Comparative genomics and pathogenicity of *Xanthomonas vasicola*

By

Nomakula Y. Zim

Submitted in partial fulfilment of the requirements of the degree

Master of Science

In the Faculty of Natural and Agricultural Sciences

University of Pretoria

Pretoria

January 2020

Supervisor: Prof Teresa A. Coutinho

Co-supervisors: Dr Emmanuel Wicker and Ms Sharon McFarlane

Declaration of Originality

I, Nomakula Y. Zim, declare that the dissertation, which I hereby submit for the degree **Master of Science** at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

SIGNATURE 

DATE 6 January 2020

Acknowledgements

First and foremost, I wish to express my sincerest gratitude to my primary supervisor, Prof Teresa Coutinho, for her continuous support and guidance throughout this research project. I want to thank my co-supervisors Dr Emmanuel Wicker and Ms Sharon McFarlane, for the invaluable advice and guidance they provided in the experiments undertaken in this research project. I wish to thank Dr Anna Yssel for the help she provided, completing the bioinformatic analyses would have been near-impossible without your help. Thanks to Dr Surendra Vikram for the initial assistance in the bioinformatic analyses.

I want to thank my colleagues in the Plant Pathogenic Bacteria Research group for the support and the invaluable feedback sessions. To Gina Shin, I can never thank you enough for always willing to help.

I wish to thank the Centre for Microbial Ecology and Genomics (CMEG), the Forestry and Agricultural Biotechnology Institute (FABI), and the South African Sugarcane Research Institute (SASRI) for providing me with the facilities to conduct my research. I also wish to thank the National Research Foundation (NRF) for providing the funds for the research.

I wish to convey my deepest gratitude to all the individuals who contributed in their unique way to this project, no matter how small it may have seemed. To my friends, Patience Modiba and Alphonse Onana, thank you for the emotional support, you made even the hardest disappointments easier to face and get over. To my brother Mashadu Zim, thank you for all the afternoons you spent taking care of plants and for all the heavy lifting you did, you made this journey a little easier. To my mother Vuyelwa Zim, thank you for your unwavering support and the motivational talks, this achievement is in many ways yours as much as it is mine. To the rest of my family, I am eternally grateful for your the calls and messages which kept me sane throughout this journey.

Last but not least, ke leboha Modimo le Badimo ba tsamayang le nna ka mehla.

Preface

In the mid-2000s an outbreak of bacterial blight and dieback caused by *Xanthomonas vasicola* pv. *vasculorum* was observed on a single *Eucalyptus grandis* clone in KwaZulu-Natal. It was suggested that this outbreak was as a result of a host jump from sugarcane. Therefore, the purpose of this dissertation is to determine the host range of *X. vasicola* strains and to identify the genetic determinants which may have played a role in this pathogen's ability to jump to a new host. This dissertation will be presented in three independent chapters.

Chapter 1 will review previous literature on the changes that take place in a bacterial genome leading to its adaptation to a changing environment. The chapter will include sections on the modifications of existing genes by mutations, gene duplications and gene rearrangements, acquisition of new genes by horizontal gene transfer, and the loss of genes as pathogens become specialised to their hosts.

In **chapter 2,** the host range of *X. vasicola* strains isolated from different hosts will be tested. Host's include the monocotyledonous plant species banana, maize, sorghum, sugarcane and the dicotyledonous plant species *Eucalyptus grandis*. The significance of the differences observed between the number of infected plants and the severity of the disease symptoms will be determined.

In **chapter 3**, the differences between the genomes of five *Xvv* isolated from *E. grandis* and *X. vasicola* isolated from other hosts will be determined. The genomes of five *Xvv* strains from *Eucalyptus* were sequenced on the Illumina HiSeq 2500 platform. The chromosomes and plasmids were assembled and annotated, and the genomes were compared to those of other *X. vasicola* strains isolated from different hosts. These genome differences may have played a role in their adaptation to a new host.

Table of Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

List of Tables

Chapter 2

Chapter 3

Chapter 1

Pathoadaptation of bacterial phytopathogens leading to host shift and host jump events

1.1 Introduction

Phytopathogens can either be generalists and infect a wide range of plant hosts, or they may be specialists and only infect a single species (Baumler & Fang, 2013). At the metacommunity scale, generalists are abundant and show high variation, whereas specialists tend to be rare and show a low degree of variation (Szekely & Langenheder, 2014). The low number of specialists compared to generalists may be as a result of the specialists being highly adapted to their specific hosts, and the maladaptation of generalists (Kniskern *et al.*, 2011). The host recognises more of the effectors from a specialist pathogen compared to a generalist pathogen (Kniskern *et al.*, 2011). This maladaptation of generalists to their hosts leads to their suppressed growth rates and limited genetic variation. In contrast, specialists are genetically better adapted to utilise the resources available to them in that environment.

Pathoadaptation is a process englobing all the genetic change that occurs after an organism occupies a new niche and adopts a pathogenic lifestyle, and these changes confer fitness to the pathogen in the new environment (Pallen & Wren, 2007). These changes can be as a result of the modification of the existing gene repertoire, loss of genes which are not under selection in the new environment and acquisition of new genes from other organisms occupying the same niche (Gal-Mor & Finlay, 2006). For these changes to be fixed in the genome and maintained in subsequent generations, there are several processes which the genes have to undergo. Following gene modification, gene acquisition or gene loss, there is a fixation phase where the genomic change spreads into the population, and in the final phase in which fixation occurs, the change in the genome is maintained in the population (Innan & Kondrashov, 2010).

Pathoadaptation may lead to host shift or host jump events. A host jump is defined as the colonisation of new host species that are phylogenetically distant from the current host of a pathogen, and a host shift is the colonisation of new host species that are phylogenetically closely related to the current host (Schulze-Lefert & Panstruga,

2011). For a pathogen to adapt to a new host, there are barriers which must be overcome. These include physical barriers such as the cuticle, cellular barriers and molecular barriers (Ochman & Moran, 2001). Host shift events occur more frequently than host jump events because crossing further than the species barrier is more complicated, limiting the number of host jumps that occur (Baumler & Fang, 2013). A pathogen infecting a host that is phylogenetically closely related to its current host has a subset of effector proteins that will remain functional in the new host (Schulze-Lefert & Panstruga, 2011) since the targets of these effector proteins are also conserved in the new host. In the case of a host jump event however, the new host may lack the targets for the effector repertoire of the pathogen. Pathogen fitness increases have been shown to occur rapidly in the early phases of adaptation and decline over time (Barrick *et al.*, 2009). This may be because the mutations that occur in the early phases of adaptation occur in regions involved in the global regulatory functions such as DNA topology (Crozat *et al.*, 2004), amino acid metabolism, and carbon catabolism (Cooper *et al.*, 2003).

Currently, there is limited information on the adaptation of bacterial plant pathogens to new hosts as a result of host shift and host jump events. The purpose of this review is to understand which factors allow bacterial pathogens to adapt to new hosts. This will include information on the factors that lead to the modification of existing genes, how new genes are acquired from closely related as well as distantly related species, how genes are lost, and how these factors together result in a host shift or host jump event. Attempts will be made to give examples among phytopathogenic bacteria where possible.

1.2 Modification of existing genes

Existing genes can undergo multiple changes due to errors which may occur during replication of the chromosome, or these changes may be programmed by the cell. Gene modifications can occur as a result of mutations within the genes, which may lead to changes in the protein structure and function, rearrangement of genes or the

duplication of the gene. These errors within the genome are corrected by the DNA repair machinery, but some are missed and may confer adaptive advantages to the host cell.

1. 2.1 Mutations

Mutations are changes that occur in the DNA sequence, and these changes can be due to errors during replication or through external causes such as ultraviolet rays and chemicals. These mutations can lead to a change in the structure of the protein encoded by the gene or may decrease or increase the expression of the gene (Lodish *et al.*, 2000). Mutations in existing genes that result in proteins losing their functions are known as null mutations (Fig. 1.1)(Sheppard *et al.*, 2018). The loss of function may be as a result of frameshift indels, nonsense mutations, or insertions in a coding region if it occurs within the open reading frame of the gene, as well as missense mutations which occur in proteins (Hottes *et al.*, 2013, Sheppard *et al.*, 2018). These mutations may also aid the survival of the partially adapted pathogen in a new environment, giving the population enough time to acquire more mutations to adapt to the new environment (Hottes *et al.*, 2013).

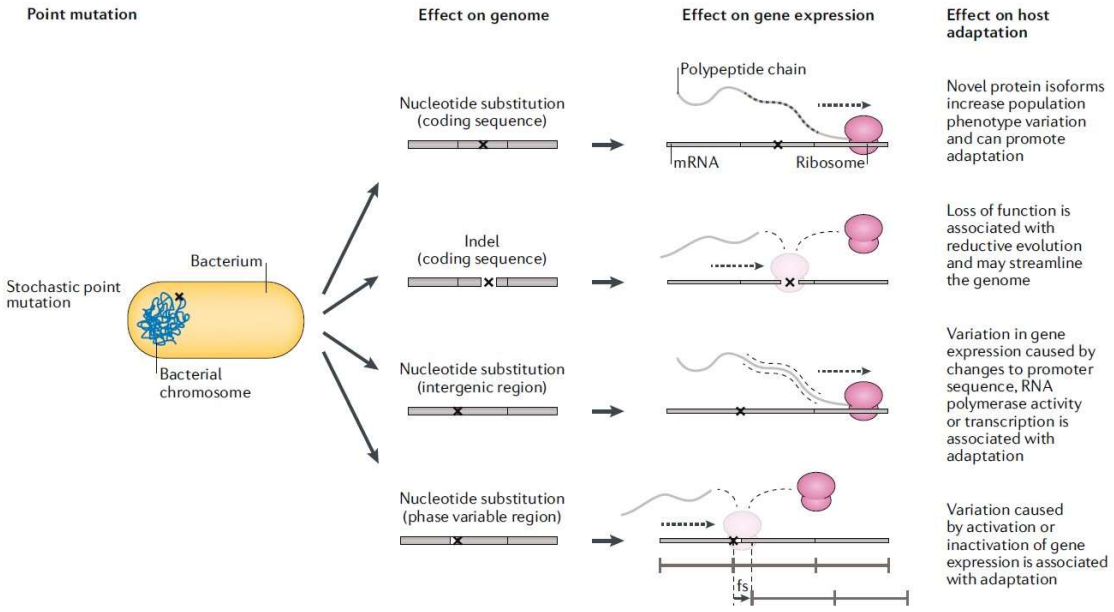


Figure 1.1: Point mutations are single nucleotide changes that occur in DNA sequences. Mutations in protein-coding sequences may lead to an increase in new protein isoforms which have unique functions. Indels and insertions in the protein-coding protein and intergenic regions disrupt the translation process, which may lead to proteins losing their functions. Mutations may also result in phase variation, which allows a pathogen to adapt rapidly to a change in the environment (Sheppard *et al*., 2018).

Mutations can occur at random or maybe a pre-emptive strategy for adaptation as is the case with single sequence repeats (SSR) contingency loci. These contingency loci are hypermutable DNA regions which mediate high frequency, stochastic, and heritable gene switching (Moxon *et al.*, 2006). Heritable gene switching is a recombination-independent process which is considered to occur through strand slippage, and the switch in these contingency loci occurs at a rate of 10^{-2} to 10^{-5} in the human pathogen *Haemophilus influenzae* (Moxon *et al.*, 2006). In a study by Jerome *et al*. (2011), the foodborne pathogen *Campylobacter jejuni* was shown to utilise its contingency loci to adapt to mice, resulting in campylobacteriosis. Significant changes were observed in contingency genes in the flagellar glycosylation locus, which encodes for the lipopolysaccharide and capsule biosynthesis (Jerome *et al.*, 2011). These changes aid in the evasion of the host's immune response, allowing the pathogen to multiply further, adapt and propagate (Alexander & Rietschel, 2001). The

open reading frames of genes for a restriction/modification enzyme and an iron-binding protein were also altered. The changes observed in the restriction/modification and iron-binding genes may enable the pathogen to increase its range of protection against phages and increase its iron uptake. Bacterial phytopathogens may also employ the same mechanism when they encounter a new host. This may be one of the many ways in which these pathogens can rapidly adapt to new hosts which are phylogenetically similar to their current host, resulting in a host shift event.

1.2.2 Gene duplications

Gene duplications are the emergence of a heritable copy of a gene resulting in paralogs which are duplicated homologous sequences within a genome ascribed to the same family (Gevers *et al.*, 2004, Innan & Kondrashov, 2010). Gene duplications can be neutral, advantageous, or deleterious. Advantageous gene duplications act by increasing the dosage of the beneficial gene product, by masking deleterious mutations and result in the emergence of a new function (Innan & Kondrashov, 2010).

Gene duplications may result in an increase in gene dosage. A beneficial gene dosage may be positively selected, resulting in the fixation of the duplicated gene (Innan & Kondrashov, 2010). Genes which have been identified as being under this model are those that mediate interactions between organisms and the environment such as stress-response genes, genes that have dosage-sensitive functions such as functioning in protein-protein interactions, and genes whose products are required in large quantities such as ribosomal genes (Innan & Kondrashov, 2010).

The selection for gene duplication may be driven by the new environment in which the pathogen evolves in. The functions resulting from the increase in the gene dosage, thus provide an advantage to the new pathogen in interacting with its host and causing disease (Innan & Kondrashov, 2010). An analysis of 106 bacterial genomes showed that genes encoding for ABC-type transporters, transcriptional regulators and

dehydrogenases represented the largest group of paralogs (Gevers *et al.*, 2004). An increase of these gene products when a pathogen infects a novel host will enable it to increase its growth rate as it has more energy to direct towards metabolic processes, allowing rapid adaptation and a host shift or host jump event. In the yeast *Saccharomyces cerevisiae*, the gene which encodes for the yeast dextrose transporter duplicates under conditions of reduced glucose conditions (Brown *et al.*, 1998). The progeny which had the gene duplication were shown to out-compete the parents in a competition experiment (Brown *et al.*, 1998). A similar event may take place in bacterial phytopathogens when they encounter novel hosts which lack certain substrates required for growth. This then enables the phytopathogens to adapt rapidly to the host, leading to a host shift or host jump event.

Gene duplications, which result in a modified gene product, may result in neofunctionalisation. Neofunctionalisation is the random acquisition of a novel function as neutral mutations accumulate and selection is relaxed in duplicated genes (Innan & Kondrashov, 2010, Remigi *et al.*, 2011). Following the duplication event, one of the genes evolves faster while the other retains its function (Remigi *et al.*, 2011). The GALA type III effector genes in *Ralstonia solanacearum* underwent a duplication event resulting in two subclades, and the genes experienced purifying and diversifying selection which is consistent with the neofunctionalisation model (Remigi *et al.*, 2011). The genes *GALA2* and *GALA5* kept their ancestral functions of promoting pathogenicity in tomato, whereas *GALA7* is essential for adaptation to novel hosts such as *Medicago truncatula* (Remigi *et al.*, 2011). Thus gene duplication leading to neofunctionalisation may occur in many phytopathogens when they encounter novel hosts, resulting in a host jump event.

1.2.3 Gene rearrangements

Gene rearrangements can be stochastic, occurring during the cell's vegetative growth, and this results in several phenotypes, some of which will be selected if the habitat changes (Haselkorn, 2001). Gene rearrangements can occur through phase variation

which is the ability of pathogens to alternate between two phenotypes in a heritable and reversible manner, and the rate at which this reversible change occurs exceeds that of random mutations in response to a fluctuating environment (van de Woude & Baumler, 2004, Vial *et al.*, 2010). This mechanism is essential in allowing pathogens to adapt to different niches and increasing their fitness when they encounter a novel host (Vial *et al.*, 2010).

Strains of the pathogen *Salmonella* can switch between expressing the H1 flagellar antigen and the H2 flagellar antigen as a result of an inversion by site-specific recombinase enzymes (Haselkorn, 2001). This phase variation occurs in one in every thousand cells per generation (Haselkorn, 2001, van de Woude & Baumler, 2004). This phase variation may help pathogens occupying a new habitat to adapt rapidly, and because the plant host has not evolved mechanisms to fight the pathogen, the disease caused is more virulent in the new host.

This phenomenon has also been observed in the soil-borne phytopathogen *Ralstonia solanacearum*. A phase variation from mucoid to non-mucoid colonies occurs as a result of changes within the regulatory protein *phcA* gene which plays a central role in the Phc confinement sensing system (Poussier *et al.*, 2003). When tandem duplications and the disruption of the gene by insertion sequences occurs, a change from the virulent wild-type to the non-virulent type occurs. A phase reversion to the wild-type mucoid colonies occurred when the pathogens were inoculated into tomato roots and stem but not in vitro experiments, and it has been suggested that this reversion by duplications and disruptions allows the pathogen to adapt to the plant environment and the phase change allows for better adaptation to the soil environment (Poussier *et al.*, 2003). This mechanism may be employed by other phytopathogens when encountering new hosts, allowing their persistence in their new host.

Gene rearrangement events have been observed in the phytopathogen *Xylella fastidiosa*. A comparison between the genomes of strains causing Pierce's disease in grapevines and citrus variegated chlorosis in citrus revealed three large regions in the

chromosome that were rearranged (van Sluys *et al.*, 2003, Brussow *et al.*, 2004). These translocated and inverted regions were flanked by a putative phage-related integrase (van Sluys *et al.*, 2003). This suggests that the chromosomal regions were rearranged as a consequence of the presence of phage-related integrases (van Sluys *et al.*, 2003, Brussow *et al.*, 2004). The rearrangements observed, together with genes acquired through horizontal gene transfer, may have helped the strains further adapt to their respective hosts.

1.3 Acquisition of new genes

New genes can be acquired through horizontal gene transfer (HGT) which is the transfer of genetic material between two bacterial cells without cell division occurring and these bacteria are usually different strains (Lawrence, 2006, Pallen & Wren, 2007). Horizontal gene transfer is likely to occur in environments that have high bacterial diversity, but it can occur between eukaryotes and bacteria (Gal-Mor & Finlay, 2006). When it occurs between distantly related organisms, the genes involved are those which encode for the simplest sets of functional networks (Jain *et al.*, 1999). Horizontal gene transfer is facilitated by natural transformation, conjugation, and transduction (Fig. 1.2)(Sheppard *et al.*, 2018). Natural transformation is the process in which bacterial cells can take up naked DNA from their environment (Gal-Mor & Finlay, 2006, Pallen & Wren, 2007). Conjugation occurs between the donor and recipient bacterial cells; this process is unidirectional and is mediated by the mating pore of the conjugative element (Gal-Mor & Finlay, 2006, Johnson & Grossman, 2015). Transduction involves the transfer of genetic material from one bacterium to another through bacteriophages, which infect them (Gal-Mor & Finlay, 2006). The importance of the type of HGT process that bacterial pathogens undertake is thought to be dependent on the characteristics of the bacterial species and the mobile genetic elements involved (Heuer & Smalla, 2012).

