Antisense transposase sequence downstream from *xopAF2* and the prophage sequence identified among *Xhv* race 1 and 3 insertion sequences suggest that an ancestor of *Xhv* race 1 and 3 strains may have acquired *xopAF2* in a phage-mediated gene transfer. Our ongoing research seeks to demonstrate whether these effectors are responsible for HR induction in ARM-09-161-10-1. Additionally, more study is necessary to identify other possible gene products involved in the incompatible interactions of *Xhv* races 1 and 2 with *Xar1*-containing lettuce cultivars Little Gem and PI491114, respectively. Furthermore, the close relatedness of the *Xhv* race 1 and 3 strains may make it difficult to identify genotypic differences that explain their different disease phenotypes on lettuce cultivar Little Gem, as none were identified in this study.

This study also provided several insights into *Xanthomonas* genetic diversity. The genome statistics for our assemblies are consistent with what is expected for *Xanthomonas* spp. Secondary metabolite and bacteriocin production varied by strain, and no pattern was found that would be useful for characterizing the *Xhv* races. All strains tested here encoded for type II and type III secretion systems, and *X. campestris* pv. *coriandri* also encoded for a type VI secretion system. These systems all appear to be encoded chromosomally, as none of the genes were identified among predicted plasmid sequences.

Understanding the genetic variation within strains of *Xhv* is crucial for developing effective disease management strategies, especially the breeding of lettuce cultivars with durable resistance against the different races of *Xhv* (Sandoya et al., 2019). Lettuce germplasm resistant to *Xhv* race 1 strains has already been improved (Hayes et al., 2014; Wang et al., 2016) and additional germplasm is being bred (Sandoya et al., 2019). These resistant cultivars encode R-loci *Xar1* and *Xcvr*, which have been mapped to lettuce chromosome two, but the precise gene sequences have not yet been determined. The presence of these R-genes suggests a possible gene-for-gene interaction between these cultivars and *Xhv* race 1, and the race-specific interactions of *Xhv* race 2 and 3 with lettuce cultivars PI491114 and ARM-09-161-10-1, respectively, may also be gene-for-gene interactions with other yet unidentified R-genes. With knowledge of the specific genes interacting with lettuce hosts to induce HR, research could proceed for the precise identification of R-gene interaction partners and subsequent engineering of lettuce cultivars expressing those R-genes.

Chapter 4

Progress toward a pathovar and race-specific detection methods for *Xanthomonas hortorum* pv. *vitians*

4.1 Introduction

United States lettuce production is valued at 4.1 billion dollars, making lettuce the highest valued vegetable crop (USDA-NASS, 2022). Large-scale production occurs throughout California, Arizona, and Florida, and small to mid-sized operations are conducted in Mid-western and Northeastern states. Bacterial leaf spot is a disease that poses a significant threat to lettuce production in the United States and worldwide (Pernezny et al., 1995; Sahin and Miller, 1997). Nearly all market lettuce is cultivated for their leaves but infection with the BLS pathogen causes tissue damage that requires removal of infected leaves, shrinking head size and value. Symptoms include water-soaking, chlorosis, and small necrotic spots on the leaves that later coalesce to form larger lesions (Bull and Koike, 2005). Cool, wet weather favors disease development and although the disease appears sporadically, outbreaks can affect up to 100% of a field crop and therefore severely reduce the sellable yield (Carisse et al., 2000).

Research showed that an integrated management approach is best for preventing outbreaks of BLS. Barak et al., 2001 demonstrated that the bacterial pathogen can persist in crop debris and weeds during fallow periods and become an inoculum source in subsequent growing seasons. This finding led to their recommendations to remove crop debris and weeds and to rotate with slow-maturing crops to give the pathogen time to die out between plantings. The pathogen may also be seedborne (Umesh et al., 1996), and the Sahin and Miller (1997) finding that seeds treated with sodium hypochlorite greatly reduced the rate of seed contamination led to the recommendation to use treated seeds for plantings. Another effective approach for disease

prevention is the use of resistant lettuce cultivars. Chemical bactericides are also commonly used, but they vary in their effectiveness in controlling the disease (Bull and Koike, 2005). They are best applied prophylactically, and the sporadic nature of the pathogen can result in a large sunken cost to growers in years that the pathogen does not appear.

The pathogen causing lettuce bacterial leaf spot is *Xanthomonas hortorum* pv. *vitians* (*Xhv*; formally *Xanthomonas campestris* pv. *vitians*). The first description of the pathogen included its formation of yellow, mucoid colonies when cultured from infected lettuce in a South Carolina field (Brown 1918). Bacterial naming conventions shifted throughout the 20th century, from more subjective classifications based on observed morphological and phenotypic traits that distinguish groups of strains to more objective methods that identify similarities between strains and use statistical analyses to group them together. The BLS pathogen underwent several naming changes, most recently to *Xhv* following new evidence based on whole genome sequence analysis and the description of a pathotype strain (Morinière et al., 2020). Classification and naming conventions and the effectiveness of detection methods are intrinsically linked; a detection method specific to a pathogen must be developed using the appropriate in-group and out-group strains, and the successful selection of these strains depends on the accuracy of their classification.

A conventional PCR-based detection method intended to distinguish *Xhv* from closely related strains was developed by Barak et al. in 2001 using an in-group of 130 *Xhv* strains (then *Xcv*) and an out-group consisting of various *Xanthomonas* (other than *Xhv*), *Pseudomonas*, *Erwinia*, *Clavibacter*, and *Agrobacterium* species. However, because *Xhv* had not yet been re-classified from *X. campestris*, their outgroup *Xanthomonas* strains included mainly *X. campestris* pathovars and lacked several of the other pathovars of *X. hortorum*, including *X. hortorum* pvs.

gardneri, *cynarae*, *taraxaci*, and *hederae*, Their B162 primers amplified a 700 bp band from only the *Xhv* strains they tested, but when we include the additional *X. hortorum* pathovars, we find that they also produce this fragment for *X. hortorum* pvs. *gardneri* and *cynarae*, revealing that it is not truly a *Xhv*-specific method. Another detection method published in 2022 uses loop—mediated isothermal amplification (LAMP) assays with specific primers to detect strains in the *X. hortorum*-*X. hydrangea* species complex (Dia et al., 2021 and 2022a). Such isothermal amplification protocols are useful, even in field settings, for rapid pathogen detection. However, their methods are also not *Xhv* specific; they can detect three clades within *X. hortorum* but not individual pathovars. Their LAMP assay for detecting *Xhv* would also detect *X. hortorum* pvs. *gardneri* and *cynarae*. These limitations in published detection methods prompted our research to evaluate the B162 primer set further and subsequently develop an *Xhv* detection method that eliminates these related *X. hortorum* non-targets.

The diversity among *Xhv* strains adds complexity to the development of a detection method. In 2003, Sahin defined two distinct groups of *Xhv* strains based on genetic and phenotypic variation. Group A strains produced local and systemic infections in lettuce, while group B strains only produced local infections in lettuce leaf tissue. Each group also had unique fatty acid methyl ester, carbon utilization, SDS-PAGE protein, and rep-PCR profiles. In 2016, Fayette et al., 2016, used multi-locus sequence typing on a larger set of *Xhv* strains, including some of the strains from Sahin (2003), and they identified three distinct genotypes that were all pathogenic. Multi-locus sequence analysis (MLSA) showed that all the group A strains from Sahin's study clustered together in one clade and that the group B strains were distributed across all three clades. Phylogenetic analysis revealed the population of *Xhv* strains to be predominantly clonal and that genetic variation between the three clades was likely the result of recombination.

Phenotypic diversity was observed again when a collection of over a hundred *Xhv* strains were inoculated into many lettuce cultivars and accessions and the lettuce varied in their resistance and susceptibility. Some strains elicited a hypersensitive response from lettuce cultivars Little Gem, Pavane, and La Brillante, while others elicited this response only from either plant introduction line PI491114 or ARM-09-161-10-1 (Bull et al., 2016). The first set of strains were designated as *Xhv* race 1, while the other two were defined as *Xhv* races 2 and 3, respectively. Race 1 strains also elicit HR in a Florida plant introduction PI358001-1 (Wang et al., 2016) and ARM-09-161-10-1. Additional MLSA revealed a pattern between the races and the sequetypes identified; all race 1 strains belonged to sequetypes B, D, E, F, or G; all race 2 strains belonged to sequetype A; and all race 3 strains belonged to sequetype C. Genetic mapping of resistance genes in a race 1 determinant cultivar revealed that resistance is likely due to a dominant locus located within linkage group two of the genome (Hayes et al., 2014; Wang et al., 2016). These findings all support the hypothesis that genetic variation between the races may be responsible for this race-level of host specificity in HR elicitation (Bull et al., 2015; Sandoya et al., 2023 in preparation).

Methods of pathovar- and race-specific detection would not only assist in research tracking these *Xhv* strains in new outbreaks worldwide, but it would speed up disease screening from seed lots. The seed industry currently relies on time-consuming lettuce grow out tests. It would also allow for clinicians to provide rapid pathogen detection from suspect *Xhv* infections and precise resistant cultivar recommendations to growers to deploy in subsequent seasons following an outbreak with a particular *Xhv* race. We developed touchdown PCR protocols that were designed to specifically detect *Xhv* strains and each of the three known races. Whole genome sequence alignments of 19 *Xhv* strains and eight closely related strains were parsed for

gene clusters that were unique to either *Xhv* or to one of the *Xhv* races. We first evaluated the specificity of our methods using a collection of 95 geographically diverse *Xhv* isolates, the other seven pathovars of *X. hortorum*, and an additional strain isolated from radicchio that is hypothesized to be *X. hortorum*. We then tested the efficacy of our protocol for detecting *Xhv* from a field lettuce sample submitted from a Pennsylvania outbreak. Our results demonstrated an improvement over previous methods designed to be *Xhv* specific, though further work is necessary to achieve true *Xhv* pathovar- and race-specific detection.

4.2 Materials and Methods

4.2.1 Bacterial strains and sample preparation methods

Bacterial strains used in this study are listed in Table 4-1 along with their location of origin, race designation, and citation. Bacteria were generally cultured using nutrient agar (NA) and nutrient broth (NB) at an incubation temperature of 28 °C. Colony suspensions, used as template for PCR, were made by selecting one colony from each culture on solid agar, adding it to 30 μL of sterile water, and vortexing to mix. Strains were stored long-term in a mixture of 50% nutrient broth and 50% glycerol, at – 80 °C. Liquid overnight cultures were made by inoculating 10 mL of NB broth media with a single colony, then incubating at 28 °C while shaking at 200 rpm. Genomic DNA extract, also used template for PCR, was obtained from the overnight cultures using the Qiagen DNeasy UltraClean Microbial Kit (Valencia, CA, USA) and by following the manufacturer's instructions. DNA concentration of these extracts were measured using the Thermo Fischer Qubit Fluorometer 3.0 and the Invitrogen dsDNA Broad Range Assay Kit (Waltham, MA, United States). Sanger sequencing was completed by the Genomics Core Facility at the Pennsylvania State University.

Several romaine, green leaf, and red leaf lettuce heads were received from a suspected BLS outbreak at a Pennsylvania lettuce farm. Nine symptomatic leaves, showing chlorosis and necrotic spots, were removed for each romaine and green leaf lettuce types and used for isolating DNA and bacteria. Four leaves were removed from the red leaf lettuce heads; fewer leaves were used because more of the red leaf lettuce heads were too far degraded. Leaf tissue was thrice washed with 10% bleach before a segment spanning the symptomatic and healthy tissue was cut out using a sterile scalpel. The segment was crushed using a pestle inside a 1.5 mL tube with 40 mL of sterile water and the resulting extract in water suspension was used both for DNA extraction using the Qiagen DNEasy Plant Mini Kit (Valencia, CA, USA) and for spread plating onto NA. Prior to DNA extraction, the plant extracts were mixed with 50% glycerol and 50% NB and stored at -80 °C. Colonies from plating these plant extracts on NA that matched the typical *Xanthomonas* morphology, yellow, mucoid, and round, were also stored in 50% glycerol, 50% NB at -80 °C.

Table 4-1. *Xanthomonas* strains included in this study.

Organism	Strain	Characteristics	Origin	Source or citation
X. hortorum pv. *vitians*	BP5172*	*Xhv* MLSA group D, *Xhv* race 1	Salinas, CA, United States	J. Barak
	BS0339*	*Xhv* MLSA group B, *Xhv* race 1	Salinas, CA, United States	J. Barak
	BS0340*	*Xhv* MLSA group B, *Xhv* race 1	Salinas, CA, United States	J. Barak
	BS0347*	*Xhv* MLSA group B, *Xhv* race 1	Salinas, CA, United States	J. Barak
	BP5176*	*Xhv* MLSA group E, *Xhv* race 1	Salinas, CA, United States	J. Barak
	BP5177*	*Xhv* MLSA group F, *Xhv* race 1	Colorado, United States	S. Koike
	BP5179*	*Xhv* MLSA group E, *Xhv* race 1	Salinas, CA, United States	J. Barak
	BP5182*	*Xhv* MLSA group G, *Xhv* race 1	Santa Maria, CA, United States	J. Barak
	NCPPB 4058*	*Xhv* MLSA group B, *Xhv* race 1	United Kingdom	H. Stanford
	CFBP 8686^{PT}*	*Xhv* MLSA group B, *Xhv* race 1	Zimbabwe	Vauterin et al., 1995; Morinière et al., 2020
	BP5191*	*Xhv* MLSA group B, *Xhv* race 1	Canada	V. Toussaint
	BP5192*	*Xhv* MLSA group D, *Xhv* race 1	CA, United States	C. T. Bull
	ICMP 1408*	*Xhv* MLSA group A, *Xhv* race 2	Ithaca, NY, United States	W. H. Burkholder
	ICMP 4165*	*Xhv* MLSA group A, *Xhv* race 2	New Zealand	H. J. Boesewinkel
	BS3127*	*Xhv* MLSA group A, *Xhv* race 2	Canada	V. Toussaint
	BP5194*	*Xhv* MLSA group A, *Xhv* race 2	OH, United States	Sahin et al., 2003
	BS2861*	*Xhv* MLSA group C, *Xhv* race 3	King City, CA, United States	S. Koike and Rianda

	BP5181*	*Xhv* MLSA group C, *Xhv* race 3	King City, CA, United States	S. Koike and Rianda
	BS0313	*Xhv* MLSA group B, hypothesized *Xhv* race 1	HI, United States	A. Alvarez
	BS0341	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Salinas, CA, United States	J. Barak
	BS0342	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Salinas, CA, United States	J. Barak
	BS0343	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Salinas, CA, United States	J. Barak
	BS0345	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Salinas, CA, United States	J. Barak
	BS0346	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Salinas, CA, United States	J. Barak
	BS2849	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Salinas, CA, United States	S. Koike
	BS2850	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Salinas, CA, United States	S. Koike
	BS2852	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Salinas, CA, United States	S. Koike
	BS2855	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Salinas, CA, United States	S. Koike
	BS2857	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Salinas, CA, United States	S. Koike
	BS2858	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Salinas, CA, United States	S. Koike
	BS2859	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Salinas, CA, United States	S. Koike
	BS2870	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Santa Maria, CA, United States	S. Koike
	BS2871	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Salinas, CA, United States	S. Koike

	BS2872	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Salinas, CA, United States	S. Koike
	BS2873	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Salinas, CA, United States	S. Koike
	BS2874	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Salinas, CA, United States	S. Koike
	BS2875	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Salinas, CA, United States	S. Koike
	BS2876	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Salinas, CA, United States	S. Koike
	BS2908	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Watsonville, CA, United States	S. Koike
	BS2994	*Xhv* MLSA group B, hypothesized *Xhv* race 1	New South Wales, Australia	R. Fitzell
	BS2996	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Palmerston North, WI, New Zealand	D. R. W. Watson
	BS2997	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Patumahoe, AK, New Zealand	D. R. W. Watson
	BS3050	*Xhv* MLSA group B, hypothesized *Xhv* race 1	New Zealand	D. R. W. Watson
	BS3051	*Xhv* MLSA group B, hypothesized *Xhv* race 1	New Zealand	D. R. W. Watson
	BS3054	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Vaucluse, France	C. Audusseau
	BS3056	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Isère, France	C. Audusseau
	BS3128	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Canada	V. Toussaint
	BS3131	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Canada	V. Toussaint
	BS3132	*Xhv* MLSA group B, *Xhv* race 1	Canada	V. Toussaint

	BS3272	*Xhv* MLSA group B, hypothesized *Xhv* race 1	FL, United States	Pernezny et al., 1995
	BS3300	*Xhv* MLSA group B, hypothesized *Xhv* race 1	CA, United States	C. T. Bull
	BS3301	*Xhv* MLSA group B, hypothesized *Xhv* race 1	CA, United States	C. T. Bull
	BS3303	*Xhv* MLSA group B, hypothesized *Xhv* race 1	CA, United States	C. T. Bull
	BS3304	*Xhv* MLSA group B, hypothesized *Xhv* race 1	CA, United States	C. T. Bull
	BS3306	*Xhv* MLSA group B, hypothesized *Xhv* race 1	CA, United States	C. T. Bull
	BS0301	*Xhv* MLSA group B, hypothesized *Xhv* race 1	HI, United States	A. Alvarez
	BS0302	*Xhv* MLSA group B, hypothesized *Xhv* race 1	HI, United States	A. Alvarez
	BS0304	*Xhv* MLSA group B, hypothesized *Xhv* race 1	HI, United States	A. Alvarez
	BS0305	*Xhv* MLSA group B, hypothesized *Xhv* race 1	HI, United States	A. Alvarez
	BS0309	*Xhv* MLSA group B, hypothesized *Xhv* race 1	HI, United States	A. Alvarez
	BS0310	*Xhv* MLSA group B, hypothesized *Xhv* race 1	HI, United States	A. Alvarez
	BS0311	*Xhv* MLSA group B, hypothesized *Xhv* race 1	HI, United States	A. Alvarez
	BS0314	*Xhv* MLSA group B, hypothesized *Xhv* race 1	HI, United States	A. Alvarez
	BS0316	*Xhv* MLSA group B, hypothesized *Xhv* race 1	HI, United States	A. Alvarez
	BS0318	*Xhv* MLSA group B, hypothesized *Xhv* race 1	HI, United States	A. Alvarez

	BS0335	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Salinas, CA, United States	J. Barak
	BS0337	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Salinas, CA, United States	J. Barak
	BS0338	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Salinas, CA, United States	J. Barak
	BS0542	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Salinas, CA, United States	J. Barak
	BS0543	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Salinas, CA, United States	J. Barak
	BS2946	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Brazil	I. M. G. Almeida
	BS2998	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Brazil	C. F. Robbs
	BS3035	*Xhv* MLSA group B, hypothesized *Xhv* race 1	United States	H. H. Thornberry
	BS3036	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Brazil	A. P. Viegas
	BS3041	*Xhv* MLSA group B, hypothesized *Xhv* race 1		J. R. Neto
	BS3042	*Xhv* MLSA group B, hypothesized *Xhv* race 1	United Kingdom	J.E. Sellwood & J.K. Wilson
	BS3047	*Xhv* MLSA group B, hypothesized *Xhv* race 1	United States	H. Anderson
	BS3049	*Xhv* MLSA group B, hypothesized *Xhv* race 1	Australia	P. Fahy
	BS3053	*Xhv* MLSA group B, hypothesized *Xhv* race 1		C. Audusseau
	BS3055	*Xhv* MLSA group B, hypothesized *Xhv* race 1	France Isère	C. Audusseau
	BS3130	*Xhv* MLSA group B, hypothesized *Xhv* race 1		V. Toussaint

	BS3271	*Xhv* MLSA group B, hypothesized *Xhv* race 1	FL, United States	C. T. Bull
	BS3526	*Xhv* MLSA group B, hypothesized *Xhv* race 1	CA, United States	Sahin et al., 2003
	BS3527	*Xhv* MLSA group B, hypothesized *Xhv* race 1	CA, United States	Sahin et al., 2003
	BS0344	*Xhv* MLSA group E, *Xhv* race 1	Salinas, CA, United States	J. Barak
	BS2851	*Xhv* MLSA group E, *Xhv* race 1	Salinas, CA, United States	S. Koike
	BS2909	*Xhv* MLSA group D, hypothesized *Xhv* race 1	Watsonville, CA, United States	S. Koike
	BS3528	*Xhv* MLSA group D, *Xhv* race 1	CA, United States	Sahin et al., 2003
	BS3034	*Xhv* MLSA group A, *Xhv* race 2	United States	W. H. Burkholder
	BS3043	*Xhv* MLSA group A, *Xhv* race 2	United States	F. Sahin
	BS3126	*Xhv* MLSA group A, *Xhv* race 2	Isère, France	C. Audusseau
	BS3529	*Xhv* MLSA group A, *Xhv* race 2	Ohio, United States	S. Miller
	BS3531	*Xhv* MLSA group A, *Xhv* race 2	Ohio, United States	Sahin et al., 2003
	BS3532	*Xhv* MLSA group A, *Xhv* race 2	Ohio, United States	Sahin et al., 2003
	BS2860	*Xhv* MLSA group C, hypothesized *Xhv* race 3	King City, CA, United States	S. Koike
	BS2863	*Xhv* MLSA group C, hypothesized *Xhv* race 3	King City, CA, United States	S. Koike
X. hortorum from radicchio	BP5178*	N/A	Salinas, CA, United States	Zacaroni et al., 2012
X. hortorum pv. *hederae*	CFBP 4925^{T}*	N/A	United States	Arnaud, 1920; Dye, 1978; Vauterin et al., 1995

X. hortorum pv. *taraxaci*	CFBP 410^{PT}*	N/A	Ithaca, NY, United States	Niederhauser, 1943; Dye, 1978; Vauterin et al., 1995
X. hortorum pv. *pelargonii*	CFBP 2533^{PT}*	N/A	Auckland, New Zealand	Brown, 1923; Dye, 1978
X. hortorum pv. *gardneri*	CFBP 8163^{PT}*	N/A	Yugoslavia	(ex Šutic´ 1957) Jones and Dangl, 2006
X. hortorum pv. *cynarae*	CFBP 4188^{PT}*	N/A	France	Trébaol et al., 2000

Strains marked with an asterisk were used in the whole genome sequence alignments for identifying *Xhv*-specific and race-specific gene clusters to target with our detection methods.

4.2.2 Pangenome analysis

Previously published whole genome sequences were used to identify DNA sequences unique to *Xhv* and each of the *Xhv* races (Rosenthal et al., 2022). Eighteen *Xhv* strains and seven closely related strains were aligned to a reference strain, *Xanthomonas hortorum* pv. *carotae* M81, using the anvi'o pangenome pipeline (Eren et al., 2015). These alignments were parsed for gene clusters unique to either *Xhv* or each of the *Xhv* races. The DNA sequence for each selected gene cluster was downloaded and evaluated using NCBI's conserved domain search tool. They were then evaluated using NCBI's BLASTx to search for potential protein hits in the non-redundant protein sequence database and BLASTn to identify possible non-target matches in the NCBI whole genome sequence database (organism set to '*Xanthomonas hortorum*'). For these sequence analyses, we used an e-value threshold of $1x10^{-5}$ and a percentage identity threshold of 60%.

4.2.3 Primer development

The *Xhv*-specific and race-specific gene clusters were entered into NCBI's online primer blast tool. The 'nr' database was selected for evaluating primer pair specificity, and '*Xanthomonas hortorum*' was entered as the organism to narrow the search. Primer pairs were selected from the tool's output that had few or no reported non-target amplification, low complementarity scores, similar melting temperatures around 60 °C, and similar GC content around 50%. The primers were ordered as custom DNA oligos from Thermo Fisher Scientific (Waltham, MA, United States). A total of 24 primer sets were evaluated using a subset of *Xhv* strains that represented all three known races and several related *X. hortorum* strains, and those primer sets that were selected demonstrated the desired *Xhv* pathovar- or race-specificity and are

listed in Table 4-3. All other primer sets tested, and their results are shown in Appendix A4-1. The B162 primer set previously described for *Xhv*-specific amplification was also evaluated.

4.2.4 PCR development

We found that touchdown PCR, a design with a slightly higher than usual annealing temperature for the early PCR cycles that is typically employed to reduce non-specific primer binding (Don et al., 1991), worked best for achieving the desired specificity of our detection methods. Each touchdown PCR was completed in a 25 µl volume with the following formulation: 9 µl of sterile water, 1.25 µl of each 10 mM primer, 12.5 µl of 2X Bioline Immomix containing IMMOLASE DNA polymerase (London, England), and 1 µl of 30 ng/µl *Xanthomonas* DNA extract or colony suspension in 30 µl of sterile water, using the conditions given in Table 4-2. The reactions were run in an MJ research PTC-100 thermocycler and the cycling conditions are shown in Table 4-2. DNA amplicons were separated via gel electrophoresis using a 1% agarose gel and VWR Life Science EZ-Vision One Dye-as-Loading Buffer 6X (Radnor, PA, United States), with 84 V applied for 2 hours. The New England Biolabs 1 kb DNA ladder was used as reference for amplicon size. Amplicons were visualized using a Bio-Rad Gel Doc XR Imaging System (Hercules, CA, United States). The specificity of the reactions was evaluated using a catalog of 95 *Xhv* strains, seven strains representing the other pathovars of *X. hortorum*, and one strain isolated from radicchio that is hypothesized to belong to *X. hortorum*. These experiments were conducted twice using colony suspensions from the same strain collection.

Table 4-2. Touchdown PCR Cycling Conditions.

STEP	INSTRUCTION	PURPOSE
1	95°C for 1 minute	Taq polymerase activation
2*	95°C for 30 seconds	Denaturation
3*	68°C for 30 seconds, -1°C every cycle	Annealing
4*	72°C for 30 seconds	Extension
5*	GOTO Step 2 10X	Cycling
6	95°C for 30 seconds	Denaturation
7	58°C for 30 seconds	Annealing
8	72°C for 30 seconds	Extension
9	GOTO Step 6 23X	Cycling
10	72°C for 5 minutes	Final Extension

The asterisk refers to the touchdown phase of our protocol when the annealing temperature is kept higher than for a typical PCR to reduce non-specific primer binding. The remaining cycling temperatures are typical of general PCR.

4.2.5 Multi-locus sequence analysis

Multi-locus sequence analysis (MLSA) was completed using a scheme originally designed by Young et al., 2008 but amended by Bull et al., 2016. Colony suspensions made from the suspect *Xhv* strains isolated in Pennsylvania were used as templates in PCRs to amplify four *Xanthomonas* housekeeping genes: *rpoD*, *fyuA*, *gyrB*, and *gap1*. Plant extractions from the infected leaf tissue were also used as templates. Amplification was confirmed by gel electrophoresis and then PCR products were cleaned using EXOSAP-IT (Thermo Fisher Scientific, Waltham, MA, USA). The four housekeeping genes were sequenced by the Genomic Core Facility at Penn State. Sequence data was analyzed using Qiagen's CLC Genomics Workbench version 21 (Valencia, CA, USA). Forward and reverse sequences assembled into consensus sequences with the *X. hortorum* pv. *hederae* strain CFBP 4925^T genes serving as reference sequences. Consensus sequences of each gene for each sample were concatenated and then added to an existing alignment of the same genes from *Xhv* strains of known races. Nucleic

acid substitution model testing was completed, and a maximum likelihood phylogeny was generated using the recommended general time reversible model with four categories of rate variation, estimated topology, and 1000 bootstraps.

4.2.6 16S rRNA subunit analysis

The genomic DNA isolated from plant extracts and colony suspensions of bacteria isolates from the PA 2022 lettuce disease outbreak were used as templates for amplification of the 16S rRNA subunit gene (Stackebrandt and Goebel, 1994). PCR products were cleaned using EXOSAP-IT (Thermo Fisher Scientific, Waltham, MA, USA) and sequencing was completed by the Genomics Core Facility at Penn State. Forward and reverse reads were assembled to a the 16S rRNA subunit gene from the *X. hortorum* pv. *hederae* pathotype strain, downloaded from NCBI. The consensus sequences were aligned, nucleic acid substitution model testing was completed, and the recommended Kimura 80 was used to generate a maximum likelihood phylogeny, with rate variation, estimated topology, and 1000 bootstraps. The consensus sequences were also used as queries for the 16S sequence database in the web interface of NCBI's BLASTn to search for top hits, which were selected to be those with the highest percent identity.

4.3 Results and Discussion

4.3.1 B162 primer set is not *Xhv*-specific.

The B162 primer set that was developed by Barak et al. 2001 was designed to amplify an approximately 700 BP region specific to *Xhv* strains. The gene region encodes for a glycosyl hydrolase of family 3 (Figure 4-1A), an enzyme that breaks down complex sugars and may play

a role in toxin degradation to promote virulence (Bradley et al., 2022). Barak (2001) demonstrated it could produce this amplicon for 130 *Xhv* strains, previously classified as *X. campestris* pv. *vitians*, but not for seven other *X. campestris* pathovars, which then included *X. campestris* pvs. *armoraciae*, *carotae*, and *pelargonii*. The latter two are now classified as pathovars of *X. hortorum*. The amplicon was also not produced by seven additional plant pathogenic bacteria tested, including *X. oryzae* pv. *oryzae*, *Pseudomonas syringae* pv. *syringae*, *P. fluorescence*, *Erwinia carotovora* subsp. *carotovora*, *Clavibacter michiganensis* subsp. *michiganensis*, and *Agrobacterium tumefaciens*.

However, four other pathovars of *X. hortorum* were not included in their analyses. Our searches for the B162 amplicon among *Xanthomonas hortorum* whole genome sequences revealed close matches to many *Xhv* genomes, as well as to *X. hortorum* pvs. *gardneri*, *cynarae*, and *taraxaci* (Figure 4-1B). We hypothesized that these *X. hortorum* non-target strains would be amplified by the B162 primers alongside with the *Xhv* strains, rendering the method not *Xhv-specific*. Other *X. campestris*, *X. arboricola*, and *X. fastidiosa* strains were also among the matches, but with considerable sequence variation that could exclude B162 primer binding and subsequent amplification from these strains.

Upon empirical testing, the PCR with the B162 primer set produced a bright 700 bp band for the strains representing the three races of *Xhv,* as we expected (Figure 4-2). The same 700 bp bands were seen for *X. hortorum* pv. *garderi* (CFBP 8163^{PT}), *X. hortorum* pv. *cynarae* (CFBP 4188^{PT}), and *X. hortorum* pv. *taraxaci* (CFBP 410^{PT}), and the hypothesized *X. hortorum* strain isolated from radicchio (BP5178). Strains that did not produce the B162 amplicon were *X. hortorum* pv. *hederae* (CFBP 4925^{T}), *X. hortorum* pv. *pelargonii* (CFBP 2533^{PT}), and *X. hortorum* pv. *carotae* (CFBP 7900). These results demonstrate that the B162 primer set cannot

be used for *Xhv*-specific detection and that further work was necessary to create a detection method specific to the lettuce BLS pathogen.

A **glycosyl hydrolase family 3, partial [Xanthomonas campestris pv. vitians]**

Sequence ID: AXG24216.1 **Length:** 220 **Number of Matches:** 1

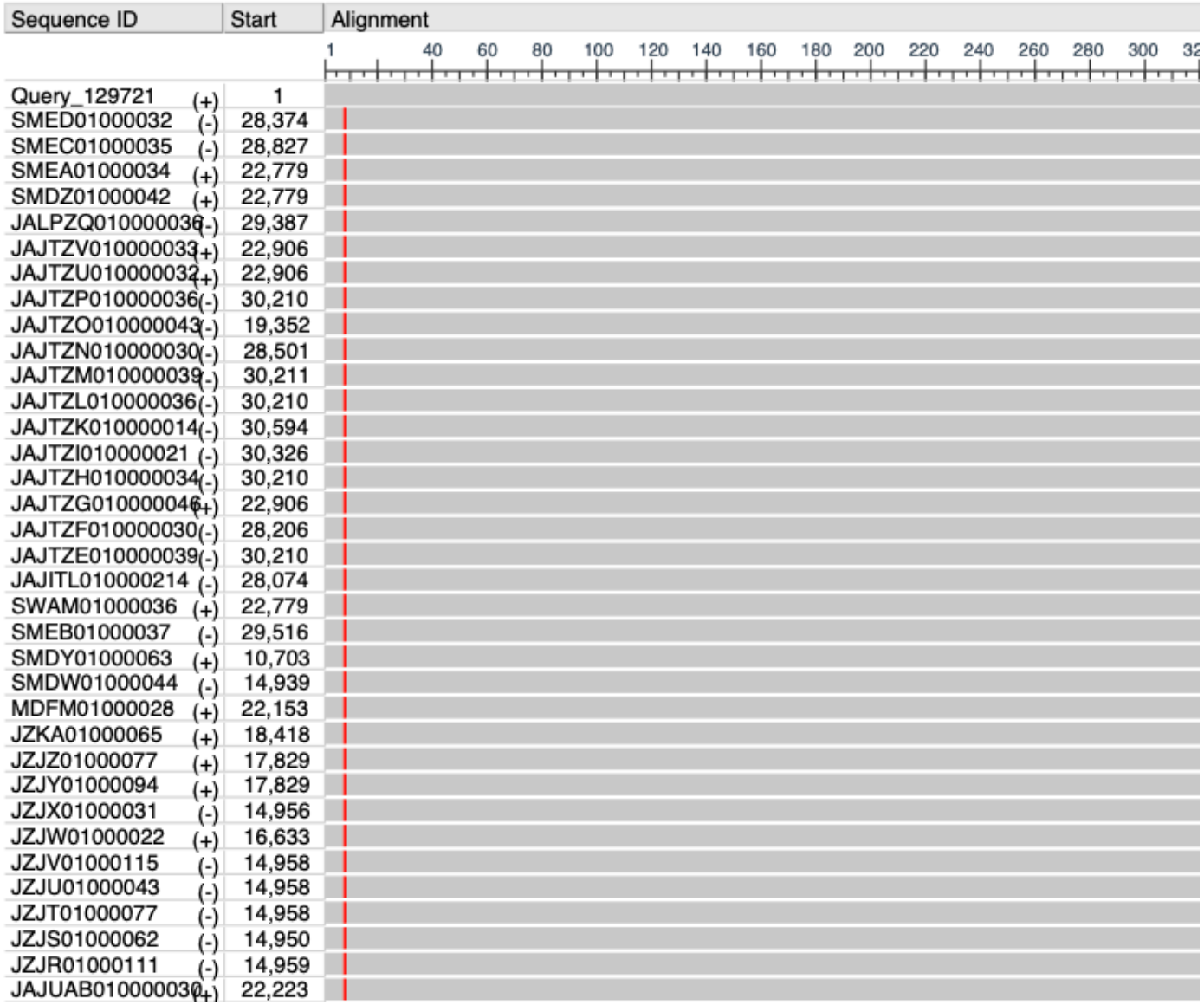
See 1 more title(s) See all Identical Proteins(IPG)

Range 1: 6 to 216 GenPept Graphics ▼ Next Match ▲ Previous Match

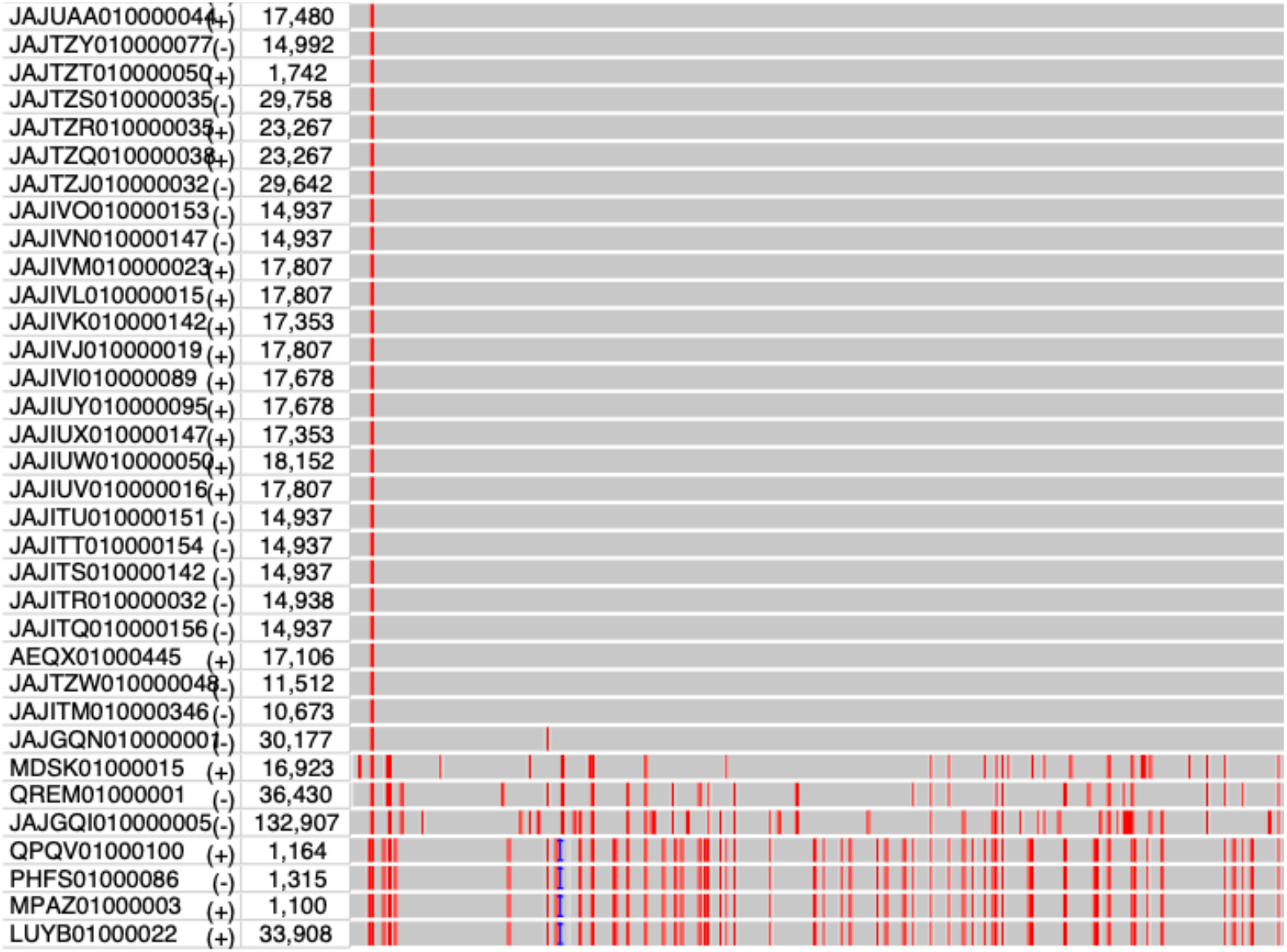
Score	Expect	Method	Identities	Positives	Gaps	Frame
434 bits(1115)	3e-153	Compositional matrix adjust.	211/211(100%)	211/211(100%)	0/211(0%)	-3

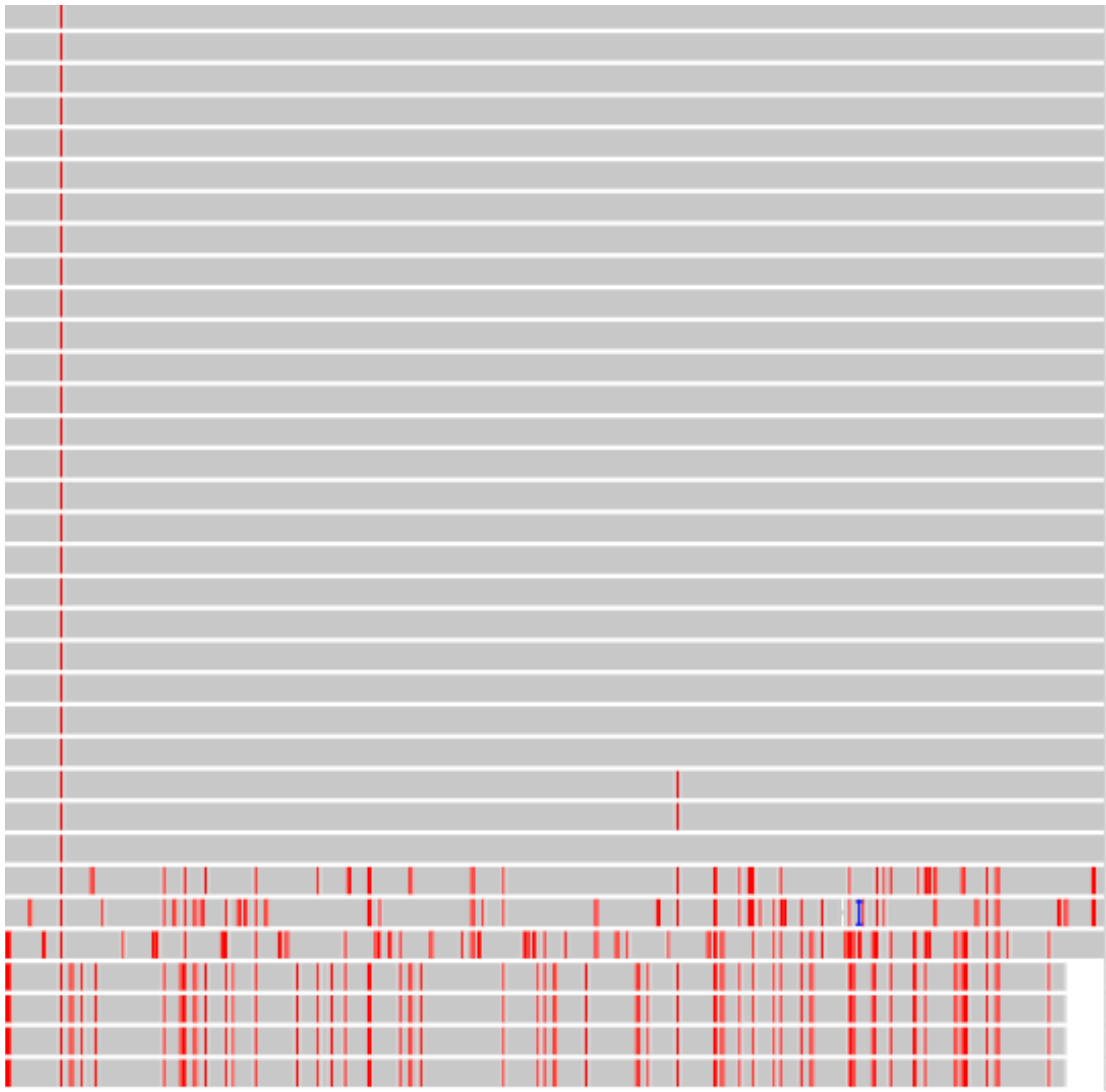
```
Query  633  SNHVGGVILFSNNIKERSQIEKLTSWYAGMESSAGVHLLIATDNEGGNVFRLPRNEYASF  454
            SNHVGGVILFSNNIKERSQIEKLTSWYAGMESSAGVHLLIATDNEGGNVFRLPRNEYASF
Sbjct  6    SNHVGGVILFSNNIKERSQIEKLTSWYAGMESSAGVHLLIATDNEGGNVFRLPRNEYASF  65

Query  453  PGNMALAAAIEGGSSEQLAFEQGRLLAQDLLALKINTNFAPVADVNANPFNPVINVRAFS  274
            PGNMALAAAIEGGSSEQLAFEQGRLLAQDLLALKINTNFAPVADVNANPFNPVINVRAFS
Sbjct  66   PGNMALAAAIEGGSSEQLAFEQGRLLAQDLLALKINTNFAPVADVNANPFNPVINVRAFS  125

Query  273  DNADVVSRLAGKIAAGMERQGLVTTYKHFPGHGSTSTDSHTGLPRVDLSRDQAFAIDIAP  94
            DNADVVSRLAGKIAAGMERQGLVTTYKHFPGHGSTSTDSHTGLPRVDLSRDQAFAIDIAP
Sbjct  126  DNADVVSRLAGKIAAGMERQGLVTTYKHFPGHGSTSTDSHTGLPRVDLSRDQAFAIDIAP  185

Query  93   YQQAISAHAAPDMVMTAHIQYPALDETLGTN  1
            YQQAISAHAAPDMVMTAHIQYPALDETLGTN
Sbjct  186  YQQAISAHAAPDMVMTAHIQYPALDETLGTN  216
```

B NCBI Multiple Sequence Alignment Viewer, Version 1.24.0



End	Organism
635	
27,740	Xanthomonas hortorum pv. vitians
28,193	Xanthomonas hortorum pv. vitians
23,413	Xanthomonas hortorum pv. vitians
23,413	Xanthomonas hortorum pv. vitians
28,753	Xanthomonas hortorum pv. vitians
23,540	Xanthomonas hortorum pv. vitians
23,540	Xanthomonas hortorum pv. vitians
29,576	Xanthomonas hortorum pv. vitians
18,718	Xanthomonas hortorum pv. vitians
27,867	Xanthomonas hortorum pv. vitians
29,577	Xanthomonas hortorum pv. vitians
29,576	Xanthomonas hortorum pv. vitians
29,960	Xanthomonas hortorum pv. vitians
29,692	Xanthomonas hortorum pv. vitians
29,576	Xanthomonas hortorum pv. vitians
23,540	Xanthomonas hortorum pv. vitians
27,572	Xanthomonas hortorum pv. vitians
29,576	Xanthomonas hortorum pv. vitians
27,440	Xanthomonas hortorum pv. gardneri
23,413	Xanthomonas hortorum pv. vitians
28,882	Xanthomonas hortorum pv. vitians
11,337	Xanthomonas hortorum pv. taraxaci
14,305	Xanthomonas hortorum pv. gardneri
22,787	Xanthomonas hortorum pv. cynarae
19,052	Xanthomonas hortorum pv. gardneri
18,463	Xanthomonas hortorum pv. gardneri
18,463	Xanthomonas hortorum pv. gardneri
14,322	Xanthomonas hortorum pv. gardneri
17,267	Xanthomonas hortorum pv. gardneri
14,324	Xanthomonas hortorum pv. gardneri
14,324	Xanthomonas hortorum pv. gardneri
14,324	Xanthomonas hortorum pv. gardneri
14,316	Xanthomonas hortorum pv. gardneri
14,325	Xanthomonas hortorum pv. gardneri
22,857	Xanthomonas hortorum pv. cynarae


JAJUAA010000044 (+) 17,480
JAJTZY010000077 (-) 14,992
JAJTZT010000050 (+) 1,742
JAJTZS010000035 (-) 29,758
JAJTZR010000035 (+) 23,267
JAJTZQ010000038 (+) 23,267
JAJTZJ010000032 (-) 29,642
JAJIVO010000153 (-) 14,937
JAJIVN010000147 (-) 14,937
JAJIVM010000023 (+) 17,807
JAJIVL010000015 (+) 17,807
JAJIVK010000142 (+) 17,353
JAJIVJ010000019 (+) 17,807
JAJIVI010000089 (+) 17,678
JAJIUY010000095 (+) 17,678
JAJIUX010000147 (+) 17,353
JAJIUW010000050 (+) 18,152
JAJIUV010000016 (+) 17,807
JAJITU010000151 (-) 14,937
JAJITT010000154 (-) 14,937
JAJITS010000142 (-) 14,937
JAJITR010000032 (-) 14,938
JAJITQ010000156 (-) 14,937
AEQX01000445 (+) 17,106
JAJTZW010000048 (-) 11,512
JAJITM010000346 (-) 10,673
JAJGQN010000001 (-) 30,177
MDSK01000015 (+) 16,923
QREM01000001 (-) 36,430
JAJGQI010000005 (-) 132,907
QPQV01000100 (+) 1,164
PHFS01000086 (-) 1,315
MPAZ01000003 (+) 1,100
LUYB01000022 (+) 33,908



18,114	Xanthomonas hortorum pv. gardneri
14,358	Xanthomonas hortorum pv. taraxaci
2,376	Xanthomonas hortorum pv. vitians
29,124	Xanthomonas hortorum pv. vitians
23,901	Xanthomonas hortorum pv. vitians
23,901	Xanthomonas hortorum pv. vitians
29,008	Xanthomonas hortorum pv. vitians
14,303	Xanthomonas hortorum pv. gardneri
14,303	Xanthomonas hortorum pv. gardneri
18,441	Xanthomonas hortorum pv. gardneri
18,441	Xanthomonas hortorum pv. gardneri
17,987	Xanthomonas hortorum pv. gardneri
18,441	Xanthomonas hortorum pv. gardneri
18,312	Xanthomonas hortorum pv. gardneri
18,312	Xanthomonas hortorum pv. gardneri
17,987	Xanthomonas hortorum pv. gardneri
18,786	Xanthomonas hortorum pv. gardneri
18,441	Xanthomonas hortorum pv. gardneri
14,303	Xanthomonas hortorum pv. gardneri
14,303	Xanthomonas hortorum pv. gardneri
14,303	Xanthomonas hortorum pv. gardneri
14,304	Xanthomonas hortorum pv. gardneri
14,303	Xanthomonas hortorum pv. gardneri
17,740	Xanthomonas hortorum ATCC 19...
10,878	Xanthomonas hortorum
10,039	Xanthomonas hortorum pv. gardneri
29,543	Xanthomonas campestris pv. nigr...
17,556	Xanthomonas arboricola pv. guiz...
35,797	Xanthomonas arboricola
132,273	Xanthomonas campestris pv. part...
1,786	Xylella fastidiosa subsp. multiplex
693	Xylella fastidiosa subsp. multiplex
1,722	Xylella fastidiosa
34,530	Xylella fastidiosa subsp. multiplex

Figure 4-1. BLASTx protein alignment for top hit (A) and BLASTn DNA alignments (B) to query B162 amplicon sequence. **(A)** The top protein match to the B162 amplicon sequence was glycosyl hydrolase family 3 from *X. hortorum* pv. *vitians* (formerly *X. campestris* pv. *vitians*; 100.0% query coverage; 100% identity; 3e-153 e-value. **(B)** Multiple sequence alignment of the B162 amplicon sequence query to top DNA hits. Red bars represent nucleotide substitutions, white bars represent gaps, and blue 'I's represent insertions.

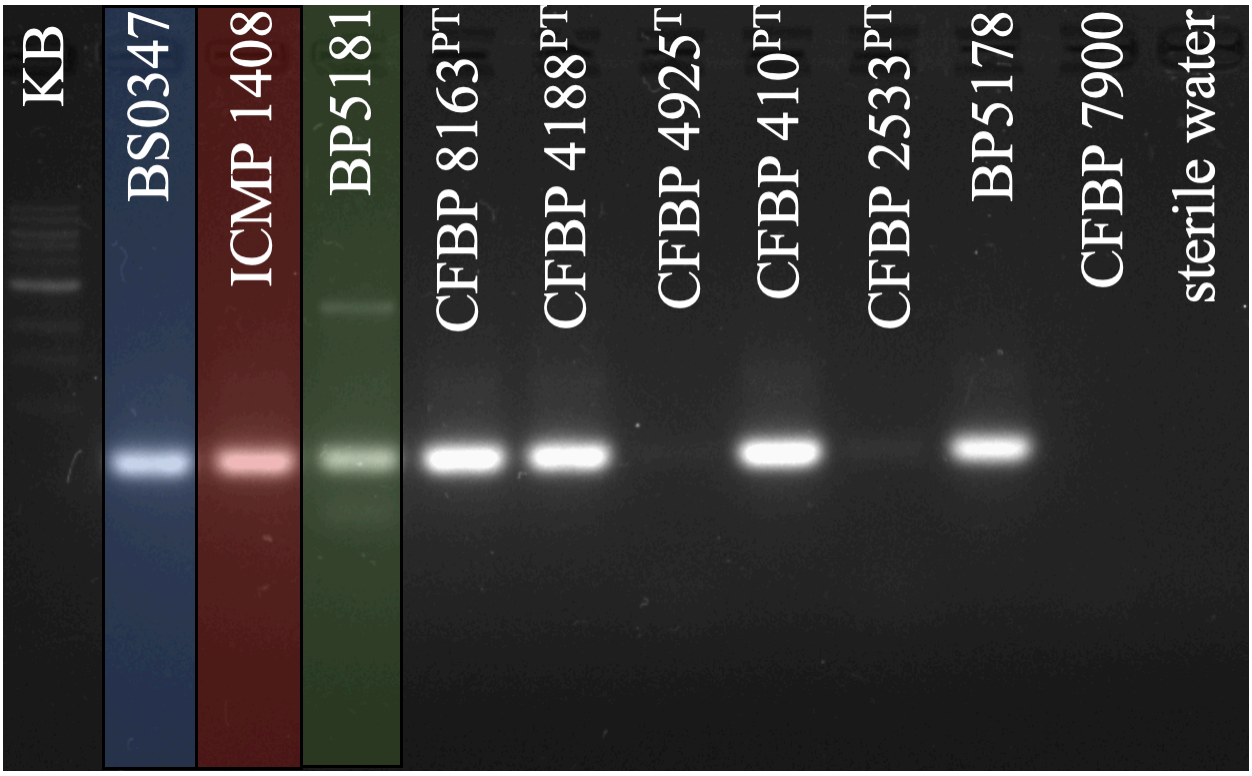


Figure 4-2. PCR Amplification using the B162 primer set. Agarose gel of PCR products created using the B162 primer set and genomic DNA of several *Xanthomonas hortorum* strains as templates. *Xanthomonas hortorum* pv. *vitians* strains B0347 (blue; race 1), ICMP 1408 (red; race 2), and BP5181 (green; race 3) are included along with *Xanthomonas hortorum* pv. *gardneri* (CFBP 8163PT), *X. hortorum* pv. *cynarae* (CFBP 4188PT), *X. hortorum* pv. *hederae* (CFBP 4925T), *X. hortorum* pv. *taraxaci* (CFBP 410PT), *X. hortorum* pv. *pelargonii* (CFBP 2533PT), *X. hortorum* from *radicchio* (BP5178), and *X. hortorum* pv. *carotae* (CFBP 7900). The 1 kb ladder was run in the first well and a sterile water negative control was run in the final well.

4.3.2 Comparative genomic analysis reveals pathovar- and race-specific gene cluster targets.

The alignment of 19 *Xhv* strains and eight related type strains revealed many gene clusters that were either present in all *Xhv* strains or were race-specific. Among those were gene cluster 3889 (GC3889), GC4021, GC4381, and GC4980, which were only present in *Xhv* strains, *Xhv* race 1 strains, *Xhv* race 2 strains, and *Xhv* race 3 strains, respectively. These gene clusters were evaluated for use in PCR-based detection methods for *Xhv* and its three pathogenic races. Gene clusters that were unique among *Xhv* strains to one of the three races were not necessarily absent from the related *X. hortorum* pathotype strains or other reference strains. The primers designed for these methods are listed in Table 4-3. Other gene clusters were investigated but were not selected for our detection protocol due to their lack of specificity (Appendix A4-1).

Table 4-3. Primer sets developed for specific detection of *Xhv* and its three known races.

Primer set	Forward	Reverse	Amplicon size	Target strains
GC3889-110	GCCAGGCCTATG GACTCAAG	ATGGTCGTTGGT GAGCATGA	110	*Xhv*
GC4021-112	GGTGGCCTACTT TCATGCGA	GAGCAAGCCCTT CACAAGGT	112	*Xhv* race 1
GC4381-178	TATGATGCGGCA CACAACCT	CGTATTGCGGTG CGAACTTT	178	*Xhv* race 2
GC4980-138	TCACTCAAAAGC CCACCCTC	ACATTCCTCGGC TATCCCCT	138	*Xhv* race 3

4.3.3 Development and evaluation of *Xhv*-pathovar specific detection method

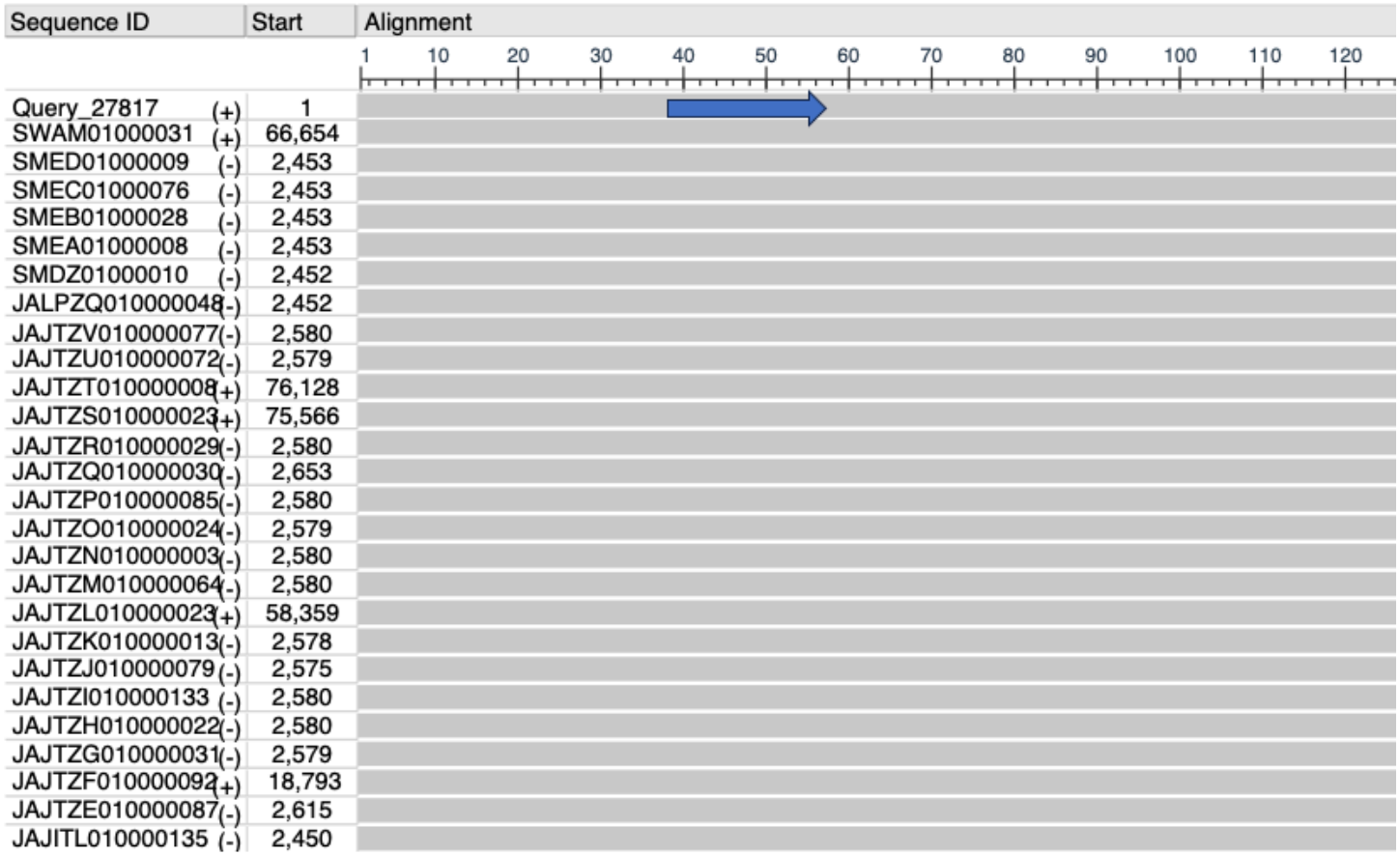
The conserved domain search using the NCBI database did not yield any results for the *Xhv*-specific gene cluster GC3889. Analysis using BLASTn revealed two hypothetical protein matches for GC3889, one from an *X. hortorum* strain and another from an *Xhv* strain, and both had 100% identity, 98% query cover, and e-values of 5e-55 (Figure 4-3A). Hypothetical proteins are predicted based on nucleic acid sequence but are not confirmed to be expressed nor do they

have any known function. The multiple DNA sequence alignment revealed identical matches to all *Xhv* strains submitted in Rosenthal et al., 2022, as well as one *X. hortorum* pv. *gardneri* (*Xhg*) strain (Figure 4-3B). These results suggested that the GC3889 primer set could be used to amplify a region within this hypothetical gene and that detection of *X. hortorum* strains would be limited to *Xhv* and possibly *Xhg* strains.

A

Query	1	MASGKTLNAEAICQAYGLKRVVELDERLQRKGEDWPLSQ
X. hortorum	11	MASGKTLNAEAICQAYGLKRVVELDERLQRKGEDWPLSQ
X. hortorum pv. *vitians*	1	MASGKTLNAEAICQAYGLKRVVELDERLQRKGEDWPLSQ

B NCBI Multiple Sequence Alignment Viewer, Version 1.24.0



NDVLMLTNDHALAERAAQRMRIQTVAIVEARERVGASWRAPR 243 Matches:
NDVLMLTNDHALAERAAQRMRIQTVAIVEARERVGASWRAPR 91 Hypothetical protein (100%)
NDVLMLTNDHALAERAAQRMRIQTVAIVEARERVGASWRAPR 81 Hypothetical protein (100%)

End	Organism
246	
66,899	Xanthomonas hortorum pv. vitians
2,208	Xanthomonas hortorum pv. vitians
2,208	Xanthomonas hortorum pv. vitians
2,208	Xanthomonas hortorum pv. vitians
2,208	Xanthomonas hortorum pv. vitians
2,207	Xanthomonas hortorum pv. vitians
2,207	Xanthomonas hortorum pv. vitians
2,335	Xanthomonas hortorum pv. vitians
2,334	Xanthomonas hortorum pv. vitians
76,373	Xanthomonas hortorum pv. vitians
75,811	Xanthomonas hortorum pv. vitians
2,335	Xanthomonas hortorum pv. vitians
2,408	Xanthomonas hortorum pv. vitians
2,335	Xanthomonas hortorum pv. vitians
2,334	Xanthomonas hortorum pv. vitians
2,335	Xanthomonas hortorum pv. vitians
2,335	Xanthomonas hortorum pv. vitians
58,604	Xanthomonas hortorum pv. vitians
2,333	Xanthomonas hortorum pv. vitians
2,330	Xanthomonas hortorum pv. vitians
2,335	Xanthomonas hortorum pv. vitians
2,335	Xanthomonas hortorum pv. vitians
2,334	Xanthomonas hortorum pv. vitians
19,038	Xanthomonas hortorum pv. vitians
2,370	Xanthomonas hortorum pv. vitians
2,205	Xanthomonas hortorum pv. gardneri

Figure 4-3. BLASTx protein alignments for top hits (A) and BLASTn DNA alignments (B) to query GC3889 sequence. (A) The top five protein matches to the GC3889 query DNA sequence; percent identity value for each match is shown in parentheses. **(B)** Multiple sequence alignment of the GC3889 DNA sequence query to top DNA hits. Arrows represent the binding sites for the forward and reverse primers in 5' to 3' and 3' to 5' orientation, respectively. In this figure, strains submitted as whole genome sequences in Rosenthal et al., 2022 include those with the sequence IDs starting with 'JA'

Upon empirical testing, the targeted 110 bp amplicon was produced for 96 of the 99 *Xhv* strains tested (Figure 4-4). The lack of any amplicon for the known *Xhv* race 1 strain BS0347 was unexpected; this strain was included in our in-silico analyses that demonstrated GC3889 to be present in eighteen *Xhv* strains tested. The strain did produce a band using our primer set for detection of *Xhv* race 1 strains. It is possible that there is sequence variation at the site of GC3889 in BS0347 that is not well represented by the genome assembly, such as long repeat sequences that do not resolve well in assemblies such as ours that are built from short reads. Such variation could prevent primer binding. The predicted *Xhv* race 1 strain BS2997 and predicted *Xhv* race 2 strain BS3126 also did not amplify the expected 110 bp band; however, the former did amplify the correct size fragment using our *Xhv* race 1 detection method and the latter did amplify using the correct size fragment using our *Xhv* race 2 detection method. It may be that these strains also have sequence variation at this gene cluster that prevents primer binding. The fragment was not produced for any of the related *Xanthomonas hortorum* pathotype strains outside of *X. hortorum* pv. *vitians*, nor for the *X. campestris* pv. *coriandri* strain, which was hypothesized to also be a member of *X. hortorum* (Zacaroni et al., unpublished). Further research is necessary to determine why this and two other *Xhv* strains, BS2997 and BS3126, were not detected using this method. One approach could be long-read sequencing of these genomes to look for sequence variation in this gene cluster.

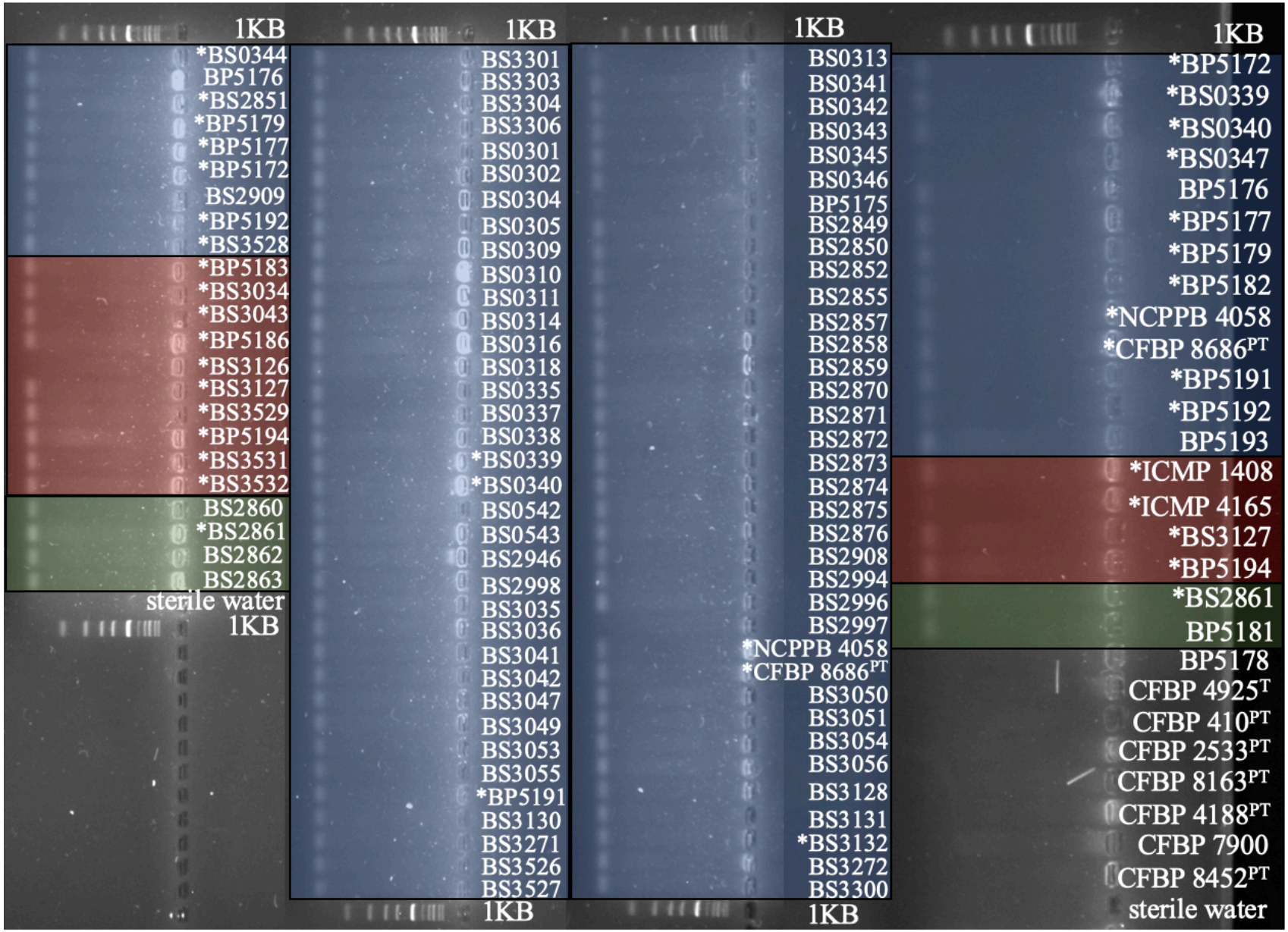

1KB
*BS0344
BP5176
*BS2851
*BP5179
*BP5177
*BP5172
BS2909
*BP5192
*BS3528
*BP5183
*BS3034
*BS3043
*BP5186
*BS3126
*BS3127
*BS3529
*BP5194
*BS3531
*BS3532
BS2860
*BS2861
BS2862
BS2863
sterile water
1KB
1KB
BS3301
BS3303
BS3304
BS3306
BS0301
BS0302
BS0304
BS0305
BS0309
BS0310
BS0311
BS0314
BS0316
BS0318
BS0335
BS0337
BS0338
*BS0339
*BS0340
BS0542
BS0543
BS2946
BS2998
BS3035
BS3036
BS3041
BS3042
BS3047
BS3049
BS3053
BS3055
*BP5191
BS3130
BS3271
BS3526
BS3527
1KB
1KB
BS0313
BS0341
BS0342
BS0343
BS0345
BS0346
BP5175
BS2849
BS2850
BS2852
BS2855
BS2857
BS2858
BS2859
BS2870
BS2871
BS2872
BS2873
BS2874
BS2875
BS2876
BS2908
BS2994
BS2996
BS2997
*NCPPB 4058
*CFBP 8686PT
BS3050
BS3051
BS3054
BS3056
BS3128
BS3131
*BS3132
BS3272
BS3300
1KB
1KB
*BP5172
*BS0339
*BS0340
*BS0347
BP5176
*BP5177
*BP5179
*BP5182
*NCPPB 4058
*CFBP 8686PT
*BP5191
*BP5192
BP5193
*ICMP 1408
*ICMP 4165
*BS3127
*BP5194
*BS2861
BP5181
BP5178
CFBP 4925T
CFBP 410PT
CFBP 2533PT
CFBP 8163PT
CFBP 4188PT
CFBP 7900
CFBP 8452PT
sterile water

Figure 4-4. Amplification of known *Xhv* strains and related *Xanthomonas* strains using the GC3889-110 primer set. Gel electrophoresis showing the amplification produced using genomic DNA extracts (first row) and colony suspensions (three bottom rows) and the GC3889 primer set. Strains tagged with an asterisk have been demonstrated by HR screening to be either *Xhv* race 1 (blue), *Xhv* race 2 (red), or *Xhv* race 3 (green), while all other *Xhv* strains are hypothesized to belong to one of the three races based on MLSA sequetype (Bull et al., 2016). Closely related strains included to test the specificity of the protocol were *X. hortorum* isolated from radicchio (BP5178), *X. hortorum* pv. *hederae* (CFBP 4925^{PT}), *X. hortorum* pv. *taraxaci* (CFBP 410^{PT}), *X. hortorum* pv. *pelargonii* (CFBP 2533^{PT}), *X. hortorum* pv. *gardneri* (CFBP 8163^{PT}), *X. hortorum* pv. *cynarae* (CFBP4188^{PT}), *X. hortorum* pv. *carotae* (CFBP 7900), and *X. campestris* pv. *coriandri* (CFBP 8452^{PT}). A 1 kb ladder from New England Biolabs was run as a size reference, and the 1% agarose gel was run for 2 hours at 84 V.

4.3.4 Development and evaluation of *Xhv* race 1-specific detection method.

The *Xhv* race 1 specific gene cluster GC4021 encoded for the conserved domain of the helix-turn-helix (HTH) xenobiotic response element (XRE) family of transcriptional regulators. This result was supported by the BLASTx results (Figure 4-5A), which revealed four out of the top five matches to be to HTH transcriptional regulators. These were identified from two *Xhv* strains (strain 1, 100% identity, 99% query cover, and 6e-79 e-value; strain 2, 100% identity, 99% query cover, and 7e-79 e-value) and two *X. hortorum* strains (strain 1, 92.3% identity, 92% query cover, and 1e-67 e-value; strain 2, 85.5% identity, 97% query cover, 1e-65 e-value). One of the top five matches was to a hypothetical protein of an *Xhv* strain (92.3% identity, 92% query cover, and 1e-67 e-value), but the other matches provide more evidence to support the hypothesis

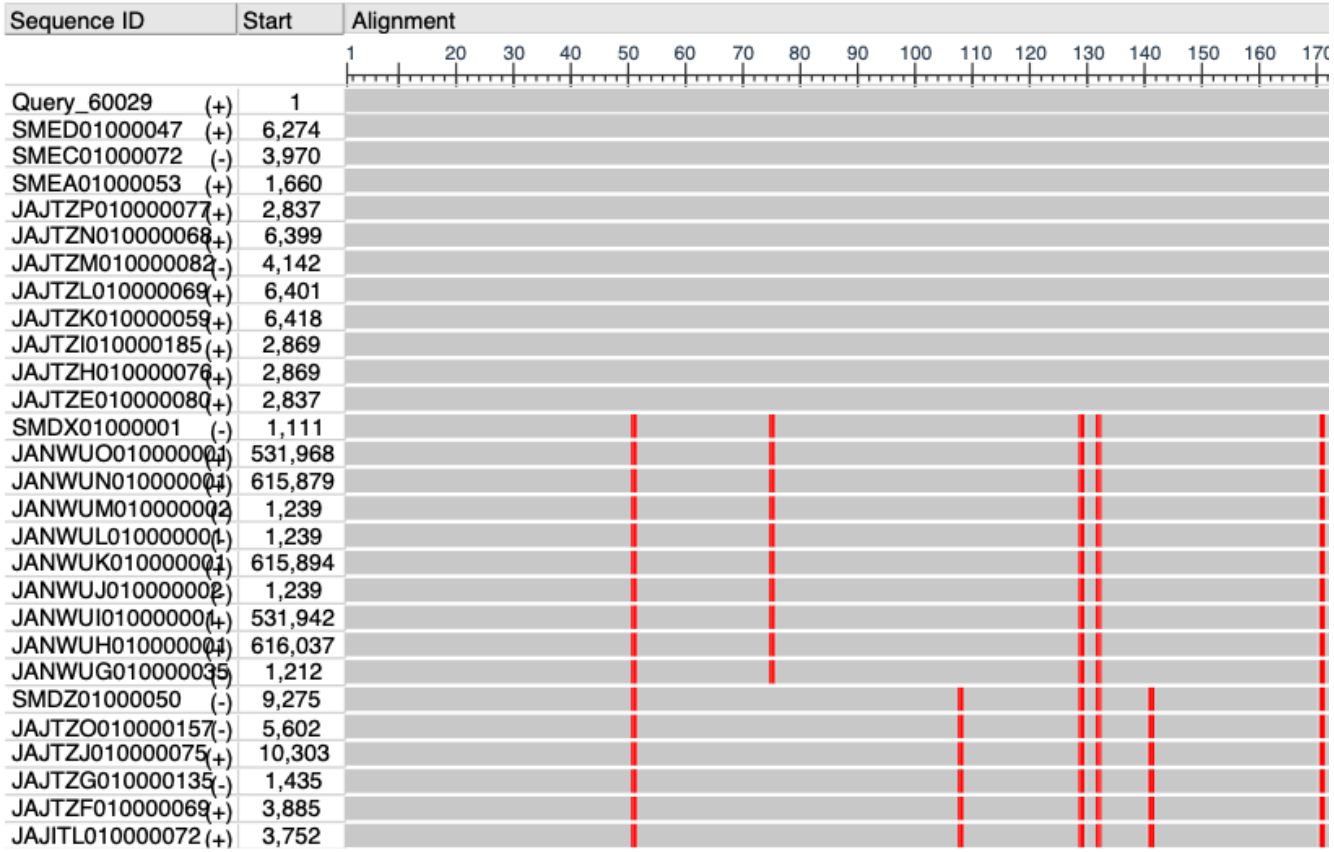
that this gene cluster encodes an HTH transcriptional regulator. These proteins are known to bind HTH regions in the genome that are upstream of XREs, and are triggered in response to chemical stress, such as the presence of reactive oxygen species (ROS) (Baldwin, 2019). XRE activity has been shown to protect bacteria against stress, but more research is necessary to determine the role of this system in *Xhv* strains.

The multiple DNA sequence alignment for GC4021 produced using BLASTn showed that this gene cluster had identical matches in all twelve *Xhv* race 1 whole genome sequences published in Rosenthal et al., 2022 (Figure 4-5B, sequence IDs 'JAJTZ…'). Many *X. hortorum* pv. *pelargonii* strains matched, but with sequence variation in where the GC4021-112F forward primer would bind and truncation that leaves no binding sequence for the GC4021-112R reverse primer to bind. The same is true for one *X. hortorum* pv. *gardneri* strain match. These results suggested that among *X. hortorum* strains, only *Xhv* race 1 strains would be targeted through amplification of the GC4021-112 fragment.