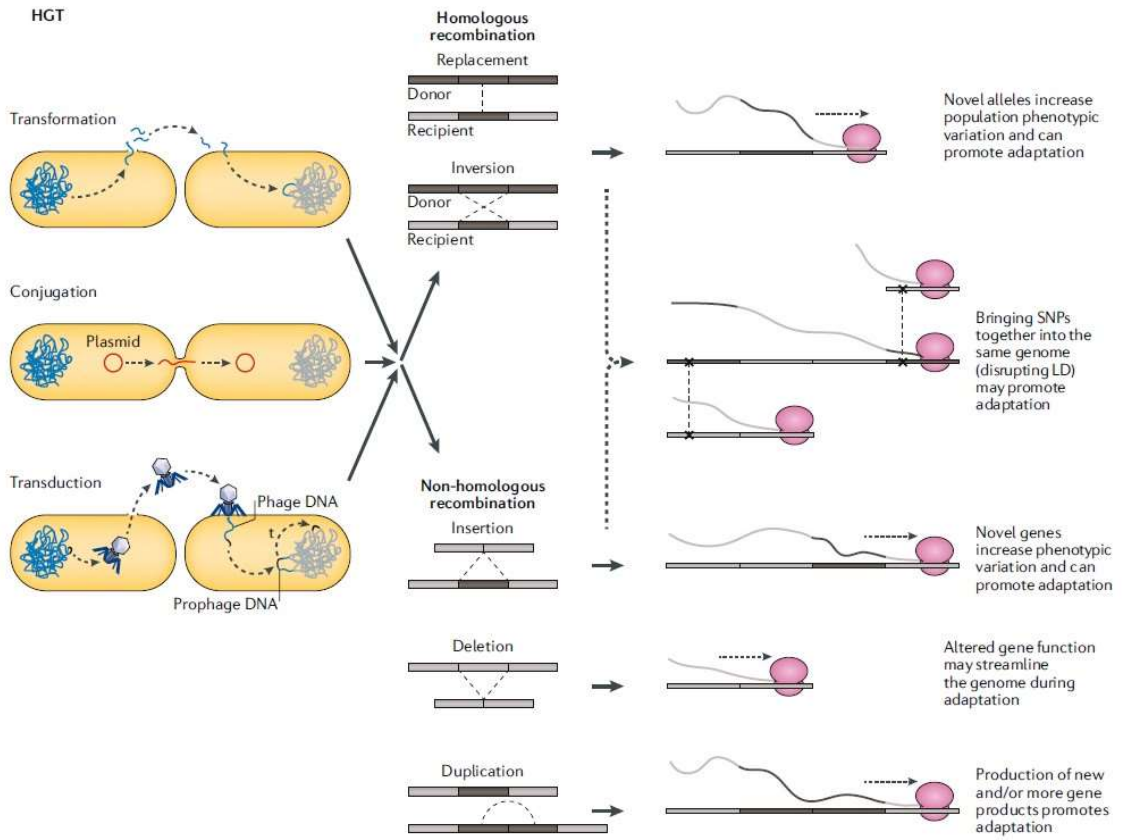


Figure 1.2: Horizontal gene transfer occurs natural transformation, conjugation, and transduction. Transformation is the uptake of naked DNA from the environment by naturally competent recipient cells, and conjugation is the transfer of DNA from the donor cell to the recipient cell using the type IV secretion system. The process of transduction is the transfer of donor cell DNA to the recipient cell by a bacteriophage where the DNA is incorporated into the recipient cell as a prophage. The DNA sequences are incorporated into the recipient's genome using homologous or non-homologous recombination, and this introduces novel functions to the cell leading to adaptation (Sheppard *et al*., 2018).

1.3.1 Natural transformation

Natural transformation is the stable uptake, integration, and expression of DNA acquired from its environment by naturally competent cells (Thomas & Nielsen, 2005, Gal-Mor & Finlay, 2006). The physiological status of competence, which is time-limited and is as a response to environmental changes, is achieved through the expression of over 50 proteins and this process is initiated by the recipient cell (Thomas & Nielsen, 2005, Johnsborg *et al.*, 2007). For natural transformation to succeed, the naked DNA released from decomposing cells, disrupted cells, viral particles, or living cells, must

be able to persist long enough for it to be taken up by the surrounding recipient cells (Thomas & Nielsen, 2005). The integration of DNA acquired through natural transformation relies on the recombinases of the host cell and can occur through homologous or illegitimate recombination (Hülter & Wackernagel, 2008). The uptake of this DNA has been suggested to occur during the phytopathogen's infection process. Following the co-infection of tomato plants with the recombinant strains *R. solanacearum* GMI1581 and GMI1000FAP2, the isolates obtained following a six-week infection period had genes from both parent recombinant strain (Bertolla *et al.*, 1999). This interaction of phytopathogens during the infection process leads to the emergence of strains which show increased aggressiveness on some hosts (Coupat-Goutaland *et al.*, 2011).

In Gram-negative bacteria, natural transformation is often mediated by the type IV secretion system (T4SS) (Johnsborg *et al.*, 2007, Seitz & Blokesch, 2013). The T4SS not only facilitates the transfer of naked DNA from the environment to the recipient but is also responsible for adhesion and twitching motility (Johnsborg *et al.*, 2007). The system also allows enhanced survival under UV light conditions, as observed in phytopathogens *Xanthomonas campestris* and *Pseudomonas syringae* (Kang *et al.*, 2002). The enhanced survival under UV light allows the pathogens to remain longer on the leaf surface until conditions allow entry into the host. The presence of the T4SS or its homologs may thus enable a phytopathogen in a novel host to acquire genetic material which allows it to utilise resources and adapt, resulting in a host shift or host jump event.

The soil inhabitant *Pseudomonas stutzeri* is one of the naturally competent members of *Pseudomonadaceae*. Mutations in the *pilA* and *pilC* genes were shown to terminate competence (Graupner *et al.*, 2000). These mutations resulted in the formation of immature pili which were unable to bind and take up naked DNA (Graupner *et al.*, 2000). This, therefore, further supports the importance of pili in the process of natural competence. The uptake of single-stranded DNA was also shown to be more efficient than the uptake of double-stranded DNA and was not dependent on a signal sequence as seen in *Haemophilus influenzae* (Graupner *et al.*, 2000, Seitz & Blokesch, 2013).

1.3.2 Conjugation and integrative and conjugative elements

Conjugation is a contact-dependent mechanism which uses a type IV secretion system pilus to transfer genetic material (Juhas *et al.*, 2008, Johnson & Grossman, 2015). During this process, conjugative plasmids and integrative and conjugative elements (ICEs) are exchanged between the donor cell and recipient cell. Conjugation may play an essential role in aiding the persistence of a pathogen's response to a rapidly changing environment (Thomas & Nielsen, 2005).

Integrative and conjugative elements (ICEs) are genetic elements which carry genes required for conjugation and can be transmitted to the next generation through lateral gene transfer following their integration into the chromosome (Johnson & Grossman, 2015, Delavat *et al.*, 2017). Following the induction of the ICE genes, the element is excised and then forms a double-stranded circular DNA molecule, a mating pore is formed, and the DNA material in its single-strand form is transferred to the recipient cell and is integrated into its genome (Johnson & Grossman, 2015). The transfer of ICEs occurs at a rate of 10^{-2} to 10^{-7} transconjugants per donor cell (Delavat *et al.*, 2017).

Genomic islands are the mobile segments of the bacterial genome, and they make up the majority of the gene pool obtained horizontally (Juhas *et al.*, 2008). These regions are 10-100kb in size and in most prokaryotes, their mechanism of mobility is yet to be established (Delavat *et al.*, 2017). These regions offer a selective advantage to bacteria as they harbour genes responsible for the metabolism of novel compounds, resistance to antibiotics, and genes responsible for virulence (Gal-Mor & Finlay, 2006, Juhas *et al.*, 2008). These genomic islands are then transferred to the recipient as part of a conjugation plasmid.

The animal pathogen *Bartonella* has been shown to be adapted to its host through conjugation. A genome comparison of *Bartonella* species revealed that the ruminant

pathogen *Bartonella tribocorum* acquired a T4SS soon after separation from *Bartonella bacilliformis*, a human pathogen (Saenz *et al.*, 2007). This acquisition allowed a host jump from ruminants to humans to occur, resulting in their speciation over time.

The human pathogen *Pseudomonas aeruginosa* has, on the other hand, been shown to have altered its conjugation machinery in response to siderophore-conjugated antibiotics (Tomaras *et al.*, 2013).

Phytopathogens encountering a novel host may acquire genes which increase their fitness in the new environment and may over time alter their conjugation machinery to become resistant to the host's response. This, over evolutionary time, may result in a host shift or host jump event.

1.3.3 Horizontal gene transfer in *Xanthomonas*

Members of the genus *Xanthomonas* are Gram-negative, straight rods, in the Gammaproteobacteria class. More than 20 *Xanthomonas* species are known to cause disease in some economically important crops such as maize, rice and sugarcane. About 20% of the gene contents of *Xanthomonas campestris* pv. *campestris* (*Xcc*) and *Xanthomonas axonopodis* pv. *citri* (*Xac*) are made up of horizontally transferred genes (Lima *et al.*, 2008). It can, therefore, be expected that other members of this genus will show a similar horizontally acquired gene content. Horizontally acquired genes are usually those responsible for pathogenicity and virulence, and mobile genetic elements. Complete clusters of genes responsible for the metabolism of NAD arginine, and cysteine, as well as the tricarboxylic acid cycle, have been identified in *Xanthomonas* and phylogenetic reconstruction show that these genes were acquired from closely related *Beta-* and *Gammaproteobacteria* as well as *Eukarya* and *Archaea* (Lima *et al.*, 2008). This, therefore, suggests that *Xanthomonas* phytopathogens may acquire genes with novel metabolic functions from both closely and distantly related organisms, providing them with an adaptive advantage and increasing their virulence.

Xanthomonas species have been shown to acquire genes horizontally from organisms which are not from the class of *Gammaproteobacteria* (Lima *et al.*, 2005, Lima *et al.*, 2008). A gene with high sequence similarity to plant natriuretic peptide (PNP)-encoding genes were found in the genome of *Xanthomonas axonopodis* (Nembaware *et al.*, 2004). The protein sequences of PNP from *Xac* and *Arabidopsis thaliana* were found to be of similar length, have N-terminal transmembrane signal peptides to direct the molecules into the extracellular space, and show significant conservation between amino acids essential for plant homeostasis. These findings suggest that this gene may have been acquired from the pathogen's eukaryotic host (Nembaware *et al.*, 2004). The expression of this gene by the pathogen may result in an increase of the water content and the cells rupturing. The rupture then releases water and nutrients to the pathogen, and this gives the characteristic wet rimmed lesions observed in the bacterial spot of citrus.

Horizontal gene transfer has also been shown to occur within the core genome. In the tomato pathogen *Xanthomonas perforans*, phage genes responsible for recombination were found within the core genome. This suggests horizontal gene transfer and recombination between the core genome and effectors of the Type III secretion system (Timilsina *et al.*, 2019).

1.4 Loss of genes

Despite the acquisition of new genes being frequent, bacterial pathogen genomes remain stable in size, and this is due to the loss of genes (Mira *et al.*, 2001). Loss of genes can be identified by a dramatic reduction of G+C content (Pallen & Wren, 2007). Genes that are not in use are usually lost over generations due to the lack of selection or due to a positive selection for gene loss resulting in pathoadaptation (Gal-Mor & Finlay, 2006, Pallen & Wren, 2007).

Koskiniemi *et al.* (2012) showed that following 1000 generations of growth, deletions increased in the genome, and this resulted in the progeny outcompeting the parental strains with bigger genomes. The increased fitness due to the deletions may be as a result of decreased energy use on making DNA, RNA, and proteins, and this energy is then directed to other essential growth-limiting processes (Koskiniemi *et al.*, 2012). Genes in the terminus region are more prone to selection for loss (Koskiniemi *et al.*, 2012).

Obligate pathogens often have reduced genomes as a result of adaptation. Mutations, as previously discussed, may lead to genes losing their functions and becoming pseudogenes (Bliven & Maurelli, 2012). These pseudogenes arise from genes involved in metabolic pathways which can be supplemented by the host. These genes become superfluous and accumulate mutations without impacting on the pathogen's fitness (Ochman & Moran, 2001, Moran, 2002). The genes that have accumulated these mutations then become inactivated and this, in the absence of selection pressure, results in their progressive loss and the reduction of the genome (Moran, 2002, Bliven & Maurelli, 2012).

The human pathogen *Yersinia pestis*, as well as many symbionts, have undergone a genome reduction. The genes that were lost encode for membrane proteins and membrane receptors, ABC transporters and some regulatory genes (Chain *et al.*, 2006, McCutcheon & Moran, 2011).

During the process of adaptation to a novel host, phytopathogens may lose some genes which become superfluous in the environment. This may then lead to them becoming specialised to the host resulting in speciation as a result of a host shift or host jump event.

1.5 Detecting changes in the genome

To detect the changes that occur in a population leading to adaptation, computational sequence composition-based and phylogenetic approaches can be used. The

sequence composition-based approaches look for deviations from the genomic average in the genome by reconstructing patterns of descent among genotypes and phylogenetic approaches identify genes which have an evolutionary history which is significantly different from that of the host species (Grunwald & Goss, 2011, Ravenhall *et al.*, 2015).

To determine the genetic changes which occurred when analysing whole genomes, identifying single nucleotide polymorphisms (SNPs) may explain the phenotypes observed (Yoshimura *et al.*, 2019). Single nucleotide polymorphisms analyses identify indels, transitions and transversions, and frameshifts which may have occurred.

To detect genes which have been acquired through HGT, parametric methods are used to identify regions in the genome which have a significantly different GC content or codon usage or phylogenetic methods which integrate the information of multiple genomes to build evolutionary models which best explain the gene history (Ravenhall *et al.*, 2015).

1.6 Conclusion

Host adaptation by phytopathogens can occur through the modification of genes in their chromosomes, by acquiring genes from their environment, host, or other microorganisms surrounding them, or by losing genes which have become superfluous in the new environment. The changes that allowed the phytopathogen to cause disease may not always be retained in the pathogen; thus, multiple adaptation events may have to occur before speciation can occur.

Understanding factors which may lead to a host shift or a host jump event are vital in coming up with preventative and management strategies.

1.7 References

Alexander C & Rietschel ET (2001) Bacterial lipopolysaccharides and innate immunity. *Journal of Endotoxin Research* **7**: 167-202.

Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE & Kim JF (2009) Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* **461**: 1243.

Baumler A & Fang FC (2013) Host Specificity of Bacterial Pathogens. *Cold Spring Harbour Perspectives in Medicine* **3**: 1-20.

Bertolla F, Frostegård Å, Brito B, Nesme X & Simonet P (1999) During infection of its host, the plant pathogen *Ralstonia solanacearum* naturally develops a state of competence and exchanges genetic material. *MPMI* **12**: 467–472.

Bliven KA & Maurelli AT (2012) Antivirulence genes: Insights into pathogen evolution through gene loss. *Infection and Immunity* **80**: 4061–4070.

Brown CJ, Todd KM & Rosenzweig RF (1998) Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. *Molecular Biology & Evolution* 931–942.

Brussow H, Canchaya C & Hardt W-D (2004) Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiology and Molecular Biology Reviews* **68**: 560–602.

Chain PSG, Hu P, Malfatti SA, Radnedge L, Larimer F, Vergez LM, Worsham P, Chu MC & Andersen GL (2006) Complete genome sequence of *Yersinia pestis* strains Antiqua and Nepal516: Evidence of gene reduction in an emerging pathogen. *Journal Of Bacteriology* **188**: 4453–4463.

Cooper TF, Rozen DE & Lenski RE (2003) Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proceedings of the National Academy of Sciences* **100**: 1072-1077.

Coupat-Goutaland B, Bernillon D, Guidot A, Prior P, Nesme X & Bertolla F (2011) *Ralstonia solanacearum* virulence increased following large interstrain gene transfers by natural transformation. *MPMI* **24**: 497*505.

Crozat E, Philippe N, Lenski RE, Geiselmann J & Schneider D (2004) Long-term experimental evolution in *Escherichia coli.* XII. DNA topology as a key target of selection. *Genetics*.

Delavat F, Miyazaki R, Carraro N, Pradervand N & van der Meer JR (2017) The hidden life of integrative and conjugative elements. *FEMS Microbiology Reviews* **41**: 512-537.
Gal-Mor O & Finlay BB (2006) Pathogenicity islands: a molecular toolbox for bacterial virulence. *Cellular Microbiology* **8**: 1707-1719.

Gevers D, Vandepoele K, Simillion C & van de Peer Y (2004) Gene duplication and biased functional retention of paralogs in bacterial genomes. *Trends in Microbiology* **12**: 148-154.

Graupner S, Frey V, Hashemi R, Lorenz MG, Brandes G & Wackernagel W (2000) Type IV pilus genes *pilA* and *pilC* of *Pseudomonas stutzeri* are required for natural genetic transformation, and *pilA* can be replaced by corresponding genes from nontransformable species. *Journal of Bacteriology* **182**: 2184–2190.

Grunwald NJ & Goss EM (2011) Evolution and population genetics of exotic and re-emerging pathogens: Novel tools and approaches. *Annual Review of Phytopathology* **49**: 249–267.

Haselkorn R (2001) Gene rearrangements, prokaryotic. Vol. 1 (Brenner S & Miller JH, eds.), p.^pp. 800-803. Academic Press.

Heuer H & Smalla K (2012) Plasmids foster diversification and adaptation of bacterial populations in soil. *FEMS Microbiology Reviews* **36**: 1083-1104.

Hottes AK, Freddolino PL, Khare A, Donnell ZN, Liu JC & Tavazoie S (2013) Bacterial adaptation through loss of function. *PLoS Genetics* **9**: e1003617.

Hülter N & Wackernagel W (2008) Double illegitimate recombination events integrate DNA segments through two different mechanisms during natural transformation of *Acinetobacter baylyi*. *Molecular Microbiology* **67**: 984-995.