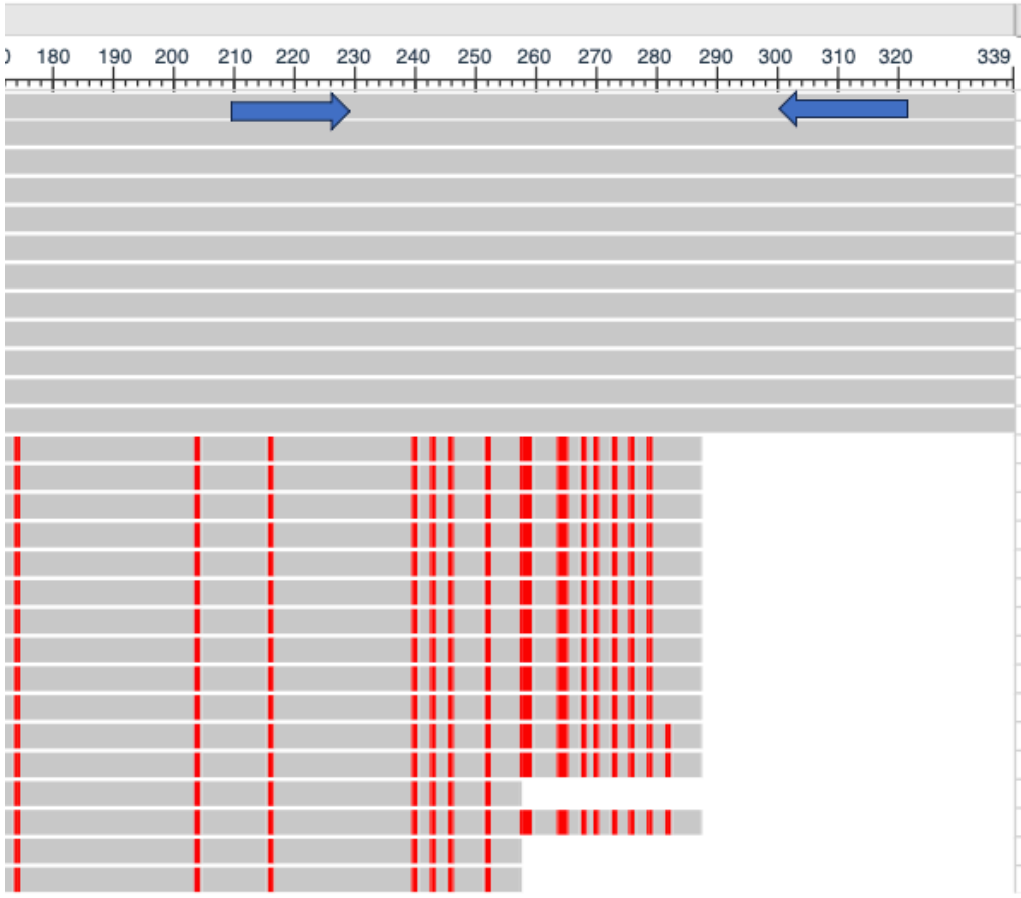
A

Query	1	MPKPKTPATVYGTRLRHARMAMGWTQAELAERIGMVD
X. hortorum pv. *vitians*	1	MPKPKTPATVYGTRLRHARMAMGWTQAELAERIGMVD
X. hortorum pv. *vitians*	21	MPKPKTPATVYGTRLRHARMAMGWTQAELAERIGMVD
X. hortorum	1	MPKPKTPATVYGTRLRHARMAMGWTQAELAERIGMVD
X. hortorum pv. *vitians*	21	MPKPKTPATVYGTRLRHARMAMGWTQAELAERIGMVD
X. hortorum	21	MPKPKTPATVYGTRLRHARMAMGWTQAELAERIGMVD

B NCBI Multiple Sequence Alignment Viewer, Version 1.24.0



SVSGATRVSRYETGQHDPDPATA	180	Matches:
SVSGATRVSRYETGQHDPDPATA	60	HTH transcriptional regulator (100%)
SVSGATRVSRYETGQHDPDPATA	80	HTH transcriptional regulator (100%)
SVSGATRVSRYETGQHDPDPATA	60	HTH transcriptional regulator (92.3%)
SVSGATRVSRYETGQHDPDPATA	80	Hypothetical protein (92.3%)
SVSGATRVSRYETGQHDPDPATA	80	HTH transcriptional regulator (85.5%)



End	Organism
339	
6,612	Xanthomonas hortorum pv. vitians
3,632	Xanthomonas hortorum pv. vitians
1,998	Xanthomonas hortorum pv. vitians
3,175	Xanthomonas hortorum pv. vitians
6,737	Xanthomonas hortorum pv. vitians
3,804	Xanthomonas hortorum pv. vitians
6,739	Xanthomonas hortorum pv. vitians
6,756	Xanthomonas hortorum pv. vitians
3,207	Xanthomonas hortorum pv. vitians
3,207	Xanthomonas hortorum pv. vitians
3,175	Xanthomonas hortorum pv. vitians
825	Xanthomonas hortorum pv. pelargonii
532,254	Xanthomonas hortorum pv. pelargonii
616,165	Xanthomonas hortorum pv. pelargonii
953	Xanthomonas hortorum pv. pelargonii
953	Xanthomonas hortorum pv. pelargonii
616,180	Xanthomonas hortorum pv. pelargonii
953	Xanthomonas hortorum pv. pelargonii
532,228	Xanthomonas hortorum pv. pelargonii
616,323	Xanthomonas hortorum pv. pelargonii
926	Xanthomonas hortorum pv. pelargonii
8,989	Xanthomonas hortorum pv. vitians
5,316	Xanthomonas hortorum pv. vitians
10,559	Xanthomonas hortorum pv. vitians
1,149	Xanthomonas hortorum pv. vitians
4,141	Xanthomonas hortorum pv. vitians
4,008	Xanthomonas hortorum pv. gardneri

Figure 4-5. BLASTx protein alignments for top hits (A) and BLASTn DNA alignments (B) to query GC4021 sequence. (A) The top five protein matches to the GC4021 query DNA sequence; percent identity value for each match is shown in parentheses. **(B)** Multiple sequence alignment of the GC4021 DNA sequence query to top DNA hits. Arrows represent the binding sites for the forward and reverse primers designed for this gene cluster in 5' to 3' and 3' to 5' orientation, respectively. Red bars represent nucleotide substitutions. Whole genomes of *Xhv* strains submitted by Rosenthal et al., 2022 include those with sequence IDs starting with 'JAJTZ.'

Empirical testing of our *Xhv* race 1 detection protocol resulted in the amplification of a 112 bp fragment for 52 out of 85 strains that were either demonstrated to be *Xhv* race 1 by HR screening on *L. sativa* cv. Little Gem or were predicted to be race 1 because they clustered with race 1 strains in MLSA (Figure 4-6) (Bull et al., 2016). Among the thirty-three strains that are not detected by this protocol, two of them are known *Xhv* race 1 strains due to their HR phenotypes on *L. sativa* cv. Little Gem. These two strains, BP5172 and BP5177, were included in the whole genome sequencing by Rosenthal et al., 2022 and were included in the anvi'o alignments, and so in silico testing had demonstrated that they should produce the 112 bp target amplicon. The two strains did amplify using the GC3889-110 primers designed for *Xhv* detection. These results suggest that these two strains may have some sequence variation in this gene cluster that is not well represented by their whole genome assemblies, such as long repetitive sequences that do not resolve well in assemblies built from short reads. This may be true, too, for the other twenty-eight predicted *Xhv* race 1 strains that are not detected using this method, all of which cluster with *Xhv* race 1 strains by MLSA. Long-read sequencing of these strains, and the *Xhv* race 1 strains that do amplify, followed by their assembly and gene cluster alignment, could reveal whether there are sequence differences that prevent successful amplification of the 112 bp fragment from all *Xhv* race 1 strains. Further, this could help identify better target sequence for the detection of all *Xhv* race 1 strains. None of the *Xhv* race 2 or 3 strains, nor any of the non-*Xhv* strains tested, produced a DNA fragment using this protocol; these results were expected because this gene cluster was not present in these strains as observed in the anvi'o alignment.


1KB
*BS0344
BP5176
*BS2851
*BP5179
*BP5177
*BP5172
BS2909
*BP5192
*BS3528
*BP5183
*BS3034
*BS3043
*BP5186
*BS3126
*BS3127
*BS3529
*BP5194
*BS3531
*BS3532
BS2860
*BS2861
BS2862
BS2863
sterile water
1KB
1KB
BS3301
BS3303
BS3304
BS3306
BS0301
BS0302
BS0304
BS0305
BS0309
BS0310
BS0311
BS0314
BS0316
BS0318
BS0335
BS0337
BS0338
*BS0339
*BS0340
BS0542
BS0543
BS2946
BS2998
BS3035
BS3036
BS3041
BS3042
BS3047
BS3049
BS3053
BS3055
*BP5191
BS3130
BS3271
BS3526
BS3527
1KB
1KB
BS0313
BS0341
BS0342
BS0343
BS0345
BS0346
BP5175
BS2849
BS2850
BS2852
BS2855
BS2857
BS2858
BS2859
BS2870
BS2871
BS2872
BS2873
BS2874
BS2875
BS2876
BS2908
BS2994
BS2996
BS2997
*NCPPB 4058
*CFBP 8686PT
BS3050
BS3051
BS3054
BS3056
BS3128
BS3131
*BS3132
BS3272
BS3300
1KB
*BP5172
*BS0339
*BS0340
*BS0347
BP5176
*BP5177
*BP5179
*BP5182
*NCPPB 4058
*CFBP 8686PT
1KB
1KB
*BP5191
*BP5192
BP5193
*ICMP 1408
*ICMP 4165
*BS3127
*BP5194
*BS2861
BP5181
BP5178
CFBP 4925T
CFBP 410PT
CFBP 2533PT
CFBP 8163PT
CFBP 4188PT
CFBP 7900
CFBP 8452PT
sterile water

Figure 4-6. Amplification of known *Xhv* strains and related *Xanthomonas* strains using the GC4021-112 primer set. Gel electrophoresis showing the amplification produced using genomic DNA extracts (first row) and colony suspensions (three bottom rows) and the GC4021 primer set. Strains tagged with an asterisk have been demonstrated by HR screening to be either *Xhv* race 1 (blue), *Xhv* race 2 (red), or *Xhv* race 3 (green), while all other *Xhv* strains are hypothesized to belong to one of the three races based on MLSA sequetype (Bull et al., 2016). Closely related strains included to test the specificity of the protocol were *X. hortorum* isolated from radicchio (BP5178), *X. hortorum* pv. *hederae* (CFBP 4925^{PT}), *X. hortorum* pv. *taraxaci* (CFBP 410^{PT}), *X. hortorum* pv. *pelargonii* (CFBP 2533^{PT}), *X. hortorum* pv. *gardneri* (CFBP 8163^{PT}), *X. hortorum* pv. *cynarae* (CFBP4188^{PT}), *X. hortorum* pv. *carotae* (CFBP 7900), and *X. campestris* pv. *coriandri* (CFBP 8452^{PT}). A 1 kb ladder from New England Biolabs was run as a size reference, and the 1% agarose gel was run for 2 hours at 84 V.

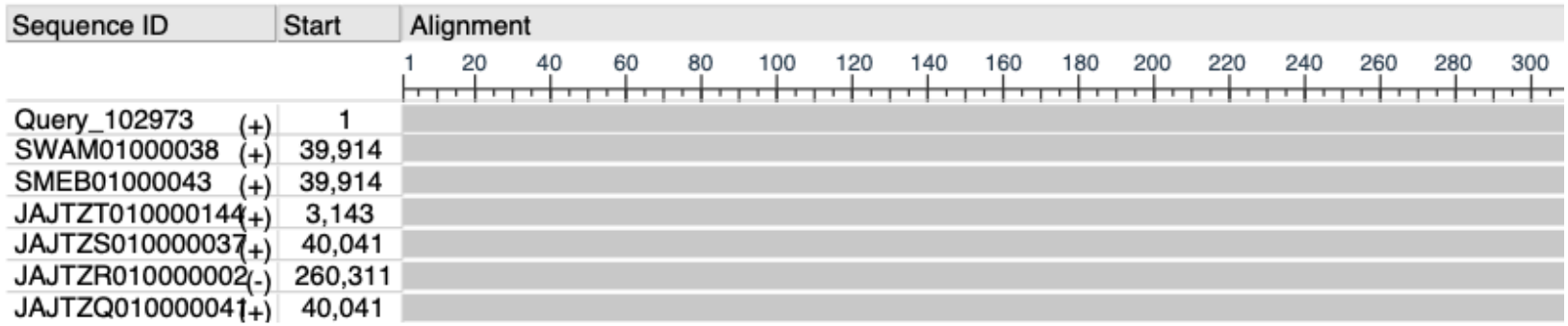
4.3.5 Development and evaluation of *Xhv* race 2-specific detection method.

For the *Xhv* race 2 specific gene cluster, GC4381, the sole protein match identified from *X. hortorum* strains was a hypothetical protein from *Xhv* (100% identity, 99% query cover, and 3e-150 e-value) (Figure 4-7A). The multiple sequence alignment showed identical matches to all four *Xhv* race 2 strains sequenced in Rosenthal et al., 2022 (sequence IDs starting with 'JAJTZ') and two additional *Xhv* strains in the NCBI whole genome sequence database that we now hypothesize to also be *Xhv* race 2 strains (Figure 4-7B). If this is true, then these results suggest that among *X. hortorum* strains, only *Xhv* race 2 strains would result in a 178 bp fragment upon PCR with the GC4381-178 primer

A

```
Query                          1    MTIRIIFLTAGLAAIFFPTLNIVMSLSVDKTFSLSDAAGIAAGIA
X. hortorum pv. vitians        1    MTIRIIFLTAGLAAIFFPTLNIVMSLSVDKTFSLSDAAGIAAGIA

Query                        451    MSTVYPGNQRGGSRPDSIKQSAATLNASARKFAPQYDEFVRYTRR
X. hortorum pv. vitians      151    MSTVYPGNQRGGSRPDSIKQSAATLNASARKFAPQYDEFVRYTRR
```

B



AGIAAIASTFLGASKEKESVRHDREIFAKFLRDLPHKPSIQLLKDHDFGGDYYKSQVQHLFDFVREW
AGIAAIASTFLGASKEKESVRHDREIFAKFLRDLPHKPSIQLLKDHDFGGDYYKSQVQHLFDFVREW

YTRRRLSMES 603 Match:
YTRRRLSMES 201 Hypothetical protein (100%)

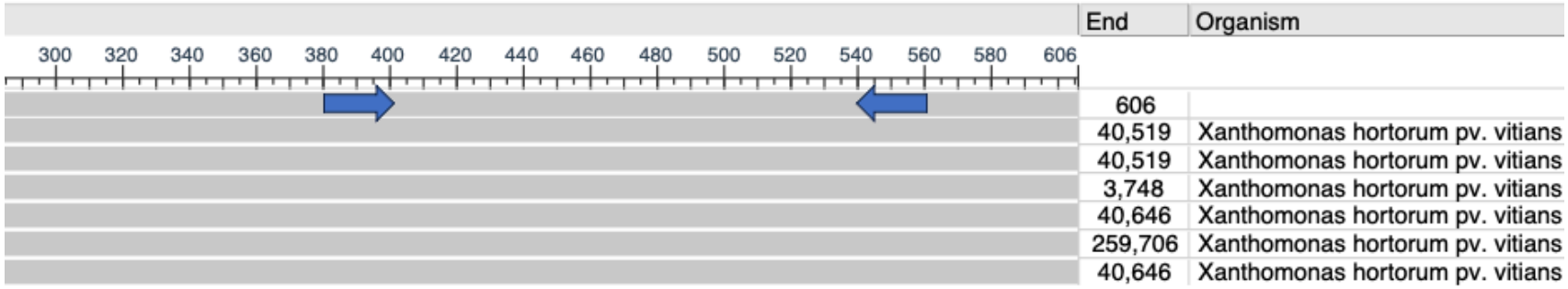


Figure 4-7. BLASTx protein alignments for top hits (A) and BLASTn DNA alignments (B) to query GC4381 sequence. (A) The singular protein match to the GC4381 query DNA sequence; percent identity value for the match is shown in parentheses. **(B)** Multiple sequence alignment of the GC4381 DNA sequence query to top DNA hits. Arrows represent the binding sites for the forward and reverse primers designed for this gene cluster in 5' to 3' and 3' to 5' orientation, respectively. Whole genomes of *Xhv* strains submitted by Rosenthal et al., 2022 include those with sequence IDs starting with 'JAJTZ.'

The *Xhv* race 2 detection protocol resulted in a 178 bp fragment for eight of the ten *Xhv* race 2 strains (Figure 4-8); the two *Xhv* race 2 strains that were not detected were BS3531 and BS3532. These two strains were not included in the anvi'o sequence analysis that we used to identify race-specific sequences because we lacked whole genome sequence data for these strains. These two strains were also not included in Rosenthal et al., 2022, and so the multiple sequence alignments in Figure 4-8 also did not include these strains. It is possible that these strains do not encode GC4381 as we had hypothesized, or that they have sufficient sequence variation in this gene cluster that prevents our primers from binding. Amplification of the region that should contain this gene cluster with external primers could help identify whether it is absent or has significant sequence variation. Two *Xhv* race 1 strains (BP5182 and CFBP 8686^{PT}), and one hypothesized race 1 strain (BS2996), produced the 178 bp fragment. Whole genome sequences for BP5182 and CFBP 8686^{PT} were included in Rosenthal et al., 2022 and they did not appear as hits in our multiple sequence alignment of this gene cluster, nor did they show GC4381 in the anvi'o alignment, so their detection using this method was unexpected. It is possible that the assemblies used in these sequence analyses had gaps where the GC4381 sequence would be because they were produced with short sequence reads; long-read sequencing could fill in some of these gaps and reveal whether these strains do encode this gene cluster.

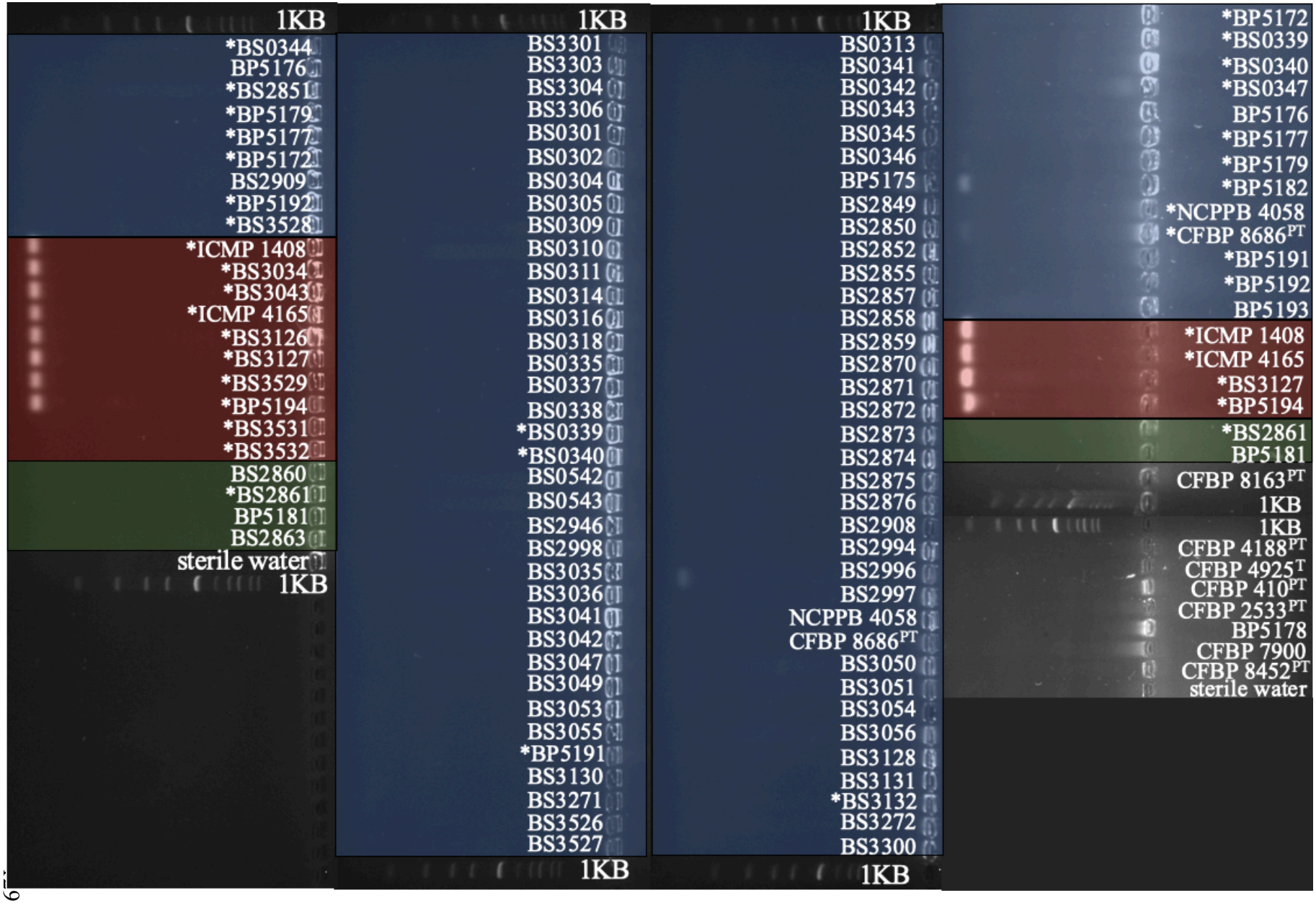

1KB
*BS0344
BP5176
*BS2851
*BP5179
*BP5177
*BP5172
BS2909
*BP5192
*BS3528
*ICMP 1408
*BS3034
*BS3043
*ICMP 4165
*BS3126
*BS3127
*BS3529
*BP5194
*BS3531
*BS3532
BS2860
*BS2861
BP5181
BS2863
sterile water
1KB
1KB
BS3301
BS3303
BS3304
BS3306
BS0301
BS0302
BS0304
BS0305
BS0309
BS0310
BS0311
BS0314
BS0316
BS0318
BS0335
BS0337
BS0338
*BS0339
*BS0340
BS0542
BS0543
BS2946
BS2998
BS3035
BS3036
BS3041
BS3042
BS3047
BS3049
BS3053
BS3055
*BP5191
BS3130
BS3271
BS3526
BS3527
1KB
1KB
BS0313
BS0341
BS0342
BS0343
BS0345
BS0346
BP5175
BS2849
BS2850
BS2852
BS2855
BS2857
BS2858
BS2859
BS2870
BS2871
BS2872
BS2873
BS2874
BS2875
BS2876
BS2908
BS2994
BS2996
BS2997
NCPPB 4058
CFBP 8686PT
BS3050
BS3051
BS3054
BS3056
BS3128
BS3131
*BS3132
BS3272
BS3300
1KB
*BP5172
*BS0339
*BS0340
*BS0347
BP5176
*BP5177
*BP5179
*BP5182
*NCPPB 4058
*CFBP 8686PT
*BP5191
*BP5192
BP5193
*ICMP 1408
*ICMP 4165
*BS3127
*BP5194
*BS2861
BP5181
CFBP 8163PT
1KB
1KB
CFBP 4188PT
CFBP 4925T
CFBP 410PT
CFBP 2533PT
BP5178
CFBP 7900
CFBP 8452PT
sterile water

Figure 4-8. Amplification of known *Xhv* strains and related *Xanthomonas* strains using the GC4381-178 primer set. Gel electrophoresis showing the amplification produced using genomic DNA extracts (first row) and colony suspensions (three bottom rows) and the GC4381 primer set. Strains tagged with an asterisk have been demonstrated by HR screening to be either *Xhv* race 1 (blue), *Xhv* race 2 (red), or *Xhv* race 3 (green), while all other *Xhv* strains are hypothesized to belong to one of the three races based on MLSA sequetype (Bull et al., 2016). Closely related strains included to test the specificity of the protocol were *X. hortorum* isolated from radicchio (BP5178), *X. hortorum* pv. *hederae* (CFBP 4925^{PT}), *X. hortorum* pv. *taraxaci* (CFBP 410^{PT}), *X. hortorum* pv. *pelargonii* (CFBP 2533^{PT}), *X. hortorum* pv. *gardneri* (CFBP 8163^{PT}), *X. hortorum* pv. *cynarae* (CFBP4188^{PT}), *X. hortorum* pv. *carotae* (CFBP 7900), and *X. campestris* pv. *coriandri* (CFBP 8452^{PT}). A 1 kb ladder from New England Biolabs was run as a size reference, and the 1% agarose gel was run for 2 hours at 84 V.

4.3.6 Development and evaluation of *Xhv* race 3-specific detection method.

The *Xhv* race 3 specific gene cluster, GC4980, matched only to a hypothetical protein encoded by *Xhv* (100% identity, query cover, and e-value) (Figure 4-9A). The multiple sequence alignment showed two identical matches, and both were known *X. hortorum* pv. *vitians* race 3 strains from Rosenthal et al., 2022 (Figure 4-9B). The absence of non-target matches supports the hypothesis that this region can be targeted for *Xhv* race 3-specific amplification that would exclude all other strains of *X. hortorum*.

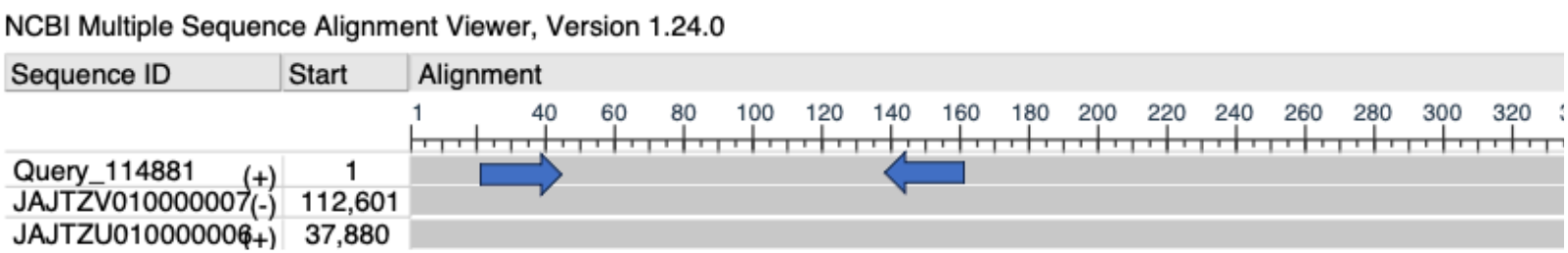
The *Xhv* race 3 detection method resulted in a 138 bp band for the known *Xhv* race 3 strain BS2861 and the predicted race 3 strains BS2860, BS2862, and BS2863 (Figure 4-10). Non-target amplification of the 138 bp fragment was also seen for *X. hortorum* pv. *carotae*

(CFBP 7900), which was unexpected because this strain did not appear to have matching sequence to GC4980 in our anvi'o alignment or in our search using NCBI's BLASTn. It is possible that sequence alignments failed to show GC4980 for this strain due to gaps in our short-read sequencing data; this could be addressed using long-read sequencing technology to produce an improved assembly with fewer gaps, and then incorporating it into the in-silico development of the *Xhv* race 3-specific primer set. The predicted *Xhv* race 1 strain BS3272 also produced a 138 bp band, and this too was unexpected due to its clustering with *Xhv* race 1 strains in MLSA and detection among other *Xhv* race 1 strains using our GC4021-112 primer set. Whole genome sequence data is not yet available for this strain, so it could not be incorporated into the in silico analyses, and HR phenotype data is not yet available for this strain, so its race has not yet been confirmed. Completing the HR screening with this strain and assigning its race will help to determine whether GC4980 is capable of also amplifying the 138 bp fragment from a *Xhv* race 1 strain

A

Query	1	MPPGKSWLTQKPTLSIMPNHADYWHVFKEPHQATEYNQNSK
X. hortorum pv. *vitians*	1	MPPGKSWLTQKPTLSIMPNHADYWHVFKEPHQATEYNQNSK
Query	451	SGFGMPGMPCLAWVSRQHSGSNVFLCYDPPTKPDNWLIKDNVVA
X. hortorum pv. *vitians*	151	SGFGMPGMPCLAWVSRQHSGSNVFLCYDPPTKPDNWLIKDNVVA

B



ARFALTGDSRGMFYVGDTLNGVLWETVLRDVEPNRRLEVDVDVNQLRDMRAVRLALMEPALPLLDLR
ARFALTGDSRGMFYVGDTLNGVLWETVLRDVEPNRRLEVDVDVNQLRDMRAVRLALMEPALPLLDLR

LDDPRTGHDLIKKALEEAGFSWAPAYYTSAI 675 Match:
LDDPRTGHDLIKKALEEAGFSWAPAYYTSAI 225 Hypothetical protein (100%)

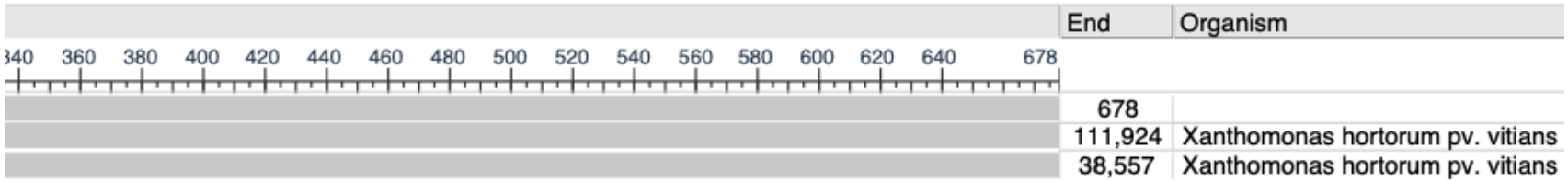


Figure 4-9. BLASTx protein alignments for top hits (A) and BLASTn DNA alignments (B) to query GC4980 sequence. (A) The top five protein matches to the GC4980 query DNA sequence; percent identity value for each match is shown in parentheses. (B) Multiple sequence alignment of the GC4980 DNA sequence query to top DNA hits. Arrows represent the binding sites for the forward and reverse primers in 5' to 3' and 3' to 5' orientation, respectively.

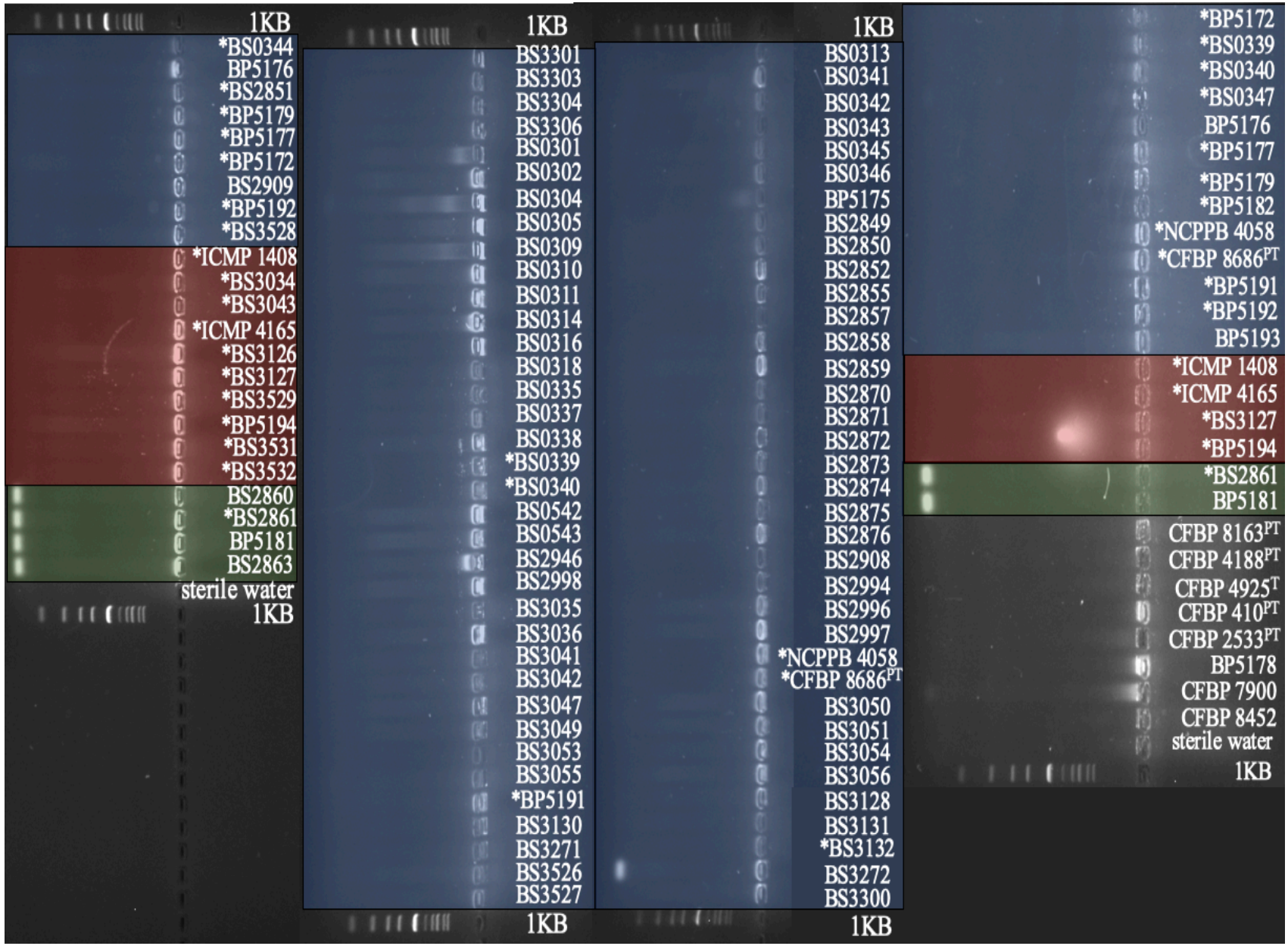

1KB
*BS0344
BP5176
*BS2851
*BP5179
*BP5177
*BP5172
BS2909
*BP5192
*BS3528
*ICMP 1408
*BS3034
*BS3043
*ICMP 4165
*BS3126
*BS3127
*BS3529
*BP5194
*BS3531
*BS3532
BS2860
*BS2861
BP5181
BS2863
sterile water
1KB
1KB
BS3301
BS3303
BS3304
BS3306
BS0301
BS0302
BS0304
BS0305
BS0309
BS0310
BS0311
BS0314
BS0316
BS0318
BS0335
BS0337
BS0338
*BS0339
*BS0340
BS0542
BS0543
BS2946
BS2998
BS3035
BS3036
BS3041
BS3042
BS3047
BS3049
BS3053
BS3055
*BP5191
BS3130
BS3271
BS3526
BS3527
1KB
1KB
BS0313
BS0341
BS0342
BS0343
BS0345
BS0346
BP5175
BS2849
BS2850
BS2852
BS2855
BS2857
BS2858
BS2859
BS2870
BS2871
BS2872
BS2873
BS2874
BS2875
BS2876
BS2908
BS2994
BS2996
BS2997
*NCPPB 4058
*CFBP 8686PT
BS3050
BS3051
BS3054
BS3056
BS3128
BS3131
*BS3132
BS3272
BS3300
1KB
*BP5172
*BS0339
*BS0340
*BS0347
BP5176
*BP5177
*BP5179
*BP5182
*NCPPB 4058
*CFBP 8686PT
*BP5191
*BP5192
BP5193
*ICMP 1408
*ICMP 4165
*BS3127
*BP5194
*BS2861
BP5181
CFBP 8163PT
CFBP 4188PT
CFBP 4925T
CFBP 410PT
CFBP 2533PT
BP5178
CFBP 7900
CFBP 8452
sterile water
1KB

Figure 4-10. Amplification of known *Xhv* strains and related *Xanthomonas* strains using the GC4980-138 primer set. Gel electrophoresis showing the amplification produced using genomic DNA extracts (first row) and colony suspensions (three bottom rows) and the GC4980 primer set. Strains tagged with an asterisk have been demonstrated by HR screening to be either *Xhv* race 1 (blue), *Xhv* race 2 (red), or *Xhv* race 3 (green), while all other *Xhv* strains are hypothesized to belong to one of the three races based on MLSA sequetype (Bull et al., 2016). Closely related strains included to test the specificity of the protocol were *X. hortorum* isolated from radicchio (BP5178), *X. hortorum* pv. *hederae* (CFBP 4925^{PT}), *X. hortorum* pv. *taraxaci* (CFBP 410^{PT}), *X. hortorum* pv. *pelargonii* (CFBP 2533^{PT}), *X. hortorum* pv. *gardneri* (CFBP 8163^{PT}), *X. hortorum* pv. *cynarae* (CFBP4188^{PT}), *X. hortorum* pv. *carotae* (CFBP 7900), and *X. campestris* pv. *coriandri* (CFBP 8452^{PT}). A 1 kb ladder from New England Biolabs was run as a size reference, and the 1% agarose gel was run for 2 hours at 84 V.

4.3.7 Evaluation of suspected *Xhv* isolates and novel detection protocols.

For the symptomatic lettuce heads that we received and were experiencing a suspected *Xhv* infection, five yellow, mucoid strains were isolated from romaine lettuce leaves (BP4469-BP4473), four were isolated from green lettuce leaves (BP4474-BP4477), and two were isolated from red lettuce leaves (BP4478 and BP4479). These isolates, along with the DNA extracted from the nine romaine and green leaves (BP4447-BP4455; BP4456-BP4464), and four red leaves (BP4465-4468), were evaluated using our touchdown PCR detection methods (Figure 4-11). Two of the nine green leaf extracts (BP4456 and BP4458) and two of the four red leaf extracts (BP4465 and BP4468) resulted in faint 110 bp bands using the *Xhv*-specific primer set GC3889. All four suspected *Xhv* bacterial isolates from green leaf lettuce resulted in bright bands using

this primer set. None of the other PA extracts or isolates produced any bands, and it was later determined that they belonged to other genera (Table 4-4). The *Xhv* race 1, 2, and 3 detection protocols did not result in detection for any of the PA extracts or isolates. The controls (*Xhv* strains representing races 1, 2, and 3) all did produce the appropriately sized bands as expected.

These results suggested that both green leaf and red leaf lettuce heads were experiencing infection with *Xhv*, but the romaine lettuce heads were not. Further work was necessary to determine the validity of our detection method results, namely whether our touchdown PCR methods produced false positives or false negatives. Regarding the strains that tested positive for *Xhv*, our detection protocols did not reveal their races. These strains might belong to a known *Xhv* race but go undetected by our methods, or these strains might belong to a yet unknown *Xhv* race and have sufficient sequence variation that allowed them to evade detection. MLSA was then completed to evaluate these hypotheses.

The MLSA results supported those of our touchdown PCR detection method. One plant extract from green leaf lettuce (BP4456) and the four Xanthomonad isolates from leaf tissue (BP4474-BP4477) clustered with known *Xhv* strains, specifically those of race 1 (Figure 4-12). The other green leaf extract (BP4458) and the two red leaf extracts (BP4465 and BP4468) produced poor quality sequence data and were excluded from this analysis. These results confirm that these strains are *Xhv*, and we hypothesize that these strains do belong to *Xhv* race 1 but are members of the group of race 1 strains that evade detection when using the GC4021 primer set (Section 4.3.3). HR screening is necessary to confirm this hypothesis; only if they produce HR in *L. sativa* cv. Little Gem can they be confirmed to be *Xhv* race 1.