Innan H & Kondrashov F (2010) The evolution of gene duplications: classifying and distinguishing between models. *Nature Reviews: Genetics* **11**: 97-108.

Jain R, Rivera MC & Lake JA (1999) Horizontal gene transfer among genomes: The complexity hypothesis. *Proceedings of the National Academy of Sciences* **96**: 3801-3806.

Jerome JP, Bell JA, Plovanich-Jones AE, Barrick JE, Brown CT & Mansfield LS (2011) Standing genetic variation in contingency loci drives the rapid adaptation of *Campylobacter jejuni* to a novel host. *PLoS ONE* **6**: e16399.

Johnsborg O, Eldholm V & Havarstein LS (2007) Natural genetic transformation: prevalence, mechanisms and function. *Research in Microbiology* **158**: 767e778.

Johnson CM & Grossman AD (2015) Integrative and Conjugative Elements (ICEs): What they do and how they work. *Annual Review of Genetics* **49**: 1-13.

Juhas M, Crook DW & Hood DW (2008) Type IV secretion systems: tools of bacterial horizontal gene transfer and virulence. *Cellular Microbiology* **10**: 2377–2386.

Kang Y, Liu H, Genin S, Schell MA & Denny TP (2002) *Ralstonia solanacearum* requires type 4 pili to adhere to multiple surfaces and for natural transformation and virulence. *Molecular Microbiology* **2**: 427-437.

Kniskern JM, Barrett LG & Bergelson J (2011) Maladaptation in wild populations of the generalist plant pathogen *Pseudomonas syringae*. *Evolution* **65**: 818-830.

Koskiniemi S, Sun S, Berg OG & Andersson DI (2012) Selection-Driven Gene Loss in Bacteria. *PLoS Genetics* **8**: e1002787.

Lawrence JG (2006) Horizontal and vertical gene transfer: The life history of pathogens. *Concepts in bacterial virulence,* Vol. 12 (Russell W & Herwald H, eds.), p.^pp. 255-271. Karger, Basel, Switzerland.

Lima WC, Van Sluys M-A & Menck CFM (2005) Non-Gamma-Proteobacteria gene islands contribute to the *Xanthomonas* genome. *A Journal of Integrative Biology* **9**: 160-172.

Lima WC, Paquola ACM, Varani AM, Van Sluys M-A & Menck CFM (2008) Laterally transferred genomic islands in *Xanthomonadales* related to pathogenicity and primary metabolism. *FEMS Microbiology Letters* **281**: 87-97.

Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D & Darnell J (2000) Genetic analysis in cell biology. p.^pp. W.H Freeman, New York.

McCutcheon JP & Moran NA (2011) Extreme genome reduction in symbiotic bacteria. *Nature Reviews: Microbiology* **10**: 13-26.

Mira A, Ochman H & Moran NA (2001) Deletional bias and the evolution of bacterial genomes. *TRENDS in Genetics* **17**: 589-596.

Moran NA (2002) Microbial minimalism: genome reduction in bacterial pathogens. *Cell* **108**: 583–586.

Moxon R, Bayliss C & Hood D (2006) Bacterial contingency loci: The role of simple sequence DNA repeats in bacterial adaptation. *Annual Review of Genetics* **40**: 307–333.

Nembaware V, Seoighe C, Sayed M & Gehring C (2004) A plant natriuretic peptide-like gene in the bacterial pathogen *Xanthomonas axonopodis* may induce hyper-hydration in the plant host: a hypothesis of molecular mimicry. *BMC Evolutionary Biology* **4**: 1-7.

Ochman H & Moran NA (2001) Genes lost and genes found: Evolution of bacterial pathogenesis and symbiosis. *Science* **292**: 1096-1098.

Pallen MJ & Wren BW (2007) Bacterial pathogenomics. *Nature* **449**: 835-842.

Poussier S, Thoquet P, Trigalet-Demery D, Barthet S, Meyer D, Arlat M & Trigalet A (2003) Host plant-dependent phenotypic reversion of *Ralstonia solanacearum* from non-pathogenic to pathogenic forms via alterations in the *phcA* gene. *Molecular Microbiology* **49**: 991-1003.

Ravenhall M, Škunca N, Lassalle F & Dessimoz C (2015) Inferring horizontal gene transfer. *PLoS Computational Biology* **11**: e1004095.

Remigi P, Anisimova M, Guidot A, Genin S & Peeters N (2011) Functional diversification of the GALA type III effector family contributes to *Ralstonia solanacearum* adaptation on different plant hosts. *New Phytologist* **192**: 976–987.

Saenz HL, Engel P, Stoeckli MC, Lanz C, Raddatz G, Vayssier-Taussat M, Birtles R, Schuster SC & Dehio C (2007) Genomic analysis of *Bartonella* identifies type IV secretion systems as host adaptability factors. *Nature Genetics* **39**: 1469-1476.

Schulze-Lefert P & Panstruga R (2011) A molecular evolutionary concept connecting nonhost resistance, pathogen host range, and pathogen speciation. *Trends in Plant Science* **16**: 117-125.

Seitz P & Blokesch M (2013) Cues and regulatory pathways involved in natural competence and transformation in pathogenic and environmental Gram-negative bacteria. *FEMS Microbiology Reviews* **37**: 336-363.

Sheppard SK, Guttman DS & Fitzgerald JR (2018) Population genomics of bacterial host adaptation. *Nature Reviews: Genetics* **19**: 549–565.

Szekely AJ & Langenheder S (2014) The importance of species sorting differs between habitat generalists and specialists in bacterial communities. *FEMS Microbial Ecology* **87**: 102-112.

Thomas CM & Nielsen KM (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature Reviews: Microbiology* **3**: 711-721.

Timilsina S, Pereira-Martin JA, Minsavage GV, Iruegas-Bocardo F, Abrahamian P, Potnis N, Kolaczkowski B, Vallad GE, Goss EM & Jones JB (2019) Multiple recombination events drive the current genetic structure of *Xanthomonas perforans* in Florida. *Frontiers in Microbiology* **10**: 1-13.

Tomaras AP, Crandon JL, McPherson CJ, Banevicius MA, Finegan SM, Irvine RL, Brown MF, O'Donnell JP & Nicolau DP (2013) Adaptation-based resistance to siderophore-conjugated antibacterial agents by *Pseudomonas aeruginosa*. *Antimicrobial agents and chemotherapy* **57**: 4197–4207.

van de Woude MW & Baumler AJ (2004) Phase and antigenic variation in bacteria. *Clinical Microbiology Reviews* **17**: 581-611.

van Sluys MA, de Oliveira MC, Monteiro-Vitorello CB*, et al.* (2003) Comparative analyses of the complete genome sequences of Pierce's disease and citrus variegated chlorosis strains of *Xylella fastidiosa*. *Journal of Bacteriology* **185**: 1018–1026.

Vial L, Groleau M-C, Lamarche MG, Filion G, Castonguay-Vanier J, Dekimpe V, Daigle F, Charette SJ & Deziel E (2010) Phase variation has a role in *Burkholderia ambifaria* niche adaptation. *International Society for Microbial Ecology* **4**.

Yoshimura D, Kajitani R, Gotoh Y, Katahira K, Okuno, Ogura Y, Hayashi T & Itoh T (2019) Evaluation of SNP calling methods for closely related bacterial isolates and a novel high-accuracy pipeline: BactSNP. *Microbial genomics* **5**: 1-8.

Chapter 2

Host range determination of *Xanthomonas vasicola* on maize, sorghum, banana, sugarcane and *Eucalyptus grandis.*

2.1 Abstract

Xanthomonas species are pathogens known to cause disease in over 400 monocotyledon and dicotyledon species. Three pathovars of *X. vasicola* have been identified, namely *X. vasicola* pv. *musacearum* causing Xanthomonas wilt of banana and ensete, *X. vasicola* pv. *vasculorum* causing bacterial blight and dieback of *Eucalyptus*, and *X. vasicola* pv. *vasculorum* and *holcicola* causing bacterial leaf streak of maize, sugarcane, and sorghum. In 2015, it was suggested that *X. vasicola* pv. *vasculorum* jumped hosts from the host sugarcane to *E. grandis*. The aim of this chapter was to determine the host range of *X. vasicola* strains isolated from maize, sugarcane, sorghum, and *E. grandis*. Two strains from each host were selected, and their identity verified by 16S rRNA and *rpoD* sequencing. The strains isolated from their specific hosts were grown on Yeast Dextrose Carbonate media at 28^{o}C for 48h. Thereafter the cells were harvested and suspended in phosphate-buffered saline and the concentration adjusted to 10^{7} CFU/ml. In the case of monocotyledonous plant species, the stem was cut, and 50µl of inoculum was pipetted onto the cut. In the case of the dicotyledon *E. grandis*, a 50µl of inoculum was infiltrated into the leaf lamina with a sterile insulin syringe. The wounds were covered with foil for 48h and observed over four weeks. *Eucalyptus grandis* and maize were susceptible to all the *X. vasicola* strains used in the study, irrespective of the host of origin of the strain. No symptoms were observed on the banana plants. The sorghum plants only showed symptoms when inoculated with sorghum- and maize-infecting *X. vasicola* strains. *X. vasicola* was successfully re-isolated from the leaves of all the hosts, including from asymptomatic banana plants. Therefore, the *X. vasicola* strains used in this study are pathogenic to four of the five hosts tested. They were not pathogenic to banana but could latently infect it.

2.2 Introduction

Xanthomonas species are Gram-negative straight rods which produce smooth, mucoid, and yellow colonies (Saddler & Bradbury, 2015). The characteristic yellow colour is due to the production of unique brominated aryl polyenes known as xanthomonadins (Saddler & Bradbury, 2015). *Xanthomonas* phytopathogens cause disease on approximately 400 monocotyledon and dicotyledon species, some of which are of economic importance (Ryan *et al.*, 2011, Jacques *et al.*, 2016). These phytopathogens usually have a restricted host range and are assigned pathovars which are designated according to tissue or host specificity (Dye *et al.*, 1980, Ryan *et al.*, 2011).

The taxonomy of plant pathogens is vital in understanding the disease and developing effective control measures. Three pathovars of *X. vasicola* are known, viz. *X. vasicola* pv. *vasculorum* (*Xvv*) isolated from maize and *Eucalyptus grandis*, *X. vasicola* pv. *holcicola* (*Xvh*) isolated from sorghum, and *X. campestris* pv. *musacearum* (*Xcm*) isolated from banana and ensete (Korus *et al.*, 2017). It has been suggested that *Xanthomonas vasicola pv. vasculorum* (*Xvv*), *Xanthomonas vasicola* pv. *holcicola* (*Xvh*), and *Xanthomonas campestris* pv. *musacearum* (*Xcm*) have a phylogenetically strong relationship based on fatty acid methyl ester analysis, genomic fingerprinting and partial nucleotide sequencing of the *gyrase* B gene (Parkinson *et al.*, 2007, Aritua *et al.*, 2008). Recent genomic data further established that *Xcm* is an evolutionary lineage belonging to the *X. vasicola* species (Studholme *et al.*, 2020).

Xanthomonas species gain entry into the host through natural openings such as hydathodes and stomata or through microscopic wounds which can be as a result of plants rubbing against each other (Boch & Bonas, 2010, Ryan *et al.*, 2011). The pathogens colonise the guttation droplets released at high humidity conditions and enter when the droplets are taken back into the plant when the humidity drops, or through the stomata during transpiration (Ryan *et al.*, 2011). Following entry, the pathogen proceeds to the vascular bundles, where it colonises the mesophyll cells or

the vascular bundles (Boch & Bonas, 2010, Ryan *et al.*, 2011). *X.vasicola* is one of the vascular *Xanthomonas* species, as it colonises the xylem of monocotyledon species such as maize, sugarcane, banana and ensete, as well as the dicotyledon *Eucalyptus grandis* (Coutinho & Wallis, 1991, Tushemereirwe *et al.*, 2004, Coutinho *et al.*, 2015). The pathogen causes bacterial leaf streak on maize and sugarcane, Xanthomonas bacterial wilt on banana and ensete, and bacterial blight and dieback on *E. grandis* where the symptoms include water-soaked lesions between the leaf veins, which may turn yellow, tan, brown or orange, followed by rapid wilting (Coutinho & Wallis, 1991, Coutinho *et al.*, 2002, Boch & Bonas, 2010).

The aim of this chapter was to compare the pathogenicity of *X. vasicola* strains isolated from different hosts the maize (*Zea mays*), sugarcane (*Saccharum* spp. hybrids), sorghum (*Sorghum bicolor*), banana (*Musa* spp.), and *Eucalyptus grandis*. The banana pathogen, *Xvm*, could not be included in the study as it is a quarantine organism in South Africa.

2.3 Materials and Methods

2.3.1 Verification of the identity of strains

2.3.1.1 Culturing of strains

Xanthomonas vasicola strains were obtained from the bacterial culture collection located at CMEG, University of Pretoria. The strains (Table 2.1) had been isolated from sugarcane, maize, sorghum and *Eucalyptus grandis*. The isolates were cultured on Yeast Dextrose Carbonate agar and incubated at 28^{o}C for 48h.

Table 2.1: *Xanthomonas vasicola* strains used in the inoculations of banana, *Eucalyptus grandis*, maize, sorghum, and sugarcane.

Host	Pathovar	Bacterial culture collection ID	LMG ID	NCPPB ID
Sugarcane	*vasculorum*	448	LMG 8744	-
Sugarcane	*vasculorum*	510	LMG 900	NCPPB 895
Maize	*vasculorum*	1689	SAM 113	SAM 113
Maize	*vasculorum*	169	LMG 8284	NCPPB 206
Sorghum	*holcicola*	516	LMG 7489	-
Sorghum	*holcicola*	777	LMG 736	NCPPB 2417
Eucalyptus	*vasculorum*	42	-	-
Eucalyptus	*vasculorum*	J5	-	-

2.3.1.2 DNA extractions

The Zymo Research DNA *Quick*-gDNA^{TM} Miniprep kit was used to extract DNA from all the bacterial strains as per manufacturer's instructions.

A 1% agarose gel (Lonza) was prepared using 1X Tris-acetate-EDTA (TAE) buffer. A mixture of 1µl GelRed and 4µl of the sample was loaded onto the gel and electrophoresis was done at 80 Volts for 30 minutes. The gel was then viewed under UV light (BioRad) to verify the presence of the DNA.

2.3.1.3 Polymerase Chain Reaction

For the 16S rRNA region, universal primers were used (Table 2.2). Each 25µl reaction was as follows: 2.5µl of 10X reaction buffer (Supertherm), 2µl (25mM) MgCl (Supertherm), 2µl of (250 µM) dNTPs (Fermentas), 0.5µl (5 pmol) of the forward and reverse primer (Inqaba Biotec), 0.3µl (5U/µl) Taq polymerase (Supertherm), 16.2µl nuclease-free water (WhiteSci), and 1µl genomic DNA (50-100ng). Amplification was done using a BioRad T1000 thermocycler. The reaction conditions were as follows: initiation at 94^0C for ten minutes followed by 30 cycles of denaturation at 92^0C for one minute, annealing at 60^0C for one minute, extension at 72^0C for one minute and a final extension step at 72°C for 7 minutes.

For the *rpoD* gene amplification, synthesised primers were used (Table 2.2). Each 25µl reaction was as follows: 2.5µl of 10X reaction buffer (Supertherm), 2µl (25mM) MgCl (Supertherm), 2µl (250 µM) dNTPs (Fermentas), 0.5µl (5 pmol) of the forward and reverse primer (Inqaba Biotec), 0.3µl (5U/µl) Taq polymerase (Supertherm), 16,2µl nuclease-free water (WhiteSci), and 1µl genomic DNA (50-100ng). Amplification was done using a BioRad T1000 thermocycler. The reaction conditions were as follows: initiation at 94^0C for 10 minutes followed by 34 cycles of denaturation at 92^0C for one minute, annealing at 55^0C for one minute and an extension at 72^0C for one minute and a final extension step at 72°C for 7 minutes.

Table 2.2: Primer sequences of 16S rRNA and *rpoD* genes.

Name	Sequence (5'→3')
16S rRNA Forward (27F)	AGAGTTTGATCMTGGCTCAG
16S rRNA Reverse (1492R)	GGTTACCTTGTTACGACTT
rpoD 01 Forward	TGGAACAGGGTATCTGACC
rpoD 02 Reverse	CATTCYAGGTTGGTCTGRTT

2.3.1.4 Sequencing Reaction

A total volume of 12µl of the sequencing reaction was made up of the following: 0.5µl of 5X sequencing buffer (Applied Biosystems), 1µl BigDye Terminator (v3.1 ABI PRISM®, Applied Biosystems), 0.5µl of 5pmol of either primer, 8µl dH_2O (WhiteSci)

and 0,5µl cleaned PCR reaction (50-100ng). The sequencing reaction conditions were as follows: initiation at 96^0C for five seconds followed by 25 cycles of denaturation at 96^0C for 10 seconds, annealing at 55^0C for 55 seconds, extension at 60^0C for four minutes and a final extension step at 60°C for 7 minutes.

2.3.1.5 Ethanol precipitation

Each ethanol precipitation reaction was made up of the following: 2µl of 3M Sodium Acetate (NAOAc, pH 4.8), 20µl absolute ethanol (Merck) and 6µl of the sequencing reaction. The samples were centrifuged using an Eppendorf 5424 R centrifuge at maximum speed for 30 minutes. The supernatant was removed, and the sample washed twice with 70% ethanol (Merck). The samples were centrifuged using an Eppendorf 5424 R centrifuge at max speed for 10 minutes and the supernatant removed. The samples were dried using a Techne Dri-Block® DB-2D heating block at 90^0C for five minutes.

2.3.1.6 Sequencing and identification

The samples were sequenced at the SeqServe sequencing facility at the University of Pretoria using an ABI PRISM® DNA Automated Sequencer (Perkin-Elmer). The sequences were edited using the BioEdit Sequence Alignment Editor v7.2.6.1 (Madden, 2002, Hall, 2011) and resulting sequences were identified using the National Centre for Biotechnology Information (NCBI) nucleotide blast (Altschul *et al.*, 1990, Madden, 2002).

2.3.2 Pathogenicity experiments

Fourteen plants of each of the five host plants (banana, *E. grandis*, maize, sorghum and sugarcane) were inoculated with each *X. vasicola* strain. Control plants were inoculated with sterile water. There were two technical replicates of each.