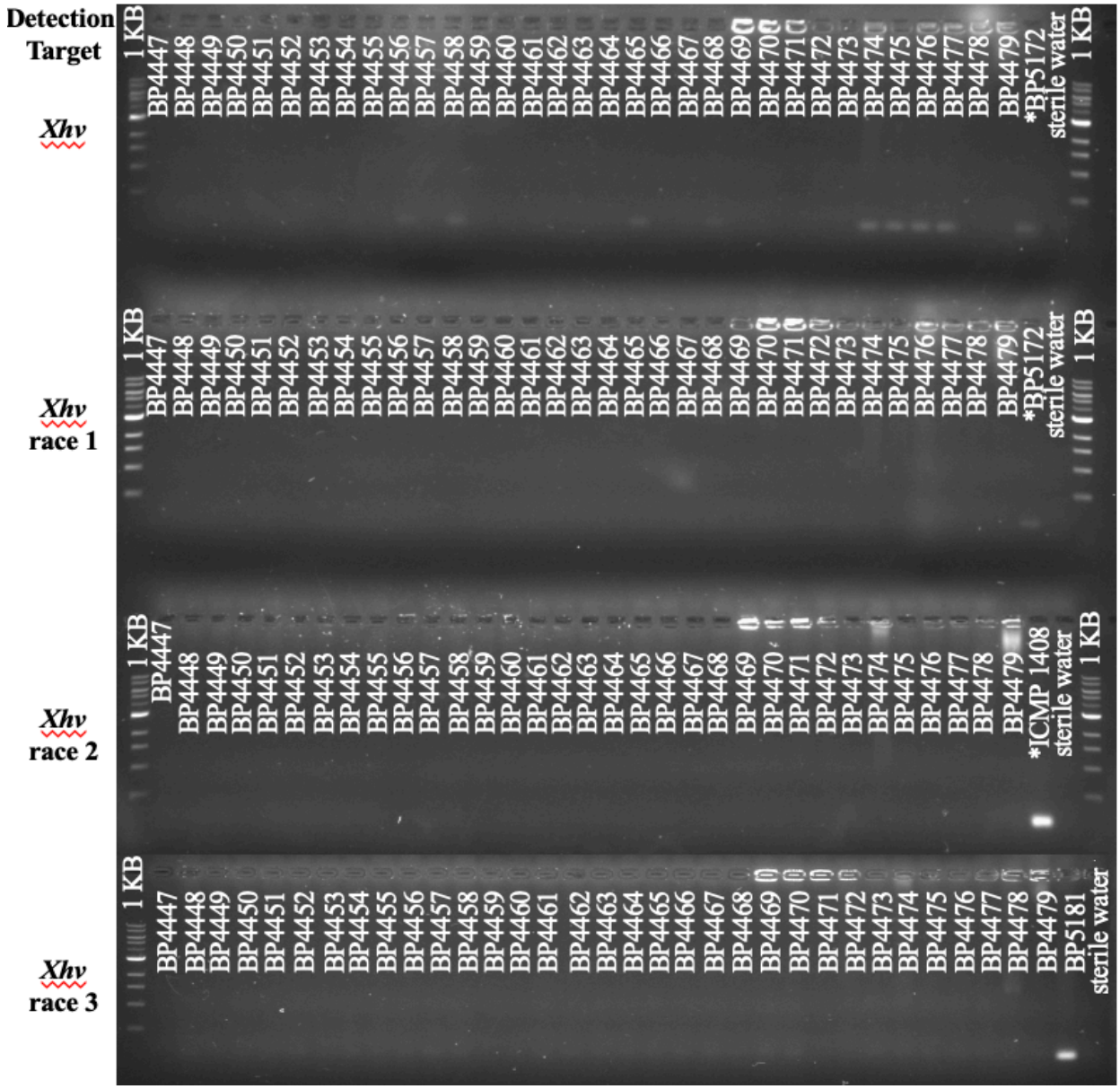


Figure 4-11. Evaluating novel detection methods with DNA extracts and bacterial isolates from a suspected BLS outbreak in Pennsylvania in 2022. Gel electrophoresis showing the amplification produced using DNA isolated from plant extracts (BP4447-BP4468) and colony suspensions (BP4469-BP4479), all from a suspected BLS outbreak in PA in 2022. The detection target is shown to the left of the row of reaction, and positive controls included BP5172 (*Xhv* race 1), ICMP 1408 (*Xhv* race 2), and BP5181 (hypothesized *Xhv* race 3). The asterisk refers to

the strains of confirmed races from HR screening on race-differential cultivars. Sterile water was included in the reactions as negative controls, and a 1 kb ladder from NEB was used as a size reference for each row. The 1% agarose gels were run for 2 hours at 84V.

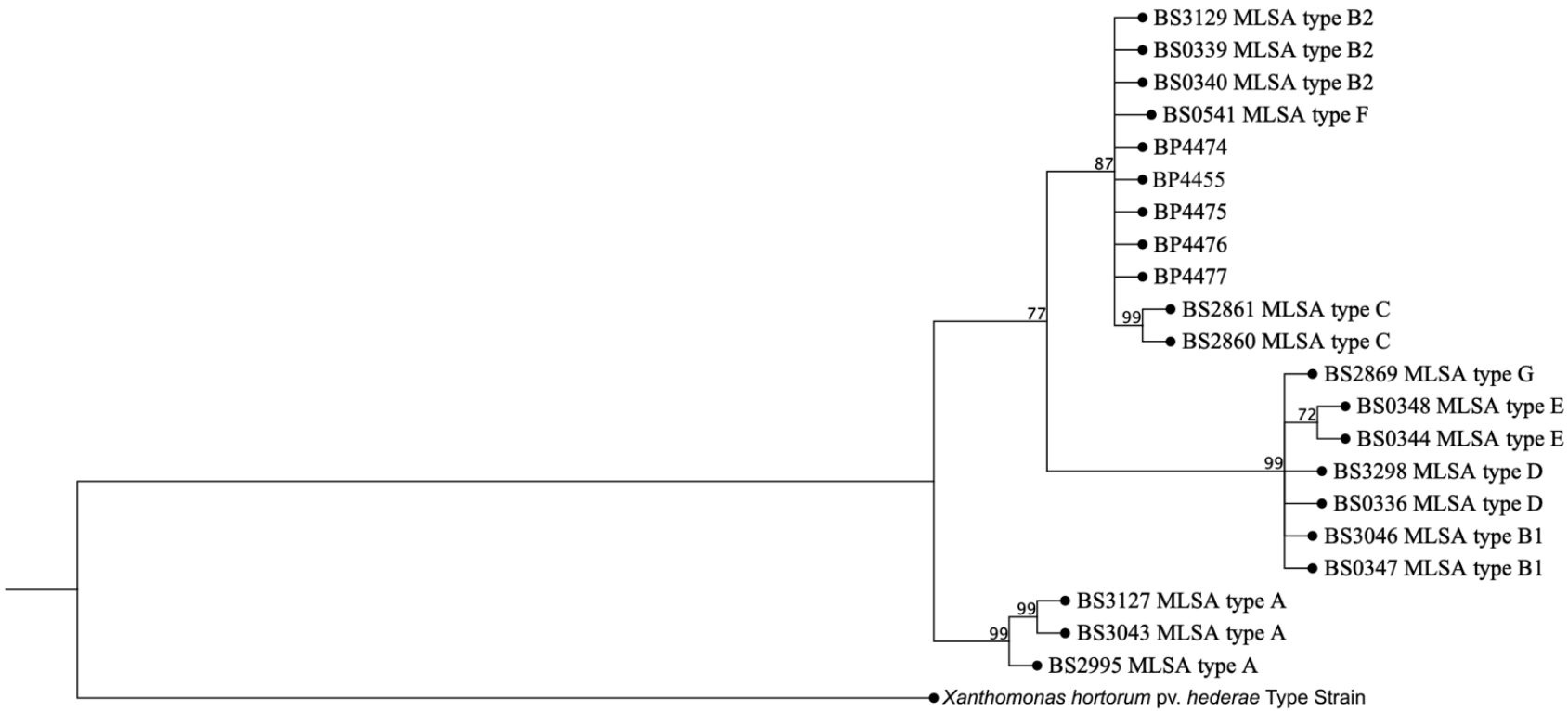


Figure 4-12. MLSA of suspected *Xhv* strains isolated in 2022 Pennsylvania outbreak. Maximum likelihood phylogeny generated from the alignment of concatenated *rpoD*, *fyuA*, *gyrB*, and *gap1* housekeeping genes using the neighbor-joining algorithm and the general time reversible nucleotide substitution model (including rate variation and estimated topology). The bootstrap threshold was set to 70% and the branch lengths represent the expected number of nucleotide substitutions per site. Strains of known *Xhv* MLSA

type were included as a reference set in order to predict the races of the PA strains that tested positive for *Xhv* detection using our touchdown PCR-based method. The type strain of *X. hortorum* pv. *hederae* was included as an outgroup.

Our 16S rRNA subunit analyses revealed there may have been some false positives using our method, which in practice could lead to unnecessary action taken to prevent a BLS outbreak should a non-*Xhv* strain be detected. However, there were no false negatives, which would have resulted in potential *Xhv* infection or seed contamination going undetected and thereby risking the development of a new outbreak. (Figure 4-13 and Table 4-4). As expected, the Xanthomonad isolates that tested positive using our detection method clustered with *X. hortorum* strains (BP4474-BP4477). Two more samples that were positive with our method (BP4456 and BP4465) clustered with *X. hortorum* strains and searching for matches to their 16S rRNA subunit sequences revealed them to be *X. hydrangeae*, which are known to be close relatives of *X. hortorum* and are pathogenic on smooth hydrangea (Dia et al., 2021). BP4458 and BP4468 also tested positive for *Xhv* with our detection method, but the former (which did not produce enough 16S rRNA gene sequence to align to the others and so was left out of the phylogeny) had a match to *Dactylococcopsis salina*, and the latter grouped with a large clade of strains all matching to *Microcoleus anatoxicus*. Both are cyanobacteria found in lakes and rivers, and while they were present in the plant DNA extract, they can be easily distinguished from *Xanthomonas* strains when grown on solid nutrient agar due to their distinct morphologies (Walsby et al., 1983; Conklin et al., 2023).

Regarding those that tested negative with our *Xhv* detection protocol, one entire clade generated in our 16S rRNA gene analysis was comprised of BP4457, BP4478, and BP4479, and all three matched to plant pathogenic *Pseudomonas* species with wide host ranges and that can cause necrotic spots on susceptible hosts (Dutta et al., 2018; Lipps and Samac, 2022). A report from Turkey confirmed that one of these *Pseudomonas* species, *P. viridiflava*, can cause bacterial leaf spot on curly lettuce (*Lactuca sativa* var. crispa), and soft rot on other lettuce

cultivars (Aksoy et al., 2018; González et al., 2003). Further characterization, such as LOPAT testing and MLSA, could help determine whether BP4478 and BP4479 are *P. viridiflava* strains. Pathogenicity assays with these strains and the *P. viridiflava* type and pathotype strains on these romaine, green leaf, and red lettuce varieties could reveal whether they and/or *P. viridiflava* can be pathogenic on these hosts and cause necrotic spots that resemble *Xhv* infection, and whether they might constitute a new pathovar of *P. viridiflava*.

Another clade drawn from 16S rRNA analysis, BP4455, BP4472, and BP4473, matched to *Stenotrophomonas nematodicola*, a bacterium isolated from a nematode species and member of the family Xanthomonadacae (Wei et al., 2021). Another clade was comprised of BP4469, BP4471, and BP4470, with top matches to various species of *Microbacterium* for the former two and *Curtobacterium* for the latter. Both are usually gram-positive, environmental microbes from the family Microbacteriaceae that commonly inhabit soil and water, however one of the top hits was *Curtobacterium alli*, an onion pathogen (Chorost et al., 2018, Khanal et al., 2023). They both can also form yellow and somewhat mucoid colonies in culture, which explains why they were selected as possible *Xanthomonas* isolates when taken from symptomatic lettuce tissue but were correctly excluded using our *Xhv* detection procedure.

Other strains that tested negative using our procedure and did not have long enough sequence to be included in our 16S rRNA-based phylogeny were BP4461, BP4466, and BP4477. These strains matched to *Janibacter hoylei*, *Sphingomonas hylomeconis,* and *Micrococcus aloeverae*, respectively. *Janibacter* species are environmental bacteria, and *J. hoylei* has only been isolated from the atmosphere at high altitudes and from clinical samples of immunocompromised patients experiencing bacteremia (Lim et al., 2017). *S. hylomeconis* are

endophytic bacteria from Japanese poppywort (Akbar et al., 2015). *M. aloverae* is also a plant endophyte, isolated from aloe vera (Prakash et al., 2014).

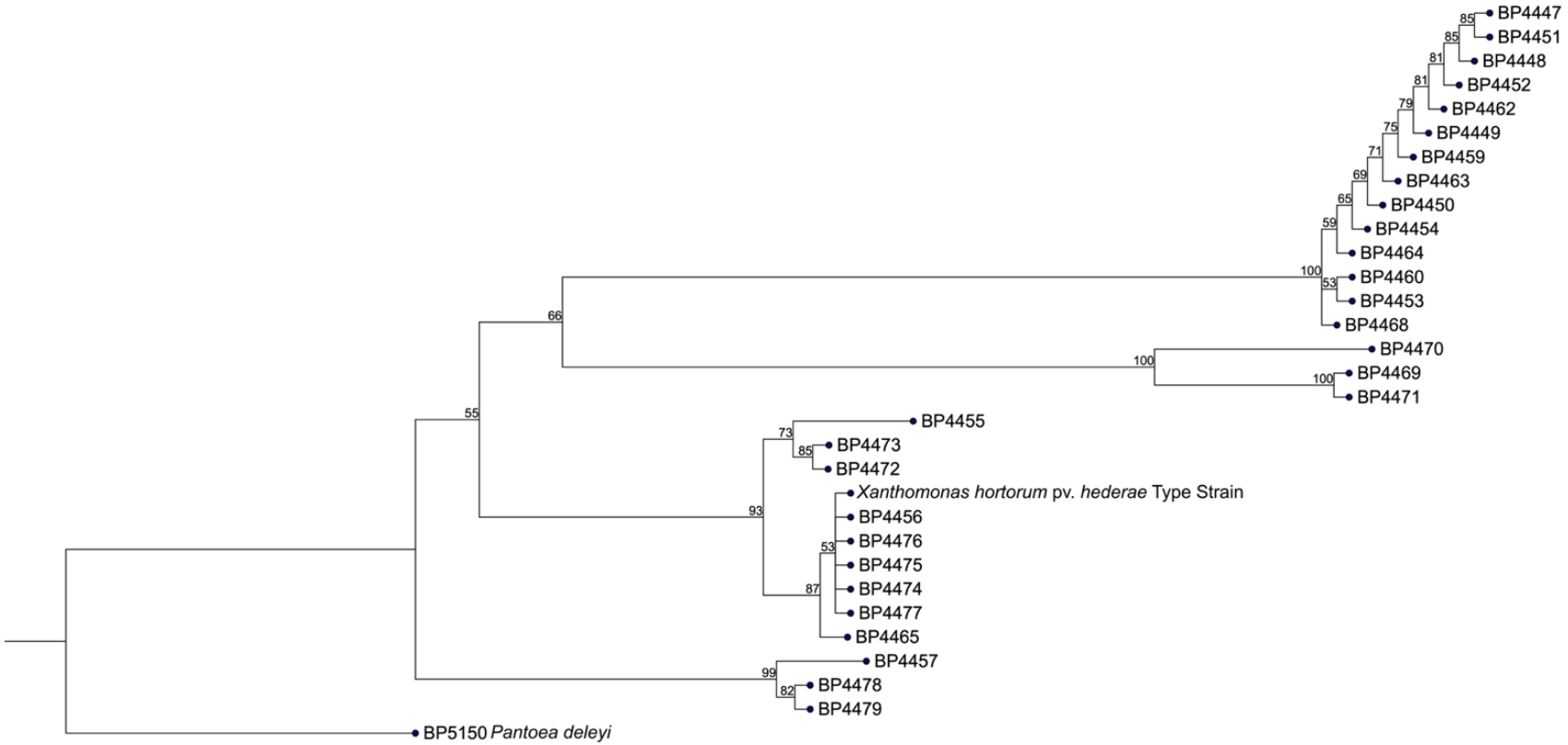


Figure 4-13. 16S rRNA subunit-based phylogeny of suspected *Xhv* strains isolated in 2022 Pennsylvania outbreak. Maximum likelihood phylogeny generated from the alignment of 16S rRNA subunit genes using the neighbor-joining algorithm and the Kimura 80 nucleotide substitution model (including rate variation and estimated topology). The bootstrap threshold was set to 50% and the branch lengths represent the expected number of nucleotide substitutions per site. The *X. hortorum* type strain was included as a reference for comparison with PA strains, along with a *Pantoea deleyi* strain that was previously isolated from symptomatic lettuce tissue (Section 2.3.2).

Table 4-4. Top matches to the 16S rRNA subunit gene sequences of strains isolated from PA 2022 lettuce disease outbreak. Top hits were determined to be those with the highest percent identity and only type strains were included in the searched database. The asterisks refer to searches that had multiple top hits.

Strain	Top BLASTn Hit	Percent Identity	Query coverage	E-value	Category
BP4447	*Microcoleus anatoxicus*	91.2%	99%	0.0	Cyanobacterium
BP4448	*Microcoleus anatoxicus*	91.2%	99%	0.0	Cyanobacterium
BP4449	*Microcoleus anatoxicus*	91.2%	99%	0.0	Cyanobacterium
BP4450	*Microcoleus anatoxicus*	91.2%	99%	0.0	Cyanobacterium
BP4451	*Microcoleus anatoxicus*	91.2%	99%	0.0	Cyanobacterium
BP4452	*Microcoleus anatoxicus*	91.2%	99%	0.0	Cyanobacterium
BP4453	*Microcoleus anatoxicus*	91.3%	100%	0.0	Cyanobacterium
BP4454	*Microcoleus anatoxicus*	91.2%	99%	0.0	Cyanobacterium
BP4455	*Stenotrophomonas nematodicola*	98.6%	100%	0.0	Nematode isolate
BP4456*	*Xanthomonas hydrangeae*	98.7%	100%	0.0	Plant pathogen
BP4457	*Pseudomonas coronafaciens; Pseudomonas viridiflava**	98.4%	100%	0.0	Plant pathogen
BP4458*	*Dactylococcopsis salina*	100%	94.1%	1e-37	Cyanobacterium
BP4459	*Microcoleus anatoxicus*	91.2%	99%	0.0	Cyanobacterium
BP4460	*Microcoleus anatoxicus*	91.3%	100%	0.0	Cyanobacterium
BP4461	*Janibacter hoylei*	96.9%	100%	1e-167	Environmental microbe
BP4462	*Microcoleus anatoxicus*	91.2%	99%	0.0	Cyanobacterium
BP4463	*Microcoleus anatoxicus*	91.2%	99%	0.0	Cyanobacterium

BP4464	*Microcoleus anatoxicus*	91.2%	99%	0.0	Cyanobacterium
BP4465*	*Xanthomonas hydrangeae*	99.4%	100%	0.0	Plant pathogen
BP4466	*Sphingomonas hylomeconis*	91.1%	53%	7e-43	Plant endophyte
BP4467	*Micrococcus aloeverae*	100%	100%	0.0	Plant endophyte
BP4468*	*Microcoleus anatoxicus*	91.2%	99%	0.0	Cyanobacterium
BP4469	*Microbacterium algeriense* and other species	100%	100%	0.0	Cyanobacterium
BP4470	*Curtobacterium alli* and other species	100%	100%	0.0	Cyanobacterium
BP4471	*Microbacterium algeriense* and other species	100%	100%	0.0	Cyanobacterium
BP4472	*Stenotrophomonas nematodicola*	99.7%	100%	0.0	Nematode isolate
BP4473	*Stenotrophomonas nematodicola*	99.6%	100%	0.0	Nematode isolate
BP4474	*Xanthomonas hortorum* and other species	100%	100%	0.0	Plant pathogen
BP4475	*Xanthomonas hortorum* and other species	100%	100%	0.0	Plant pathogen
BP4476	*Xanthomonas hortorum* and other species	99.8%	100%	0.0	Plant pathogen
BP4477	*Xanthomonas hortorum* and other species	100%	100%	0.0	Plant pathogen
BP4478	*Pseudomonas viridiflava*	99.9%	100%	0.0	Plant pathogen
BP4479	*Pseudomonas allivorans*	100%	100%	0.0	Plant pathogen

The asterisk refers to strains that tested positive with our *Xhv* detection method.

4.4 Conclusions

Our set of detection methods provides an improvement over previous methods for *Xhv*-specific detection. The B162 primers designed by Barak et al., 2001 were intended for *Xhv* detection only, we found that they also detected *X. hortorum* pv. *garderi*, *X. hortorum* pv. *cynarae, X. hortorum* pv. *taraxaci*, and an *X. hortorum* strain isolated from radicchio. We found that our detection method using the GC3889 primer set successfully detected 97% of our *Xhv* strains in our collection and did not detect any of the other six *X. hortorum* pathovars.

However, when we evaluated genomic DNA isolated from twenty-two plant extracts, all taken from symptomatic romaine, green, or red leaf lettuce tissue collected during a 2022 outbreak, we found that four of them were positive using our hypothesized *Xhv*-specific detection method, and two of those contained 16S rRNA sequences that matched to *X. hydrangea*e. This is a smooth hydrangea pathogen first described in 2021 that is a close relative of *X. hortorum* strains that also causes leaf necrotic spots (Dia et al., 2021). The other two DNA isolates matched to two species of cyanobacteria, which are easily distinguished in culture from *Xhv*. *X. hydrangeae* is not known to infect lettuce hosts, but represented a false positive hit in our analyses and may do so when our procedure is applied in the future. To rectify this, our whole genome sequence alignments could be repeated, with the addition of the *X. hydrangeae* type strain genome sequence. Then one could search for target sequences that includes *Xhv* but excludes all other strains, then develop primers to screen that would produce amplicons within these target sequences.

Alternatively, this PCR could be used in tandem with a LAMP-assay designed by Dia et al. in 2022. This method involves selective, isothermal amplification that targets groups of strains within the *X. hortorum*-*X. hydrangea* species complex (Dia et al., 2021). These include group A1 (*X. hortorum* pv. *gardneri*, *X. hortorum* pv. *cynarae*, and *Xhv*), group A2 (*X. hortorum* pv.

taraxaci), group B1 (*X. hortorum* pv. *hederae*), group B2 (*X. hortorum* pv. *carotae*), group B3 (*X. hortorum* pv. *pelargonii*), and group C (*X. hydrangeae*). Positive *Xhv*-*X. hydrangeae* detection using our touchdown PCR method could be combined with the group A1-targeted targeted LAMP assay to confirm whether an unknown isolate is *Xhv*, instead using the group C-targeted LAMP assay would confirm it to be *X. hydrangeae*. Using these two methods in tandem would provide rapid, molecular-based detection.

More study is necessary to develop methods capable of distinguishing the *Xhv* races. Our method for *Xhv* race 1 detection using primer set GC4021-112 detected fourteen out of sixteen (88%) known *Xhv* race 1 strains tested and thirty-nine out of sixty-seven (58%) of predicted *Xhv* race 1 strains. Completing the HR screening for these predicted strains is necessary to fully evaluate the effectiveness of our detection method designed for *Xhv* race 1 detection. The *Xhv* race 2 detection method we developed with primer set GC4381-178 detected eight out of ten (80%) known *Xhv* race 2 strains, but also two known race 1 strains and one hypothesized race 1 strain. It did not detect any of the other *X. hortorum* pathotype strains. Our *Xhv* race 3 detection method with primer set GC4980 detected the one known *Xhv* race 3 strain and the three hypothesized *Xhv* race 3 strains, but also it detected *X. hortorum* pv. *carotae* and a one *Xhv* race 1 strain. Long-read sequencing and including more *Xhv* strains could improve our whole genome alignments and allow for more accurate targeting of race-specific sequences. For now, race is best predicted by MLSA and confirmed by HR screening.

These detection methods are intended to serve researchers, clinicians, and seed companies. Once fully developed, they could be used by researchers to track the distribution of *Xhv* and its pathogenic races globally. Clinicians would be able to offer quick answers to farmers who suspect their crops are experiencing a BLS outbreak, as well as to provide recommendations of cultivars

to use in subsequent planting that would be resistant to the *Xhv* races detected in a current outbreak. These methods could be useful in industry for quickly testing seed lots for the presence of the pathogen; however, the efficacy of these methods for detecting *Xhv* from seed samples remains to be evaluated. Currently, we offer to those with lettuce plants presenting BLS symptoms a *Xhv*-*X. hydrangeae* detection method that works with DNA extract and colony suspension, and that can be paired with a published LAMP assay to distinguish *Xhv* from *X.hydrangeae* (Dia et al., 2022a).

Chapter 5

Investigating the role of *X. hortorum* pv. *vitians*'s race-specific effectors in lettuce HR induction.

5.1 Introduction

Lettuce bacterial leaf spot is a devastating disease caused by the bacterial plant pathogen *Xanthomonas hortorum* pv. *vitians* (*Xhv*). The disease affects lettuce production worldwide, including in the United States where lettuce is the highest valued vegetable crop at 4.1 billion dollars in production in 2022 (USDA-NASS, 2022). *Xhv* causes water-soaking of the lettuce leaves, along with chlorosis and small necrotic spots that later coalesce to form large lesions. Although BLS appears sporadically, it poses a significant threat because severe outbreaks can cause 100% crop loss (Lu and Raid, 2013). Even minor outbreaks are costly; lettuce head infections require the removal of the outer leaves prior to sale, which causes a reduction of the sellable yield. Integrated disease management strategies include the elimination of possible inoculum sources, like weeds, debris, and fast-maturing crops in rotation that could harbor the pathogen between lettuce plantings (Barak et al. 2001). Copper bactericides can be used to limit *Xhv* spread, but they vary in their effectiveness; additionally, because they are most effective when applied prophylactically, they result in a large sunken cost in years that the pathogen does not appear (Bull et al., 2005).

Another solution is the use of resistant cultivars, such as *Lactuca sativa* cvs. La Brillante, Little Gem, Pavane, and Reine des Glaces (RG) (Bull et al., 2007). The first cultivar has demonstrated resistance to *Xhv* infection based on a single, dominant gene called *Xanthomonas resistance* 1 (*Xar1*) located within linkage group 2 of the lettuce genome (Hayes et al., 2014). Little Gem and Pavane are hypothesized to also encode *Xar1* due to the similarity of their

resistance response to the same subset of *Xhv* strains. In this interaction, *Xhv* infection induces a lettuce hypersensitive response (HR), or programmed cell death, to limit the spread of bacteria to healthy tissues. Another lettuce introduction, PI358001-1, demonstrated a similar interaction with *Xhv* strains isolated from Florida outbreaks, and resistance was also mapped to a single, dominant gene on linkage group 2, named *Xcvr* (Wang et al., 2016). The cultivar RG encodes only partial resistance in this same locus, named qXCR2.1 for this cultivar (Sandoya et al., 2019). *Xar1* and *Xcvr* make great candidates for breeding programs looking to engineer plants with stable resistance and they may be supplemented by the partial resistance offered by qXCR2.1. However, the mechanism by which *Xhv* interacts with these resistance genes is yet unknown.

Molecular plant-pathogen interactions are often modeled as an arms race; bacterial genes that aid in colonization and pathogenicity provide selective pressure for plants to develop pathogen recognition, which in turn selects for bacteria that can evade recognition. Primary recognition of pathogen associated molecular patterns (PAMPs) occurs at the plant cell surface and leads to a PAMP-triggered immune (PTI) response, or the activation of defense-related genes to fight infection (Elstner, 1991). Many bacterial pathogens encode effector genes that help them overcome host basal defenses in various ways, such as the disruption of signaling pathways or the degradation of host defense proteins and hormones (Macho et al, 2014; Zhou et al, 2014; Xiang et al, 2008; Rosebrock et al, 2007). Most effectors are delivered through the type III secretion system, a transmembrane, needle like structure that injects them directly from the bacterial cell to the plant cell cytoplasm (Hueck, 1998). *Xanthomonas* species usually encode between fifteen and thirty different effectors, and along with *Ralstonia* species encode a special class of effector genes known as transcription activator-like effectors (TALEs) that localize to

the plant cell nucleus and trigger the expression of susceptibility genes (Boch et al., 2009). When plants evolve to recognize bacterial effectors via resistance genes, the effectors are described as avirulence genes (Zipfel and Felix, 2005; Dangl and Jones, 2001). Avirulence gene recognition by R-genes can lead to the hypersensitive response that kills infected tissue cells and limits bacterial spread.

Research into *Xhv* diversity increased following multiple outbreaks in the late 20th century. In 2003, Sahin identified two groups of *Xhv* strains, one that caused both local and systemic infections and the other caused only local infections of the leaves. In another study, researchers compared the sequences of core housekeeping genes, and determined that the *Xhv* strains cluster into three groups based on genetic relatedness (Fayette et al., 2016). A similar analysis was conducted with a larger collection of *Xhv* strains, and in tandem with HR screening on many lettuce cultivars and accessions (Bull et al., 2016). They identified seven sequetypes, A-G, and that the sequetype corresponded to the compatible or incompatible reaction on three lettuce lines. Sequetype B and D-G strains elicited HR in *L. sativa* cv. Little Gem and *L. serriola* ARM-09-161-10-1; sequetype A strains elicited HR in *L. serriola* PI491114; and sequetype C strains elicited HR in *L. serriola* ARM-09-161-10-1 only. These reactions define the *Xhv* races 1, 2, and 3, respectively. Further research showed that *Xhv* race 1 strains also elicited HR in *L. sativa* PI358001-1 (Wang et al., 2016).

While the interaction between an *Xhv* avirulence gene and a lettuce R-gene has not yet been confirmed, such a gene-for-gene interaction has been shown for the bacterial leaf spot disease complex of tomato and pepper (Potnis et al., 2015). These strains, (from *X. euvesicatoria*, *X. gardneri*, *X. perforans*, and *X. vesicatoria*), include four tomato races and at least six pepper

races that are defined by their ability to trigger HR in specific pepper and tomato cultivars. The avirulence genes *avrRxv*, *avrXv3*, *xopJ4*, and *avrBs4* interact with R-genes *Rx*[*1*, *2*, and *3*], *Xv3*, *RXopJ4*, and *Bs4*, respectively, to trigger HR in tomato; avirulence genes *avrBs1*, *avrBs2*, *avrBs3*, *avrBs4C*, *avrBs7*, and *avrBsT* interact with R-genes *Bs1*, *Bs2*, *Bs3*, *Bs4C*, *Bs7*, and *BsT* respectively to trigger HR in pepper. The *Xhv* races may similarly encode avirulence genes that are responsible for eliciting HR in the differential lettuce cultivars and accessions.

While *Xhv* populations are predominantly clonal, strain diversity can be attributed to recombination (Fayette et al., 2016). In Rosenthal et al., 2022, we hypothesized that different recombination events may have led to each of the *Xhv* races having a distinct effector repertoire. Through comparative genomic analysis, we showed that the effectors *xopAQ* and *xopAF2* were encoded by *Xhv* races 1 and 3 strains, which elicit HR in *L. serriola* ARM-09-161-10-1, and are not encoded by *Xhv* race 2 strains, which cause disease in *L. serriola* ARM-09-161-10-1. *xopAQ* homologs in *X. arboricola*, *X. citri*, and *X. euvesicatoria* are thought to be involved in pathogen recognition and HR (Barak et al., 2016; and Thieme et al., 2007). *xopAF2* is a homolog of *xopAF*/*avrXv3*, first described as the avirulence gene required for *X. euvesicatoria* pv. *perforans* tomato race 3 strains' race-specific HR induction in tomato cultivar Hawaii 7981 (Astua-Monge et al., 2000). Homologs of both *xopAQ* and *xopAF2* were also found in other *Xanthomonas* species to impact virulence, but with no effect on pathogenicity.

With this study we sought to investigate the hypothesis that *xopAQ* and *xopAF2* in *Xhv* race 1 and 3 strains function as avirulence genes in *L. serriola* ARM-09-161-10-1, interacting with a yet unknown R-gene. Deletion of these avirulence genes from the *Xhv* race 1 and 3 strains was expected to make them pathogenic, while expression of these avirulence genes in an *Xhv* race 2 background and back in our deletion constructs was expected to confer or re-establish the

ability to induce HR. We designed the deletion and expression constructs for both effector genes but have thus far completed transformation and transconjugation experiments with just the *xopAQ* constructs to establish a working set of protocols. We have several candidate *xopAQ* deletion mutants, but the movement of the expression vector into a *Xhv* race 2 strain has not yet succeeded. Further work is necessary to generate this *xopAQ* expression mutant and to generate *xopAF2* deletion and expression mutants. Then, HR screens with these mutants could demonstrate whether *xopAQ* or *xopAF2* play a role in HR induction in *L. serriola* ARM-09-161-10-1; when deleted from a *Xhv* race 1 strain, we expect to see a loss of HR induction and where plasmid-expressed in a *Xhv* race 2 strain, we expect to see gain of HR induction.

5.2 Materials and Methods

5.2.1 Bacterial strains

The bacterial strains used in this study are listed in Table 5-1. *Xanthomonas* strains were routinely cultured on nutrient agar or broth media at 28 °C, and *E. coli* strains on Luria Bertani agar or broth media at 37 °C. Yeast dextrose carbonate agar (YDC) and 1/10 tryptic soy media were used during the generation of rifampicin resistant *Xanthomonas* strains. Antibiotic resistant variants were cultured on the above media amended with rifampicin (100 μg/ml), kanamycin (50 μg/ml), spectinomycin (60 μg/ml), gentamycin (15 μg/ml), or ampicillin (50 μg/ml). Glycerol stocks were prepared with a mixture of 50% glycerol 50% nutrient broth, and these were stored at -80 °C.

Table 5-1. Bacterial strains and plasmid used in this study.

	Strain Name(s)	Characteristics	Source or citation
Xanthomonas hortorum* pv. *vitians	BP5186; ICMP 4165, BS3048	*Xhv* MLSA group A, race 2.	H. J. Boesewinkel.
	BP5169	Rif^r variant of BP5186	This study.
	BS0347	*Xhv* MLSA group B, *Xhv* race 1	J. Barak
	BS2885	*Xhv* MLSA group B, *Xhv* race 1, Rif^r variant of BS0340.	Bull strain collection
	BS2861	*Xhv* MLSA group C, *Xhv* race 3.	S. Koike and Rianda
Escherichia coli	DH5α	Host for pDSK519	N. Potnis collection
		Host for pRK2073	N. Potnis collection
Plasmids	pUC19	Amp^r	Norander et al., 1983
	pDSK519	Kan^r	Keen et al., 1988
	pDSK519::*xopAQ*	Kan^r, *xopAQ* effector gene following *lacZ* promoter	This study
	pDSK519::*xopAF2*	Kan^r, *xopAF2* effector gene following *lacZ* promoter	This study
	pRK2073	Sp^r Tra^+, helper plasmid	Better and Helinski, 1983
	pk18mobsacB	Kan^r	Schäfer et al., 1994
	pk18mobsacB::*xopAQ* fusion fragment	Kan^r, merged flanking regions of *xopAQ* gene in MCS, donor strain	This study
	pk18mobsacB::*xopAQ* fusion fragment with gm^R cassette	Kan^r, merged flanking regions of *xopAQ* gene, with Gm^r cassette between, in MCS, donor strain	This study
	pk18mobsacB::*xopAF2* fusion fragment	Kan^r, merged flanking regions of *xopAF2* gene in MCS, donor strain	This study
	pk18mobsacB::*xopAF2* fusion fragment with gm^R cassette	Kan^r, merged flanking regions of *xopAF2* gene, with Gm^r cassette between, in MCS, donor strain	This study

Rif^r = rifampicin resistant, Gm^r = gentamycin resistant, Amp^r = ampicillin resistant, Sp^r = spectinomycin resistant, MCS = multicloning site.

5.2.2 Generation of a rifampicin resistant *Xhv* race 2 variant

Overnight cultures of *Xhv* race 2 BP5186 in 50 mL nutrient broth were centrifuged at 4,000 rpm for 10 minutes at 4 °C, and the cell pellets were resuspended in 1 mL of sterile water. Aliquots of 100 µL of cell suspension were spread onto NA amended with (30%) rifampicin. Resistant variants then were passed onto NA amended with 60%, followed by passage onto 100% rifampicin. To be sure that the metabolic activity of the variants was not significantly affected by the *rpoB* mutation, the growth rates of the variants were compared to that of the wild type BP5186 strain; variants were streaked for single colonies alongside BP5186 onto NA and YDC agar and incubated for 3 days at 28 °C. Those that did not lag behind BP5186 in growth were used to create overnight cultures in nutrient broth; the next day 5 mL of culture was used to inoculate 125 mL of sterile nutrient broth and growth rate was measured as optical density (OD_{600}) during the logarithmic growth phase over time. Variant growth rates that did not differ significantly from the wild type using a student's T test were further assessed using the same method, except this time with 1/10 tryptic soy agar and 1/10 tryptic broth. Variants that still showed no significant difference in growth rate compared to BP5186 were preserved in glycerol stocks. Pathogenicity testing was completed with these variants and among these, the variant designated BP5169 retained pathogenicity in *L. serriola* ARM-09-161-10-1 and was thus selected for downstream application.

5.2.3 Development of *xopAQ* and *xopAF2* deletion constructs.

Two methods for designing the deletion constructs were used to determine which could more efficiently generate deletion mutants (Figure 5-1). For the scarless deletion method, upstream and downstream regions of the effector genes were identified and merged to create a

fusion fragment to be inserted into the nonreplicative plasmid pk18mobsacB within the multicloning site (MCS). Upon movement of this construct into the *Xhv* race 1 strain, it was expected that recombination between the homologous sites of the plasmid and the *Xhv* race 1 genomic DNA (gDNA) would generate double recombinants at an observable frequency on selective media. The desired end products would be the gDNA lacking the effector gene and the plasmid containing the effector gene in the multicloning site. Counterselection on media amended with 10% sucrose would leave only colonies that no longer contain the plasmid; those colonies would then need to be evaluated via PCR and sequencing to determine whether the procedure had resulted in the desired effector gene deletion, rather than simply the uptake and expulsion of the plasmid without recombination taking place. The marker exchange method was identical except for the inclusion of a gentamycin resistance cassette between the upstream and downstream regions in the fusion fragment. Rather than creating a scarless deletion, this would leave the final gDNA product with this cassette at the site that previously encoded the effector gene. This design would allow for the screening of the transformants on media amended with both 10% sucrose and gentamycin, so that all resulting colonies should be double recombinants with the plasmid vector cleared. While the scarless deletion method was expected to have a higher transconjugation efficiency, the marker exchange method was designed to be easier to select for double recombinants.

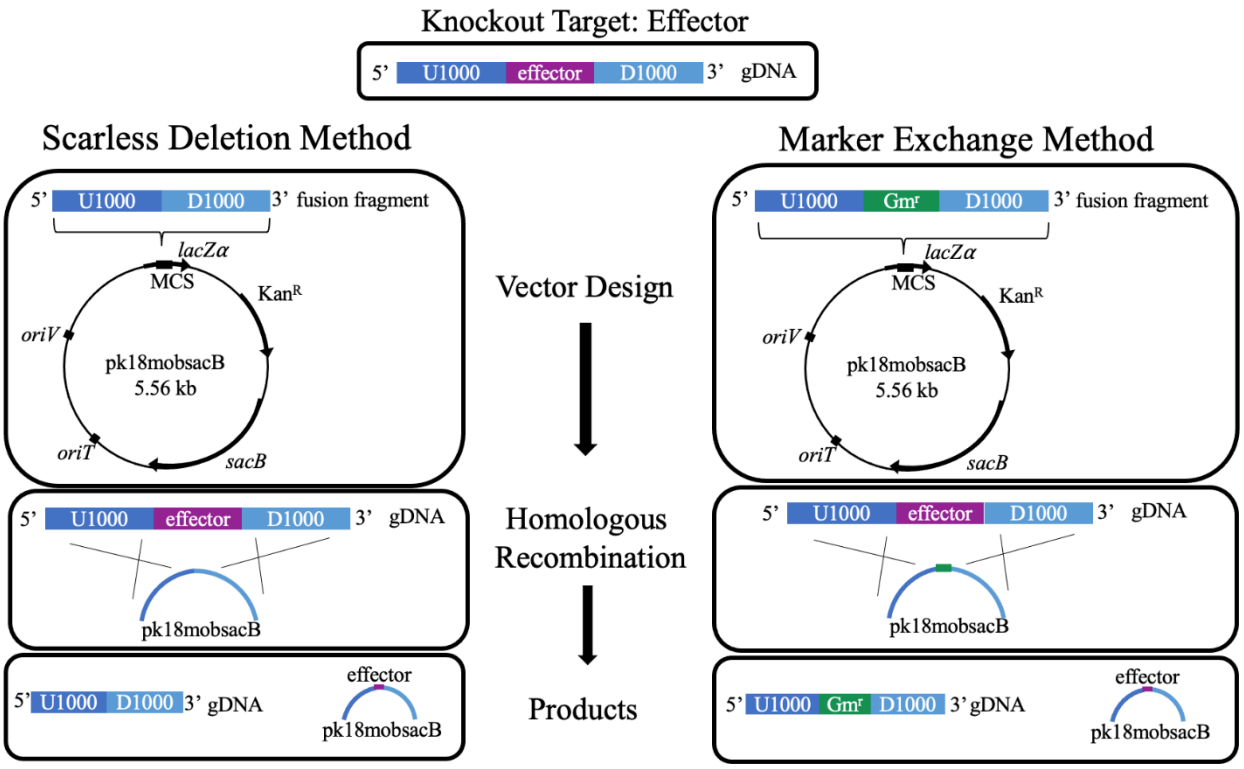


Figure 5-1 Strategies employed to generate effector gene knockouts. The 1 kb upstream and downstream regions of the effector genes are labelled as U1000 and D1000. Vector design: the merging of these regions creates a fusion fragment using the scarless deletion method, and the insertion of the gentamycin resistance cassette (Gm^r) between them creates the fusion fragment for the marker exchange method. Homologous recombination: once transformed into the *Xhv* race 1 or 3 host strain, the desired result is recombination at both sites of homology, the upstream and downstream regions. Products: the genomic DNA (gDNA) loses the effector gene to the plasmid, which will be cleared from the cell, but is otherwise unchanged (scarless deletion method); or the gDNA receives the Gm^r cassette and the plasmid receives the effector gene (marker exchange method).

To create both the *xopAQ* and *xopAF2* deletion constructs (shown in Figure 5-2), the *Xhv* race 1 strain BS0347 whole genome sequence nodes containing these genes (Rosenthal et al., 2022) were opened in SnapGene Viewer (from Dotmatics; available at snapgene.com) and approximately 1 kb regions flanking each gene were identified. For the scarless deletion method, these flanking regions were each merged in a standard text editor to create a fusion fragment for each gene. The fusion fragment for the marker exchange method was designed with a gentamycin resistance cassette inserted between the up- and downstream regions. The pK18mobsacB FASTA file was downloaded from NCBI, and the fusion fragment sequences were then copied and pasted into the multicloning site between 5' BamHI and 3' HindIII.

The deletion vector pK18mobsacB was isolated from their host *E. coli* strain using the PureYield Plasmid Midiprep System according to manufacturer's instructions (Promega, Madison, WI). Plasmid size was confirmed via gel electrophoresis. The isolated plasmid and the sequence files for the desired constructs were sent to Synbio Technologies (Monmouth Junction, NJ) for gene synthesis and custom cloning. Once the constructs[4] were received from Synbio Tech., the contents of the multicloning sites were confirmed via PCR amplification and gel electrophoresis. The 25 µL PCRs consisted of 1X ImmoMix (Meridian Bioscience, Cincinnati, OH), 0.5 uM of each forward and reverse primer, and 1 µL of colony suspensions. The cycling steps included ten minutes at 95 °C, thirty seconds at 95 °C, one minute at 58 °C, one minute at 72 °C, cycling thirty-four times to step two, and six minutes at 72 °C. These reaction conditions were the same for testing the gene inserts to the expression vectors.

[4] Fabrication of the deletion constructs was attempted in our lab; however, these attempts were not successful. These results are described in Appendix A5-1.

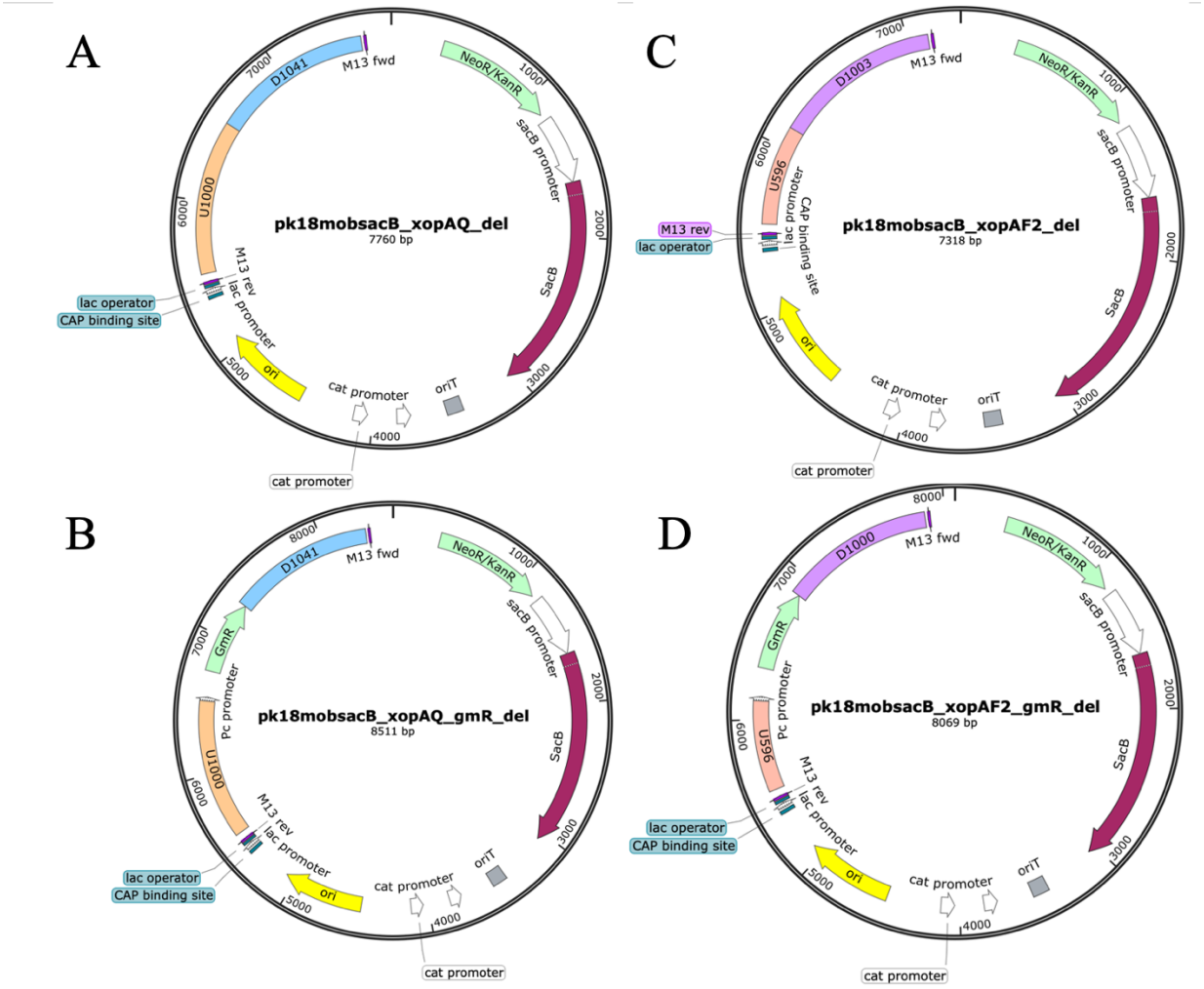


Figure 5-2. *xopAQ* and *xopAF2* deletion construct designs. (A) pk18mobsacB::*xopAQ* fusion fragment; (B) pk18mobsacB::*xopAQ* fusion fragment with Gm^r cassette; (C) pk18mobsacB::*xopAF2* fusion fragment; (D) pk18mobsacB::*xopAF2* fusion fragment with Gm^r cassette. Insertions into the multicloning site disrupted the *lacZ*α gene.

5.2.4 Development of *xopAQ* and *xopAF2* expression constructs.

For designing the expression vectors (shown in Figure 5-3), the pDSK519 FASTA file was downloaded from NCBI and the *xopAQ* and *xopAF2* gene sequences from the BS0347 genome were inserted into the multicloning sites 5' BamHI and 3' EcoRI using a standard text editor. The expression vector pDSK519 was isolated from its host *E. coli* strain using the PureYield Plasmid Midiprep System according to manufacturer's instructions (Promega, Madison, WI). Plasmid size was confirmed via gel electrophoresis. The isolated plasmid and the sequence files for the desired constructs were sent to Synbio Technologies (Monmouth Junction, NJ) for gene synthesis and custom cloning. Once the constructs were received from Synbio Tech., the contents of the multicloning sites were confirmed via PCR amplification and gel electrophoresis.

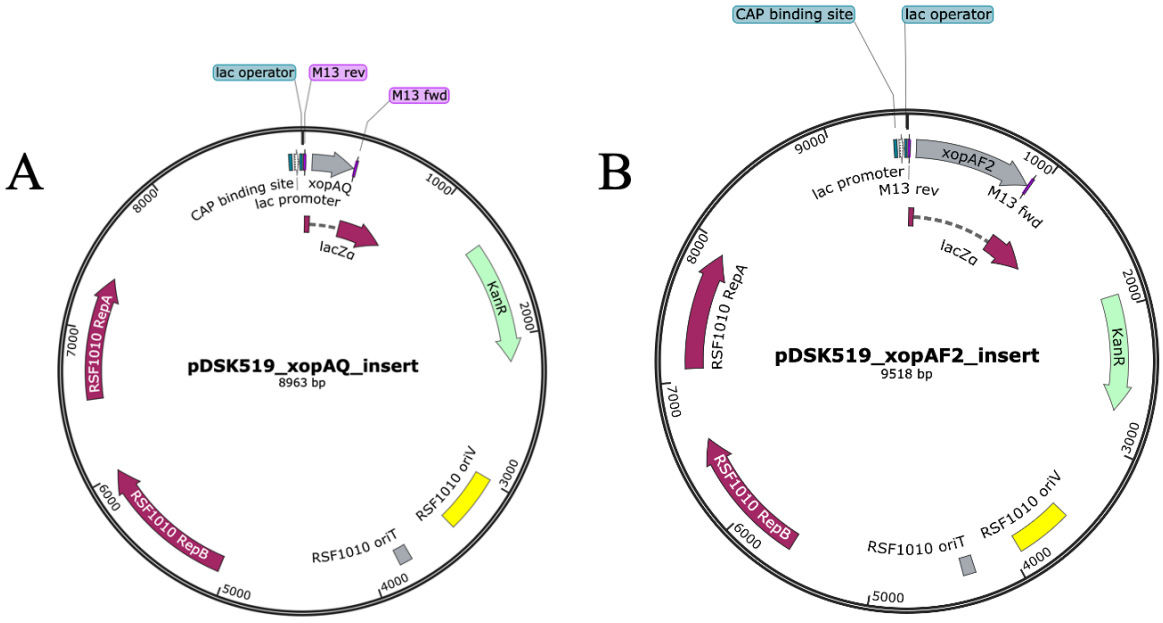


Figure 5-3. *xopAQ* and *xopAF2* expression construct designs. (A) pDSK519::*xopAQ* (B) pDSK519::*xopAF2.* Insertions into the multicloning site disrupt the *lacZα* gene.

5.2.5 Transformation of deletion and expression constructs into *E. coli* DH5a host.

E. coli DH5a chemically competent cells from Intact Genomics (St. Louis, MO) were used as hosts for the *xopAQ* gene deletion and expression constructs received by Synbio Technologies. Competent cells were removed from -80 C storage and thawed on ice for 10 minutes. Each raw vector, suspended in sterile water, was added to each tube of competent cells in 3 µl aliquots and mixed by pipetting. Another plasmid vector encoding ampicillin resistance, pUC19, was used as a positive control and so also aliquoted into a tube of *E. coli* DH5a competent cells. After five minutes of incubation on ice, 950 mL aliquots of IG recovery medium (Intact Genomics, St. Louis, MO) were added to the tubes, and they were set to incubate at 37 °C and shaking at 200 rpm for one hour. 100 µL aliquots of the resulting cultures were spread plate onto LB agar plates amended with the appropriate antibiotic for each plasmid construct. Plates were incubated at 37 °C for two days, and colonies that appeared were selected for storage in glycerol at -80 °C. These transformants served as hosts for maintaining these plasmid constructs.

5.2.6 Creation of *Xhv* race 1 strains with *xopAQ* deletions.

Our goal was to generate *xopAQ* deletions in the *Xhv* race 1 strain BS2885, a rifampicin resistant mutant of the *Xhv* race 1 strain BS0340. The two *E. coli* DH5a donor strains containing the *xopAQ* deletion constructs (scarless and marker exchange), the *E. coli* strain containing the helper plasmid pRK2073, and *Xhv* race 1 strain BS2885 were grown on solid agar media amended with the appropriate selective antibiotic. Single colonies of the *E. coli* strains were used to inoculate 5 mL of sterile LB with appropriate antibiotics, and overnight cultures were grown,

shaking at 200 rpm and 37 °C. One mL of overnight culture was sub-cultured in another 5 mL of sterile LB broth with appropriate antibiotics and incubated until cloudy, for about six hours shaking at 200 rpm and 37 °C. Subcultures were centrifugated at top speed for 5 minutes and resuspended in 1 mL of new sterile LB; this step was repeated once and the suspension was adjusted to approximately $1x10^8$ CFU/ml (OD_{600} = 0.6+\- 0.05). The rifampicin resistant *Xhv* race 1 strain BS2885 was scraped from solid NA into 5 mL of sterile LB and adjusted to the same concentration as the *E. coli* strains. The donor, helper, and recipient strains were mixed 1:1:1, vortexed, and plated in 25 µL drops on to NA. After one to two days of incubation at 28 °C, these mating reactions were suspended in 1 mL NB broth each, and plated onto NA plates amended with rifampicin, to select for transformants with the scarless deletion construct, or rifampicin and gentamycin, to select for transformants the marker exchange construct. The resulting rifampicin resistant transformants were plated on NA amended with 10% sucrose and rifampicin, to select for the transformants that have cleared the scarless deletion construct, or NA amended with 10% sucrose, rifampicin, and gentamycin, to select for transformants that had cleared the plasmid and incorporated the gentamycin resistance cassette at the *xopAQ* gene site in the genome. The resulting colonies were suspended in 100 µl of sterile NB broth, spread plated on NA, and grown for one to two days at 28 °C. Glycerol stocks were made from the spread plates and stored at -80 °C. Primers were designed to amplify a gene fragment internal to *xopAQ* at 257 bp and an external fragment containing *xopAQ* at 559 bp; these were used to evaluate the presence or absence of the effector gene in our raw vectors as well as the *E. coli* and *Xhv* strains transformed with these vectors.

5.2.7 Creation of an *Xhv* race 2 strain expressing *xopAQ*.

Xhv race 2 strain BP5169 (Rif^R)was made electrocompetent (EC) using a previously published method proven for *Xanthomonas campestris* pv. *campestris* strains (Wang et al., 2016). The pDSK519 *xopAQ* expression vectors were rehydrated in sterile water and 3 μL of each was added to a separate 50 μL aliquot of EC BP5169 cells. pUC19 DNA was included as a positive control, and a tube of EC BP5169 with no addition was included as a negative control. These were all incubated on ice for 30 minutes, transferred to a cuvette, and pulsed at 15 kV/cm using a BioRad MicroPulser Electroporator (Hercules, CA). 1 mL of LB broth was added immediately to each cuvette, and the suspension was transferred to a 1.5 mL tube for a 3-hour incubation at 28 °C. To select for transformants, the suspensions were then spread onto LB agar plates amended with kanamycin and rifampicin in 100 μL aliquots. Resulting colonies were screened for *xopAQ* gene sequence via PCR amplification and gel electrophoresis.

5.3 Results and Discussion

5.3.1 Confirmation of *xopAQ* deletion constructs and their transformations into *E. coli* DH5α.

The primer sets designed to amplify fragments from within the *xopAQ* gene region and from outside the gene region are shown in Table 5-1. These were first used to confirm the contents of the vectors we received from Synbio Technologies and transformation of those vectors into our *E. coli* strains that would later serve as donors in triparental mating. Transformations with the *xopAQ* deletion constructs resulted in hundreds of colonies for each different vector transformation, and Figure 5-4 shows a subset of those that were evaluated via PCR and gel electrophoresis for the presence of desired the vector sequence.

Table 5-2. Primers sets designed for evaluating *xopAQ* gene presence and absence. Reaction conditions are listed in Section 5.2.3.

Primer Set	Forward 5' to 3'	Reverse 5' to 3'	Amplicon Size (bp)	
			***xopAQ* present**	***xopAQ* absent**
xopAQ-559 (outer)	TGCCGACCCTTAATTTC AACTGTG	GTCAACAAGCGTGA GGCTTTCTATG	559	257
xopAQ-266 (inner)	ATGGGATGCTTCAACG TCAC	GCGATCAGTTCCTTC TGCTC	266	N/A

In Figure 5-4, the band sizes were all the result of amplification with the outer *xopAQ*-559 primer set. Where the *xopAQ* gene was intact, the product was predicted to be 559 bp; where only the fusion fragment was present, the product was predicted to be 257 bp; and where fusion fragment with the inserted gentamycin resistance cassette was present, the product was predicted to be 1008 bp. The wild type *Xhv* race 1 strain BP5172 had the expected 559 bp band indicating the native presence of *xopAQ*. The scarless and marker exchange deletion constructs in sterile water suspension had the expected 257 bp and 1008 bp bands, respectively. *E. coli* DH5α donor strain transformed with either the scarless deletion construct or the maker exchange construct showed the expected 257 bp or 1008 bp bands, respectively, indicating the presence of the vectors. The negative control with sterile water as the template showed two bands, at about 257 and 1008 bp, indicating contamination in the sample. A repetition of this experiment showed no contamination in the negative control and the other results were the same (data not shown). These results indicated that the *xopAQ* scarless and marker-assisted deletion vectors likely contain the desired constructs within their multicloning sites. They also suggested that the transformations of these vectors into the *E. coli* donor strain were successful.

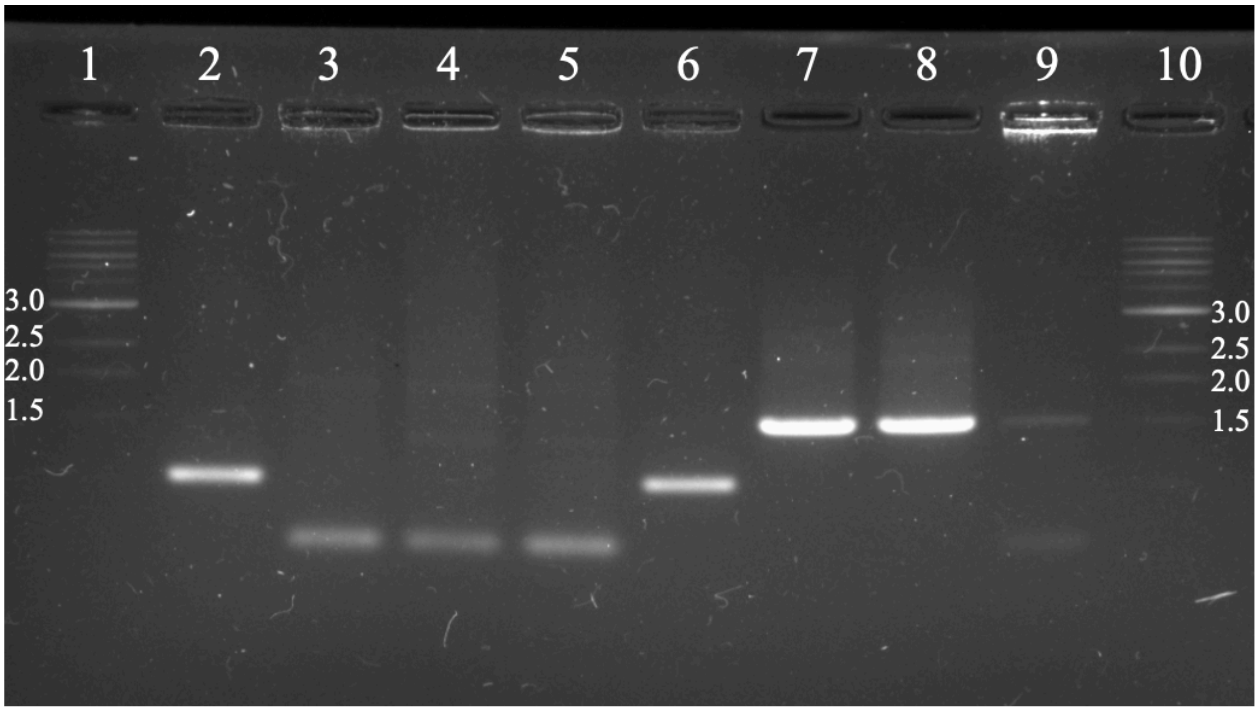


Figure 5-4 Gel electrophoresis to confirm *xopAQ* deletion vector sequence and transconjugation into *E. coli* DH5α. All PCRs were run with the *xopAQ*-559 primer set, which binds sequences outside the *xopAQ* gene and within the flanking regions. Lanes 1 and 10: 1 kb ladder from NEB; lane 2: BP5172 (*Xhv* race 1); lane 3: scarless *xopAQ* deletion construct (raw); lane 4: *E. coli* DH5α donor strain transformed with scarless *xopAQ* deletion construct; lane 5: repeat of lane 4; lane 6: repeat of lane 2; lane 7: marker exchange *xopAQ* deletion construct (raw); lane 8: *E. coli* DH5α donor strain transformed with marker exchange *xopAQ* deletion construct; lane 9: sterile water. The 1% agarose gel was run for 20 minutes at 100V.

5.3.2 Evaluation of candidate *Xhv* transconjugants with *xopAQ* scarless deletion vector

The triparental mating procedure using the scarless deletion construct resulted in over 300 colonies that were rifampicin resistant and sucrose tolerant and were possible double recombinants (DRs). There were no rifampicin and gentamycin resistant candidate DRs observed

from the triparental mating using the marker exchange constructs. Twenty-six of the DRs from the transconjugation with the scarless deletion construct were evaluated for the desired deletion of the *xopAQ* gene. These twenty-six colonies were suspended in 100 µL of sterile water to be used as template for PCR. The outer *xopAQ*-559 primer set was designed to amplify a 559 bp band where *xopAQ* is present and a single 257 bp band where there was as successful scarless *xopAQ* gene deletion (Figure 5-5). The wild type *Xhv* race 1 BS0347 showed the expected 559 bp band for the presence of *xopAQ*. Rifampicin resistant, sucrose tolerant candidates in lanes 5, 11, 21, and 24 showed the single 257 bp band where a successful scarless *xopAQ* gene deletion and plasmid loss may have occurred. Another candidate showed a single band at 559 bp, indicating it likely cleared the plasmid but retained the *xopAQ* gene. All the other candidates showed three bands, at 257 bp, 559 bp, and approximately 1 kb, which may represent transconjugants with both intermediates and final products from the cross-over events. The sterile water control showed no bands, as expected. Regarding the possible deletion mutants, sequencing the PCR products of those strains from lanes 5, 11, 21, and 24 will reveal whether they saw a successful scarless *xopAQ* gene deletion.

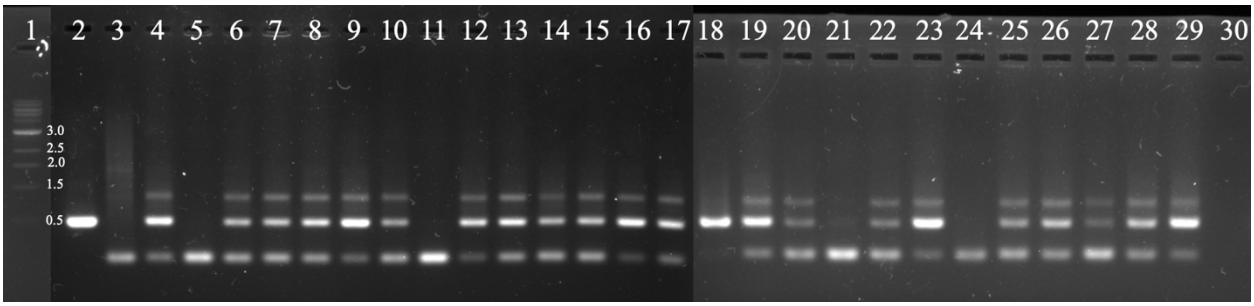


Figure 5-5 Gel electrophoresis of rifampicin resistant and sucrose tolerant *Xhv* race 1 BP2885 strains following transconjugation with *E. coli* DH5α containing the *xopAQ* scarless deletion construct. All PCRs were run with the *xopAQ*-559 primer set, which binds target sequences flanking the *xopAQ* gene. Lane 1: 1 kb ladder from NEB; lane 2: *Xhv* race 1 strain BS0347; lanes 3-29: *Xhv* race 1 strains following transconjugation with the scarless deletion construct; lane 30: sterile water. The 1% agarose gel was run for 20 minutes at 100V.

5.3.3 Confirmation of *xopAQ* expression construct and its attempted transformation into *Xhv* race 2 strain BP5169.

The electroporation procedure using the pUC19 positive control vector resulted in five to ten colonies per plate across ten plates amended with ampicillin. Using the *xopAQ* expression vector, we saw eight colonies across ten plates amended with kanamycin that were possible transformants. However, the presence of the *xopAQ* gene could not be confirmed via PCR with colony suspensions and the inner *xopAQ* primer set, which should amplify a 257 bp band when the effector gene is present. None of the colonies produced a visible band in gel electrophoresis. Figure 5-6 shows this result with one representative of these eight colonies in lane 4, while the *Xhv* race 1 strain BS2869 (lane 2) and the expression vector suspended in sterile water (lane 3) demonstrated the expected 257 bp band indicating *xopAQ* gene presence. The sterile water control also did not produce any bands, and this was expected (lane 5). These results seemed to

suggest that while the plasmids likely did contain the desired *xopAQ* gene, the transformants did not, and they may have not contained the plasmid at all. This result could have occurred due to mutation in the primer binding sites as the plasmids propagated in the transformed cells, and additional screening using new primers designed to amplify the whole multicloning site of pDSK519 will be completed to try to resolve this issue. It is unlikely that the kanamycin resistance observed in these colonies evolved naturally; spread plating the *Xhv* race 2 strain alone on NA amended with kanamycin did not result in any Kan^r colonies.

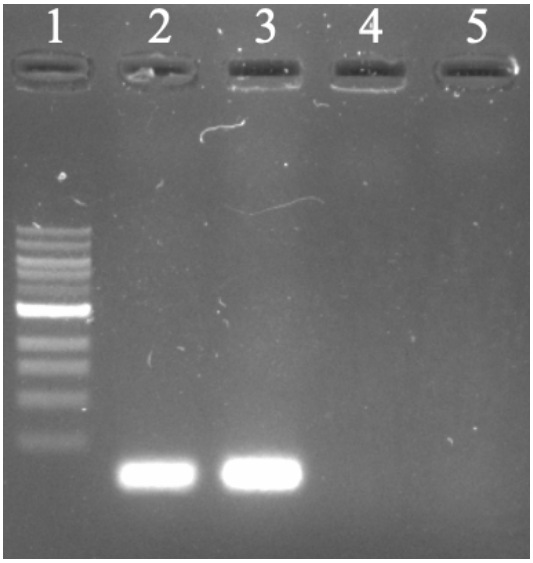


Figure 5-6 Gel electrophoresis to confirm *xopAQ* expression vector sequence and electroporation into *Xhv* race 2 strain BP5169. All reactions were run with the *xopAQ*-266 primer set, which binds to DNA sequence internal to the *xopAQ* gene. Lane 1: 1 kb ladder from NEB; lane 2: *Xhv* race 1 strain BS2869; lane 3: *xopAQ* expression construct (raw); lane 4: *Xhv* race 2 BP5169 strain transformed with *xopAQ* expression construct; lane 5: sterile water. The 1% agarose gel was run for 20 minutes at 100V.

5.4 Conclusions

This chapter demonstrates significant progress toward effector gene deletion from the *Xhv* race 1 strain BP2885 and effector gene expression in the *Xhv* race 2 strain BP5169. Testing two different methods of construct design revealed the scarless deletion method to be more effective in producing transconjugants, as none were observed using the marker exchange deletion method. Using the scarless deletion method, we identified four possible candidates that may have experienced the *xopAQ* gene deletion from their chromosomes. To continue this work, we will first evaluate these candidates by sequencing the PCR products to confirm the absence of the *xopAQ* sequence between the flanking regions. If the gene deletion is confirmed, then a deletion mutant will be inoculated into *L. serriola* ARM-09-161-10-1 to determine whether the deletion disrupted the elicitation of HR. As negative controls, we would also inoculate this lettuce accession with sterile phosphate buffer and an *Xhv* strain carrying an empty vector; as a positive control to visualize normal HR induction, we would inoculate the deletion mutant into *L. sativa* cv. Little Gem.

Further work is necessary to generate the *Xhv* race 2 strain expressing *xopAQ*. Electroporation of the expression vector directly into *Xhv* race 2 strain BP5169 was attempted as a faster alternative to the triparental mating method, at a time when we were still troubleshooting the use of the latter method for effector gene deletion and having difficulty producing transformants. Demonstrating that electroporation can be successfully used to transform an *X. hortorum* strain would be a novel finding. For this reason, this author suggests first repeating the PCR confirmation with the candidate transformants created from electroporation, but with primers to amplify the whole multicloning site to be sure real transformants were missed using *xopAQ*-specific primers. Should electroporation continue to prove unsuccessful, transformation

of the pDSK519 *xopAQ* expression vector may be achieved using the same triparental mating procedure as described for generating the scarless deletion mutants and should be attempted next.

These procedures for generating and confirming the *xopAQ* gene deletion from *Xhv* race 1 strain BP2885 and expression in *Xhv* race 2 strain BP5169 should then be repeated for the *xopAF2* deletion and expression and to create empty vector controls. Both genes were determined to be encoded by *Xhv* race 1 and 3 strains, but not by *Xhv* race 2 strains. *Xhv* strains are known to be predominantly clonal (Fayette et al., 2016), and effectors are proven to be host-limiting in other *Xanthomonas*-plant pathogen interactions such as those in tomatoes and peppers, and so we hypothesize that this *Xhv* race 1 and 3-specific variation in effector repertoire is responsible for these strains' HR elicitation in the *L. serriola* accession ARM-09-161-10-1. When one of the effector genes is deleted from the *Xhv* race 1 strain, the usual HR elicitation observed for the race 1 strain in *L. serriola* accession ARM-09-161-10-1 is expected to be lost. When one of the effector genes is expressed in the *Xhv* race 2 strain, the usual disease symptoms that are seen in *L. serriola* accession ARM-09-161-10-1 are expected to be replaced by the gain of HR elicitation. This would suggest that *xopAQ* or *xopAF2* from the *Xhv* race 1 and 3 strains interacts with this particular lettuce host in a gene-for-gene interaction that results in disease resistance.

There are several other possible outcomes. One possibility is that both effector genes are targets of lettuce R-genes and therefore both elicit HR. This would be observed in our HR screen as all deletion and expression mutants having the ability to trigger HR. To investigate this hypothesis, double deletion mutants for both *xopAQ* and *xopAF2* could be created in an *Xhv* race 1 or 3 strain. HR screens with these mutants would be expected to reveal their loss of HR elicitation in *L. serriola* accession ARM-09-161-10-1.

Another possibility is that the deletion and expression mutants will not change the resistant or disease phenotypes upon inoculation into *L. serriola* accession ARM-09-161-10-1. This would suggest that there are other, yet unknown gene(s) expressed by *Xhv* races 1 and 3 that are recognized by the plant and lead to HR induction in this accession. To explore this hypothesis, one could use a forward genetic approach, such as the creating of a cosmid library that contains fragments of the entire *Xhv* race 1 or 3 genome, move this library into an *Xhv* race 2 strain, and then conduct HR screens with groups of clones until one triggers HR in *L. serriola* accession ARM-09-161-10-1. Then, that cosmid could be sequenced to identify the genome fragment responsible for HR elicitation. Smaller gene targets within the cosmid-borne genome fragment could be cloned into an expression vector to narrow down the gene sequence triggering HR.

The continuation of this study should reveal what role, if any, the *xopAQ* and *xopAF2* effector genes play in HR elicitation in the wild lettuce accession ARM-09-161-10-1, but further research would be necessary to determine what mechanism underlies the other race-specific reactions that lettuce cultivars and accessions have to *Xhv* strains. The discovery of single, dominant lettuce genes that regulate HR elicitation to *Xhv* race 1 strains, suggested that those could be R-genes that recognize a specific gene expressed by only *Xhv* race 1 strains. Comparative genomic analysis did not reveal *Xhv* race 1-specific genes, but a forward genetics approach could also be applied here, such as the cosmid approach described above, with screening of the library in the race 1 determinant cultivars. While a potential R-gene has not been identified for the *Xhv* race 2 determinant lettuce line, *L. serriola* PI491114, the same approach could be taken to identify a bacterial gene responsible for HR induction in this line.

Understanding the genetic components underlying *Xhv* HR elicitation has important applications. First, the identification of the effector genes at play can allow for their tracking in *Xhv* populations around the world, through the PCR amplification of the genes from isolates taken from disease outbreaks. With this, we might better understand how these bacterial strains gain pathogenicity in the production setting. Second, the avirulence gene sequences can be useful for predicting their interaction partners in lettuce, the R-genes. This could be accomplished by determining the expected avirulence gene binding sequence in the lettuce chromosome or by co-immunoprecipitation to isolate the resistance protein. Lastly, the identification of the R-gene sequences could improve the efficiency of lettuce breeding programs: descendants of breeding crosses could be screened for resistance gene sequences via PCR, without the need to conduct an HR screen for every descendant.

Chapter 6

Summary and Future Directions

Bacterial leaf spot caused by *Xanthomonas hortorum* pv. *vitians* represents a significant threat to lettuce industries worldwide. With our studies, we sought to characterize the phenotypic and genetic diversity of *Xhv* strains (Chapters 2 and 3), develop rapid, PCR-based detection methods suitable for the detection of *Xhv* strains from plant tissue and bacterial culture and the differentiation of the *Xhv* races (Chapter 4), and investigate the race-specific trigger for the hypersensitive response in lettuce cultivars (Chapter 5). The results of these studies offer new insights into the incompatible reactions between *Xhv* and its lettuce host, and important progress in the development of *Xhv* detection methods. Below, we share a summary of the major findings and avenues for future work. Additional work conducted during this degree resulted in the renaming of the BLS pathogen of lettuce (Morinière et al., 2020) identification of a *Xanthomonas* pathogen of arugula (Rosenthal et al., 2017), novel breeding material, and will result in publication of race differential cultivars (Sandoya et al., in preparation).

The three races of *Xhv* were characterized by their resistance reactions on three lettuce lines. *Xhv* race 1 strains cause HR on *Lactuca sativa* cultivar Little Gem (and *L. serriola* ARM-09-161-10-1), *Xhv* race 2 strains cause HR on *L. serriola* PI491114, and *Xhv* race 3 strains cause HR on *L. serriola* ARM-09-161-10-1 only. These definitions correspond to sequetypes identified by multilocus sequence analysis, as demonstrated first by Bull et al., 2016, and we confirm that MLSA can be used as a predictive tool for determining *Xhv* race but assert that HR screening must be done using the differential cultivars to confirm the resistance reaction. Regarding suspect *Xhv* strains we received from Florida and France: the FL strains appear cluster by MLSA with known race 1 strains, and the French strains group with *Xhv* race 1 and 2 strains, and

additional HR screening is necessary to confirm these predictions. One of the FL strains evaluated we evaluated presented atypical HR upon inoculation into *L. sativa* cv. Little Gem. Additional HR screening with our collection of over 100 *Xhv* strains may reveal more strains with this atypical HR phenotype, or possibly other races or differential lettuce hosts. Then, genome sequencing of those with the atypical HR phenotype may be compared to those with the typical to see if any hypotheses can be generated about the origin of this odd response. Finally, two of the suspected *Xhv* strains from FL were identified as *Pantoea* species; the question of whether these strains are pathogenic on lettuce and capable of producing similar symptoms to BLS may be addressed through pathogenicity screening on lettuce.

Our second objective was to use genome sequence comparison as a tool to identify genetic differences between the *Xhv* races. We found two bacterial effector genes that were predicted in *Xhv* race 1 and 3 genomes, but not in those of *Xhv* race 2. Effectors proteins that are translocated from bacterial to plant cell cytoplasm and while they typically enable pathogenic bacteria to overcome host primary immunity, they can also serve as signals for HR elicitation in host. From this study, we developed the hypothesis that one of these two effectors, *xopAQ* or *xopAF2* (a novel variant related to *hopAF1* from *Pseudomonas syringae* pv. *tomato* strain DC3000), was responsible for triggering HR in *L. serriola* ARM-09-161-10-1. Next, we predicted the core effector repertoire of *Xhv* strains, consisting of fifteen effectors, as well as the bacteriocin genes, secondary metabolite gene clusters, and secretion system genes of all strains tested. This data can be used for further study to identify pathovar-specific components in the *X. hortorum* strains related to *Xhv* that are responsible for their host-specificity. Lastly, a WGS-based phylogeny we built from SNP data allowed us to hypothesize that *Xhv* race 3 strains are derived from *Xhv* race 1; this can be further investigated by annotating the whole genome

sequences of these strains and comparing loci that differentiate them to see where genes have been gained or lost.

This whole genome sequence data enabled us to find *Xhv* pathovar- and race-specific gene clusters that could be used as amplicons in PCR-based detection methods. We developed a touchdown PCR method and primer set suitable for the detection of *Xhv* and *X. hydrangea* strains from plant tissue and bacterial culture. This represents an improvement over the previously developed B162 primers, which would pick up several other *X. hortorum* relatives in addition to *Xhv*. However, further work could be done to eliminate the detection of *X. hydrangea* strains using our method. For this, we recommend incorporating genome sequences of *X. hydrangea* strains into the whole genome alignments that helped us identify *Xhv*-specific sequences and using those as targets for *Xhv* specific detection. Any additional sequencing, of X. hydrangea or *Xhv* strains, should be completed with emerging long-read sequencing technologies to ensure better genome coverage and fewer gaps in assembly. One could also investigate the host range of *X. hydrangea* using pathogenicity testing; this could demonstrate whether those strains are also capable of inducing bacterial leaf spot disease on lettuce in addition to their known host, smooth hydrangea. Finally, further study necessary to evaluate our race-specific detection methods. Our methods produce several exceptions, including the low rate of false negatives with our *Xhv* race 1- and 2- specific protocols and false positives with our *Xhv* race 2 and 3 specific protocols. These strains producing false positives and negatives should be evaluated for their race by HR to determine that they were not mischaracterized previously; if those results are consistent with their previous descriptions, then sequencing the gene clusters could help identify more precise race-specific targets. Ultimately, the pathovar- and race-specific detection methods are expected to be important tools for clinicians and seed companies

seeking to provide clients with plant disease diagnoses or clean seed, and researchers aiming to evaluate the diversity of *Xhv* strains in future BLS outbreaks.