The *X. vasicola* strains were grown on YDC agar at 28ºC for 48h and resuspended in 1X phosphate-buffered saline (PBS), and the concentration was adjusted to 10^7 CFU/ml. For the inoculation of monocotyledon plant species, the stem was cut, and 50µl was pipetted onto the cut and covered with foil for 48h. In the case of *E. grandis*, a 50µl of inoculum was infiltrated into the leaf lamina with a sterile insulin syringe. The plants were placed in a tunnel at 28ºC and watered daily. The experiments were terminated after four weeks. The number of plants showing symptoms was determined by visually inspecting plants showing symptoms and counting the number infected plants and severity of these observed symptoms was scored using the symptoms severity scales by Karamura *et al*., (2015) (Table 2.3).

Table 2.3: The symptoms severity scales by Karamura *et al*., (2015) used to score the severity of symptoms on the plants at the end of the pathogenicity trials.

Plant	Severity scale	Definition
Banana	0	No visible symptoms
	1	Slight wilting/ folding of leaves
	2	Pronounced wilting/ yellowing of most leaves
	3	Pronounced necrosis of the whole plant
	4	Complete death, rotting of the plant
Eucalyptus	0	No visible symptoms
	1	Necrosis of leaves
	2	Necrosis and wilting
	3	Discolouration and death of leaves
Maize	0	No visible symptoms
	1	White-soaked like streaks
	2	Yellow, brown or white streaks
	3	Brown lesions
	4	Deformation of plant/stunted growth
Sugarcane and sorghum	0	No visible symptoms
	1	White streaks/ lesions
	2	Reddish-brown streaks/lesions
	3	Yellow streaks/lesions

At the end of the four weeks, the leaves were collected, including leaves of the negative control plants. Leaves representing each isolate from each plant host were selected, and six 5mm x 5mm pieces of the leaf tissue including the leading edges were crushed using a mortar and pestle in 1X PBS. The maceration was allowed to stand for 10min at room temperature, and 100µl was spread plated onto YDC agar and grown at 28^{o}C for 48h.

To confirm the identity of the re-isolated strains, the *rpoD* gene was sequenced as described above.

To determine the significance of the differences observed, a 2-way ANOVA was performed for the monocotyledon species, and a one-way t-test for *Eucalyptus grandis*. The statistical tests were performed using GraphPad Prism version 8.2.1 for Windows, GraphPad Software, San Diego, California USA (www.graphpad.com).

2.4 Results

2.4.1 Verification of the identity of strains

The strains BCC448, 510,1689,169, 42, and J5 had a 16S rRNA similarity of ≥99% identity and *rpoD* similarity of ≥99% identity to *X. vasicola* pv. *vasculorum*. The strains BCC516 and 777 had a 16S rRNA similarity of ≥99% identity and *rpoD* similarity of ≥99% identity to *X. vasicola* pv. *holcicola* (refer to Chapter 3 for whole-genome phylogeny).

2.4.2 Pathogenicity experiments

2.4.2.1 Banana (cv. Cavendish)

The banana plants showed no symptoms throughout the four weeks of observation (Fig. 2.1). The strains of *X. vasicola* pathovars that were inoculated into the plants were reisolated from the asymptomatic leaves at the end of the observation period.

Figure 2.1: Banana plants at the end of four-week observation. The plants showed no symptoms and remained healthy over the period.

2.4.2.2 *Eucalyptus grandis* (cv. Tag14)

Because of leaf infiltration difficulties, only one replicate resulted in symptoms. All inoculated *E. grandis* plants showed symptoms. The control plants were asymptomatic. The percentage of plants showing symptoms was >60% for all the plants inoculated with the different *X. vasicola* strains (Fig 2.3). The *X. vasicola* strains significantly differed in their ability to induce symptoms on *E. grandis* ($P<0.0001$), with J5 and BCC42 being the most virulent and aggressive strains (Fig 2.3 and 2.4).The least virulent strains were the *Xvh* BCC516 and 777, and the *Xvv* BCC169 (isolated from maize). The plants inoculated with *Xvv* strains from *E. grandis* showed the most severe disease symptoms with a severity score of 3 (Figs 2.2B and 2.4), with leaves becoming discoloured, drying and falling off. The symptoms observed on the plants inoculated with *Xvv* from maize, sugarcane and *Xvh* were necrotic lesions on the leaves without wilting with a symptoms severity score of 1 (Figs 2.2A and 2.4). The variation observed between the severity of the symptoms on the plants inoculated with the different *X. vasicola* strains showing symptoms was significant ($P=0,0039$).

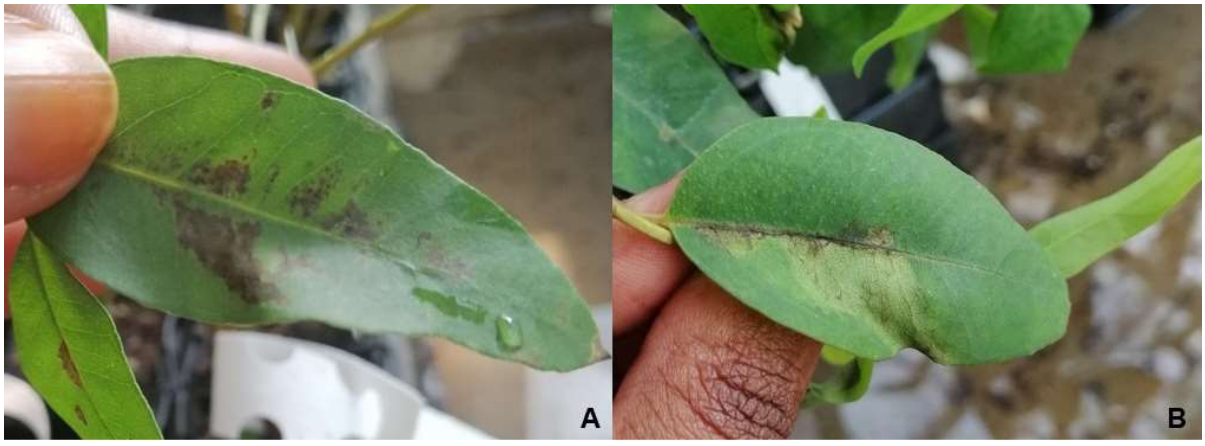


Figure 2.2: The symptoms observed on *E. grandis* plants following inoculation with *Xvv* and *Xvh* strains. A) The necrotic lesions observed on plants inoculated with *Xvv* strains from maize, sugarcane, and *Xvh* from sorghum. B) The whitening and drying out of plants inoculated with *Xvv* from *E. grandis*.

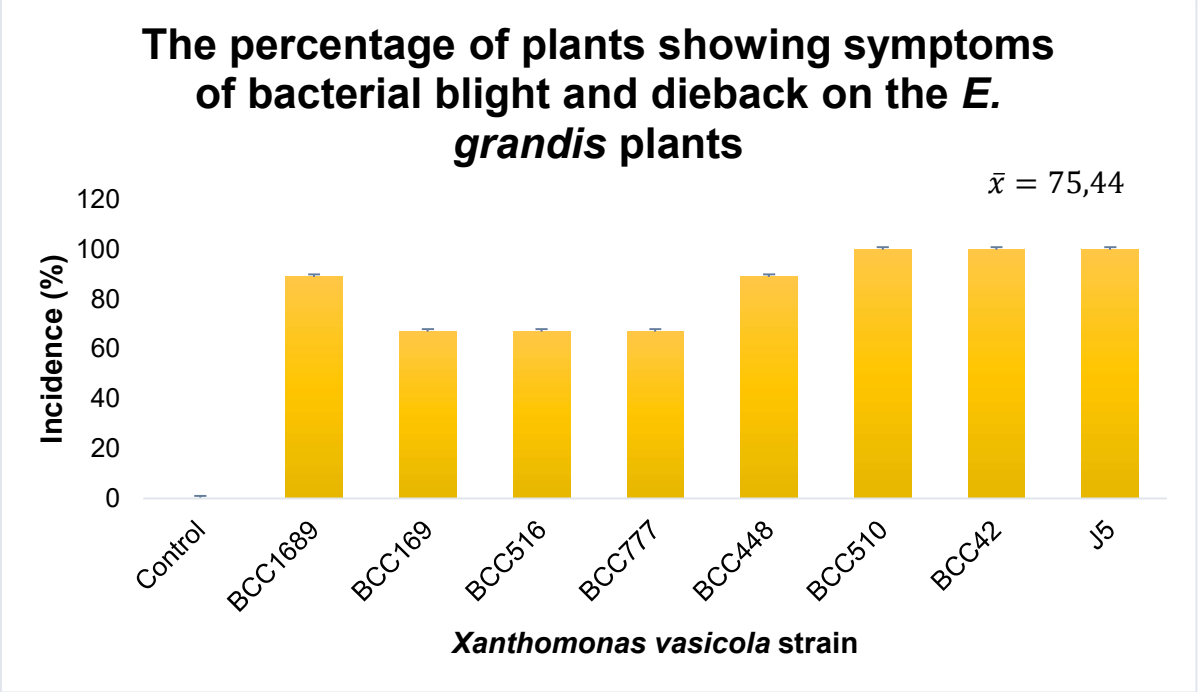


Figure 2.3: The percentage of plants showing symptoms of bacterial blight and dieback caused by *X. vasicola* strains. Fourteen *Eucalyptus grandis* plants were inoculated with each *X. vasicola* strain. The variation of the bacterial blight and dieback incidences observed between the *X. vasicola* strains was significant (P<0,0001). The error bars represent the standard deviation of each replicate. The strains BCC516 and 777 are *Xvh* whereas all others are *Xvv*.

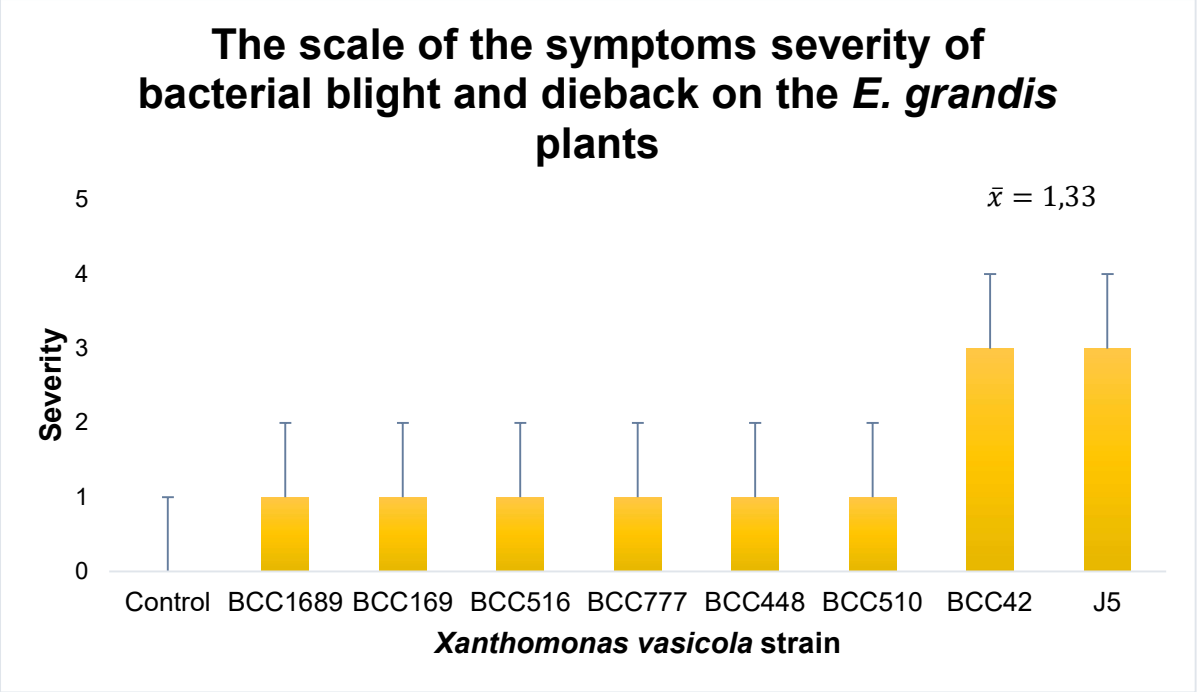


Figure 2.4: The severity of bacterial leaf streak and dieback symptoms on *Eucalyptus grandis* plants. The variation of the severity of bacterial blight and dieback symptoms observed between plants inoculated with the different *X. vasicola* strains was significant (P=0,0039). The error bars represent the standard deviation of each replicate.

2.4.2.3 Maize (cv. Zama star)

All inoculated maize plants showed symptoms. The control plants were asymptomatic. The percentage of plants showing symptoms was >55% (Fig 2.6). The symptoms observed on plants inoculated with *Xvh* and *Xvv* from sugarcane were yellow/brown streaks along the veins and a symptoms severity score of 2 (Figs 2.5A,B and 2.7). The symptoms observed on the plants inoculated with *Xvv* from maize and *E. grandis* were brown streaks along the veins, and stunting and a symptoms severity score of 4 (Figs 2.5C,D and 2.7). The variation observed between the number of infected plants between the two replicates was not significant (P=0.7428), and the different *X. vasicola* strains differed significantly in their ability to induce symptoms (P<0.0001). The variations of the severity of symptoms on the plants between the two replicates were not significant (P>0.9999), and the different *X. vasicola* strains differed significantly in their severity on maize (P<0.0001).

Figure 2.5: The symptoms observed on the maize plants at the end of the four-week observation. A & B) The yellow/brown streaks observed on *Xvv* from sugarcane and *Xvh*. C & D) The brown streaks and stunting/death of the plants inoculated with *Xvv* from maize and *E. grandis*.

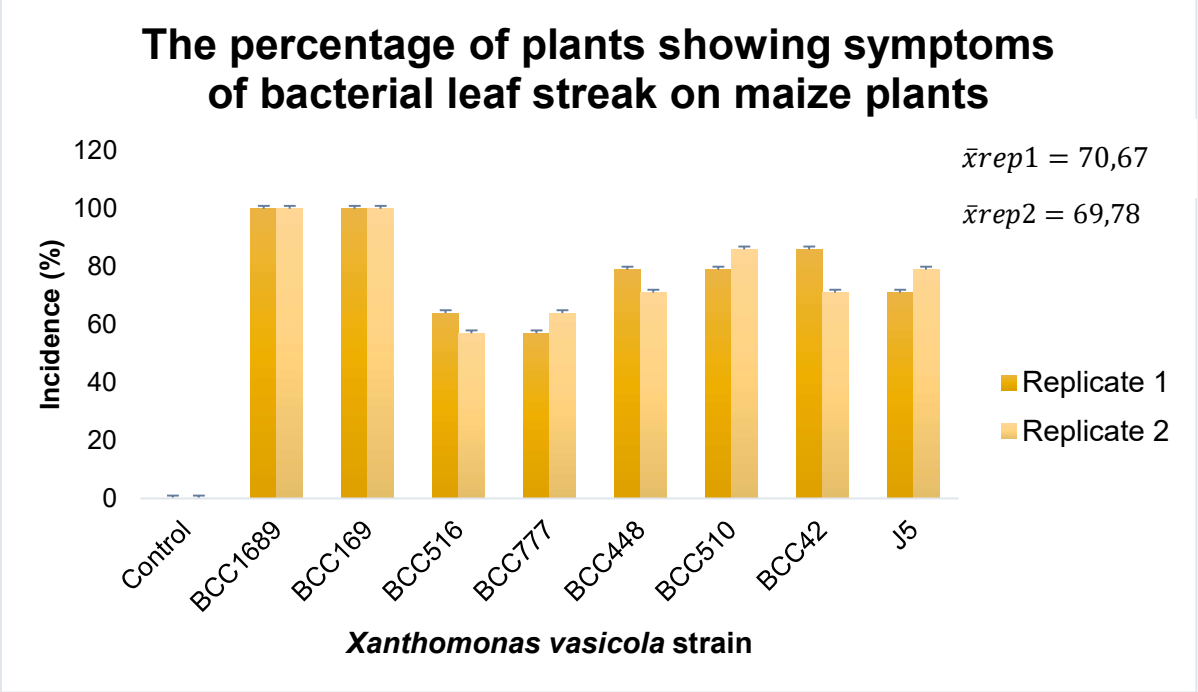


Figure 2.6: The percentage of plants showing symptoms of bacterial leaf streak of maize caused by *X. vasicola* strains. Fourteen maize plants were inoculated with each *X. vasicola* strain. Two technical replicates were done. The variation observed between the technical replicates was not significant (P=0.7428). The variation of the bacterial leaf streak incidences observed between the *X. vasicola* strains was significant (P<0,0001). The error bars represent the standard deviation of each replicate.

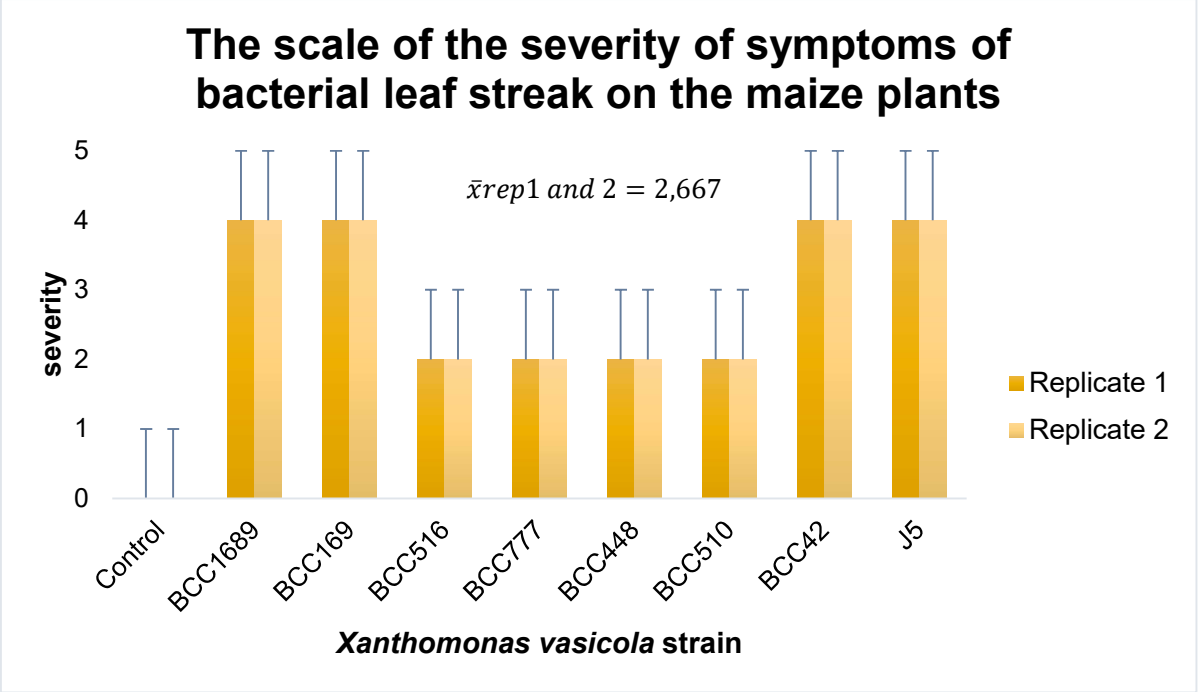


Figure 2.7: The severity of bacterial leaf streak symptoms on maize plants. The variation of the severity of bacterial leaf streak symptoms observed between the two replicates was not significant (P > 0,9999). The variation of the severity of bacterial leaf streak symptoms observed between plants inoculated with the different *X. vasicola* strains was significant (P<0,0001). The error bars represent the standard deviation of each replicate.