The fourth objective was to evaluate the *Xhv* race 1 and 3 specific effectors, *xopAQ* and *xopAF2*, for their role in HR induction in *L. serriola* ARM-09-161-10-1. Our goal was to create deletion mutants for these genes out of *Xhv* race 1 or 3 strains, and expression mutants for the genes in *Xhv* race 2 strains. Our studies determined that scarless gene deletion may be preferable to the marker exchange method, as we consistently saw the growth of transconjugants with the former but not with the latter. We appear to have several successful deletion mutants for the *xopAQ* gene from an *Xhv* race 1 strain, and amplicon sequencing can confirm this finding. Our transformations of the expression constructs into *Xhv* race 2 were so far unsuccessful, but in Chapter 5 we outline additional steps that will be taken to generate *Xhv* race 2 strains expressing *xopAQ* and *xopAF2*. which may have caused the observed growth of colonies that did not contain the expression vector. The procedures for creating *xopAQ* deletion and expression mutants should then be completed for the *xopAF2* gene, and all the mutants and empty vector controls should be tested along with wild type strains in HR screens with *L. serriola* ARM-09-161-10-1. The results of the HR screen might reveal whether these genes elicit HR in the *L. serriola* host. If necessary, double mutants could be created to evaluate whether the two genes have redundant avirulence functions and therefore both trigger HR in the same host.

Two remaining gaps in our knowledge regarding the *Xhv*-lettuce interaction are the possible genetic sources of HR elicitation from *Xhv* race 1 and 2 strains on *L. sativa* cv. Little Gem and *L. serriola* PI491114, respectively, and the identification of R-gene sequences in the differential lettuce cultivars. For the former, one could generate a cosmid library with coverage of a whole *Xhv* race 1 or 2 genome, and clones inserted into a normally pathogenic strain could

be inoculated into the appropriate lettuce cultivar to see whether the clones demonstrate a gain of HR induction ability. Then, the genome fragment in the cosmid clone could be sequenced and used to make predictions of possible proteins that trigger HR. These could then be investigated by cloning these predicted genes into expression vectors and again screening for gain of HR elicitation by a typically disease-causing strain. Ultimately, this approach could reveal the genes responsible for race-specific HR elicitation. This could assist in another major area, the identification of lettuce R-genes. Avirulence genes sometimes bind plant genomes directly, so one could use the *Xhv* avirulence gene sequences to predict the lettuce genome binding sequences, which could then be searched for in lettuce whole genome sequences that have been previously published. Alternatively, a method such as co-immunoprecipitation may be completed to isolate the binding partner of the avirulence gene from the lysis of tissue undergoing a HR. Then, peptide sequencing would reveal the lettuce R-gene. With this knowledge, lettuce germplasm quickly screened for R-gene presence by PCR, and only those descendants encoding the parental R-gene undergo confirmation via HR screening. Lastly, new cultivars could be more rapidly created with stacked R-genes for broad resistance to *Xhv* strains.

References

Abbasi, P.A., Khabbaz, S.E., Weselowski, B. & Zhang, L. (2015) Occurrence of copper-resistant strains and a shift in *Xanthomonas* spp. causing tomato bacterial spot in Ontario. *Canadian Journal of Microbiology*, **61**, 753–761.

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Arnaud, M. G. (1920). Une maladie bactcriennc du lierre (I fe-dera helix L.). *C. R. Lebdomadaires Seances Acad. Sci.* 171, 121–122.

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., & Wishart, D. S. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic acids research*, *44*(W1), W16–W21. https://doi.org/10.1093/nar/gkw387

Astua-Monge, G., Minsavage, G. V., Stall, R. E., Davis, M. J., Bonas, U., & Jones, J. B. (2000). Resistance of tomato and pepper to T3 strains of *Xanthomonas campestris* pv. *vesicatoria* is specified by a plant-inducible avirulence gene. *Molecular plant-microbe interactions: MPMI*, *13*(9), 911–921. https://doi.org/10.1094/MPMI.2000.13.9.911

Ayala, A., Muñoz, M. F., & Argüelles, S. (2014). Lipid peroxidation: production, metabolism, and signaling mechanisms of malondialdehyde and 4-hydroxy-2-nonenal. *Oxidative medicine and cellular longevity*, *2014*, 360438. https://doi.org/10.1155/2014/360438

Balaji, V., Gibly, A., Debbie, P., & Sessa, G. (2007). Transcriptional analysis of the tomato resistance response triggered by recognition of the Xanthomonas type III effector AvrXv3. *Functional & integrative genomics*, *7*(4), 305–316. https://doi.org/10.1007/s10142-007-0050-y

Baldwin W. S. (2019). Phase 0 of the Xenobiotic Response: Nuclear Receptors and Other Transcription Factors as a First Step in Protection from Xenobiotics. *Nuclear receptor research*, *6*, 101447. https://doi.org/10.32527/2019/101447

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tessler G, Alekseyev MA, Pevzner PA (2012). "SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing." Journal of Computational Biology, 19(5), 455–477. http://doi.org/10.1089/cmb.2012.0021

Barak, J. D., Koike, S. T., and Gilbertson, R. L. 2001. The role of crop debris and weeds in the epidemiology of bacterial leaf spot of lettuce in California. Plant Dis. 85:169-178.

Barak, J. D., Koike, S. T., and Gilbertson, R. L. 2002. Movement of Xanthomonas campestris pv. vitians in the stems of lettuce and seed contamination. Plant Pathology. 51, 506-512.

Barak, J. D., Vancheva, T., Lefeuvre, P., Jones, J. B., Timilsina, S., Minsavage, G. V., Vallad, G. E., & Koebnik, R. (2016). Whole-Genome Sequences of *Xanthomonas euvesicatoria* Strains Clarify Taxonomy and Reveal a Stepwise Erosion of Type 3 Effectors. *Frontiers in plant science*, *7*, 1805. https://doi.org/10.3389/fpls.2016.01805

Bernal, P., Llamas, M. A., & Filloux, A. (2018). Type VI secretion systems in plant-associated bacteria. *Environmental microbiology*, *20*(1), 1–15. https://doi.org/10.1111/1462-2920.13956

Better, M., & Helinski, D. R. (1983). Isolation and characterization of the recA gene of Rhizobium meliloti. *Journal of bacteriology*, *155*(1), 311–316. https://doi.org/10.1128/jb.155.1.311-316.1983

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., Medema, M. H., & Weber, T. (2019). antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic acids research*, *47*(W1), W81–W87. https://doi.org/10.1093/nar/gkz310

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., & Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science (New York, N.Y.)*, *326*(5959), 1509–1512. https://doi.org/10.1126/science.1178811

Boch, J., & Bonas, U. (2010). Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annual review of phytopathology*, *48*, 419–436. https://doi.org/10.1146/annurev-phyto-080508-081936

Bogdanove, A. J., Schornack, S., & Lahaye, T. (2010). TAL effectors: finding plant genes for disease and defense. *Current opinion in plant biology*, *13*(4), 394–401. https://doi.org/10.1016/j.pbi.2010.04.010

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bradley, E. L., Ökmen, B., Doehlemann, G., Henrissat, B., Bradshaw, R. E., & Mesarich, C. H. (2022). Secreted Glycoside Hydrolase Proteins as Effectors and Invasion Patterns of Plant-Associated Fungi and Oomycetes. *Frontiers in plant science*, *13*, 853106. https://doi.org/10.3389/fpls.2022.853106

Brady, C. L., Venter, S. N., Cleenwerck, I., Engelbeen, K., Vancanneyt, M., Swings, J., & Coutinho, T. A. (2009). Pantoea vagans sp. nov., Pantoea eucalypti sp. nov., Pantoea deleyi sp. nov. and Pantoea anthophila sp. nov. *International journal of systematic and evolutionary microbiology*, *59*(Pt 9), 2339–2345. https://doi.org/10.1099/ijs.0.009241-0

Brim, C. A. (1966). A Modified Pedigree Method of Selection in Soybeans. *Crop Science*, 6(1): 95.

Brown, J.K.M. and Rant, J.C. (2013), Fitness costs and trade-offs of disease resistance and their consequences for breeding arable crops. Plant Pathol, 62: 83-95. https://doi.org/10.1111/ppa.12163

Brown, N. A. (1918). Some bacterial diseases of lettuce. J. Agric. Res. 13, 367-388.

Brown, N. A. (1923). Bacterial leaf spot of geranium in the eastern United States. *J. Agric. Res.* 23, 361-372.

Bull CT, Koike ST (2005) Evaluating the efficacy of commercial products for management of bacterial leaf spot on lettuce. Plant Health Prog. https://doi.org/10.1094/PHP-2005-1121-01-RS

Bull, C. T., Goldman, P. H., Hayes, R., Madden, L. V., Koike, S. T., and Ryder, E. 2007. Genetic diversity of lettuce for resistance to bacterial leaf spot caused by *Xanthomonas campestris* pv. *vitians*. Online. Plant Health Progress. https://doi.org/10.1094/PHP-2007-0917-02-RS.

Bull, C. T., Gebben, S. J., Goldman, P. H., Trent, M. A., Hayes, R. J. 2015. Host genotype and hypersensitive reaction influence population levels of *Xanthomonas campestris* pv. *vitians* in lettuce. Phytopathology 105:316-324.

Bull, C., Trent, M., Hayes, R., 2016. Three races of *Xanthomonas campestris* pv. *vitians* causing bacterial leaf spot on lettuce identified. Phytopathology S4:100.

Büttner, D. (2016). Behind the lines-actions of bacterial type III effector proteins in plant cells. *FEMS microbiology reviews*, *40*(6), 894–937. https://doi.org/10.1093/femsre/fuw026

Carisse, O., Ouimet, A., Toussaint, V., & Philion, V. (2000). Evaluation of the Effect of Seed Treatments, Bactericides, and Cultivars on Bacterial Leaf Spot of Lettuce Caused by *Xanthomonas campestris* pv. *vitians*. *Plant Disease*, *84*(3), 295–299. https://doi.org/10.1094/PDIS.2000.84.3.295.

Cesbron, S., Briand, M., Essakhi, S., Gironde, S., Boureau, T., Manceau, C., Fischer-Le Saux, M., & Jacques, M. A. (2015). Comparative Genomics of Pathogenic and Nonpathogenic Strains of *Xanthomonas arboricola* Unveil Molecular and Evolutionary Events Linked to Pathoadaptation. *Frontiers in plant science*, *6*, 1126. https://doi.org/10.3389/fpls.2015.01126

Charkowski, A. O., Alfano, J. R., Preston, G., Yuan, J., He, S. Y., & Collmer, A. (1998). The *Pseudomonas syringae* pv. *tomato* HrpW protein has domains similar to harpins and pectate lyases and can elicit the plant hypersensitive response and bind to pectate. *Journal of bacteriology*, *180*(19), 5211–5217. https://doi.org/10.1128/JB.180.19.5211-5217.1998

Chen, L.-Q., Hou, B.-H., Lalonde, S., Takanaga, H., Hartung, M.L., Qu, X.-Q. et al. (2010) Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature*, **468**, 527–532.

Chorost, M. S., Smith, N. C., Hutter, J. N., Ong, A. C., Stam, J. A., McGann, P. T., Hinkle, M. K., Schaecher, K. E., & Kamau, E. (2018). Bacteraemia due to *Microbacterium paraoxydans* in a patient with chronic kidney disease, refractory hypertension and sarcoidosis. *JMM case reports*, *5*(11), e005169. https://doi.org/10.1099/jmmcr.0.005169

Conklin KY, Stancheva R, Otten TG, Fadness R, Boyer GL, Read B, Zhang X, Sheath RG. Molecular and morphological characterization of a novel dihydroanatoxin-a producing Microcoleus species (cyanobacteria) from the Russian River, California, USA. Harmful Algae. 2020 Mar;93:101767. doi: 10.1016/j.hal.2020.101767. Epub 2020 Mar 4. PMID: 32307065.

Croucher, N. J., Page, A. J., Connor, T. R., Delaney, A. J., Keane, J. A., Bentley, S. D., Parkhill, J., & Harris, S. R. (2015). Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic acids research*, *43*(3), e15. https://doi.org/10.1093/nar/gku1196

Damerum, A., Chapman, M. A., & Taylor, G. (2020). Innovative breeding technologies in lettuce for improved post-harvest quality. *Postharvest biology and technology*, *168*, 111266. https://doi.org/10.1016/j.postharvbio.2020.111266

Dangl, J. L., & Jones, J. D. (2001). Plant pathogens and integrated defence responses to infection. *Nature*, *411*(6839), 826–833. https://doi.org/10.1038/35081161

Darrasse, A., Carrère, S., Barbe, V., Boureau, T., Arrieta-Ortiz, M. L., Bonneau, S., Briand, M., Brin, C., Cociancich, S., Durand, K., Fouteau, S., Gagnevin, L., Guérin, F., Guy, E., Indiana, A., Koebnik, R., Lauber, E., Munoz, A., Noël, L. D., Pieretti, I., … Jacques, M. A. (2013). Genome sequence of *Xanthomonas fuscans* subsp. *fuscans* strain 4834-R reveals that flagellar motility is not a general feature of xanthomonads. *BMC genomics*, *14*, 761. https://doi.org/10.1186/1471-2164-14-761

Dia, N. C., Van Vaerenbergh, J., Van Malderghem, C., Blom, J., Smits, T. H. M., Cottyn, B., & Pothier, J. F. (2021). *Xanthomonas hydrangeae* sp. nov., a novel plant pathogen isolated from *Hydrangea arborescens*. *International journal of systematic and evolutionary microbiology*, *71*(12), 10.1099/ijsem.0.005163. https://doi.org/10.1099/ijsem.0.005163

Dia, N. C., Cottyn, B., Blom, J., Smits, T. H. M., and Pothier, J. F. (2022a) Differentiation of the *Xanthomonas hortorum – Xanthomonas hydrangeae* Species Complex Using Sensitive and Selective LAMP Assays. Front. Agron. 4(898778). doi: 10.3389/fagro.2022.898778

Dia, N. C., Morinière, L., Cottyn, B., Bernal, E., Jacobs, J. M., Koebnik, R., Osdaghi, E., Potnis, N., & Pothier, J. F. (2022b). *Xanthomonas hortorum* - beyond gardens: Current taxonomy, genomics, and virulence repertoires. *Molecular plant pathology*, *23*(5), 597–621. https://doi.org/10.1111/mpp.13185

Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K., & Mattick, J. S. (1991). 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic acids research*, *19*(14), 4008. https://doi.org/10.1093/nar/19.14.4008

Dowson WJ, (1939) “On the systematic position and generic names of the Gram negative bacterial plant pathogens.” Zentralblatt fiir Bakteriologie, Parasitenkunde und lnfektionskrankheiten 100: 177-193.

Dowson, W. J. 1943. On the generic names Pseudomonas, *Xanthomonas* and Bacterium for certain bacterial plant pathogens. Trans. Br. MycoL Soc. 26:4-14.

Durrant, M. G., Li, M. M., Siranosian, B. A., Montgomery, S. B., & Bhatt, A. S. (2020). A Bioinformatic Analysis of Integrative Mobile Genetic Elements Highlights Their Role in Bacterial Adaptation. *Cell host & microbe*, *27*(1), 140–153.e9. https://doi.org/10.1016/j.chom.2019.10.022

Dye D. W. (1978). Genus IX *Xanthomonas*. In a proposed nomenclature and classification for plant pathogenic bacteria. *N. Z. J. Agric. Res.* 21, 153–177. 10.1080/00288233.1978.10427397

Dye DW, Bradbury JF, Coto M, Hayward AC, Lelliott RA, and Schroth MN. (1980) “International standards for naming pathovars of phytopathogenic bacteria and a list of pathovar names and pathotypes.” Review of Plant Pathology 59: 153-168.

Elstner EF. (1991) “Mechanisms of oxygen activation in different compartments of plant cells.” Active Oxygen/Oxidative Stress in Plant Metabolis, eds. EJ Pelland, KL Steffen. pp.13–25. Rockville, MD: Am. Soc. Plant Physiol.

Eren, A. M., Esen, Ö. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., & Delmont, T. O. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ*, *3*, e1319. https://doi.org/10.7717/peerj.1319

Fayette, J., Raid, R., Roberts, P. D., Jones, J. B., Pernezny, K., Bull, C. T., and Goss, E. M. (2016) “Multilocus sequence typing of strains of bacterial spot of lettuce collected in the United States.” Phytopathology http://dx.doi.org /10.1094/PHYTO-11-15-0302-R.

Furutani A, Takaoka M, Sanada H, Noguchi Y, Oku T, Tsuno K, Ochiai H, and Tsuge S. (2008) “Identification of Novel Type III Secretion Effectors in Xanthomonas oryzae pv. oryzae) MPMI. 22(1):96-106.

Gentzel, I., Giese, L., Ekanayake, G., Mikhail, K., Zhao, W., Cocuron, J. C., Alonso, A. P., & Mackey, D. (2022). Dynamic nutrient acquisition from a hydrated apoplast supports biotrophic proliferation of a bacterial pathogen of maize. *Cell host & microbe*, *30*(4), 502–517.e4. https://doi.org/10.1016/j.chom.2022.03.017

Gimenez-Ibanez S, Boter M, Fernandez-Barbero G, Chini A, Rathjen JP, and Solano R. (2014). “The bacterial effector HopX1 targets JAZ transcriptional repressors to activate jasmonate signaling and promote infection in Arabidopsis.” PLoS Biol. 12(2): e1001792.

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics (Oxford, England)*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Guttman, D. S., Vinatzer, B. A., Sarkar, S. F., Ranall, M. V., Kettler, G., & Greenberg, J. T. (2002). A functional screen for the type III (Hrp) secretome of the plant pathogen *Pseudomonas syringae*. *Science (New York, N.Y.)*, *295*(5560), 1722–1726. https://doi.org/10.1126/science.295.5560.1722

Hajri, A., Brin, C., Hunault, G., Lardeux, F., Lemaire, C., Manceau, C., et al. (2009). A "repertoire for repertoire" hypothesis: repertoires of type three effectors are candidate determinants of host specificity in Xanthomonas. PLoS One 4:e6632. doi: 10.1371/journal.pone.0006632

Hayes, R. J., Trent, M. A., Truco, M. J., Antonise, R., Michelmore, R. W., and Bull, C. T. (2014) "The inheritance of resistance to bacterial leaf spot of lettuce caused by *Xanthomonas campestris* pv. *vitians* in three lettuce cultivars." Horticulture Research. 66.

Heid, C. A., Stevens, J., Livak, K. J., & Williams, P. M. (1996). Real time quantitative PCR. *Genome research*, *6*(10), 986–994. https://doi.org/10.1101/gr.6.10.986

Hegemann, J. D., Zimmermann, M., Zhu, S., Klug, D., & Marahiel, M. A. (2013). Lasso peptides from proteobacteria: Genome mining employing heterologous expression and mass spectrometry. *Biopolymers*, *100*(5), 527–542. https://doi.org/10.1002/bip.22326

Hilbi, H., Nagai, H., Kubori, T., & Roy, C. R. (2017). Subversion of Host Membrane Dynamics by the *Legionella* Dot/Icm Type IV Secretion System. *Current topics in microbiology and immunology*, *413*, 221–242. https://doi.org/10.1007/978-3-319-75241-9_9

Holtsmark, I., Eijsink, V. G., & Brurberg, M. B. (2008). Bacteriocins from plant pathogenic bacteria. *FEMS microbiology letters*, *280*(1), 1–7. https://doi.org/10.1111/j.1574-6968.2007.01010.x

Hou, Z., Fink, R. C., Radtke, C., Sadowsky, M. J., & Diez-Gonzalez, F. (2013). Incidence of naturally internalized bacteria in lettuce leaves. *International journal of food microbiology*, *162*(3), 260–265. https://doi.org/10.1016/j.ijfoodmicro.2013.01.027

Hueck C. J. (1998). Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiology and molecular biology reviews : MMBR*, *62*(2), 379–433. https://doi.org/10.1128/MMBR.62.2.379-433.1998

Jha, G., Rajeshwari, R., & Sonti, R. V. (2005). Bacterial type two secretion system secreted proteins: double-edged swords for plant pathogens. *Molecular plant-microbe interactions: MPMI*, *18*(9), 891–898. https://doi.org/10.1094/MPMI-18-0891.

Johansson, O. N., Nilsson, A. K., Gustavsson, M. B., Backhaus, T., Andersson, M. X., & Ellerström, M. (2015). A quick and robust method for quantification of the hypersensitive response in plants. *PeerJ*, *3*, e1469. https://doi.org/10.7717/peerj.1469.

Khanal, S., Hind, S.R. & Babadoost, M. (2020) Occurrence of copper-resistant *Xanthomonas perforans* and *X. gardneri* in Illinois tomato fields. *Plant Health Progress*, **21**, 338–344.

Khanal, M., Bhatta, B. P., Timilsina, S., Ghimire, S., Cochran, K., & Malla, S. (2023). Curtobacterium allii sp. nov., the actinobacterial pathogen causing onion bulb rot. *Antonie van Leeuwenhoek*, *116*(2), 83–96. https://doi.org/10.1007/s10482-022-01775-z

Lane, D. J., Pace, B., Olsen, G. J., Stahl, D. A., Sogin, M. L., & Pace, N. R. (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proceedings of the National Academy of Sciences of the United States of America*, *82*(20), 6955–6959. https://doi.org/10.1073/pnas.82.20.6955.

Lapage SP, Sneath PHA, Lessel EF, et al., editors. "International Code of Nomenclature of Bacteria: Bacteriological Code, 1990 Revision." Washington (DC): ASM Press; 1992. Available from: https://www.ncbi.nlm.nih.gov/books/NBK8817/.

La Rosa, G., Pourshaban, M., Iaconelli, M., & Muscillo, M. (2010). Quantitative real-time PCR of enteric viruses in influent and effluent samples from wastewater treatment plants in Italy. *Annali dell'Istituto superiore di sanita*, *46*(3), 266–273. https://doi.org/10.4415/ANN_10_03_07

Lim, Y. K., Kweon, O. J., Kim, H. R., Kim, T. H., & Lee, M. K. (2017). First case of bacteremia caused by Janibacter hoylei. *APMIS : acta pathologica, microbiologica, et immunologica Scandinavica*, *125*(7), 665–668. https://doi.org/10.1111/apm.12693

Liyanapathiranage, P., Jones, J. B., & Potnis, N. (2021). A mutation of a single core gene, tssM, of Type VI secretion system of *Xanthomonas perforans* influences virulence, epiphytic survival and transmission during pathogenesis on tomato. Phytopathology, 10.1094/PHYTO-02-21-0069-R. Advance online publication. https://doi.org/10.1094/PHYTO-02-21-0069-R

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, Y. R., Zou, H. S., Che, Y. Z., Cui, Y. P., Guo, W., Zou, L. F., Chatterjee, S., Biddle, E. M., Yang, C. H., & Chen, G. Y. (2011). A novel regulatory role of HrpD6 in regulating *hrp-hrc-hpa* genes in *Xanthomonas oryzae* pv. *oryzicola*. *Molecular plant-microbe interactions: MPMI*, *24*(9), 1086–1101. https://doi.org/10.1094/MPMI-09-10-0205

Lowe, H. H. (1927) Dosage: Pedigree and Performance. The Russell Meerdink Company Ltd.

Lu H, Raid R (2013) A novel screening method for evaluation of lettuce germplasm for bacterial leaf spot resistance. HortScience 48(2):171–174

Macho AP, Schwessinger B, Ntoukakis V, Brutus A, Segonzac C, Roy S, Kadota Y, Oh MH, Skenlar J, Derbyshire P, Lozano-Durán R, Malinovsky FG, Monaghan J, Menke FL, Huber SC, He SY, and Zipfel C. (2014) "A bacterial tyrosine phosphatase inhibits plant pattern recognition receptor activation." Science. 343:1509–12.

Maiden, M.C., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M., and Spratt, B.G. (1998) "Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms." Proc. Natl. Acad. Sci. U. S. A. 95(6): 3140-5.0027-8424

Maksimov, M. O., Pan, S. J., & James Link, A. (2012). Lasso peptides: structure, function, biosynthesis, and engineering. *Natural product reports*, *29*(9), 996–1006. https://doi.org/10.1039/c2np20070h

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., Zdobnov, E. M. (2021). BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Molecular Biology and Evolution [Internet]. Available from: https://doi.org/10.1093/molbev/msab199

McCann, H. C., and Guttman, D. S. (2008). Evolution of the type III secretion system and its effectors in plant-microbe interactions. New Phytol. 177, 33-47.

Merda, D., Briand, M., Bosis, E., Rousseau, C., Portier, P., Barret, M., et al. (2017). Ancestral acquisitions, gene flow and multiple evolutionary trajectories of the type three secretion system and effectors in Xanthomonas plant pathogens.Mol.Ecol.26, 5939–5952. doi: 10.1111/mec.14343

Migula, W. (1900). System der bakterien: Handbuch der morphologie, entwicklungsgeschichte und systematik der bakterien.

Minsavage, G. V., Dahlbeck, D., Whalen, M. C., Kearney, B., Bonas, U., Staskawicz, B. J., Stall, R. E. (1990). Gene-For-Gene Relationships Specifying Disease Resistance in *Xanthomonas campestris* pv. *vesictoria*-Pepper interactions. *Molecular Plant-Microbe Interactions*: *MPMI*, *3*(1), 41-47. DOI: 10.1094/MPMI-3-041

Morel, JB., Dangl, J. The hypersensitive response and the induction of cell death in plants. *Cell Death Differ* **4**, 671–683 (1997). https://doi.org/10.1038/sj.cdd.4400309

Morinière L., Burlet A., Rosenthal E. R., Nesme X., Portier P., Bull C. T., Lavire C., Le Saux M. F., Bertolla F. (2020) "Clarifying the taxonomy of the causal agent of bacterial leaf spot of lettuce through a polyphasic approach reveals that *Xanthomonas Cynara* Trébol et al. 2000 emend. Timilsina et al. 2019 is a later heterotypic synonym of *Xanthomonas hortorum* Vauterin

et al. (1995). Systematic and Applied Microbiology. https://doi.org/10.1016/j.syapm.2020.126087

Mukaihara, T., Tamura, N., & Iwabuchi, M. (2010). Genome-wide identification of a large repertoire of Ralstonia solanacearum type III effector proteins by a new functional screen. *Molecular plant-microbe interactions : MPMI*, *23*(3), 251–262. https://doi.org/10.1094/MPMI-23-3-0251

Newberry, E. A., Bhandari, R., Minsavage, G. V., Timilsina, S., Jibrin, M. O., Kemble, J., Sikora, E. J., Jones, J. B., & Potnis, N. (2019). Independent Evolution with the Gene Flux Originating from Multiple *Xanthomonas* Species Explains Genomic Heterogeneity in Xanthomonas perforans. *Applied and environmental microbiology*, *85*(20), e00885-19. https://doi.org/10.1128/AEM.00885-19

Norrander, J., Kempe, T., & Messing, J. (1983). Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene*, *26*(1), 101–106. https://doi.org/10.1016/0378-1119(83)90040-9

Pammel, L. H. (1895) "Bacteriosis of Rutabaga." *Microscopical Journal*, 145-151.

Pandey, P., Irulappan, V., Bagavathiannan, M. V., & Senthil-Kumar, M. (2017). Impact of Combined Abiotic and Biotic Stresses on Plant Growth and Avenues for Crop Improvement by Exploiting Physio-morphological Traits. *Frontiers in plant science*, *8*, 537. https://doi.org/10.3389/fpls.2017.00537

Pernezny K, Raid RN, Stall RE, Hodge NC, and Collins J. (1995). "An outbreak of bacterial spot of lettuce in Florida caused by *Xanthomonas campestris* pv. *vitians.*" *Plant disease*, *79*, 359-360.

Petnicki-Ocwieja, T., Schneider, D. J., Tam, V. C., Chancey, S. T., Shan, L., Jamir, Y., Schechter, L. M., Janes, M. D., Buell, C. R., Tang, X., Collmer, A., & Alfano, J. R. (2002). Genomewide identification of proteins secreted by the Hrp type III protein secretion system of *Pseudomonas syringae* pv. *tomato* DC3000. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(11), 7652–7657. https://doi.org/10.1073/pnas.112183899

Potnis, N., Krasileva, K., Chow, V., Almeida, N. F., Patil, P. B., Ryan, R. P., Sharlach, M., Behlau, F., Dow, J. M., Momol, M., White, F. F., Preston, J. F., Vinatzer, B. A., Koebnik, R., Setubal, J. C., Norman, D. J., Staskawicz, B. J., & Jones, J. B. (2011). Comparative genomics reveals diversity among xanthomonads infecting tomato and pepper. *BMC genomics*, *12*, 146. https://doi.org/10.1186/1471-2164-12-146

Potnis, N., Timilsina, S., Strayer, A., Shantharaj, D., Barak, J. D., Paret, M. L., Vallad, G. E., & Jones, J. B. (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. https://doi.org/10.1111/mpp.12244

Qiu, J., & Luo, Z. Q. (2017). *Legionella* and *Coxiella* effectors: strength in diversity and activity. *Nature reviews. Microbiology*, *15*(10), 591–605. https://doi.org/10.1038/nrmicro.2017.67

Rajagopal, L., Sundari, C. S., Balasubramanian, D., & Sonti, R. V. (1997). The bacterial pigment xanthomonadin offers protection against photodamage. *FEBS letters*, *415*(2), 125–128. https://doi.org/10.1016/s0014-5793(97)01109-5

Roach, R., Mann, R., Gambley, C. G., Chapman, T., Shivas, R. G., & Rodoni, B. (2019). Genomic sequence analysis reveals diversity of Australian *Xanthomonas* species associated with bacterial leaf spot of tomato, capsicum and chilli. *BMC Genomics*, 20(310). https://doi.org/10.1186/s12864-019-5600-x

Rosebrock TR, Zeng L, Brady JJ, Abramovitch RB, Xiao F, and Martin GB. (2007) "A bacterial E3 ubiquitin ligase targets a host protein kinase to disrupt plant immunity." Nature. 448:370–4.

Rosenthal, E. R., Sebastian, A., Potnis, N., Albert, I., Bull, C. T. (2018) Investigating effector diversity as a source of race-specific pathogenicity for the lettuce bacterial leaf spot pathogen. Boston, Massachusetts. The American Phytopathological Society.

Rosenthal, E.R., Potnis, N., Bull, C.T. (2022). "Comparative genomic analysis of the lettuce bacterial leaf spot pathogen, *Xanthomonas hortorum* pv. *vitians,* to investigate race specificity." Frontiers in Microbiology, 13. https://doi.org/10.3389/fmicb.2022.840311.

Rossier, O., Wengelnik, K., Hahn, K., & Bonas, U. (1999). The *Xanthomonas* Hrp type III system secretes proteins from plant and mammalian bacterial pathogens. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(16), 9368–9373. https://doi.org/10.1073/pnas.96.16.9368.

Russell, A. B., Peterson, S. B., & Mougous, J. D. (2014). Type VI secretion system effectors: poisons with a purpose. *Nature reviews. Microbiology*, *12*(2), 137–148. https://doi.org/10.1038/nrmicro3185

Sahin, F., and Miller, S. A. 1997. Identification of the bacterial leaf spot pathogen of lettuce, *Xanthomonas campestris* pv. *vitians*, in Ohio, and assessment of cultivar resistance and seed treatment. Plant Dis. 81:1443-1446.

Sahin, F., Abbasi, P. A., Lewis Ivey, M. L., Zhang, J., and Miller, S. A. (2003). Diversity among strains of *Xanthomonas campestris* pv. *vitians* from lettuce. Phytopathology 93:64-70.

Sandoya, G. V., Maisonneuve, B., Truco, M. J. Bull, C.T., Simko, I., Trent, M., Hayes, R. J., Michelmore, R. W. (2019). Genetic analysis of resistance to bacterial leaf spot in the heirloom lettuce cultivar Reine des Glaces. *Mol Breeding* **39,** 160. https://doi.org/10.1007/s11032-019-1072-6

Sandoya, G.V., Rosenthal, E.R., Rodrigues-Porto, L., Wadlington, W.H., Bull, C.T., Simko, I., Carroll, A. (2022). "Germplasm Resistant to Race 1 of *Xanthomonas hortorum pv. vitians* (Brown, 1918) Moriniére et al. 2020 Causing Bacterial Leaf Spot of Lettuce (*Lactuca sativa L.*)." Journal of Plant Pathology. https://doi.org/10.1007/s42161-022-01123-0.

Sarkar, S. F., Gordon, J. S., Martin, G. B., and Guttman, D. S. (2006). Comparative genomics of host-specific virulence in Pseudomonas syringae. Genetics 174,1041–1056. doi: 10.1534/genetics.106.060996

Schornack, S., Minsavage, G. V., Stall, R. E., Jones, J. B., & Lahaye, T. (2008). Characterization of AvrHah1, a novel AvrBs3-like effector from *Xanthomonas gardneri* with virulence and avirulence activity. *The New phytologist*, *179*(2), 546–556. https://doi.org/10.1111/j.1469-8137.2008.02487.x

Schroth, M., Thompson, J., Bardin, R., Greathead, A. (1964). Bacterial Spot of Lettuce. Calif. Agr. 18(11):2-3.

Scofield, S. R., Tobias, C. M., Rathjen, J. P., Chang, J. H., Lavelle, D. T., Michelmore, R. W., & Staskawicz, B. J. (1996). Molecular Basis of Gene-for-Gene Specificity in Bacterial Speck Disease of Tomato. *Science (New York, N.Y.)*, *274*(5295), 2063–2065. https://doi.org/10.1126/science.274.5295.2063

Seemann, T. (2015). Snippy: fast bacterial variant calling from NGS reads. https://github.com/tseemann/snippy.

Sgro, G. G., Oka, G. U., Souza, D. P., Cenens, W., Bayer-Santos, E., Matsuyama, B. Y., Bueno, N. F., Dos Santos, T. R., Alvarez-Martinez, C. E., Salinas, R. K., & Farah, C. S. (2019). Bacteria-Killing Type IV Secretion Systems. *Frontiers in microbiology*, *10*, 1078. https://doi.org/10.3389/fmicb.2019.01078

Shatilov, M. V., Razin, A. F., Ivanova, M. I. (2019). Analysis of the world lettuce market. IOP Conf. Series: Earth and Environmental Science 395. doi:10.1088/1755-1315/395/1/012053.

Simmonds, R. S., Pearson, L., Kennedy, R. C., & Tagg, J. R. (1996). Mode of action of a lysostaphin-like bacteriolytic agent produced by *Streptococcus zooepidemicus* 4881. *Applied and environmental microbiology*, *62*(12), 4536–4541. https://doi.org/10.1128/aem.62.12.4536-4541.1996

Stackebrandt, E., and B. M. Goebel. 1994. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int. J. Syst. Bacteriol. **44:**846-849.

Stice, S. P., Thao, K. K., Khang, C. H., Baltrus, D. A., Dutta, B., & Kvitko, B. H. (2020). Thiosulfinate Tolerance Is a Virulence Strategy of an Atypical Bacterial Pathogen of Onion. *Current biology : CB*, *30*(16), 3130–3140.e6. https://doi.org/10.1016/j.cub.2020.05.092

Svec, D., Tichopad, A., Novosadova, V., Pfaffl, M. W., & Kubista, M. (2015). How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments. *Biomolecular detection and quantification*, *3*, 9–16. https://doi.org/10.1016/j.bdq.2015.01.005

Swanson, J., Kearney, B., Dahlbeck, D., & Staskawicz, B. J. (1988). Cloned avirulence gene of *Xanthomonas campestris* pv. *vesicatoria* complements spontaneous race-change mutants. *Mol. Plant-Microbe Interact*, *1*, 5-9.

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular biology and evolution*, *38*(7), 3022–3027. https://doi.org/10.1093/molbev/msab120

Tavaré S.Miura R.M. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences*. Providence: American Mathematical Society; p. 57–86.

Thieme, F., Szczesny, R., Urban, A., Kirchner, O., Hause, G., & Bonas, U. (2007). New type III effectors from *Xanthomonas campestris* pv. *vesicatoria* trigger plant reactions dependent on a conserved N-myristoylation motif. *Molecular plant-microbe interactions: MPMI*, *20*(10), 1250–1261. https://doi.org/10.1094/MPMI-20-10-1250

Toussaint, V. (1999). Bacterial leaf spot, a new disease of lettuce in Quebec caused by *Xanthomonas campestris* pv. *vitians*. Phytoprotection. 80:121-125. https://doi.org/10.7202/706187ar.

Tzfira, T., & Citovsky, V. (2006). Agrobacterium-mediated genetic transformation of plants: biology and biotechnology. *Current opinion in biotechnology*, *17*(2), 147–154. https://doi.org/10.1016/j.copbio.2006.01.009

Umesh, K. C., Koike, S. T., and Gilbertson, R. L. 1996. Association of Xanthomonas campestris pv. vitians with lettuce seed. (Abstr.) Phytopathology. 86:S3.

USDA National Agricultural Statistics Service (2022) State Agricultural Overview. https://quickstats.nass.usda.gov/results/3D366F3F-8B5A-3F99-8270-950C58D38DAA. Accessed 19 April 2023.

Vauterin, L., Hoste, B., Kersters, K., Swings, J. (1995) Reclassification of *Xanthomonas*. *International Journal of Systematic Biology*. *45*(3), 472-489. https://doi.org/10.1099/00207713-45-3-472

Verdier, V., Triplett, L.R., Hummel, A.W., Corral, R., Cernadas, R.A., Schmidt, C.L. et al. (2012) Transcription activator-like (TAL) effectors targeting *OsSWEET* genes enhance virulence on diverse rice (*Oryza sativa*) varieties when expressed individually in a TAL effector-deficient strain of *Xanthomonas oryzae*. *The New Phytologist*, **196**, 1197–1207.

Versalovic, J., M. Schneider, F. J. de Bruijn, and J. R. Lupski. 1994. Genomic fingerprinting of bacteria using repetitive sequence-based polymerase chain reaction. *Meth. Mol. Cell Biol.* 5:25–40.

Wang, Y., Lu, H., Hu, J. (2016) Molecular mapping of high resistance to bacterial leaf spot in lettuce PI 358001-1. Phytopathology 106(11):1319–1325

Wang, X., Zhang, L., Ji, H., Mo, X., Li, P., Wang, J., & Dong, H. (2018). Hpa1 is a type III translocator in *Xanthomonas oryzae* pv. *oryzae*. *BMC microbiology*, *18*(1), 105. https://doi.org/10.1186/s12866-018-1251-3

Wang, X., Zheng, D. and Liang, R. (2016) An Efficient Electro-Competent Cells Generation Method of *Xanthomonas campestris pv. campestris*: Its Application for Plasmid Transformation and Gene Replacement. *Advances in Microbiology*, **6**, 79-87. doi: 10.4236/aim.2016.62008.

War A. R., Paulraj M. G., War M. Y., Ignacimuthu S. (2011). Jasmonic acid-mediated-induced resistance in groundnut (*Arachis hypogaea* L.) against *Helicoverpa armigera* (Hubner) (Lepidoptera: Noctuidae). *J. Plant Growth Regul.* 30 512–523. 10.1007/s00344-011-9213-0

Weber, E., Ojanen-Reuhs, T., Huguet, E., Hause, G., Romantschuk, M., Korhonen, T. K., Bonas, U., & Koebnik, R. (2005). The type III-dependent Hrp pilus is required for productive interaction of *Xanthomonas campestris* pv. *vesicatoria* with pepper host plants. *Journal of bacteriology*, *187*(7), 2458–2468. https://doi.org/10.1128/JB.187.7.2458-2468.2005

Wei, X., Lei, S., Deng, Y., Zhang, Y., Zhang, L., & Niu, Q. (2021). Stenotrophomonas nematodicola sp. nov., isolated from Caenorhabditis elegans. *Archives of microbiology*, *203*(10), 6197–6202. https://doi.org/10.1007/s00203-021-02586-y

White, F. F., Potnis, N., Jones, J. B., & Koebnik, R. (2009). The type III effectors of *Xanthomonas*. *Molecular plant pathology*, *10*(6), 749–766. https://doi.org/10.1111/j.1364-3703.2009.00590.x

Xie, Z., & Tang, H. (2017). ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics (Oxford, England)*, *33*(21), 3340–3347. https://doi.org/10.1093/bioinformatics/btx433

Young, J. M., Dye, D. W., Bradbury, J. F., Panagopoulos, C. G., and Robbs, C. F. 1978. A proposed nomenclature and classification for plant pathogenic bacteria. N.Z. J. Agric. Res. 21:153-177.

Young J. M. (2008). An overview of bacterial nomenclature with special reference to plant pathogens. *Systematic and applied microbiology*, *31*(6-8), 405–424. https://doi.org/10.1016/j.syapm.2008.09.005

Xiang T, Zong N, Zou Y, Wu Y, Zhang J, Xing W, Li Y, Tang X, Zhu L, Chai J, and Zhou JM. (2008) "Pseudomonas syringae effector AvrPto blocks innate immunity by targeting

receptor kinases". Curr Biol. 18:74–80

Zacaroni, A. B., Koike, S. T., de Souza, R. M., & Bull, C. T. (2012). Bacterial Leaf Spot of Radicchio (*Cichorium intybus*) is caused by *Xanthomonas hortorum*. *Plant disease*, *96*(12), 1820. https://doi.org/10.1094/PDIS-07-12-0672-PDN

Zhou J, Wu S, Chen X, Liu C, Sheen J, Shan L, and He P. (2014) "The Pseudomonas syringae effector HopF2 suppresses Arabidopsis immunity by targeting BAK1." Plant Journal. 77:235–45.

Zipfel, C., & Felix, G. (2005). Plants and animals: a different taste for microbes?. *Current opinion in plant biology*, *8*(4), 353–360. https://doi.org/10.1016/j.pbi.2005.05.004

Appendix A

Supplemental Data

A3 Chapter 3 Supplemental Data

A3-1: Quality control of raw sequence reads using Trimmomatic (trimming.sh).

```
#this code will run through the quality control step for processing raw sequence data, to remove adaptor sequences and reads shorter than 50 nt, as well as groups of four consecutive nucleotides with quality scores lower than 20.

#defines where the script will run and where it will print files to
SCRIPTPATH="$( cd "$(dirname "$0")" ; pwd -P )"

#makes the Raw_Data directory to house your raw sequence data
mkdir -p $SCRIPTPATH/Raw_Data

#moves that data into the new directory
mv $SCRIPTPATH/*.fastq $SCRIPTPATH/Raw_Data

#create a directory for the fastqc reports
mkdir -p $SCRIPTPATH/Raw_Data/fastqc_reports

#generate the qc report for the raw sequence data
for file in $SCRIPTPATH/Raw_Data/*.fastq; do fastqc $file; done

#move all fastqc reports to a single directory
mv $SCRIPTPATH/Raw_Data/*.html $SCRIPTPATH/Raw_Data/*.zip
$SCRIPTPATH/Raw_Data/fastqc_reports

#trim away the poor quality data and output to new folder
for R1 in $SCRIPTPATH/Raw_Data/*R1_001.fastq
do
R2=${R1//R1_001.fastq/R2_001.fastq}
R1paired=${R1//.fastq/_paired.fq}
R1unpaired=${R1//.fastq/_unpaired.fq}
R2paired=${R2//.fastq/_paired.fq} R2unpaired=${R2//.fastq/_unpaired.fq}
trimmomatic PE $R1 $R2 $R1paired $R1unpaired $R2paired $R2unpaired
ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 SLIDINGWINDOW:4:20 MINLEN:50;
done

#make directory for outputs to go into
mkdir -p $SCRIPTPATH/Trimmed_Data

#move trimmed files to Trimmed_Data
```

```
mv $SCRIPTPATH/Raw_Data/*paired.fq $SCRIPTPATH/Trimmed_Data

#generate the qc report for the trimmed sequence data
for file in $SCRIPTPATH/Trimmed_Data/*paired.fq; do fastqc $file; done

#create a directory for these fastqc reports
mkdir -p $SCRIPTPATH/Trimmed_Data/fastqc_reports

#move all the fastqc reports over
mv $SCRIPTPATH/Trimmed_Data/*.html $SCRIPTPATH/Trimmed_Data/*.zip
$SCRIPTPATH/Trimmed_Data/fastqc_reports
```

A3-2: Genome assembly using SPAdes (spades.sh).

```
#this code is to complete a de novo assembly for each strain using spades for paired end
reads.
#Input: trimmed paired end reads. Output: single scaffold.fa assembly file.
#this code is the second step of the WGS procedure, starting with quality control and the
trimming.sh.

#defines where the script will run and where it will print files to
SCRIPTPATH="$( cd "$(dirname "$0")" ; pwd -P )"

#defines variables for the two input files
FILE_1=$1
FILE_2=$2

#Makes the directory for the output files
mkdir -p $SCRIPTPATH/Assemblies

#runs the assembly using spades and upends the results to a new directory within
Assemblies.
spades.py -1 $1 -2 $2 -o $SCRIPTPATH/Assemblies/$(basename ${1%%.fq}) -t 40 -m
64

#Comment out lines 13 and 16 and uncomment the two below one to run plasmid spades
instead of denovo spades.
#mkdir -p $SCRIPTPATH/Plasmid
#spades.py --plasmid -1 $1 -2 $2 -o $SCRIPTPATH/Plasmid/$(basename ${1%%.fq}) -t
40 -m 64
```

A3-3: Script for running SPAdes genome assembly tool in a loop for many samples (spades_loop.txt)

```
#This is the file that will run the WGS_spades_denovo.sh script in a loop
```

```
#Input: all paired end output files of the QCtrimming.sh script; Output: assembly scaffold.fa file
#to run the assembly script, use $ cat WGS_loop.txt | bash

#defines where the script will run and where it will print files to
SCRIPTPATH="$( cd "$(dirname "$0")" ; pwd -P )"

#runs the assembly script in loop for the 25 strains included in this study.
for i in {1..25}; do r1=*_S${i}_L001_R1_001_paired.fq;
r2=*_S${i}_L001_R2_001_paired.fq; bash $SCRIPTPATH/denovo_spades.sh $r1 $r2;
done

#runs the plasmid assembly script in loop - comment out line 10 and uncomment below to run plasmid assembly in a loop.
#for i in {1..27}; do r1=*_S${i}_L001_R1_001_paired.fq;
r2=*_S${i}_L001_R2_001_paired.fq; bash $SCRIPTPATH/plasmid_spades.sh $r1 $r2;
done
```

A3-4: Mining whole genome sequences for effector genes using translated nucleotide BLAST (effectorgene_mining.sh).

```
#effector mining script: input includes protein fasta file of type and pathotype strain effector genes and scaffolds from whole genome sequencing.
#to run change the pident or qcovhsp, alter line 19 and re-run a new output folder listed in lines 23 and 24.
#ouputs:
   #alignments_[filename].txt
   #rawtable_[filename].tsv
   #tidytable_[filename].tsv

# Stop if any error is found
set -ueox pipefail

#defines where the script will run and where it will print files to
SCRIPTPATH="$( cd "$(dirname "$0")" ; pwd -P )"

for file in $SCRIPTPATH/BS*.fasta
do
  makeblastdb -in $file -dbtype nucl -title -parse_seqids -out $(basename ${file//.fasta/"_DB"})
  tblastn -db  $(basename ${file//.fasta/"_DB"}) -query all_effectors_removedups.faa > alignments_$(basename ${file//.fasta/".txt"})
  tblastn -db $(basename ${file//.fasta/"_DB"}) -query all_effectors_removedups.faa -outfmt "6 qseqid sseqid length pident mismatch qstart qend sstart send qcovs qcovhsp evalue bitscore"  > rawtable_$(basename ${file//.fasta/".tsv"})
```

```
  cat rawtable_$(basename ${file//.fasta/".tsv"}) | awk '$4 >= 60 && $11 >= 40 && $12
< 0.00001 {print;}' | sort | uniq > tidytable_$(basename ${file//.fasta/".tsv"})
  cat tidytable_$(basename ${file//.fasta/".tsv"}) | cut -f 1 | uniq | sort >
effectorlist_$(basename ${file//.fasta/".txt"})
done

mkdir -p $SCRIPTPATH/effectormining03242021_pident60_qcovhsp40
mv rawtable* alignments* tidytable* effectorlist*
$SCRIPTPATH/effectormining03242021_pident60_qcovhsp40
```

A3-5: Mining for bacterial secretion system genes using nucleotide blast (SSgene_mining.sh)

```
#secretion gene mining script: input includes protein fasta file of IMG-JGI-listed
secretion system genes and scaffolds from whole genome sequencing project.

#ouputs:
   #alignments_[filename].txt
   #rawtable_[filename].tsv
   #tidytable_[filename].tsv

# Stop if any error is found
  set -ueox pipefail

#defines where the script will run and where it will print files to
SCRIPTPATH="$( cd "$(dirname "$0")" ; pwd -P )"

for file in $SCRIPTPATH/BS*.fasta
do
  makeblastdb -in $file -dbtype nucl -title -parse_seqids -out $(basename
${file//.fasta/"_DB"})
  blastn -db  $(basename ${file//.fasta/"_DB"}) -query
Potnis_secretion_system_genes.fasta > alignments_$(basename ${file//.fasta/".txt"})
  blastn -db $(basename ${file//.fasta/"_DB"}) -query
Potnis_secretion_system_genes.fasta -outfmt "6 qseqid sseqid length qlen pident
mismatch qstart qend sstart send qcovs qcovhsp evalue bitscore" | sort  >
rawtable_$(basename ${file//.fasta/".tsv"})
  cat rawtable_$(basename ${file//.fasta/".tsv"}) | awk '$5 >= 65 && $12 >= 65 && $13
< 0.00001 {print;}' | sort | uniq > tidytable_$(basename ${file//.fasta/".tsv"})
done

mkdir -p $SCRIPTPATH/secretion_genes2021_08_11-65
mv rawtable* alignments* tidytable* $SCRIPTPATH/secretion_genes2021_08_11-65
```

Table A3-6 Type 3 effectors found on plasmids. The outlined box contains *Xhv* strains: columns BP5172 to BP5192 are *Xhv* race 1 strains, ICMP 1408 to BP5194 are *Xhv* race 2 strains, and BS2861 and BP5181 are *Xhv* race 3 strains. Effectors highlighted in green are common to all *Xhv* strains tested. BP5178 is *X. hortorum* from radicchio, CFBP 4925^{T} is *X. hortorum* pv. *hederae*, CFBP 410^{PT} is *X. hortorum* pv. *taraxaci*, CFBP 2533^{PT} is *X. hortorum* pv. *pelargonii*, CFBP 8163^{PT} is *X. hortorum* pv. *gardneri*, CFBP 4188^{PT} is *X. hortorum* pv. *cynarae*, and CFBP 8452^{PT} is *X. campestris* pv. *coriandri*. The effectors *avrHah1*, *xopAQ*, *avrBsT*, and *xopH* are known to be plasmid-borne in other *Xanthomonas* strains. Plasmid sequences were predicted by plasmidSPAdes and used to build sequence databases; translated nucleotide BLAST was used to find effector matches within the plasmid sequences and matches were called at greater than 60% identity, greater than 40% qcovhsp, and less than 0.00001 e-value. The effector gene database was received from Dr. N. Potnis (Auburn University).

BP5172	BS0339	BS0340	BS0347	BP5176	BP5177	BP5179	BP5182	NCPPB 4054	CFBP 8686^{T}	BP5191	BP5192	ICMP 1408	ICMP 4165	BS3127	BP5194	BS2861	BP5181	BP5178	CFBP 4925^{T}	CFBP 410^{PT}	CFBP 2533^{PT}	CFBP 8163^{PT}	CFBP 4188^{PT}	CFBP 8452^{PT}
AvrBs1	AvrBs1	AvrBs1	AvrBs1	AvrBs1	AvrBs1	AvrBs1	AvrBs1	AvrBs1	AvrBs1	AvrBs1	AvrBs1	AvrBs1	AvrBs1	AvrBs1	AvrBs1	AvrBs1	AvrBs1	AvrBs1		AvrBs1			AvrBs1	
				AvrBs2		AvrBs2	AvrBs2		AvrBs2								AvrBs2	AvrBs2		AvrBs2				
	XCC4186_LRR			XCC4186_LRR																				
avrHah1	avrHah1	avrHah1	avrHah1	avrHah1	avrHah1	avrHah1	avrHah1	avrHah1	avrHah1	avrHah1	avrHah1	avrHah1	avrHah1	avrHah1	avrHah1	avrHah1	avrHah1			avrhah1	avrHah1	avrHah1	avrHah1	avrHah1
avrXccB		avrXccB	avrXccB		avrXccB				avrXccB		avrXccB	avrXccB	avrXccB	avrXccB	avrXccB	avrXccB	avrXccB							
				xopAQ					xopAQ												xopAQ			xopAQ
	xopB					xopB	xopB								xopB					xopB				
	xopC					xopC	xopC								xopC			xopC		xopC				
xopD		xopD	xopD	xopD		xopD	xopD	xopD																
xopE1	xopE1	xopE1	xopE1	xopE1	xopE1	xopE1	xopE1	xopE1	xopE1	xopE1	xopE1	xopE1	xopE1	xopE1	xopE1	xopE1	xopE1					xopE1		xopE1
xopE2	xopE2	xopE2	xopE2	xopE2	xopE2	xopE2	xopE2	xopE2	xopE2	xopE2	xopE2	xopE2	xopE2	xopE2	xopE2	xopE2	xopE2					xopE2		xopE2
		xopG																						
xopH–avrB	xopH–avrBs1.1	xopH–avrB	xopH–avrB	xopH–avrBs1.1	xopH–avrB	xopH–avrB	xopH–avrB	xopH–avrB	xopH–avrB	xopH–avrB	xopH–avrB	xopH–avrB	xopH–avrB	xopH–avrB	xopH–avrB	xopH–avrB	xopH–avrB	xopH–avrBs1.1		xopH–avrBs1.1			xopH–avrBs1.1	
						xopZ2																		
																		HpaFprotein						
																		avrA_family_Xg101						
																		xopAE						
Known plasmid-borne																						xopAO	xopAO	
avrHah1																					early chlorosis factor			
xopAQ																					holPsyAE/Xoo3022			
avrBsT																					xopAI_XAC3230			
xopH																								hopPmaA
																								AvrBsT
																								HopAJ1
																			xopN					

A4 Chapter 4 Supplemental Data

Table A4-1 Additional primers tested for their specificity to *Xhv* and each of the *Xhv* races. These primer sets were designed to target gene clusters that were shown by whole genome sequence alignments to be *Xhv* pathovar- or race-specific. The 25 µL PCRs consisted of 1X ImmoMix (Meridian Bioscience, Cincinnati, OH), 0.5 uM of each forward and reverse primer, and 1 µL of colony suspensions. The cycling steps included ten minutes at 95 °C, thirty seconds at 95 °C, one minute at 58 °C, one minute at 72 °C, cycling thirty-two times to step two, and six minutes at 72 °C. Most these primers produced the predicted fragments in non-target strains, including related *Xanthomonas* strains: *X. hortorum* pv. *hederae* (*Xhh*), *X. hortorum* pv. *gardneri* (*Xhg*), *X. hortorum* pv. *cynarae* (*Xhcy*), *X. hortorum* pv. *pelargonii* (*Xhp*), *X. hortorum* pv. *taraxaci* (*Xht*), and *X. campestris* pv. *coriandri*.

Gene Cluster	Forward Primer 5' to 3'	Reverse Primer 5' to 3'	Amplicon Size	Target Strains	Result	Lab Notebook Reference
GC3950	AACGCGGACT GTATTGACCA	TAGTCTTGCG TCGTGTCGTC	120	*Xhv* only	*Xhv* race 1, 2, and 3; multiple bands in *Xhh* and *Xhg*	ERR3-10
GC3950	ACGACACGAC GCAAGACTAC	AAGTGCCACA GCCTTCTCAG	254	*Xhv* only	*Xhv* race 1, 2, and 3; *X campestris* pv. *coriandri*	ERR3-10, 13, 15, 18-20, 26, 28
GC5699	AGCGCCACTTT AGCTCACTT	TCTGCTTGGC GCTAGATACG	388	*Xhv* race 3 only	*Xhv* race 3 strains and *Xhv* race 1 strains from Hawaii; *Xht*, and *X. hortorum* from radicchio	ERR3-10, 13, 15, 18-20, 26, 28; ERR4-6
GC5699	ATTAGGTGCTT GGGTGAGGC	TGACCTGTGC GGGAGTTTTT	191	*Xhv* race 3 only	Many bands in all *X. hortorum* strains; bright band in *Xhv* race 3 strain.	ERR3-10
GC5699	CCAGCCTTAGC GCCACTTTA	CCGTCTCGCG CTTACTAGAC	219	*Xhv* race 3 only	*Xhv* race 3	ERR3-10
GC3879	CAAGAGAAGA CGAGGAGCGT T	CTCAGGCGCA GGTGTAAGAC	97	*Xhv* only	*Xhv* race 1, 2, and 3; *Xht* and *Xhcy*	ERR4-8

GC3887	GCGGTACAAG AAGGCTGACT	TCTTCGGGGT CAAGCAACTC	429	*Xhv* only	*Xhv* races 1, 2, and 3; *Xhcy* and *X. campestris* pv. *coriandri*	ERR4-5
GC3980	ATGGCTACTAC GCACAATCCA CTC	GGGCCCACAC CAACAGAAT G	208	*Xhv* race 1 only	*Xhv* races 1 and 3; *Xhg*, *Xhcy*, and *X. campestris* pv. *coriandri*	ERR4-5
GC4381	GAAAGTGTGC GACATGACCG	CAAGGTTGCG GCACTTTGTT	351	*Xhv* race 2 only	*Xhv* race 2; *X. campestris* pv. *coriandri**	ERR4-4
GC4381	TCAAGGACCA TGACTTCGGC	CGTATTGCGG TGCGAACTTT	309	*Xhv* race 2 only	*Xhv* race 2; *X. campestris* pv. *coriandri**	ERR4-4
GC4980	CCCAACACCG ATAAGGCAGT	TGCACTGGTG TAGTAAGCCG	318	*Xhv* race 3 only	*Xhv* race 3; *Xht**	ERR4-4
GC4980	ACTCAAAAGC CCACCCTCAG	AGTCGAGCA ATGGAAGAG CC	295	*Xhv* race 3 only	*Xhv* race 3; *Xht**	ERR4-4
GC4980	CTTCCATTGCT CGACTTGCG	GGCTTTGTAG GCGGGTCATA	245	*Xhv* race 3 only	*Xhv* race 3; *Xht**	ERR4-4
GC4980	CTTCCATTGCT CGACTTGCG	TGCACTGGTG TAGTAAGCCG	369	*Xhv* race 3 only	*Xhv* race 3; *Xht, X. hortorum* from radicchio, *X. campestris* pv. *coriandri*	ERR4-8
GC3997	TTGAGCAGGA AGCGTATCGG	CTGTGACTGG AAGGTGCCAT	83	*Xhv* only	*Xhv* races 1 and 3; *Xhcy*, *Xht*, *X. campestris* pv. *coriandri*	ERR4-11
GC3887	CAGCCTCAAA TCCCCGCTTA	GATAACCTCC CAGACCGCTG	265	*Xhv* only	*Xhv* races 1 and 2, *Xhcy*, *Xht*, *X. hortorum* from radicchio (some dye smearing, but bright bands visible)	ERR4-9
GC3887	GGTCGCTTTCG TTTACGGTG	GGACTGCAGC CATTCGTAGA	83	*Xhv* only	Inconclusive due to dye smearing, but some visible bands for *Xhv* race 2, *Xhg*, *Xhcy*, *X. hortorum* from radicchio	ERR4-9

GC3889	GCCAGGCCTA TGGACTCAAG	ATGGTCGTTG GTGAGCATGA	110	*Xhv* only	*Xhv* races 1, 2, and 3; *Xhcy*, *Xht*, *X. campestris* pv. *coriandri*	ERR4-11
GC4021	TTAGCAAAGG CGCTGGACTT	GCAAGCCCTT CACAAGGTTC	133	*Xhv* race 1 only	*Xhv* race 1; 1 kb band for *Xht*	ERR4-11
GC4021	CTTTCATGCGA CGTCAGACC	TGGGACACCA CAAGAATCAC C	51	*Xhv* race 1 only	*Xhv* race 1	ERR4-11

The asterisk refers to primer sets that were only evaluated in silico using our anvi'o alignment and NCBI's Primer BLAST, with primer pair specificity checking with the nonredundant protein database and the organism set to '*Xanthomonas hortorum'*.

A4-2 Progress toward qPCR development for *Xhv* pathovar- and race-specific detection.

Rationale: Chapter 4 details our progress in the development of endpoint, conventional detection methods for *Xhv* pathovar- and race- specific detection. Additional qPCR methods are in development that offer the following advantages over conventional methods: 1) results can be observed more rapidly and in real time, negating the need for a gel electrophoresis setup in addition to the PCR thermocycler, 2) fluorescence-based detection provides greater sensitivity 3) the number of target cells present in an unknown sample can be calculated and used to determine the minimum level of *Xhv* cells necessary to cause infection and 4) probe-based qPCR can be used to facilitate multiplexing reactions.

Methods: We began screening our conventional PCR primer sets using SYBR green technology to evaluate their specificity, including only those that met the qualifications for qPCR primers such as an amplicon size of under 200 nucleotides. This is important because with longer amplicons, the reaction efficiency may drop so low as to render any attempts at quantification inaccurate (Svec et al., 2015); the targeted reaction efficiency should be 90-110%. The efficiency may be slightly greater than 100% with the absence of inhibitors in the reaction matrix. Our goal was to determine whether any of the primers that we had already developed could be useful for *Xhv* pathovar- or race-specific qPCR detection.

The 20 µL reaction mixes contained 7.0 µL of sterile water, 10 µL of 2x SsoAdvanced Universal SYBR Green Supermix (which includes the SYBR Green I dye that would bind to any double-stranded DNA, ROX normalization dyes, dNTPs, MgCl2, and Sso7d fusion polymerase; Bio-Rad, Hercules, CA, United States), 1 µL of each 10 µM primer (forward and reverse), and 1 µL of DNA extract (30 ng/µL). The Sso7d fusion polymerase was chosen due to its ability to

overcome PCR inhibitors and therefore have higher reaction efficiency than other polymerases; its use was deemed necessary after reactions with the standard Taq polymerase generated poor, noisy results in a preliminary trial (data not shown). The cycling steps included three minutes at 98 °C, ten seconds at 98 °C, thirty seconds at 60 °C, and cycling thirty-nine times to step two. A melting curve was drawn by ramping up from 60 °C to 98 °C and measuring the fluorescence of SYBR green at 0.5 °C intervals. Amplification and melting curves were analyzed using CFX Maestro software from Bio-Rad. DNA standards were run for each reaction at concentrations of $5E^{-4}$, $5E^{-5}$, $5E^{-6}$, and $5E^{-7}$ ng/μL, serially diluted from an original concentration of 5 ng/μL, and with DNA amplicons from *Xhv* race 1 strain BS0347 being used for the *Xhv* pathovar- and race 1-specific reactions, from *Xhv* race 2 strain BS2995 being used for the *Xhv* race 2-specific reaction, and from *Xhv* race 3 strain BS2862 being used for the *Xhv* race 3-specific reaction. These amplicons were prepared using touchdown PCR as described in Chapter 4, and were subsequently cleaned using EXO-SAPIT, quantified using a Qubit 3.0 fluorometer and the Invitrogen dsDNA Broad Range Assay Kit, and adjusted to 5 ng/μL and then diluted in series using Ultrapure Distilled Water (Thermo Fisher Scientific/Invitrogen, Waltham, MA, USA).

Results and Discussion:

***Xhv* strains can be distinguished from other *X. hortorum* strains via SYBR Green qPCR**

In qPCR, the C_T value (sometimes referred to as Cq value) represents the cycle at which the accumulation of fluorescent signal due to DNA amplification surpasses that of background fluorescence levels; using templates with a greater number of transcripts (more total DNA) would lower the C_T value as it would take less time for the amplification curve to reach the exponential growth phase (Heid et al., 1996). All of our templates were run with the same

amount of DNA to enable our comparison of the C_T values and make inferences about the relative abundances of our targets. The *Xhv* strains had low C_T values: BS0347 (C_T =15), BS2995 (C_T =15), and BS2862 (C_T =20), indicating the abundant presence of amplification target (Figure A4-2-1A). The related *Xanthomonas* strains had higher C_T values, all at around C_T =27, which suggested that either the target amplicon was also present in the template, though in much lower quantity, or that there was non-specific amplification. The negative control saw amplification similar to that of the *Xanthomonas* controls.

The R^2 value of the standard curve represents the fit of a linear regression relating the starting template concentration and the C_T value; an R^2 value > 0.99 would indicate a strong correlation between these two values (La Rosa et al., 2010). For the *Xhv*-specific reaction, the R^2 value was 0.996 (Figure 4-2-1B), indicating a strong correlation between the starting template concentration and the C_T value, and thus the C_T value may be used to accurately predict the concentration of the starting template for each unknown sample (copy number = [amount of DNA (ng) x 6.022 x 1023]/[amplicon length (bp) x 109 x 660]). These values are shown in Table A4-2-1, with nearly 740,000 copies in the starting template for the *Xhv* race 1 and 2 strains, each, and nearly 33, 200 copies in the starting template for *Xhv* race 3 strain. There were about 450 copies in each of the starting templates for the non-target strains and for the negative control. This preliminary data demonstrates that the GC3889 amplicon targeted for touchdown PCR (Chapter 4) may be useful for distinguishing *Xhv* from the other members of *X. hortorum* by qPCR given an unknown sample.

There are some changes that should be incorporated into future iterations of this experiment. One should undertake additional efforts to eliminate the contamination that was present in our negative control, and possibly our target and non-target strains, and this can be

accomplished by using all freshly prepared reagents, preparing the reactions in a non-bacteriology lab to reduce the risk of contamination by bacterial aerosols in our laboratory, and using 10% bleach to sterilize the new work environment and all appropriate tools. Next, our reaction efficiency was lower than desired at 85% (target 90-110%; La Rosa et al., 2010), which may suggest the occurrence of non-specific primer binding or the overconcentration of DNA in the unknown samples. These issues may be resolved by increasing the annealing/extension temperature in the reaction cycles or testing dilutions of the gDNA to find which provides the best reaction efficiency. Additionally, other gene cluster targets should be investigated as targets for amplification due to the discovery (Chapter 4) that the GC3889 primers can also detect *X. hydrangeae* from lettuce samples. One should incorporate *X. hydrangeae* whole genome sequence into the *X. hortorum* and related strain alignment, again search for *Xhv*-specific gene clusters, and develop and test primers to target that new sequence. Lastly, pathogenicity tests on susceptible lettuce lines using various concentrations of *Xhv* should be completed to determine the minimum bacterial concentration necessary to produce disease. This would help to establish a cutoff C_T value for this qPCR that one would expect to see in tandem with the appearance of BLS symptoms.

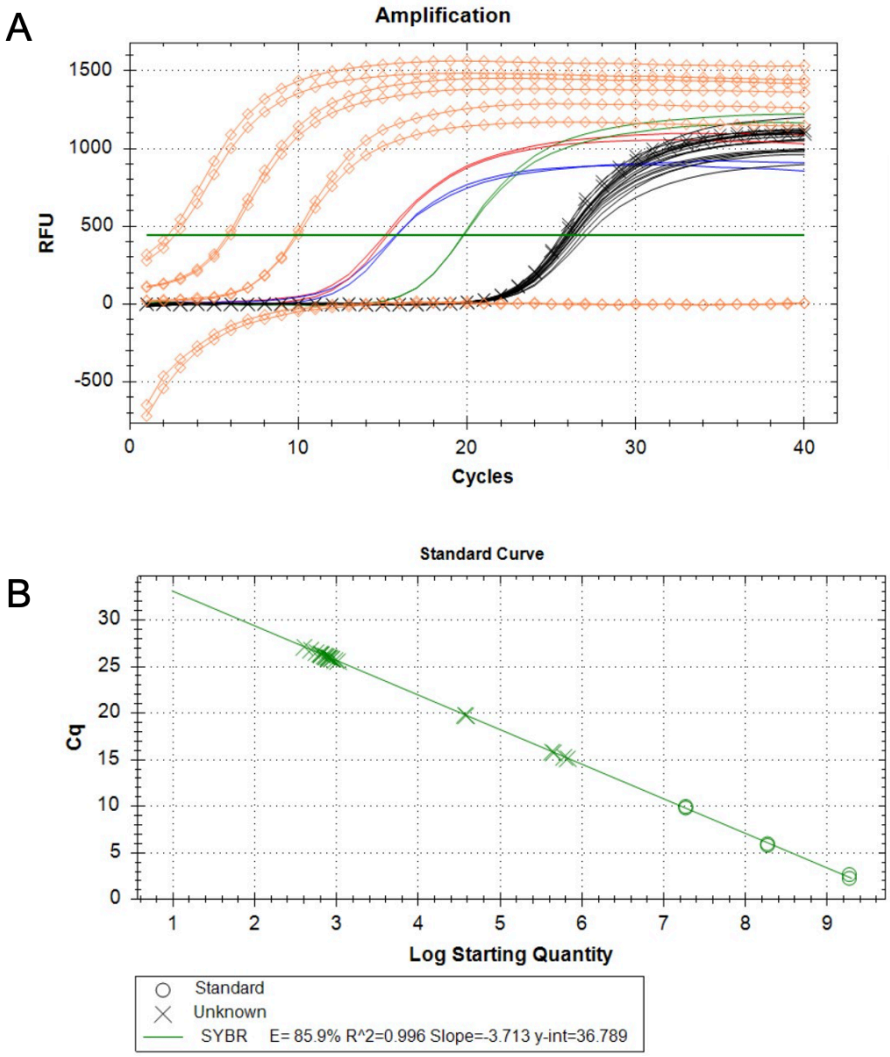


Figure A4-2-1. Real-time PCR (SYBR Green) amplification (A) and standard (B) curves for *Xhv*-specific detection. (A) All amplifications were completed in duplicate. Amplification of the dilution series of the GC3889 amplicon described in Chapter 4 is shown as orange lines with circles. Amplification of *Xhv* genomic DNA is shown in color: race 1 BS0347 (blue), race 2 BS2995 (red), and race 3 BS2862 (green). Amplification of the genomic DNA for non-target, related strains are in black straight lines, and include *X. hortorum* pvs. *gardneri* (BS3600), *cynarae* (BS3601), *pelargonii* (BS3110), *taraxaci* (BS3108), *hederae* (BS3107), *carotae* (BS3785); *X. hortorum* isolated from radicchio (BS0953); and *X. campestris* pv. *coriandri* (BS3602). The negative control is shown as black lines with 'X' marks. **(B)** The 'X' marks here

refer to the reactions run with the gDNA of *Xanthomonas* strains. The circles refer to the dilution series of the standard.

Table A4-2-1. Predicted copy numbers of GC3889 amplicon, designed for *Xhv*-specifc detection, present in *Xanthomonas* templates based on C_T value from qPCR. All templates used in qPCR were adjusted to 30 ng/μL. Three *Xhv* strains representing the three races were tested as targets, along with non-target strains from *X. hortorum* pv. *gardneri* (*Xhg*), *X. hortorum* pv. *cynarae* (*Xhcy*), *X. hortorum* pv. *hederae* (*Xhh*), *X. hortorum* pv. *pelargonii* (*Xhp*), *X. hortorum* pv. *taraxaci* (*Xht*), *X. hortorum* pv. *carotae* (*Xhca*), *X. hortorum* from radicchio, and *X. campestris* pv. *coriandri*.

Strain(s)	C_T value	Copy number in original template*
Xhv BS0347 (race 1) and BS2995 (race 2)	15	7.38×10^5
Xhv BS2862 (race 3)	20	3.32×10^4
Xhg, *Xhcy*, *Xhh*, *Xhp*, *Xht*, *Xhca*, X. *hortorum* from radicchio, *X. campestris* pv. *coriandri*	27	4.33×10^2

*Copy number (CN) was calculated using the formula: CN = 10^([C_T value – y-intercept]/slope), with the y-intercept and slope derived from the standard curve.

Distinguishing the races of *Xhv* using SYBR Green qPCR

The *Xhv* race 1, 2 and 3 primer sets described for touchdown PCR in Chapter 4 were also useful in the development of a SYBR Green qPCR protocol. For the *Xhv* race 1 qPCR reaction, the race 1 BS0347 strain had a lower C_T value than all the other strains, at C_T =16. The non-

target strains, including the other *Xhv* strains and the set of related *Xanthomonas* strains, all resulted in values of C_T =26 or higher (Figure A4-2-2A). Again, the negative control saw amplification comparable to the non-target strains, but these preliminary results suggest a method could be developed for detecting *Xhv* race 1 strains from unknown samples based on their low C_T value.

For the *Xhv* race 1 qPCR reaction, the R^2 value was lower than 0.99 (at 0.932; Figure 4-2-2B), which indicates that there is not a strong relationship between the starting copy number and the C_T value. This limits our ability to predict the starting copy number of DNA from these templates. This is likely due to inaccurate concentrations in the dilution standards, which can be improved by careful pipetting. Additionally, the reaction efficiency is calculated at 217.5%; a number this high indicates inhibition of the DNA polymerase. This can occur due to overconcentration of the DNA or contamination in the starting template and can be remedied by using cleaned (perhaps using an additional procedure/kit to remove inhibitors often found in extractions from plant tissue) and diluted DNA as template. A dilution series of the gDNA templates would likely show what DNA concentration is necessary for an optimal reaction efficiency (90-110%).

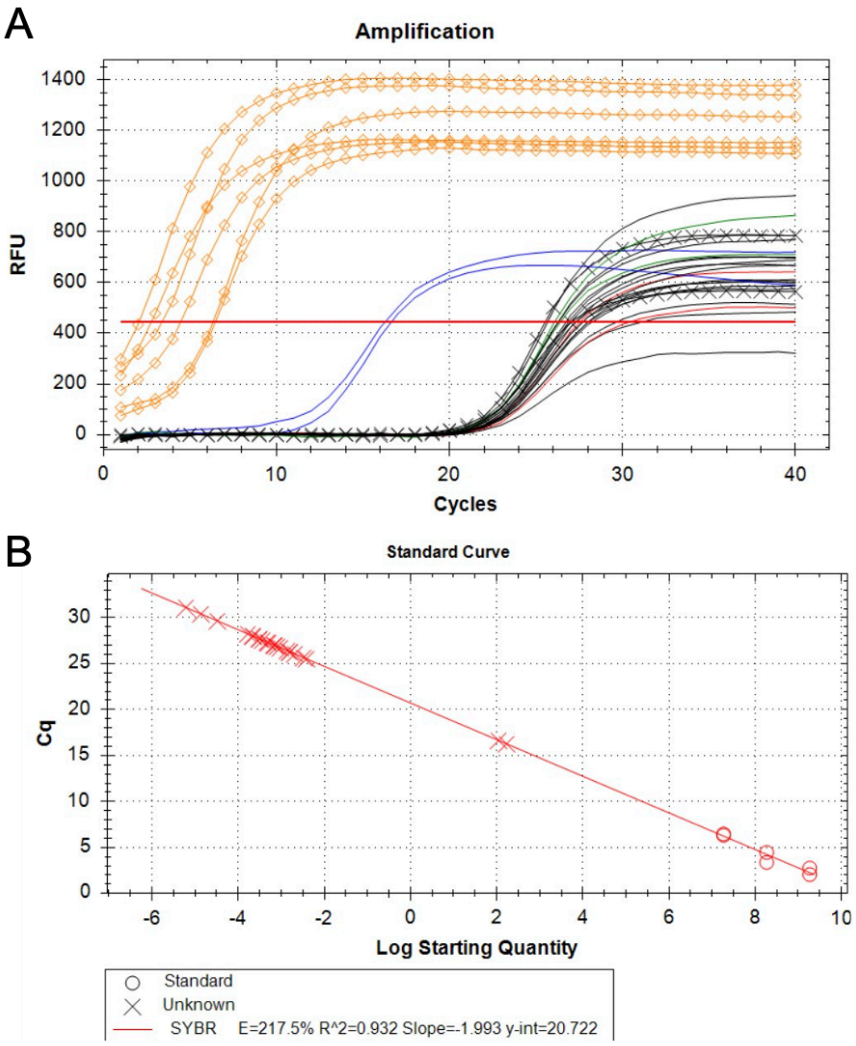


Figure A4-2-2. Real-time PCR (SYBR Green) amplification (A) and standard (B) curves for *Xhv* race 1-specific detection. (A) All amplifications were completed in duplicate. Amplification of the dilution series of the GC4021 amplicon described in Chapter 4 is shown as orange lines with circles. Amplification of *Xhv* genomic DNA is shown in color: race 1 BS0347 (blue), race 2 BS2995 (red), and race 3 BS2862 (green). Amplification of the genomic DNA for non-target, related strains are in black straight lines, and include *X. hortorum* pvs. *gardneri* (BS3600), *cynarae* (BS3601), *pelargonii* (BS3110), *taraxaci* (BS3108), *hederae* (BS3107), *carotae* (BS3785); *X. hortorum* isolated from radicchio (BS0953); and *X. campestris* pv.

coriandri (BS3602). The negative control is shown as black lines with 'X' marks. **(B)** The 'X' marks here refer to the reactions run with the gDNA of *Xanthomonas* strains. The circles refer to the dilution series of the standard.

The *Xhv* race 2-specific reaction produced similar results. The C_T value for the *Xhv* race 2 strain BS2995 was C_T =15, which was much lower than the other *Xhv* strains or the set of other related *Xanthomonas* strains at C_T =28 (averaged). The negative control again saw amplification similar to the non-target strains, but the difference we see between *Xhv* race 2 strain C_T values and all others tested may be useful for *Xhv* race 2-specific detection.

Again, several procedural improvements are necessary. The R^2 value of 0.896 indicates a weak relationship between the starting copy number of DNA and the C_T value, meaning we cannot accurately calculate the starting copy number from this data. The same recommendations for improving R^2 value for the *Xhv* race 1 reactions are appropriate here. The reaction efficiency is acceptable because it (E = 112%) is close to the target of 90-110%.