2.4.2.4 Sugarcane (cv. N29 and NCo376)

Only one replication (14 plants per *X. vasicola* strain) of each cultivar was available for the inoculations. All inoculated sugarcane plants showed symptoms. The control plants were asymptomatic. With the cultivar N29 cultivar plants, the percentage of plants showing symptoms was >75%, whereas the percentage of plants inoculated with *Xvh* showing symptoms was <25% (Fig 2.9). The cultivar NCo376 plants showed a similar trend, but there were fewer plants showing symptoms. The symptoms observed on the NCo376 plants inoculated with *Xvh* and *Xvv* from *E. grandis* and one maize strain (BCC169) were white streaks along the veins and a symptoms severity score of 1 (Figs 10A and 12). The symptoms observed on the N29 plants inoculated with *Xvh* and *Xvv* were reddish/brown streaks along the veins and a symptoms severity score of 2 (Figs 2.8B,C and 2.10). These plants inoculated with *Xvv* from sugarcane showed the most severe symptoms of yellow streaks and the death of

some plants and a severity score of 3 (Figs 2.8D and 2.10). The cultivar N29 was significantly more susceptible than NCo376, based on the percentage of symptomatic plants (P=0.0331) and the severity of symptoms observed (P=0.0222). The *X. vasicola* strains differed significantly in their ability to induce symptoms (P=0.0222), and the severity of the symptoms observed (P=0.0062).

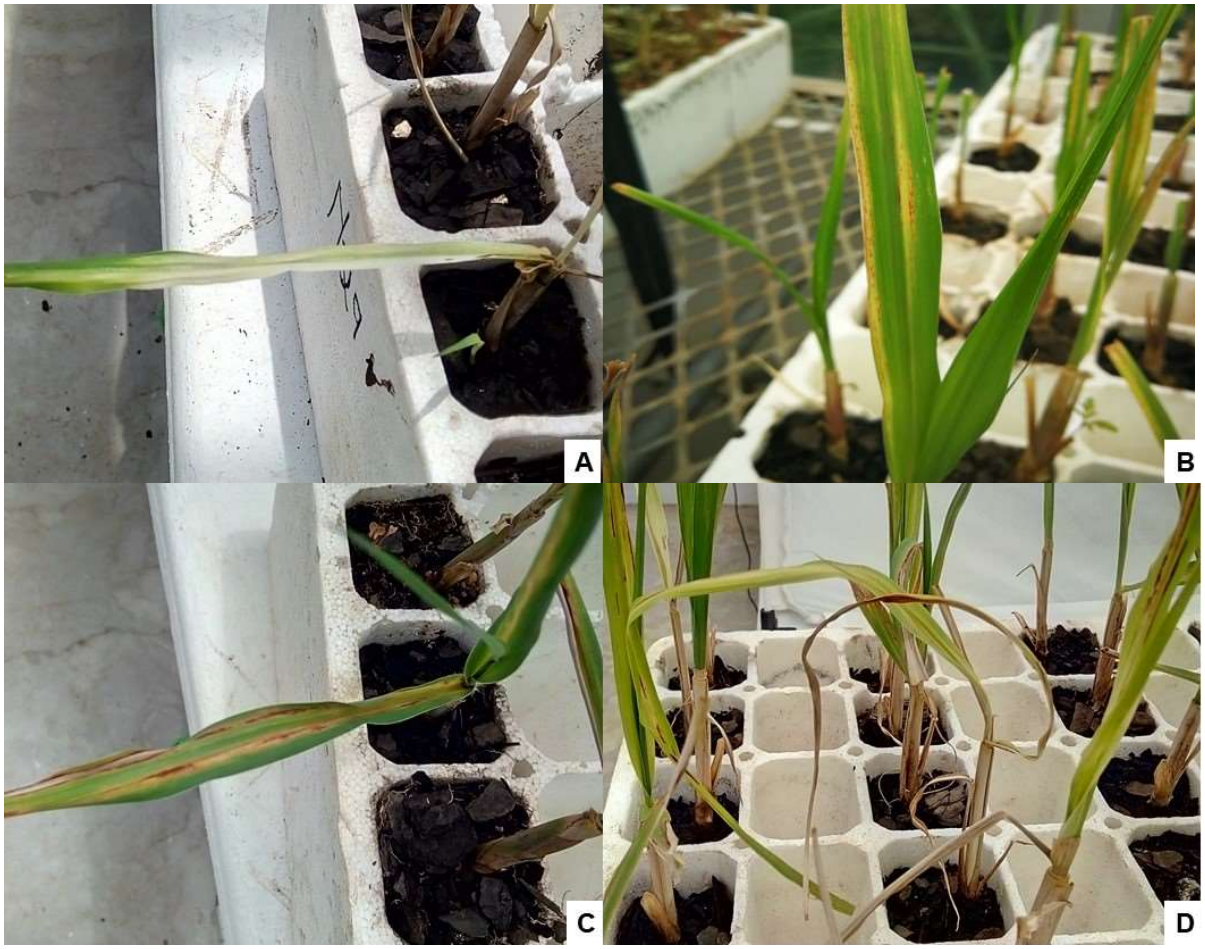


Figure 2.8: Disease symptoms on sugarcane plants four weeks post-inoculation. A) White streaks observed on the NCo376 plants inoculated with *Xvh* and *Xvv* from *E. grandis* and the maize strain BCC169. B & C) Reddish-brown streaks observed on N29 plants inoculated with *Xvv* and *Xvh*. D) Yellow streaks and death of plants inoculated with *Xvv* from sugarcane and maize (BCC1689).

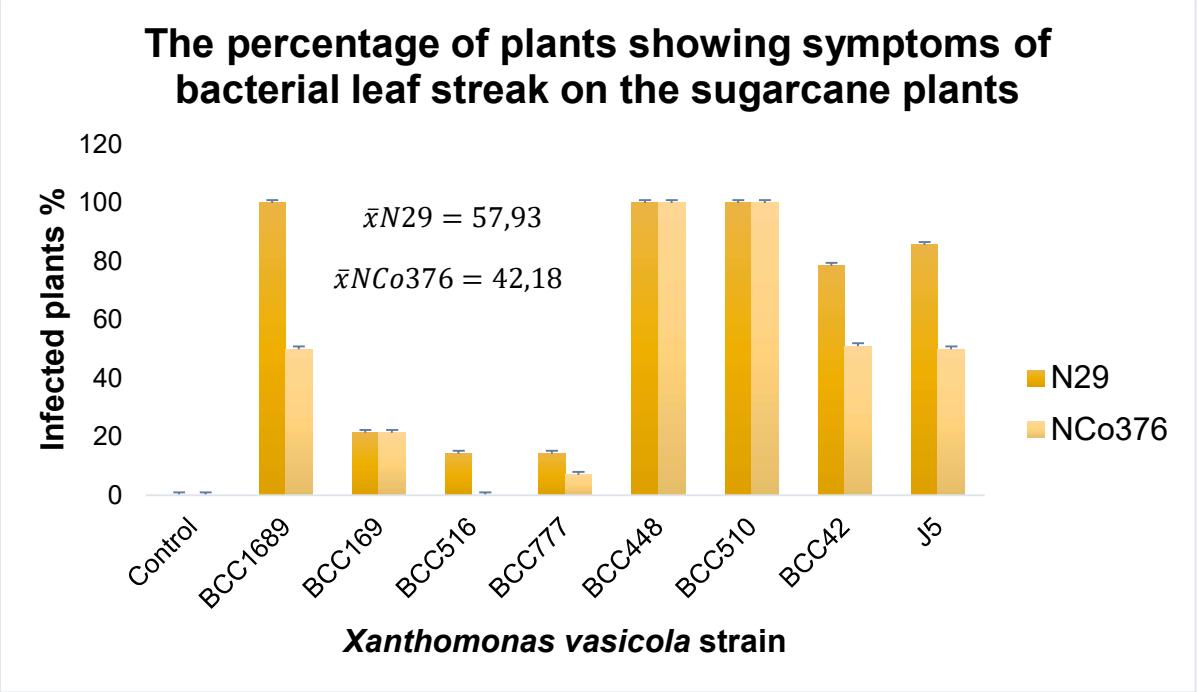


Figure 2.9: The percentage of plants showing symptoms of bacterial leaf streak of sugarcane caused by *X. vasicola* strains. Fourteen sugarcane plants were inoculated per *X. vasicola* strain. The strains were inoculated on two sugarcane cultivars (N29 and NCo376). The variation observed between the two cultivars was significant (P=0,0331). The variation of the number of plants showing symptoms observed between the *X. vasicola* strains was significant (P=0,0002). The error bars represent the standard deviation of each replicate.

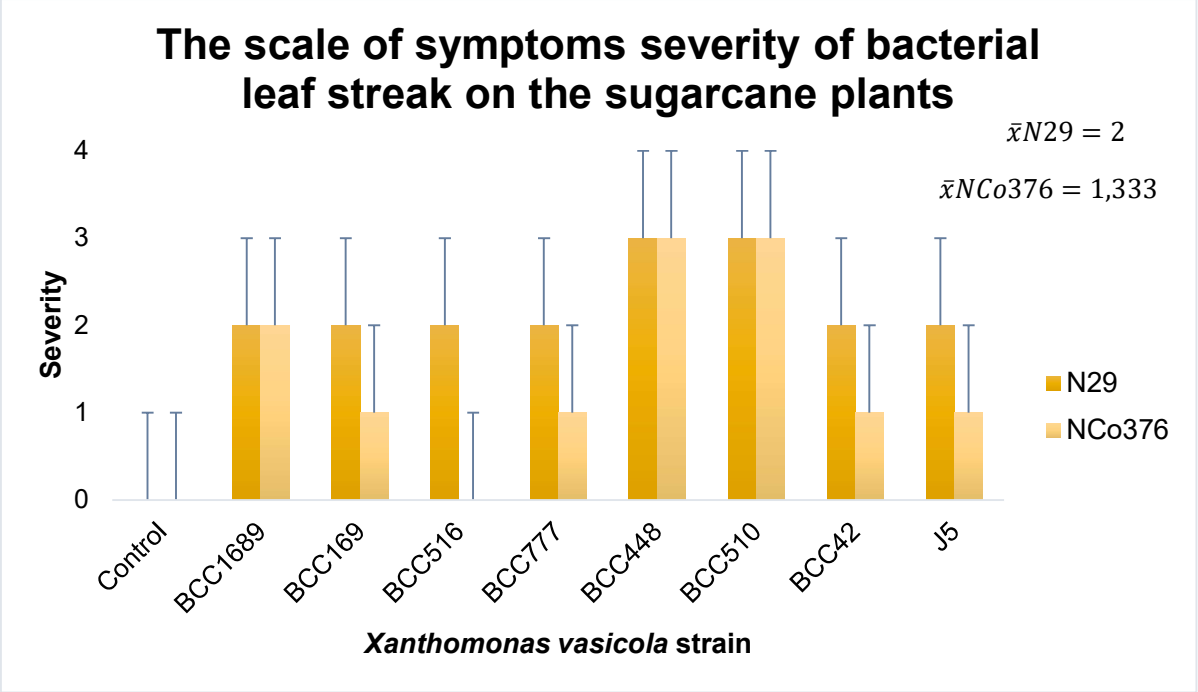


Figure 2.10: The severity of bacterial leaf streak symptoms on sugarcane plants. The variation of the severity of bacterial leaf streak symptoms observed between the two cultivars (N29 and NCo376) was significant (P=0,0222). The variation of the severity of bacterial leaf streak symptoms observed between plants inoculated with the different *X. vasicola* strains was significant (P=0,0062). The error bars represent the standard deviation of each replicate.

2.4.2.5 Sorghum (cv. Mr Buster)

The percentage of symptomatic plants inoculated with *Xvv* from maize and *Xvh* was >50% (Fig 2.12), whereas the other plants inoculated with *Xvv* from sugarcane and *E. grandis* showed no symptoms. The symptoms observed on the plants inoculated with *Xvv* from maize were reddish/brown streaks along the veins and a symptoms severity score of 2 (Figs 2.11A and 13). The plants inoculated with *Xvh* showed the most severe symptoms of yellow streaks and the deformation of the plants and a severity score of 3 (Figs 2.11B and 2.13). The variation observed between the number of plants showing symptoms between the two replicates was not significant (P=0.3466), and the variation observed on the plants showing symptoms inoculated with the different *X. vasicola* strains was significant (P<0.0001). The variation of the severity of symptoms on the plants between the two replicates was not significant (P>0.9999), and the variation of the severity of symptoms between the plants inoculated with the

different *X. vasicola* strains was significant (P<0.0001).). As expected, the *Xvh* strains BCC516 and 777 were the most virulent and aggressive, followed by the *Xvv* BCC1689, then 169. The sugarcane and *E. grandis Xvv* strains were not pathogenic on sorghum.

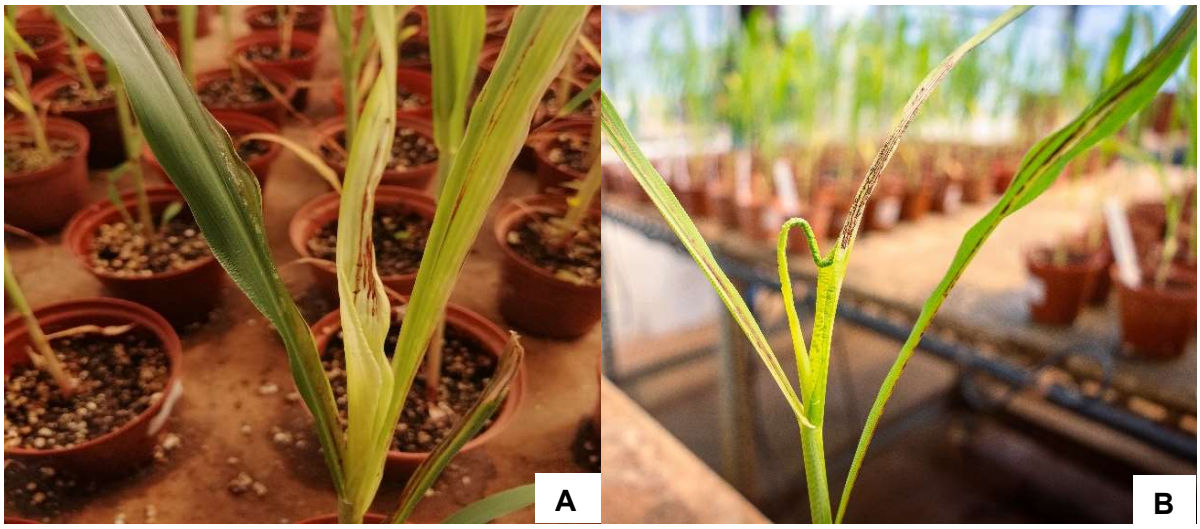


Figure 2.11: The disease symptoms observed at the end of the four-week observation. A) The reddish-brown streaks observed on the plants inoculated with *Xvv* from maize. B) The reddish-brown and deformation on the plants inoculated with *Xvh*.

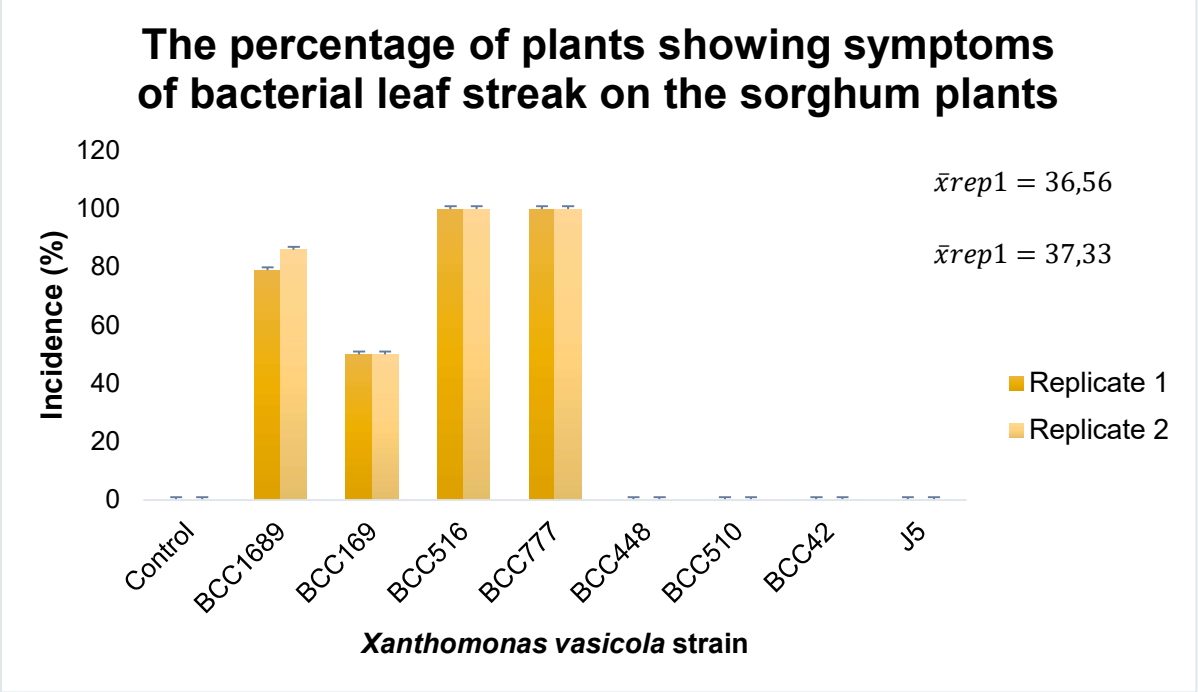


Figure 2.12: The percentage of plants showing symptoms of bacterial leaf streak of maize caused by *X. vasicola* strains. Fourteen sorghum plants were inoculated per *X. vasicola* strain. Two technical replicates were done. The variation observed between the technical replicates was not significant (P=0,466). The variation of the number of plants showing symptoms observed between the *X. vasicola* strains was significant (P<0,0001). The error bars represent the standard deviation of each replicate.

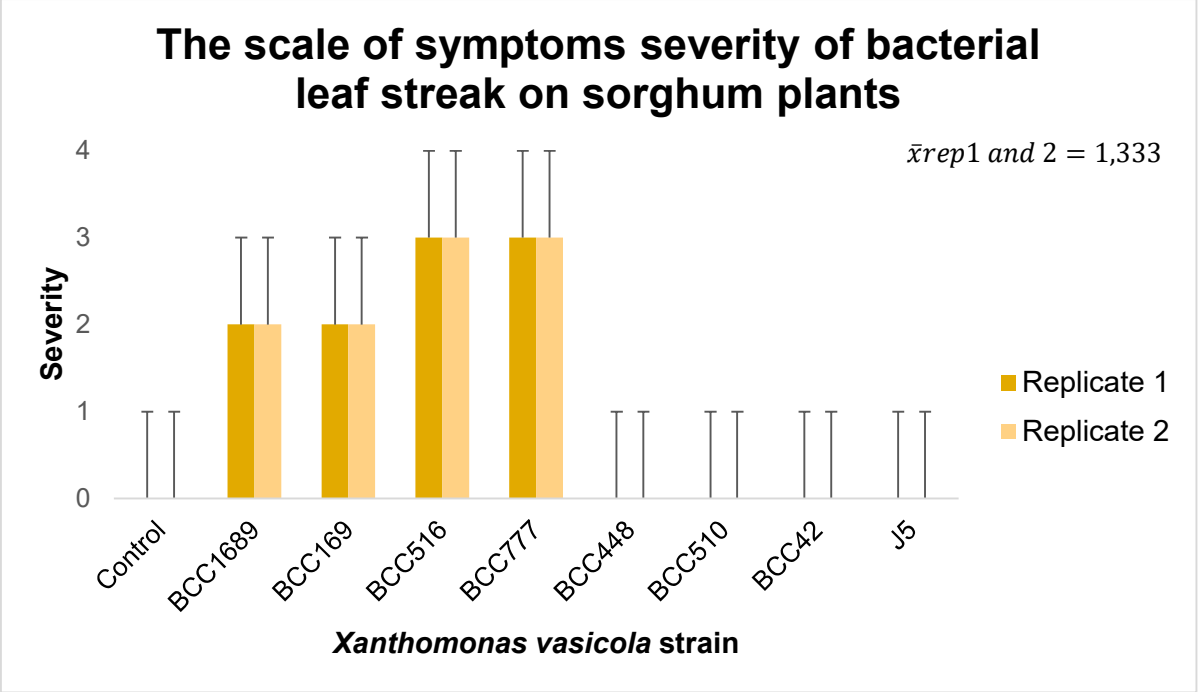


Figure 2.13: The severity of bacterial leaf streak symptoms on sorghum plants. The variation of the severity of bacterial leaf streak symptoms observed between the two replicates was not significant (P > 0,9999). The variation of the severity of bacterial leaf streak symptoms observed between plants inoculated with the different *X. vasicola* strains was significant (P<0,0001). The error bars represent the standard deviation of each replicate.

2.4 Discussion

In this study, the aim was to determine the pathogenicity of *X. vasicola* strains isolated from different hosts in an attempt to determine how wide this host-range was. The three known pathovars of *X. vasicola*, viz. *Xvv*, *Xvh* and *Xvm* were named based on their tissue or host specificity (Ryan *et al.*, 2011, Studholme *et al.*, 2020). The results obtained from this study showed that the *X. vasicola* isolated from the different hosts is able to cross-infect and cause similar symptoms, although each strain was most aggressive on its host of isolation.

The inoculation of banana plants with *X. vasicola* pathovars *vasculorum (Xvv)* and *holcicola (Xvh)* isolated from *E. grandis*, maize, sugarcane, as well as sorghum respectively did not result in disease occurrence. No symptoms were visible, but *Xvv* and *Xvh* were reisolated from the asymptomatic leaves, indicating that these strains can spread within the plant through vessels from the inoculation point. This observation was previously reported when banana plants were inoculated with *Xvm* and *Xvv*. The plants inoculated with *Xvm* developed necrotic leaf symptoms as well as wilting after three weeks, with those inoculated with *Xvv* remaining asymptomatic and healthy (Karamura *et al.*, 2015). The persistence of *X. vasicola* in the asymptomatic leaves, therefore, suggests a latent infection of the plants, which has been observed on banana plants even when infected with *Xvm* (Karamura *et al.*, 2015, Ocimati *et al.*, 2015). This may be due to an equilibrium state reached between the host and pathogen, and this state will remain as is until a significant change in the environmental conditions which weaken the host's defences enough for the pathogen to reestablish itself and cause disease occurs (Kado, 2010).

The inoculation of *E. grandis* plants with *Xvv* and *Xvh* isolates resulted in symptoms being observed on all the plants. The presence of symptoms shows that *E. grandis* is susceptible to *Xvv* strains from maize and sugarcane, and *Xvh* strains from sorghum. The symptoms on the plants inoculated with *Xvv* strains from sugarcane confirm the observation made by Coutinho *et al*., (2015) and it was suggested that a host jump

event might have occurred from the surrounding sugarcane fields to *E. grandis* plantation. The novel finding in this study is that *E. grandis* is susceptible to *Xvv* and *Xvh* from maize and sorghum respectively, but the severity of the symptoms was lower than that observed with the strains from *E. grandis*. This finding suggests that these crops may act as a source of inoculum. The susceptibility of *E. grandis* to non-host *Xvv* and *Xvh,* which infect monocotyledons suggests that the strains gained genes which enabled them to overcome the *E. grandis*' defence repertoire (refer to chapter 3).

The pathogen *Staphylococcus aureus* is a pathogen which has the ability to infect multiple hosts such as humans, livestock and birds. This ability to jump hosts can be attributed to the presence of host-specific accessory genes, a gain or a loss of certain gene functions and a modification of biological pathways to better respond to the change in environment (Richardson *et al.*, 2018). A pan-genome-wide analysis revealed accessory genome clusters specific for hosts such as birds and pigs, as well as a plasmid which confers resistance to heavy-metal ions which is unique to *S. aureus* isolated from pigs (Richardson *et al.*, 2018). The *Xvv* and *Xvh* strains may have acquired similar genetic elements which allowed them to adapt to a dicotyledon host like *E. grandis,* which provides a more complex environment than those of their monocotyledon hosts.

The maize plants were susceptible to all *X. vasicola* strains. Interestingly, the plants inoculated with *Xvv* strains from sugarcane and *Xvh* showed similar symptoms of yellow/brown lesions and this observation is similar to that previously made by Karamura *et al*., (2015). The second type of symptoms consisted of brown streaks and stunting were observed on maize plants inoculated with *Xvv* strains from maize and *E. grandis.* This was contrary to previous observations where stunting was observed on maize plants inoculated with strains from sorghum (*Xvh*), and no deformation was observed on plants inoculated with *Xvv* from sugarcane (Karamura *et al.*, 2015).

These differences may be attributed to differing virulence levels of the different strains in the two studies, and this may be due to minute differences in their genomes. In a study on the potato pathogen, *Dikeya solani,* the differences observed in the different

degree of virulence of two strains were shown to be due to small but significant differences in the production of virulence factors such as pectinases and proteases, as well as the presence of prophages (Golanowska *et al.*, 2018). Similarly, the symptomatology differences between Karamura *et al*., (2015) and this study, could be explained by subtle differences in the *Xvv* gene content, notably in genetic factors involved in stunting and leaf deformation. The novel finding was the severe stunting and death of the plants caused by *Xvv* from *E. grandis.* These severe symptoms may be due to additional virulence factors obtained by the strains, which allowed them to cause disease on a more complex dicotyledon, *E grandis*, compared to maize which is a monocotyledon.

The two cultivars of the sugarcane plants showed symptoms after being inoculated with *Xvv* and *Xvh*. The cultivar N29 appeared to be globally more susceptible than NCo376. This, therefore suggests that the NC0376 cultivar is more tolerant of both *Xvv* and *Xvh* strains. The symptoms caused by *Xvv* and *Xvh* on the sugarcane plants are similar to each other, and this may be due to the similarities present within the genomes of these pathogens (Dookun *et al.*, 2000, Karamura *et al.*, 2015, Studholme *et al.*, 2020). The differences in the aggressiveness of the maize *Xvv* strains BCC1689 and BCC169, as well as *Xvv* strains from *Eucalyptus,* may be due to changes which may have occurred in the genome such as the acquisition of virulence factors such as pectinases and proteases through horizontal gene transfer (Golanowska *et al.*, 2018).

Only the sorghum plants inoculated with *Xvv* from maize and *Xvh* from sorghum showed symptoms. The symptoms caused by the *Xvv* strain and *Xvh* were similar, and this may be attributed to their genetic similarities (Aritua *et al.*, 2008, Studholme *et al.*, 2020). Artificial inoculation of sorghum with *Xvv* and *Xvh* has been unsuccessful in other pathogenicity experiments. When inoculated with *Xvv* from sugarcane, *Xvm* from banana and *Xvh*, symptoms were not observed even when different inoculation techniques were used (Navi *et al.*, 2002, Karamura *et al.*, 2015).

2.5 Conclusion

In this study, the aim was to determine the pathogenicity of *X. vasicola* pathovars *vasculorum* and *holcicola* on banana, *E. grandis*, maize, sorghum and sugarcane. The results of the pathogenicity of *Xvv* and *Xvh* strains on banana, maize, sorghum and sugarcane are broadly consistent with previous studies. Banana plants remained healthy but could carry latent infections of *Xvv* and *Xvh*, and maize plants are susceptible to *Xvv* and *Xvh*. Sugarcane plants are highly susceptible to *Xvv* and mildly susceptible to *Xvh*, and sorghum is highly susceptible to *Xvh* strains and maize-isolated *Xvv* strains while not susceptible to other *Xvv* strains. The most important finding of the study is the susceptibility of *E. grandis* to *Xvv* from maize and sugarcane, as well as *Xvh* from sorghum. This suggests that the *X. vasicola* pathovars are able to infect hosts which are phylogenetically distinct to their native hosts and that other crops may act as reservoirs for the pathogens.

The potential ability of *X. vasicola* pathovars to jump or shift hosts is due to genetic changes that may have occurred. These genetic changes will be studied in the next chapter.

Future studies should consider including more *Xanthomonas* species which infect plants which commonly occur in the vicinity of crops of economic importance. This will pre-empt possible disease outbreaks, allowing preventative measures to be taken. It is essential to determine whether the *X. vasicola* strains remain virulent and aggressive following the inoculation of the banana plants. More inoculation techniques should be explored for *E. grandis* and sorghum to increase the successful infection of *Xvv* and *Xvh.*

2.6 References

Altschul SF, Gish W, Miller W, Meyers EW & Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.

Aritua V, Parkinson N, Thwaites R, Heeney JV, Jones DR, Tushemereirwe W, Crozier J, Reeder R, Stead DE & Smith J (2008) Characterization of the *Xanthomonas* sp. causing wilt of enset and banana and its proposed reclassification as a strain of *X. vasicola*. *Plant Pathology* **57**: 170-177.

Boch J & Bonas U (2010) *Xanthomonas* AvrBs3 family-type III effectors: Discovery and function. *Annual Review of Phytopathology* **48**: 419-436.

Coutinho TA & Wallis FM (1991) Bacterial Streak Disease of Maize (Zea mays L.) in South Africa. *Journal of Phytopathology* **133**: 112-112.

Coutinho TA, van der Westhuizen L, Roux J, McFarlane SA & Venter SN (2015) Significant host jump of *Xanthomonas vasicola* from sugarcane to a *Eucalyptus grandis* clone in South Africa. *Plant Pathology* **64**: 576-581.

Coutinho TA, Preisig O, Mergaert J, Cnockaert MC, Riedel K-H, Swings J & Wingfield MJ (2002) Bacterial blight and dieback of *Eucalyptus* species, hybrids, and clones in South Africa. *Plant Disease* **86**: 20-25.

Dookun A, Stead DE & Autrey LJ-C (2000) Variation among strains of *Xanthomonas campestris* pv. *vasculorum* from Mauritius and other countries based on fatty acid analysis. *Systematic and applied microbiology* **23**: 148-155.

Dye D, Bradbury J, Goto M, Hayward A, Lelliott R & Schroth M (1980) International standards for naming pathovars of phytopathogenic bacteria and a list of pathovar names and pathotype strains. *Review of plant pathology* **59**: 153-168.

Golanowska M, Potrykus M, Motyka-Pomagruk A*, et al.* (2018) Comparison of highly and weakly virulent *Dickeya solani* strains, with a view on the pangenome and panregulon of this species. *Frontiers in Microbiology* **9**: 1-19.

Hall T (2011) BioEdit: An important software for molecular biology. *GERF Bulletin of Biosciences* **2**: 60-61.

Jacques M-A, Arlat M, Boulanger A*, et al.* (2016) Using ecology, physiology, and genomics to understand host specificity in *Xanthomonas*. *Annual Review of Phytopathology* **54**: 163-187.

Kado CI (2010) *Plant bacteriology*. The American Phytopathological Society, St. Paul, Minnesota 55121, USA.

Karamura G, Smith J, Studholm D, Kubiriba J & Karamura E (2015) Comparative pathogenicity studies of the *Xanthomonas vasicol*a species on maize, sugarcane and banana. *African Journal of Plant Sciences* **9**: 385-400.

Korus K, Lang JM, Adesemoye AO, Block CC, Pal N, Leach JE & Jackson-Ziems TA (2017) First report of *Xanthomonas vasicola* causing bacterial leaf streak on corn in the United States. *Plant Disease* **101**: 1030.

Madden T (2002) The BLAST sequence analysis tool. *The NCBI Handbook,*(McEntyre J & Ostell J, eds.), p.^pp. National Center for Biotechnology Information (US), Bethesda (MD).

Navi S, Bandyopadhyay R, Devi K & Reddy D (2002) Bacterial leaf streak of sorghum-a new report from India. *International Sorghum and Millets Newsletter* **43**: 61-63.

Ocimati W, Nakato GV, Fiaboe KM, Beed F & Blomme G (2015) Incomplete systemic movement of *Xanthomonas campestris* pv. *musacearum* and the occurrence of latent infections in Xanthomonas wilt-infected banana mats. *Plant Pathology* **64**: 81-90.

Parkinson N, Aritua V, Heeney J, Cowie C, Bew J & Stead D (2007) Phylogenetic analysis of *Xanthomonas* species by comparison of partial gyrase B gene sequences. *International Journal of Systematic and Evolutionary Microbiology* **57**: 2881–2887.

Richardson EJ, Bacigalupe R, Harrison EM*, et al.* (2018) Gene exchange drives the ecological success of a multi-host bacterial pathogen. *Nature Ecology & evolution* **2**: 1468–1478.

Ryan RP, Vorhölter F-J, Potnis N, Jones JB, Van Sluys M-A, Bogdanove AJ & Dow JM (2011) Pathogenomics of *Xanthomonas*: understanding bacterium–plant interactions. *Nature Reviews: Microbiology* **9**: 344-355.

Saddler GS & Bradbury JF (2015) Xanthomonas. *Bergey's Manual of Systematics of Archaea and Bacteria*.

Studholme DJ, Wicker E, Abrare SM*, et al.* (2020) Transfer of *Xanthomonas campestris* pv. *arecae*, and *Xanthomonas campestris* pv. *musacearum* to *Xanthomonas vasicola* (Vauterin) as *Xanthomonas vasicola* pv. *arecae* comb. nov., and *Xanthomonas vasicola* pv. *musacearum* comb. nov. and description of *Xanthomonas vasicola* pv. *vasculorum* pv. nov. *Phytopathology* 1-19.

Tushemereirwe W, Kangire A, Ssekiwoko F, Offord L, Crozier J, Boa E, Rutherford M & Smith J (2004) First report of Xanthomonas campestris pv. musacearum on banana in Uganda. *Plant Pathology* **53**: 802-802.

Chapter 3

Comparative genomics of *Xanthomonas vasicola* isolates from maize, sugarcane, sorghum, banana and *Eucalyptus*

3.1 Abstract

Host specificity is the preference that a pathogen has for a group of defined hosts or tissues as a result of host specificity determinants. Changes in the host specificity determinants can be ascribed to point-mutations, indels, horizontal gene transfer and loss of redundant genes. The aim of this study was to determine the changes which occurred in the genomes of five *X. vasicola* strains which allowed them to switch hosts from the monocotyledon plants sugarcane to the dicotyledon *Eucalyptus grandis*. The whole genomes were sequenced on the Illumina HiSeq 2500 platform and assembled using the reference genome *X. vasicola* pv. *vasculorum* SAM119. The *de novo* assembly was done using Unicycler to separate the chromosomes and plasmids. Then a reference-assisted assembly was performed using Ragout to rearrange contigs and generate a scaffold. The chromosomes and plasmids were annotated using RAST and analysed and edited in Geneious. The average nucleotide identity (ANI), whole-genome phylogeny, orthologous clusters and the presence/absence of genes were determined to identify changes which may have taken place in genomes of the five *Xvv* strains that enabled a change in host specificity. Genes specific to the five genomes of *Xvv* strains from *Eucalyptus* were involved in adhesion, chemotaxis, tannase production, the type III secretion system effectors, and the type IV secretion system. The acquisition of these essential virulence determinants may have played a role in promoting the adaptation of *Xvv* to the dicotyledonous plant species.

3.2 Introduction

Host specificity is the preference of a pathogen to infect one or a group of defined hosts, and this is achieved through the use of host specificity factors required for an association between a pathogen and its specific host (Kirzinger & Stavrinides, 2012). A change in the host specificity determinants may result in the pathogen attaining the ability to infect a new host. This change may be as a result of mutations and indels within the existing gene repertoire, the acquisition of new genetic material through horizontal gene transfer, and a loss of genes which are redundant following host specialisation (refer to Chapter 1), or genomic rearrangement creating a new gene combination cluster.