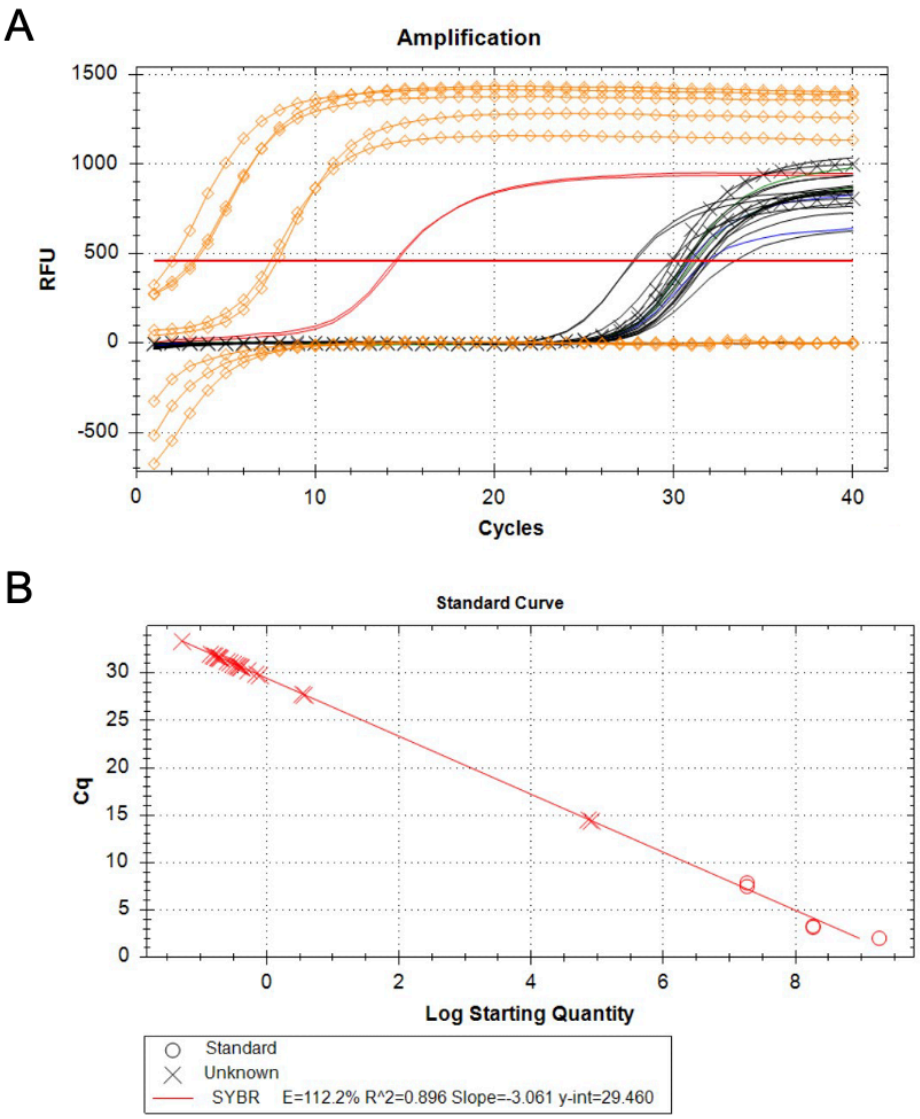


Figure A4-2-3. Real-time PCR (SYBR Green) amplification (A) and standard (B) curves for *Xhv* race 2-specific detection. (A) All amplifications were completed in duplicate. Amplification of the dilution series of the GC4381 amplicon described in Chapter 4 is shown as orange lines with circles. Amplification of *Xhv* genomic DNA is shown in color: race 1 BS0347 (blue), race 2 BS2995 (red), and race 3 BS2862 (green). Amplification of the genomic DNA for non-target, related strains are in black straight lines, and include *X. hortorum* pvs. *gardneri* (BS3600), *cynarae* (BS3601), *pelargonii* (BS3110), *taraxaci* (BS3108), *hederae* (BS3107), *carotae* (BS3785); *X. hortorum* isolated from radicchio (BS0953); and *X. campestris* pv.

coriandri (BS3602). The negative control is shown as black lines with 'X' marks. **(B)** The 'X' marks here refer to the reactions run with the gDNA of *Xanthomonas* strains. The circles refer to the dilution series of the standard.

Finally, the *Xhv* race 3-specific reactions were similar to those of the *Xhv* race 1 and 2 reactions. The *Xhv* race 3 strain BS2862 was distinguishable from the other *Xhv* strains and the related *Xanthomonas* strains by its lower C_T value of 19; all others had an average C_T value of 25. The negative control produced amplification and thus warrants the repetition of the experiment to eliminate contamination, but the results thus far indicate a possible method for *Xhv* race 3-specific detection by C_T value.

The same improvements to the reactions described above are recommended here. The R^2 value of 0.767 is very low (target > 0.99) and indicates a weak relationship between the starting copy number and the C_T value, so this data cannot be used to predict the starting copy number from the templates. The reaction efficiency was calculated to 151.5%, suggesting the presence of polymerase inhibitors. These issues can be remedied in the same ways as described above for the *Xhv* race 1 and 2 reactions (described above).

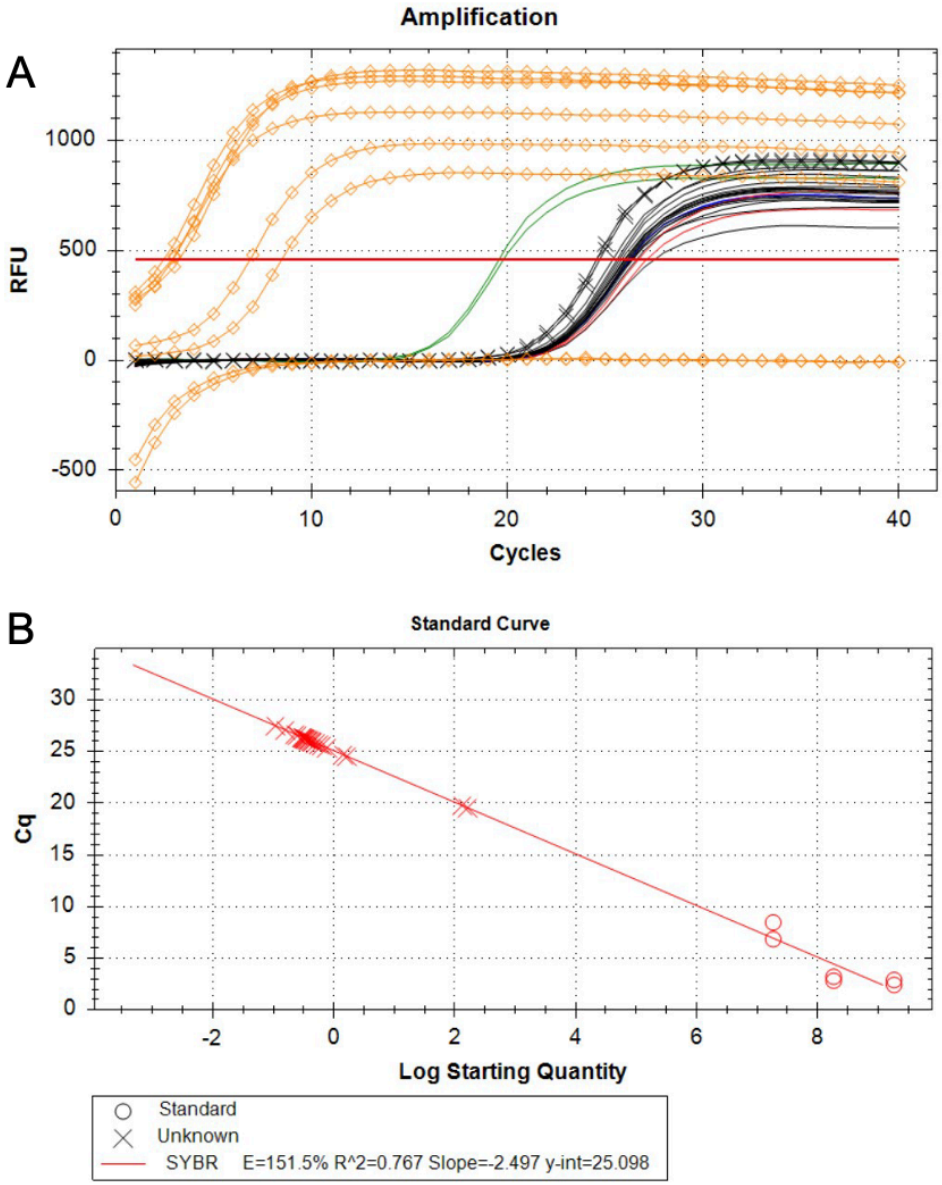


Figure A4-2-4. Real-time PCR (SYBR Green) amplification (A) and standard (B) curves for *Xhv* race 3-specific detection. (A) All amplifications were completed in duplicate. Amplification of the dilution series of the GC4980 amplicon described in Chapter 4 is shown as orange lines with circles. Amplification of *Xhv* genomic DNA is shown in color: race 1 BS0347 (blue), race 2 BS2995 (red), and race 3 BS2862 (green). Amplification of the genomic DNA for non-target, related strains are in black straight lines, and include *X. hortorum* pvs. *gardneri* (BS3600), *cynarae* (BS3601), *pelargonii* (BS3110), *taraxaci* (BS3108), *hederae* (BS3107),

carotae (BS3785); *X. hortorum* isolated from radicchio (BS0953); and *X. campestris* pv. *coriandri* (BS3602). The negative control is shown as black lines with 'X' marks. **(B)** The 'X' marks here refer to the reactions run with the gDNA of *Xanthomonas* strains. The circles refer to the dilution series of the standard.

Progress Summary and Next Steps:

These preliminary results revealed that *Xhv* pathovar- and race-specific detection may be achieved via SYBR Green qPCR, should one troubleshoot the protocols describe above. If starting templates are always adjusted to 30 ng/μL, then one could consider a cutoff of $C_T < 20$ for *Xhv* pathovar-or race-specific detection because in every case, the non-target strain amplification occurred at $C_T \geq 25$. However, qPCR cutoffs are typically $C_T > 30$ or 35, so reducing the gDNA concentration in the starting template, which would increase all the C_T values, could allow for the use of a more widely accepted cutoff. Several procedural improvements are necessary and listed below:

- To remove contamination from the negative control – Use fresh aliquots of ultrapure water and all reagents, prepare reactions in a bleach-sterilized hood with sterile pipets, practice careful sterile technique.
- To improve the correlation coefficient between copy number and C_T value – improve the accuracy of the dilution standards for the race-specific reactions through careful pipetting.
- To improve the reaction efficiency and establish a C_T value cutoff– use a dilution series of the gDNA templates to determine what concentration provides the target reaction efficiency and C_T values (90-110%).

- To reduce off-target amplification – increase the annealing/extension temperature.

These steps should be completed to troubleshoot these reactions, and then suspect *Xhv* samples should be evaluated to confirm the specificity of the reactions and their usefulness as detection methods. The establishment of a set of *Xhv* pathovar- and race-specific qPCR protocols would provide all the same benefits offered by our touchdown PCR method, but with the added advantages of being faster, more sensitive, and able to quantify the number of starting copies present in a given template. This would be beneficial to seed companies, who could then screen their seed lots more quickly, clinicians, who could then evaluate suspected BLS samples and provide grower recommendations more rapidly, and researchers who could then study new aspects of *Xhv* pathogenesis on lettuce, such as the bacterial load necessary to produce disease and whether different races of *Xhv* can be present in a mixed infection.

A5 Chapter 5 Supplemental Data

A5-1 Attempted creation of deletion constructs.

We attempted to create the *xopAQ* and *xopAF2* scarless deletion constructs in our lab but were unsuccessful. The constructs we used in Chapter 5 were designed in-silico and then synthesized by Synbio Technologies Inc. (Monmouth, NJ). Here, we show our strategy for generating the constructs, evidence of the amplification of the regions flanking the effector genes, the creation of the *xopAQ* fusion fragment, and the attempted creation of the *xopAF2* fusion fragment. The reaction to make the *xopAQ* fusion fragment with complementary overhangs to the multicloning site of the plasmid vector did not result in any amplification visible by gel electrophoresis. The reaction to generate the *xopAF2* fusion fragment was also unsuccessful. Many repetitions of the failed reactions and redesigns of the primers did not change our results. Eventually, we decided that synthesis of both deletion and expression vectors through Synbio Technologies would be a more efficient strategy.

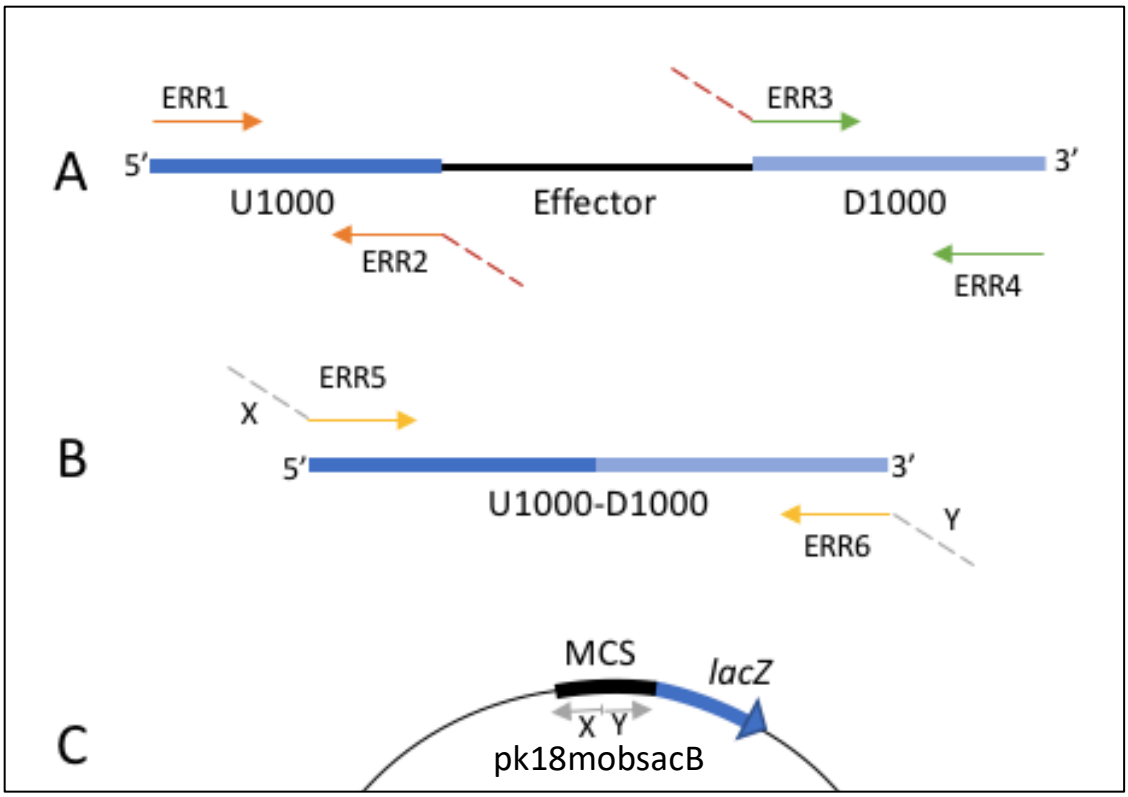


Figure A5-1-1 Strategy for creating a scarless deletion construct. (A) The approximately 1 kb regions upstream (U1000) and downstream (D1000) to the effector gene, *xopAQ* or *xopAF*, were amplified from *Xhv* race 1 strain BS0347 using primers ERR1/ERR2 and ERR3/ERR4, respectively. Primers ERR2 and ERR3 were designed to have complementary sequence overhangs; in a PCR with the first two products (U1000 and D1000) and these two primers, the fusion fragment U1000-D1000 was created. **(B)** The fusion fragment was amplified using primers ERR5 and ERR6, each with overhangs that were designed to be complementary to **(C)** adjacent sequences in the multicloning site of the non-replicative plasmid pk18mobsacB (Kan^R). This would produce a similar scarless deletion construct to the one seen in Figure 5-1.

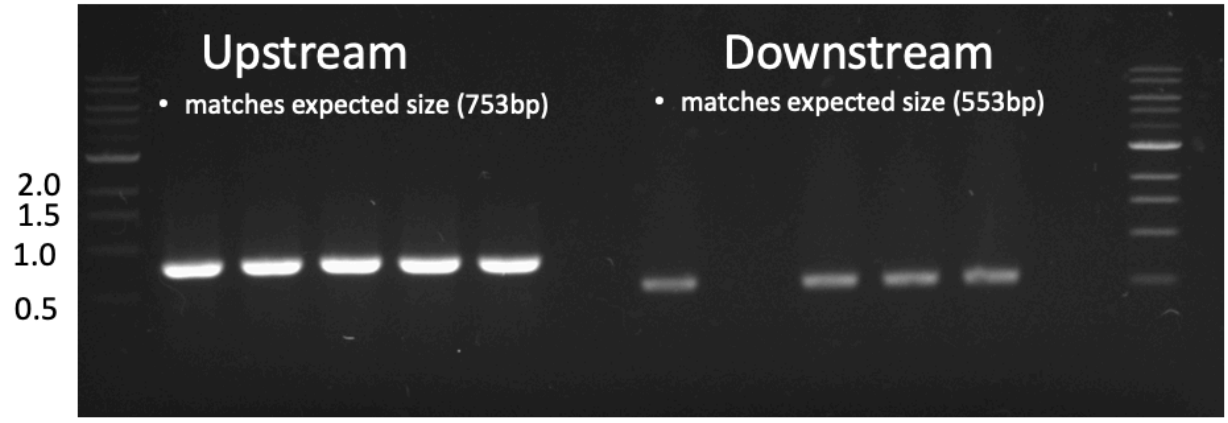


Figure A5-1-2 Amplification of the flanking regions to *xopAQ* from *Xhv* race 1 strain BS0347. The first and last lanes contained a 1 kb DNA ladder from New England Biolabs, with measurements listed on the left side in kbps. Five replicates were run for each flanking region; single bands of the expected sizes, 753 bp and 553 bp, respectively, were seen for all five of the upstream and four of the downstream replicates. Sterile water controls produced no amplification (lanes 7 and 13).

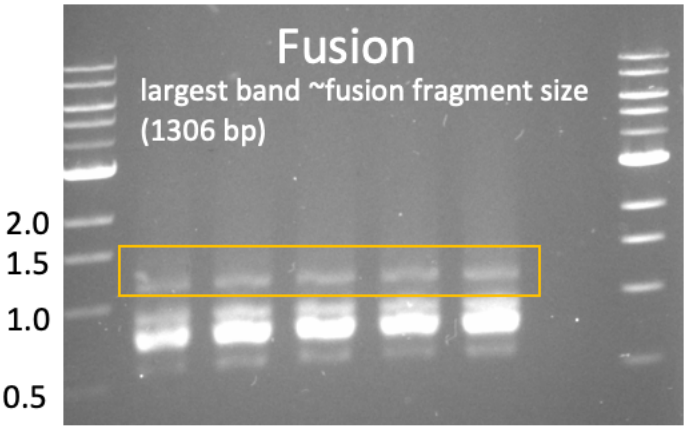


Figure A5-1-3 Creation of the fusion fragment of the *xopAQ* scarless deletion construct, from *Xhv* race 1 strain BS0347. The first and last lanes contained a 1 kb DNA ladder from New England Biolabs, with measurements listed on the left side in kbps. Five replicates of the reaction described in Figure A5-1-1B were run, and bands shown likely represented the upstream

and downstream fragments (753 bp and 533 bp, respectively), as well as the fusion fragment (1306 bp, in yellow box). The sterile water control produced no amplification (lane 7).

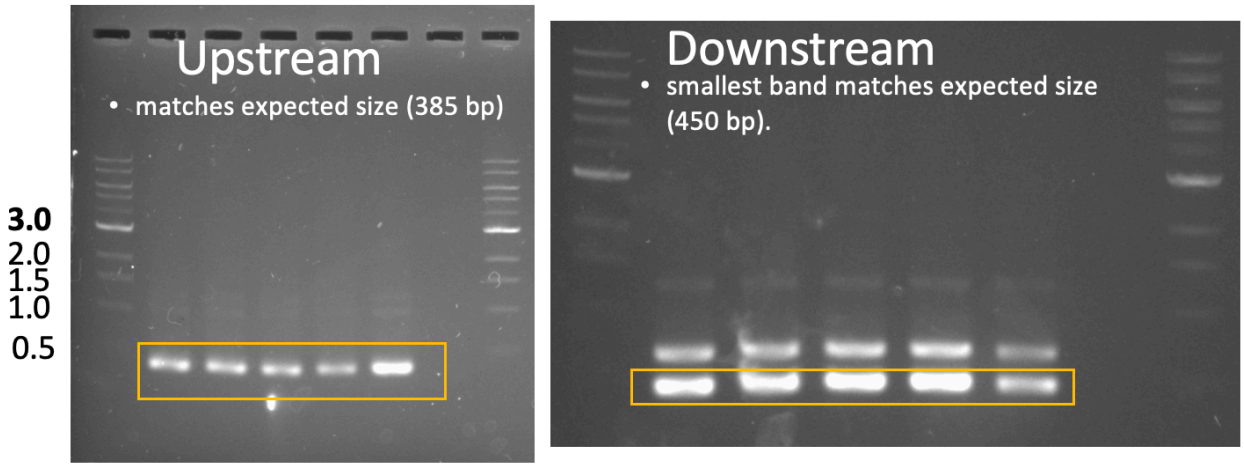


Figure A5-1-4 Amplification of the flanking regions to *xopAF2* from *Xhv* race 1 strain BS0347. The first and last lanes of each gel contained a 1 kb DNA ladder from New England Biolabs, with measurements listed on the left side in kbps. Five replicates were run for each flanking region; bands of the expected sizes, 385 bp and 450 bp, respectively, were seen for all five of the upstream and downstream replicates (in yellow box). The downstream fragment amplifications resulted in two larger bands as well. Sterile water controls produced no amplification (lane 7 in each gel).

Figure A5-1-5 Attempted creation of the fusion fragment of the *xopAF2* scarless deletion construct, from *Xhv* race 1 strain BS0347. The first lane contained a 1 kb DNA ladder from New England Biolabs, with measurements listed on the left side in kbps. Five replicates of the reaction described in Figure A5-1-1B were run, and while two of the band might represent the two flanking regions (385 bp and 450 bp), none of the bands are at the expected size for the fusion fragment (835 bp). The sterile water control produced no amplification (lane 7).

Appendix B

B1-1 First Report of Black Rot Caused by *Xanthomonas campestris* on Arugula in California.

E. R. Rosenthal, L. Ramos Sepulveda, and C. T. Bull, The Pennsylvania State University, University Park, 16802; and S. T. Koike, University of California Cooperative Extension, Salinas 93901.

Arugula (*Eruca sativa*) is a leafy crucifer that is grown in coastal California as a fresh market commodity in bagged salad mixes. In January 2016, a foliar disease developed on commercially produced arugula in the Salinas Valley (Monterey County). Initial leaf symptoms consisted of chlorotic lesions with diffuse margins and black veins; lesions tended to originate from leaf margins and have an angular V shape. As the disease progressed, lesions became tan and dry, resulting in a papery leaf texture. Disease incidence in the affected field reached approximately 5%; however, owing to quality standards, sections of the crop that exhibited any symptoms were not harvested, resulting in greater than 5% loss. Bright yellow, mucoid bacterial colonies were recovered when surface sterilized, macerated leaf tissue was streaked onto sucrose peptone agar. Five representative isolates originally from five different plants produced yellow mucoid colonies on yeast extract dextrose calcium carbonate agar and were gram negative, oxidase negative, nonfluorescent on King's Medium B agar, and positive for starch hydrolysis on SX agar (Schaad and White 1974). These five isolates and 21 *Xanthomonas* type or pathotype strains were sequenced at four housekeeping loci for multilocus sequence analysis (MLSA): *gyrB*, *dnaK*, *rpoD*, and *fyuA* (Young et al. 2008; accessions MG637405 to MG637424). Shortened gene fragments were used for improved efficiency of MLSA (Zacaroni et al. 2012). The program CLC Genomics Workbench was used to concatenate the sequences to a total length

of 2,664 kb and to generate a phylogram showing genetic relatedness among the isolates and reference strains. The sequence analysis revealed the isolates to be identical to each other and within the species *Xanthomonas campestris*. Pathogenicity was tested in two independent experiments using four arugula cultivars (Dragon Tongue, Rocket Salad, Rocket Salad Wild, and Rocky). Inocula of five isolates were prepared by growing the bacteria in nutrient broth for 48 h. Suspensions (~10 5 cfu/ml) were sprayed until runoff onto 30 plants per isolate per cultivar. Control plants were sprayed with sterile distilled water. Plants were enclosed in clear plastic bags for 48 h and then maintained in a greenhouse (20 to 22°C). After 14 days, chlorotic lesions developed on all inoculated plants; lesions later turned tan and were similar to symptoms observed in the field. Bacteria reisolated from inoculated plants were found to be identical to the original isolates by using repetitive extragenic palindromic sequence PCR with the BOXA1R primer, thereby confirming Koch's postulates. Plants treated with water developed no symptoms. Inoculations were repeated and results were the same. Brussels sprouts (*Brassica oleracea* var. *gemmifera* 'Genius') inoculated with the arugula strains likewise developed black rot symptoms. To our knowledge, this is the first report of black rot, caused by *X. campestris*, on arugula in California. Black rot of arugula has been reported from Argentina (Romero et al. 2008). Because arugula is rotated with the extensive cole crop acreage in the Salinas and other coastal California valleys, diseased arugula could be a source of inoculum that could cause black rot in these plantings.

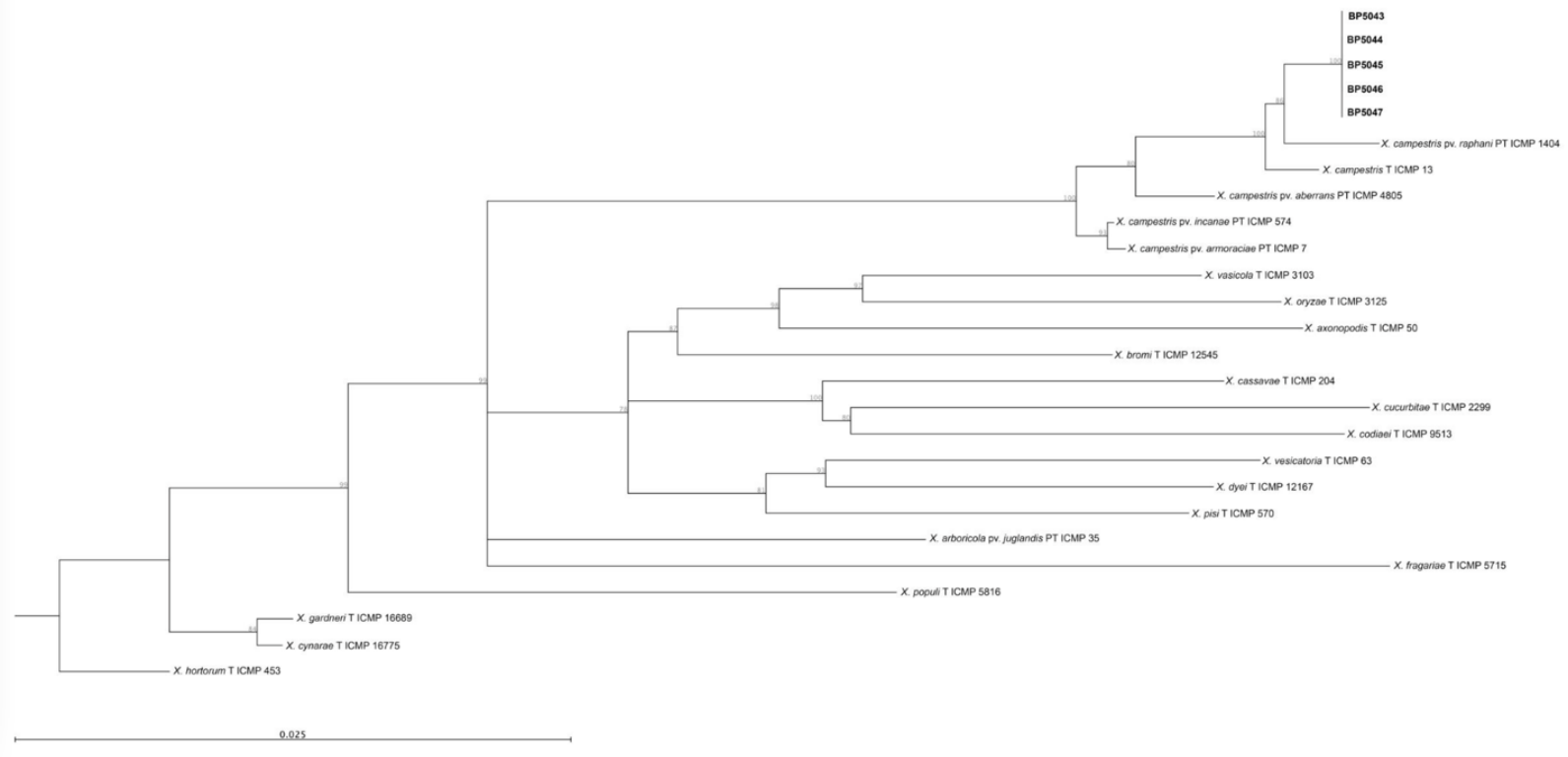


Figure B1-2. MLSA of suspect *Xanthomonas* isolates and *Xanthomonas* reference strains.

This is a maximum-likelihood phylogeny of *Xanthomonas* strains collected in Salinas Valley (Monterey County) California, shown in bold, and reference strains received from the International Collection of Microorganisms from Plants (ICMP). General time reversible model was used to generate this phylogeny. Bootstrap values are shown above branches, with the threshold set to 70%. The branch length value represents the expected number of changes per nucleotide.

Appendix C

C1-1 Clarifying the taxonomy of the causal agent of bacterial leaf spot of lettuce through a polyphasic approach reveals that *Xanthomonas cynarae* Trébaol et al. 2000 emend. Timilsina et al. 2019 is a later heterotypic synonym of *Xanthomonas hortorum* Vauterin et al. 1995.

Morinière L., Burlet A., **Rosenthal E.R.**, Nesme X., Portier P., Bull C.T., Lavire C., Le Saux M.F., Bertolla F. (2020) "Clarifying the taxonomy of the causal agent of bacterial leaf spot of lettuce through a polyphasic approach reveals that *Xanthomonas Cynara* Trébol et al. 2000 emend. *Timilsina* et al. 2019 is a later heterotypic synonym of *Xanthomonas hortorum* Vauterin et al. 1995. **Systematic and Applied Microbiology**. https://doi.org/10.1016/j.syapm.2020.126087.

C1-2 Abstract

Assessment of the taxonomy and diversity of *Xanthomonas* strains causing bacterial leaf spot of lettuce (BLSL), commonly referred to as *Xanthomonas campestris* pv. *vitians*, has been a long-lasting issue which held back the global efforts made to understand this pathogen. In order to provide a sound basis essential to its study, we conducted a polyphasic approach on strains obtained through sampling campaigns or acquired from collections. Results of a multilocus sequence analysis crossed with phenotypic assays revealed that the pathotype strain does not match the description of the nomenspecies provided by Brown in 1918. However, strain LMG 938 = CFBP 8686 does fit this description. Therefore, we propose that it replaces LMG 937 = CFBP 2538 as pathotype strain of *X. campestris* pv. *vitians.*

Then, whole-genome based phylogenies and overall genome relatedness indices calculated on taxonomically relevant strains exhibited the intermediate position of *X. campestris* pv. *vitians* between closely related species *Xanthomonas hortorum* and *Xanthomonas cynarae*. Phenotypic profiles characterized using Biolog microplates did not reveal stable diagnostic traits legitimizing their distinction. Therefore, we propose that *X. cynarae* Trébaol *et al.* 2000 emend. Timilsina *et al.* 2019 is a later heterotypic synonym of *X. hortorum*, to reclassify *X. campestris* pv. *vitians* as *X. hortorum* pv. *vitians* comb. nov. and to transfer *X. cynarae* pathovars in *X. hortorum* as *X. hortorum* pv. *cynarae* comb. nov. and *X. hortorum* pv. *gardneri* comb. nov. An emended description of *X. hortorum* is provided, making this extended species a promising model for the study of *Xanthomonas* quick adaptation to different hosts.

Appendix D

D1-1 Lettuce (*Lactuca sativa* L.) germplasm resistant to bacterial leaf spot caused by race 1 of *Xanthomonas hortorum* pv. *vitians* (Brown 1918) Morinière et al. 2020.

Sandoya, G.V., **Rosenthal, E.**, Simko, I. *et al.* Lettuce (*Lactuca sativa* L.) germplasm resistant to bacterial leaf spot caused by race 1 of *Xanthomonas hortorum* pv. *vitians* (Brown 1918) Morinière et al. 2020. *J Plant Pathol* 104, 993–1008 (2022). https://doi.org/10.1007/s42161-022-01123-0.

D1-2 Abstract

Bacterial Leaf Spot (BLS) is a disease that threatens lettuce (*Lactuca sativa* L.) production in warmer temperatures and high relative humidity, conditions that are conducive to unpredictable disease outbreaks and thus potentially resulting in total crop loss. BLS also causes problems in other major lettuce producing areas in the US and worldwide. To date, three races of the pathogen, *Xanthomonas hortorum* pv. *vitians* (Brown in J Agric Res 13:367–388, 1918; Morinière et al. in Syst App Microbiol 43:1260–1287, 2020 formerly known as *X. campestris* pv. *vitians*, were described in the US and Canada; though only race 1 has been confirmed thus far in Florida. Lettuce germplasm with a diverse set of resistance alleles is needed to improve resistance to BLS utilized in current commercial cultivars. In this study, plant introductions (PIs) from the National Plant Germplasm System (NPGS) and heirloom lettuce cultivars were tested against three isolates of the pathogen in Florida and two isolates in California. PI 667690 and PI 667709 were identified to be resistant to isolates of race 1 from both Florida and California. Allelism tests indicate that resistance found in these PIs is allelic to *Xcvr* (*Xanthomonas campestris vitians resistance*), a previously identified resistance gene

detected in PI 358001–1. The two new sources of resistance offer a different set of morphological characteristics that can be useful in breeding programs than those found in PI 358001–1.

Appendix E

E1 PSU Extension Engagement with PA Latinx Communities

The Latinx population in the United States is growing in both size and strength. Latinxs, like myself, are impacting the future of American society with our increasing cultural and economic influence. Between 2010 and 2020, the Latinxs accounted for 51% of the nation's population growth. By 2030, it is expected that a third of the U.S. population will identify as Latinx. Our communities have long been significant contributors to the American workforce, especially in agriculture. Latinx individuals represent 75% of the nation's agricultural workforce. In Pennsylvania (PA), Latinx communities make up 8% of the state's population and contribute significantly to our agricultural industries. With this growing demographic comes the need for targeted and intentional engagement on behalf of those who wish to equitably serve our communities in areas such as extension and outreach, research, and education.

Since 2010, there have been successful ad hoc efforts within Penn State Extension (PSE) to serve the growing Latinx communities in our state. In 2019, I worked with fellow graduate student Ilse Huerta Arredondo to start the Penn State Latinx Agricultural Network (LAN). We invited faculty, staff, students, and extension agents to share their expertise, assess the needs of the Latinx communities of PA, and work with us to support new and expanding programming for Latinx audiences. In 2020, I led a meeting to summarize our findings and co-wrote a strategic plan for addressing Latinx community needs, which I presented to the Deans of the College of Agricultural Sciences.

We are now a Community of Practice following this strategic plan, with the mission of implementing avenues of access and support to promote active participation and inclusion of Latinx communities in PSE programming, research, and teaching activities. We successfully acquired funding from several sources, including a commitment from Penn State Extension. Our network holds regular meetings broadly with the Latinx community to share news of programs and identify new program targets. We were honored with the National Extension Diversity

Award offered by the Extension Committee on Organization and Policy (ECOP) and USDA-NIFA. I was honored with the Evans Family Award for Graduate Student Extension Achievement by the College of Agricultural Sciences in 2022. Our original 9-person board has grown to a 110-member network in 2023, and we are excited to continue serving the Latinx communities of PA.

Emma Rosenthal – Curriculum Vitae

EDUCATION

Ph.D., Plant Pathology and Environmental Microbiology
Penn State, University Park, PA.
Lab of Dr. Carolee Bull – January 2017 to December 2023

B.S., Biological Sciences, Distinction in Research
Cornell University, Ithaca, NY
Lab of Dr. Gregory Martin – January 2014 to May 2015

PEER-REVIEWED PUBLICATIONS

Rosenthal, E.R., Potnis, N., Bull, C.T. (2022). "Comparative genomic analysis of the lettuce bacterial leaf spot pathogen, *Xanthomonas hortorum* pv. *vitians*, to investigate race specificity." **Frontiers in Microbiology,** 13.

Sandoya, G.V., **Rosenthal, E.R.**, Rodrigues-Porto, L., Wadlington, W.H., Bull, C.T., Simko, I., Carroll, A. (2022). "Germplasm Resistant to Race 1 of *Xanthomonas hortorum pv. vitians* (Brown, 1918) Moriniére et al. 2020 Causing Bacterial Leaf Spot of Lettuce (*Lactuca sativa L.*)." **Journal of Plant Pathology**.

Morinière L., Burlet A., **Rosenthal, E.R.**, Nesme X., Portier P., Bull C.T., Lavire C., Le Saux M.F., Bertolla F. (2020) "Clarifying the taxonomy of the causal agent of bacterial leaf spot of lettuce through a polyphasic approach reveals that *Xanthomonas Cynara* Trébol et al. 2000 emend. *Timilsina* et al. 2019 is a later heterotypic synonym of *Xanthomonas hortorum* Vauterin et al. 1995. **Systematic and Applied Microbiology**.

Rosenthal E.R., Ramos L., Bull C.T., and Koike S.T. (2017) "First Report of Black Rot Caused by *Xanthomonas campestris* on Arugula in California." **Plant Disease**.

MANUSCRIPTS IN PREPARATION

Rosenthal, E. R., Bull, C. T. (expected 2023). "Development of a race-specific detection method for the lettuce bacterial leaf spot pathogen." Submitting to **Plant Disease**.

Rosenthal, E. R., Bull, C. T. (expected 2024). "Examining the role of *X. hortorum* pv. *vitians*'s race-specific effectors in lettuce HR induction." Submitting to **Molecular Plant-Microbe Interactions.**

Sandoya, G.V., Trent, M., Hayes, R. J., Lebeda, A., **Rosenthal, E.R.**, Simko, I., Bull, C. T. (expected 2023). "*Lactuca serriola* is a valuable source of resistance against three races of *Xanthomonas hortorum* pathovar *lactucae* the pathogen causing Bacterial Leaf Spot of lettuce." Submitting to **Plant Disease**.

Stockham, E., **Rosenthal, E.R**., Koike, S.T., Alger, E.I., Bull, C.T. (expected 2024). "Bacterial Leaf Streak caused by a novel pathovar of *Xanthomonas translucens* on Chives." Submitting to **Plant Disease**.

AWARDS & HONORS

National
2022-08 National Extension Diversity Award; ECOP and USDA-NIFA for Penn State Extension's Latinx Ag. Outreach Network, Co-founder

2018-06 Graduate Research Fellowship Program; NSF

PSU College of Ag. Sciences
2022-03 Evans Family Award for Graduate Student Extension Achievement

2021-06 William Henson Diversity Achievement Award

2019-03 Gamma Sigma Delta Agricultural Honor Society Inductee

2017-06 NASA PA Space Grant Consortium Graduate Research Fellowship

2016-04 Bunton Waller Graduate Assistantship

SELECT PRESENTATIONS

2022-08 Research talk, APS Plant Health, Pittsburgh, PA

2022-07 Research talk, Conference on Plant Pathogenic Bacteria, Assisi, Italy

2021-01 Research poster, APS Plant Health 2020, Virtual

2019-06 Research talk, USDA Symposium on Phytobiomes and Population Biology, USDA-ARS, Beltsville MD

2018-08 Research poster, Intl. Congress of Plant Pathology, Boston, MA

ORCID: 0000-0002-6397-9696