The change of host specificity determinants may result in a host shift when the infection of a new host which is phylogenetically close to the current host, or a host jump when the infection of a new host which is phylogenetically distant to the current host occurs (Schulze-Lefert & Panstruga, 2011). The reduced rate of functional evolution in bacteria reduces their ability to adapt to new hosts, and multiple encounters with the new host increase their chances to adapt. Also, an infection which is not aggressive is more beneficial for the pathogen as it allows enough time for it to adapt and specialise (Bonneaud *et al.*, 2019). Therefore, this suggests that the changes which result in a host shift or jump occur in the accessory genome of the pathogen before being incorporated into their core genome, which leads to speciation.

Phytopathogenic species of the genus *Xanthomonas* are known to display a high level of host specificity and tissue specificity, but a few such as *X. campestris* pv. *campestris* have a broad host range in the Brassicaceae family (Vicente & Holub, 2013). *Xanthomonas* species which display a narrow host range include *X. citri* pv. *citri,* which causes citrus bacterial canker (Jalali *et al.*, 2017).

Xanthomonas vasicola is a known pathogen of the maize (*Zea mays*), sorghum (*Sorghum bicolor*), sugarcane (*Saccharum* spp. hybrids), and banana (Musa sp.)

(Korus *et al.*, 2017). However, the host range of the *X. vasicola* pv. *vasculorum* (*Xvv*) recently broadened from sugarcane, resulting in the emergence of bacterial blight and dieback on the *Eucalyptus grandis* (Coutinho *et al.*, 2015). This emergence may be the result of a change in the gene repertoire involved in host-specific interactions such as type III secretion system effectors and changes in genes involved in iron uptake, cell wall degrading enzymes, and lipopolysaccharides (Jacques *et al.*, 2016).

In this chapter, we considered the hypothesis that the *Xvv* host shift from sugarcane to *Eucalyptus* was explained by a change in the gene content of these strains. We thus aimed to determine the genetic changes which occurred in the genomes of the *X. vasicola* strains from *Eucalyptus,* which enabled the pathogen to jump from sugarcane.

3.3 Materials and Methods

3.3.1 DNA extraction and whole-genome sequencing

Five *Xvv* strains were obtained from the bacterial culture collection at the Centre for Microbial Ecology and Genomics, University of Pretoria (Table 3.1). The strains were cultured on Yeast Dextrose Carbonate agar and incubated at 28^{o}C for 48h.

Table 3.1: The five *Xvv* strains isolated from *Eucalyptus grandis* selected for whole-genome sequencing

Strain	Host and cultivar	Place collected	Year collected
BCC42	*Eucalyptus grandis* Tag14	South Africa	2005
BCC931	*Eucalyptus grandis* GN055	South Africa	2007
BCC932	*Eucalyptus grandis* GN055	South Africa	2007
J5	*Eucalyptus grandis* Tag14	South Africa	2017
J12	*Eucalyptus grandis* Tag14	South Africa	2017

The bacterial DNA was isolated using the cetyltrimethylammonium bromide (CTAB) using a protocol modified by E. van Zyl and S.N Venter (Cleenwerck *et al.*, 2002). Following culturing, between 1000-1250mg bacterial cells (±12 9cm agar plates) were collected in 15ml Greiner tubes and resuspended in 5ml RS buffer (0.15 M NaCl, ten mM EDTA, pH 8.0). The suspension was then centrifuged at 11000 rpm for 10 min at 5^{o}C, and the supernatant was removed using a pipette. The pellet was resuspended in 3,75ml buffer (10mM tris-100mM EDTA (Merck), pH 8.0), and 200µl RNase A (10 mg/ml) (ZymoResearch) was added to the suspension and mixed by inversion. One millilitre of 10% sodium dodecyl sulphate (Merk) was added to the suspension and mixed by inverting the tube. A 100µl Proteinase K (20 mg/ml) (ZymoResearch) was added, and the suspension was mixed by inversion. The suspension was then incubated at 65^{o}C in a water bath for 1h. The CTAB was kept in the water bath for the

duration of the incubation. Following incubation, 1,5ml NaCl (5 M) (Merk) was added to the suspension and the tube inverted to mix the suspension. One millilitre CTAB/NaCl (10% w/v CTAB in 5 M NaCl) was then added and the tube inverted to mix the suspension. The suspensions were incubated in the water bath for 20min. Following incubation, 5ml chloroform/isoamylalcohol (24:1) was added into each tube. The suspension was mixed by inverting, and the lid was carefully opened briefly to allow the gas to escape. The tubes were then closed and shook vigorously, and the lids were again opened to allow gas to escape. The centrifuged was then cooled to 5^{o}C, and the suspension was then centrifuged at 11000 rpm for 30 min. The upper phase (±7ml) was then carefully transferred to 50ml plastic containers, and 4,2ml isopropanol was then added and mixed by carefully swirling. A sterile Pasteur pipette with the tip closed was used to collect the DNA by swirling it carefully in the solution. The DNA was then washed with an ethanol range (70%, 80% and 90%) to remove excess CTAB, and the DNA was air-dried until it became clear. The DNA was suspended in 1ml 0,1 x TE buffer (pH 8.0) in 15ml Greiner tubes. After the DNA loosened from the rod, 50µl RNase A (10 mg/ml) (ZymoResearch) was added and mixed by inverting the tube. The suspension was then incubated for 1h at 37^{o}C in a water bath. Following incubation, 560µl acetate-EDTA (510µl3M NaAc + 50µl 0,5M EDTA) was added and mixed by inverting. Then, 3500µl chloroform/isoamylalcohol (24:1) was added to the suspension and mixed by inverting. The suspension was then centrifuged at 11000 rpm for 20min. The upper phase (±1000µl) was then transferred to a 50 ml plastic pot, and 3300µl isopropanol was added to precipitate the DNA. The DNA was collected with a Pasteur pipette with a closed tip by swirling it carefully in the solution. The DNA was washed in an ethanol range (70%, 80%, and 90) and air-dried. The DNA was suspended in 500µl 0,1 M saline-sodium citrate buffer (pH 8.0). The DNA was kept in a fridge for three days to dissolve, and the DNA concentration and quality were checked using a spectrophotometer (ThermoFisher Scientific).

The genomes were sequenced by MrDNA (Shallowater, TX, USA) on the Illumina HiSeq 2500 platform (http://www.mrdnalab.com/).

3.3.2 Genome assembly and annotation

3.3.2.1 BLASTn

With the paired-end Illumina reads, a BLASTn search (Altschul *et al.*, 1990) was performed against the National Centre for Biotechnology Information (NCBI) plasmid database to determine the presence of plasmids with an 80% identity cutoff (NCBI, 2008, Brooks *et al.*, 2019).

3.3.2.2 Assembly

The *de novo* assembly of the chromosomes and plasmids was done using Unicycler (Wick *et al.*, 2017). The reference-assisted assembly was done using Ragout with *Xvv* SAM119 as a reference, wherein the contigs were rearranged, and the scaffolds of the chromosomes and plasmids were built (Kolmogorov *et al.*, 2014). The contamination and completeness of the genomes were determined using CheckM (Parks *et al.*, 2015) and BUSCO (Simão *et al.*, 2015).

3.3.2.3 Annotation

The chromosome and plasmid scaffolds were annotated using RAST (Aziz *et al.*, 2008, Overbeek *et al.*, 2014, Brettin *et al.*, 2015). The annotations were visualised, verified, and edited in Geneious Prime 2020.0.2 (https://www.geneious.com); predicted hypothetical genes were subjected to a BLAST search to identify possible annotated homologs and their annotations were manually updated. The genomes were circularised and the images obtained from Geneious Prime 2020.0.2 (https://www.geneious.com).

3.3.3 Average nucleotide identity

The orthologous average nucleotide identity between the genomes of *X. vasicola* isolated from different hosts genomes different hosts was measured with the OrthoANI Tool 0.93.1 (Lee *et al.*, 2016). The complete genomes of *Xvv* SAM119 (PRJNA439013) and Xv1601 (PRJNA421835) isolated from maize, *X. vasicola* pv. *holcicola* (*Xvh*) NCPPB1060 (PRJNA264730) isolated from sorghum and *X. vasicola* pv. *musacearum* (*Xvm*) NCPPB4379 (PRJNA73877) isolated from banana were obtained from the NCBI RefSeq database (O'Leary *et al.*, 2015) on Geneious Prime 2020.0.2 (https://www.geneious.com).

3.3.4 Whole-genome phylogeny

The maximum likelihood whole-genome was generated using Realphy 1.1.2 with 100 trials for bootstrap support (Bertels *et al.*, 2014). The whole-genome sequences were obtained from the NCBI RefSeq database (O'Leary *et al.*, 2015). The genomes were obtained from bioprojects (Table 3.2). The whole-genome sequences were mapped against the reference sequence of *Xvv* SAM119 with Bowtie2 (Langmead & Salzberg, 2012). The whole-genome phylogenetic tree was inferred with PhyML (Guindon *et al.*, 2010), and a consensus tree visualised on Geneious 2020.0.2 (https://www.geneious.com).

Table 3.2: Summary of the data of the genomes used for whole-genome phylogeny.

Bioproject	ID	Host	Place of isolation	Year of isolation
PRJNA264730	NCPPB1060	Sorghum	Ethiopia	1961
PRJNA421835	Xv1601	Maize	Kansas, USA	2016
PRJNA439013	SAM119	Maize	Klerksdrop, SA	1988
PRJNA264615	NCPPB2417	Sorghum	New Zealand	1969
PRJNA413069	NE744	Maize	Nebraska, USA	2015
	SAM113	Maize	South Africa	2013
PRJNA73879	NCPPB4380	Banana	Uganda	2012
PRJNA73881	NCPPB4384	Banana	Uganda	2012
PRJNA73871	NCPPB206	Maize	SA	1948
PRJNA73869	NCPPB1381	Sugarcane	Zimbabwe	1962
PRJNA73853	NCPPB1326	Sugarcane	Zimbabwe	1962
PRJNA163307	NCPPB895	Sugarcane	Madagascar	1960
PRJNA163305	NCPPB890	Sugarcane	South Africa	1960

3.3.5 Orthology

To compare and annotate orthologous gene clusters between the *X. vasicola* from different hosts, the web platform OrthoVenn2 was used with the following parameters: e-value 1e-5 and inflation value 1.5 (Xu *et al.*, 2019). The protein sequences of *Xvv* SAM119 (PRJNA439013), *Xvh* NCPPB1060 (PRJNA264730), and *Xvm* NCPPB4379 were obtained from the NCBI Refseq database (O'Leary *et al.*, 2015) on Geneious 2020.0.2 (https://www.geneious.com).

3.3.6 Presence/absence of type III secretion system effectors and type IV secretion system pili

A list of type III secretion system (T3SS) effectors was obtained from Xanthomonas.org. The list of the type IV secretion system (T4SS) pili was obtained from (Mhedbi-Hajri *et al.*, 2011). The protein sequences of *Xvv* SAM119 (PRJNA439013), *Xvh* NCPPB1060 (PRJNA264730), *Xv* Xvv1601 (PRJNA421835) and *Xvm* NCPPB4379 (PRJNA73877) were obtained from the NCBI Refseq database

(O'Leary *et al.*, 2015). The protein sequences were searched for in Geneious Prime 2020.0.2 (https://www.geneious.com), and the sequences were confirmed by BLASTp with a >95% identity similarity (Madden, 2002).

3.4 Results

3.4.1 DNA extraction and whole-genome sequencing

The DNA extracted was of acceptable quality and the concentrations obtained varied from 1998ng/ml to 4400ng/ml (Table 3.3). The variation observed between the DNA concentrations may be due to the amount of cells harvested, as some bacterial cultures grew faster. There were ≈10 million 2 X 250 paired-end reads for each genome, with 100x coverage.

Table 3.3: Quality and concentration of the DNA extracted from *X. vasicola* strains from *Eucalyptus grandis*.

Culture collection ID	DNA concentration (ng/ml)	DNA purity (260/280)
BCC42	1998.2	1.9
BCC931	4191	1.8
BCC932	4453	1.8
J5	3755	1.8
J12	2587.8	1.8

3.4.2 Genome assembly and annotation

The *de novo* genome assembly resulted in ≥80 contigs and a median fragment N_{50} value of 141820bp. The reference-based assemblies had a median N_{50} value of 4793818 (Table 3.4).

Table 3.4: The summary of the *de novo* assembly using Unicycler and subsequent reference-guided assembly using Ragout. *X. vasicola* SAM119 was used as a reference genome for the reference-guided assembly.

Culture collection ID	Used contigs from *de novo* assembly	Scaffold length (bp)	Unplaced contigs (not used in scaffold)	Unplaced contigs length (bp)	Introduced Ns length (bp)	Contigs N50 (bp)	Assembly N50 (bp)
BCC42	87	4788671	64	231281 (4.72%)	120629 (2.52%)	141820	4788671
BCC931	86	4848761	65	182016 (3.71%)	130679 (2.70%)	166106	4848761
BCC932	82	4855107	62	181454 (3.70%)	131958 (2.72%)	166099	4855107
J5	84	4793818	64	286049 (5.79%)	136929 (2.86%)	141820	4793818
J12	85	4784825	67	287810 (5.81%)	118518 (2.48%)	141820	4784825

The five *Xvv* strains from *Eucalyptus* have genome sizes ranged from 4.79Mbp to 4.87Mb (Fig 3.1), median G+C content 63.5%, and median protein count of 4513. A circular 5kb bacteriophage was present in all the genomes (Table 3.5).

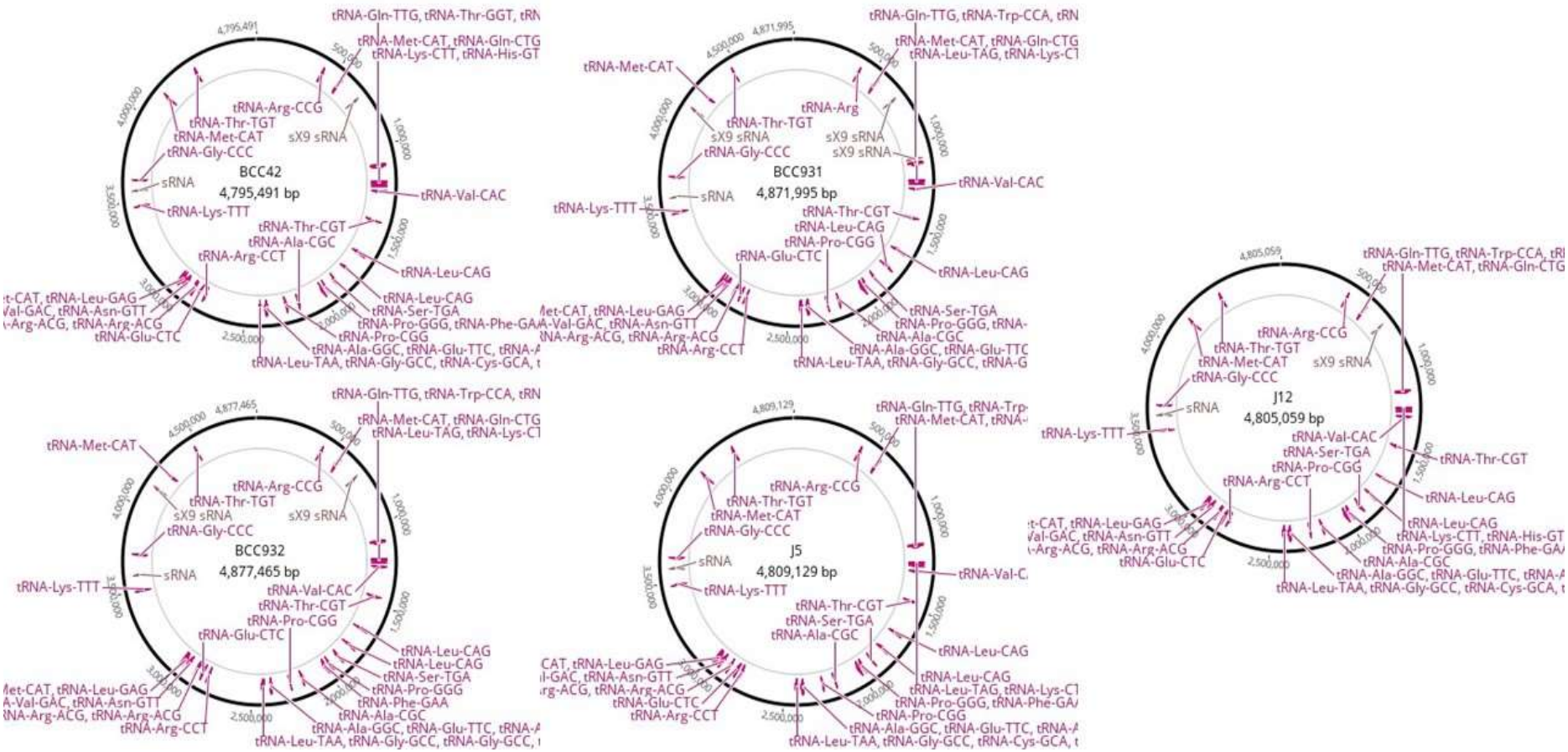


Figure 3.1: The circular genomes of the five *Xvv* strains isolated from *Eucalyptus*. The genome BCC42 isolated in 2005 had the smallest genome of 4.79Mbp, and BCC932 isolated in 2007 had the largest genome of 4.87Mbp.

Table 3.5: The summary of the genomic data of the *Xvv* isolated from *Eucalyptus*.

Culture collection ID	Sequence length (bp)	G+C (%)	No.of plasmids	No. of phages	No. of proteins
BCC 42	4909766	63.2	2	1	4491
BCC 931	4795491	63.5	1	1	4518
BCC 932	4871995	63.4	1	1	4584
J5	4809129	63.5	3	1	4495
J12	4805059	63.5	3	1	4513

The five *Xvv* strains from *Eucalyptus* carry at least one plasmid, and the strains J5 and J12 have a unique 54kbp unidentified plasmid (Table 3.6). The strains all have a plasmid that has a nucleotide sequence similarity of 86% with the tomato *X. vesicatoria* plasmid pLMG911.2 CP01872 with 64 protein-coding genes, including genes encoding for the T4SS and tannase. Tannase breaks down gallotannins, complex tannins and gallic acid esters by hydrolysing ester bonds, releasing glucose, gallic acid, and galloyl esters (Aguilar-Zarate *et al.*, 2014). The T4SS is a multifunctional system which transports macromolecules between the environment and the bacterial cell, and this secretion system is used by pathogens to transport virulence factors into the host cells (Wallden *et al.*, 2010). The majority of the genes on this plasmid (CP01872) encode for proteins of unknown function.

Table 3.6: The plasmids present in the genomes of the five *Xvv* strains isolated from *Eucalyptus*. The plasmid pLMG911.2 is present in all the *Xvv* genomes and strains J5 and J12 have a unique 54kbp plasmid.

	Plasmid	Plasmid accession no.	BIAST Hit	Sequence length	G+C %	No. of proteins
BCC42	1	CP022264	*Xanthomonas citri* pv. *vignicola* CFBP7111 plasmid plA	113877	63.3	128
	2	CP018727	*Xanthomonas vesicatoria* ATCC 35937 LMG911 plasmid pLMG911.2	47220	61.7	64
BCC931	1	CP018727	*Xanthomonas vesicatoria* ATCC 35937 LMG911 plasmid pLMG911.2	47244	61.7	64
BCC932	1	CP018727	*Xanthomonas vesicatoria* ATCC 35937 LMG911 plasmid pLMG911.2	47244	61.7	64
J5	1	CP022264	*Xanthomonas citri* pv. *vignicola* CFBP7111 plA	113877	63.3	128
	2	-	-	54118	62.5	61
	3	CP018727	*Xanthomonas vesicatoria* ATCC 35937 LMG911 plasmid pLMG911.2	47226	61.7	64
J12	1	CP022264	*Xanthomonas citri* pv. *vignicola* CFBP7111 plA	113878	63.3	128
	2	-	-	54118	62.5	61
	3	CP018727	*Xanthomonas vesicatoria* ATCC 35937 LMG911 plasmid pLMG911.2	47220	61.7	64

3.4.3 Average nucleotide identity

The *X. vasicola* strains isolated from different hosts were closely related and shared ≥98% orthologous ANI (Fig 3.2), which is above the species delimitation threshold of 95%. All *Xvv* shared more than 99% ANI, whereas the other pathovars (*holcicola* and *musacearum*) were a little more distant.

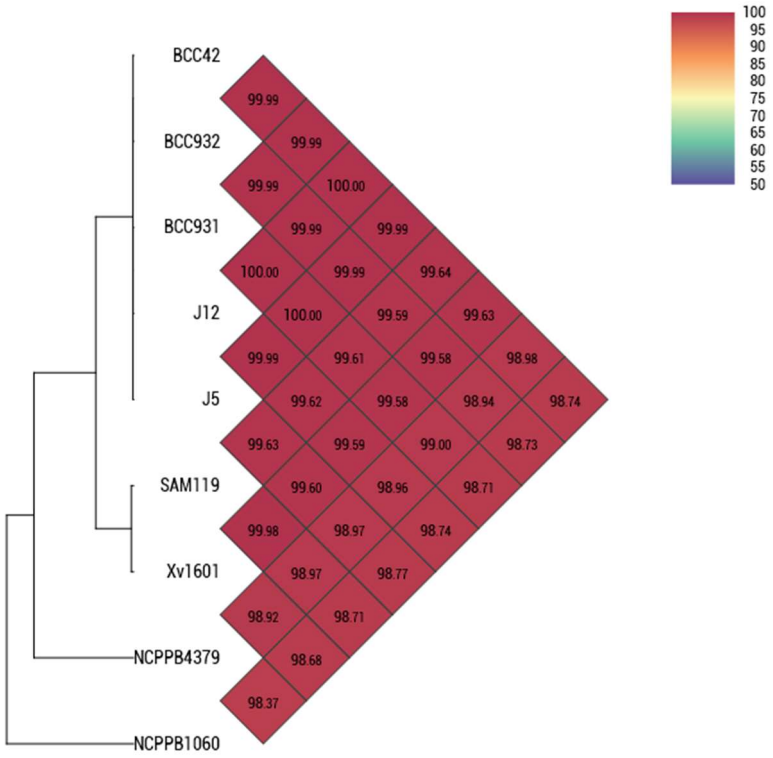


Figure 3.2: The orthologous average nucleotide identity between the five *Xvv* strains isolated from *Eucalyptus* and other *X. vasicola* isolated from maize (SAM119 and Xv1601, pathovar *vasculorum*), sorghum (NCPPB1060, pathovar *holcicola*) and banana (NCPPB4379, pathovar *musacearum*).

3.4.4 Whole-genome phylogeny

The sorghum strains *Xvh* NCPPB1060 and NCPPB2417 grouped in the most basal and distant clade (pathovar *holcicola*), but this branch was not well supported. The *X. vasicola* strains isolated from maize, sugarcane, banana, and sorghum formed monophyletic clades according to their host, except for the sugarcane *Xvv* NCPPB895, and the maize *Xvv* NCPPB206. The sugarcane strains which formed a monophyletic clade were isolated from Southern Africa (Zimbabwe and South Africa), whereas the strain NCPPB895 was isolated from the African island Madagascar. The maize strains which formed a monophyletic clade were isolated from South Africa and USA post-1998, whereas the strain NCPPB206 was isolated in South Africa in 1948. The five *Xvv* strains from *Eucalyptus* form a clade with *Xvv* NCPPB895 from sugarcane (Fig 3.3).

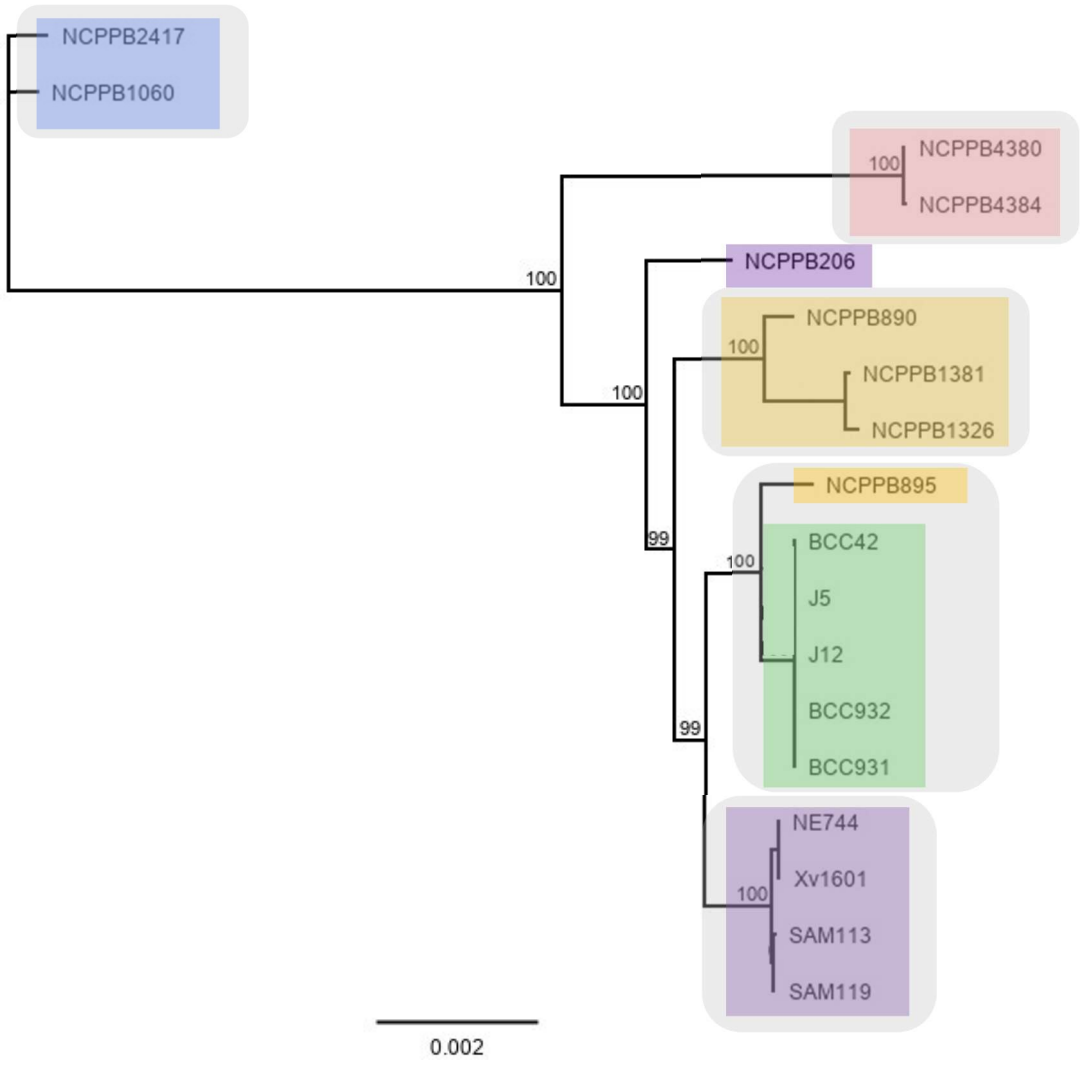
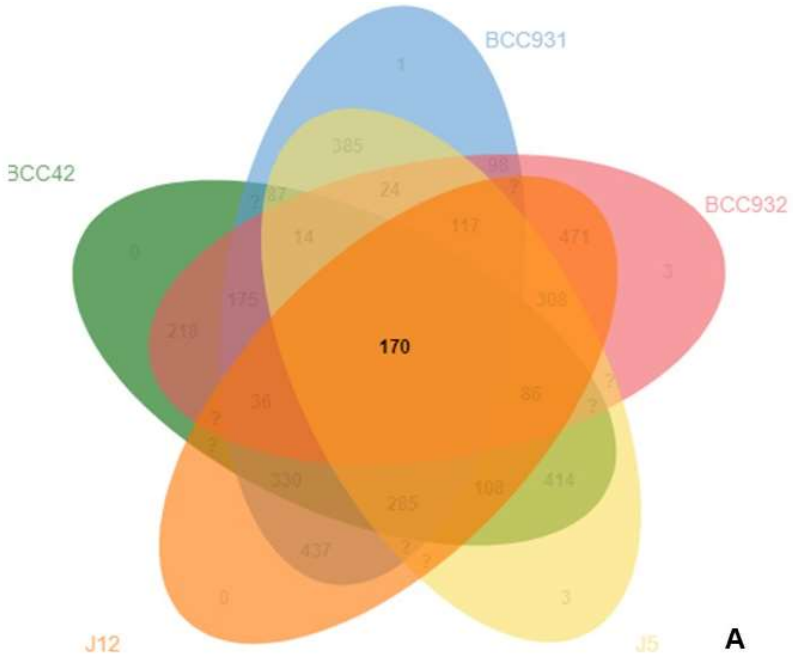


Figure 3.3: An unrooted maximum-likelihood phylogenetic tree with 100 trials for bootstrap support. The coloured blocks represent the hosts as follows: green represents *Xvv* from *Eucalyptus*, the orange block represents *Xvv* from sugarcane, the purple block represents *Xvv* from maize, the pink block represents *Xvm* from banana, and the blue block represents *Xvh* from sorghum. Most of the *X. vasicola* strains grouped according to their host (grey blocks), except the maize *Xvv* NCPPB206 and the sugarcane NCPPB895. The five *Xvv* strains from *Eucalyptus* grouped with the sugarcane *Xvv* NCPPB895.

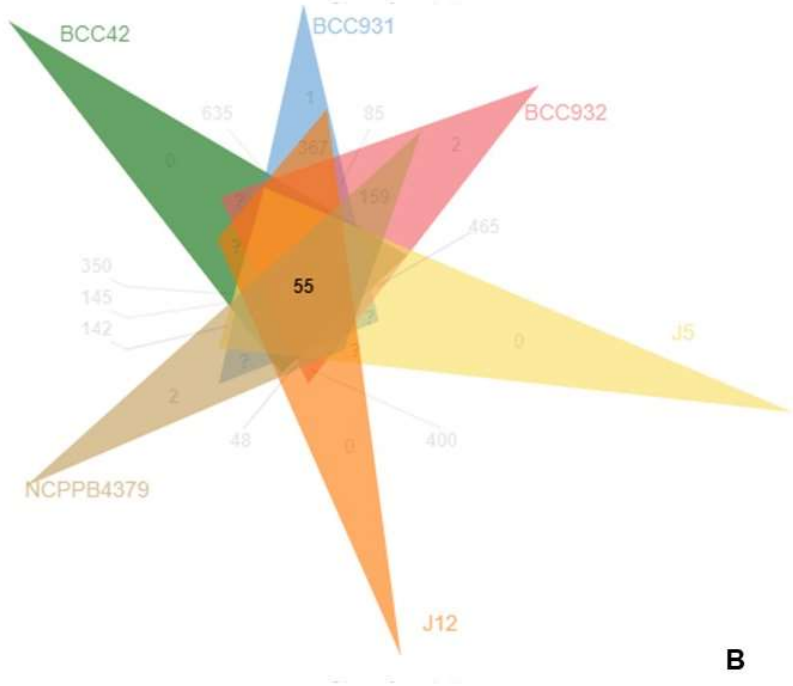
3.4.5 Orthology

3.4.5.1 Orthologous clusters between the *Eucalyptus Xvv* strains and strains from banana, maize, and sorghum

The analysis between the five *Xvv* strains from *Eucalyptus* and *Xvv* SAM119 from maize identified the most orthologous clusters with 8140, and the least number of clusters were between the five *Xvv* strains from *Eucalyptus* with 6830 orthologous clusters. The analysis between *Xvh* NCPPB1060 and the five *Xvv* strains from *Eucalyptus* identified 7585 orthologous clusters, and the analysis between *Xvm* NCPPB4379 and the *Xvv* strains from *Eucalyptus* identified 7605 clusters. The *Eucalyptus Xvv* strains shared the most core genome orthologous clusters (170 core genome clusters) relative to the total number of orthologous clusters identified (Fig 3.4A), and *Xvm* NCPPB4379 and *Xvv* from *Eucalyptus* shared the least (55 core genome clusters) (Fig 3.4B). Taken together, these data indicated that the *Eucalyptus* strains shared fewer genes with *Xvm* than with *Xvh*, and more genes with the maize *Xvv* strain SAM119 (81 core genome clusters) (Fig 3.4D).


BCC931
BCC42
BCC932
170
J12
J5

A


BCC42
BCC931
BCC932
55
J5
NCPPB4379
J12

B

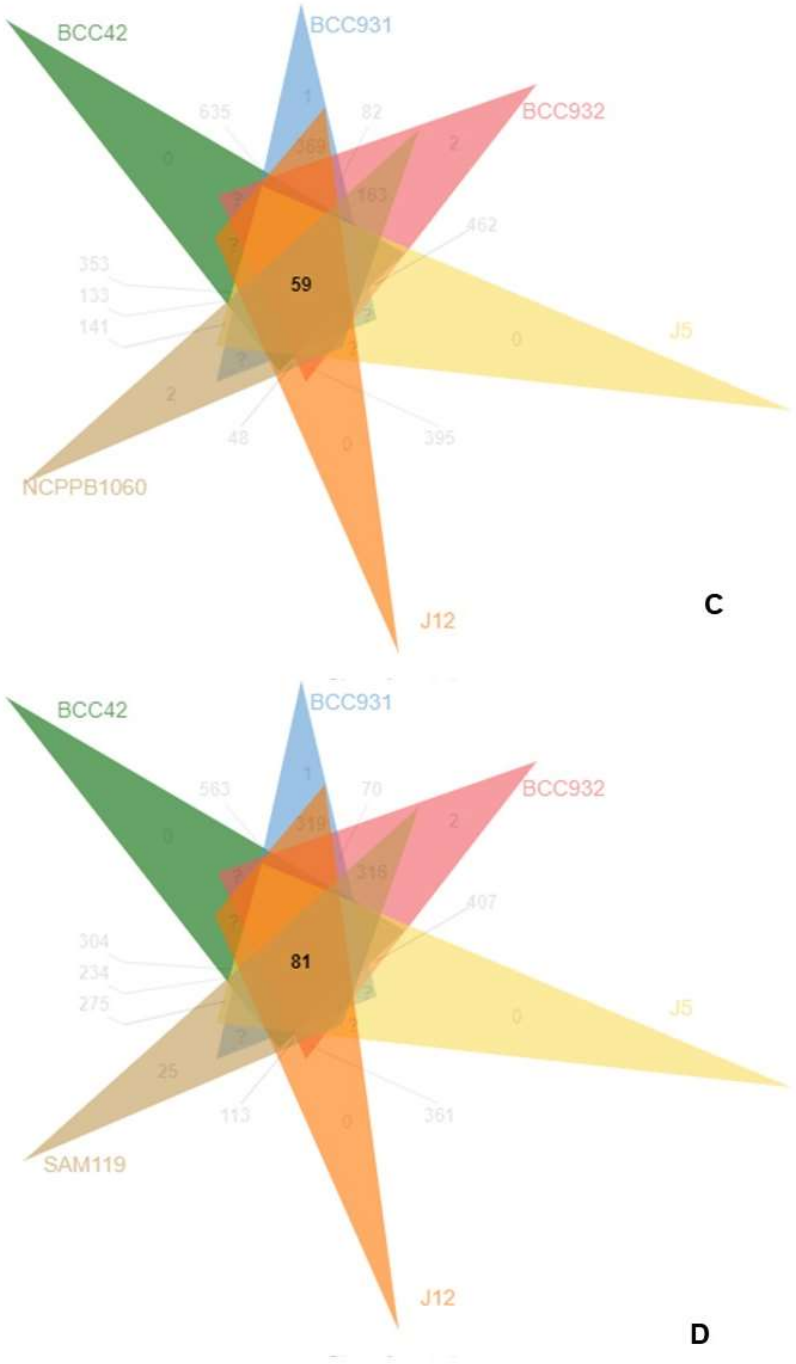


Figure 3.4: The Venn diagrams showing the core genome orthologous clusters. A) The core genome orthologous clusters shared between the five *Xvv* isolated from *Eucalyptus*. B) The core genome orthologous clusters shared between the *Xvm* NCPPB4379 isolated from banana and the *Xvv* strains from *Eucalyptus*. C) The core genome orthologous clusters shared between the *Xvh* NCPPB1060 isolated from sorghum and *Xvv* strains from *Eucalyptus*. D) The core genome orthologous clusters shared between the *Xvv* SAM119 isolated from maize and *Xvv* strains from *Eucalyptus*.

3.4.5.2 Genes unique to the *Eucalyptus* strains

The analyses between the *X. vasicola* from different hosts identified orthologous clusters unique to the five *Xvv* isolated from *Eucalyptus* (Fig 3.5). The five *Xvv* strains had 78 unique clusters when compared to the *Xvv* SAM119 strain isolated from maize (Fig 3.5A), 108 unique clusters when compared to the *Xvh* NCPPB1060 strain isolated from sorghum (Fig 3.5B), and 106 unique clusters when compared to *Xvm* NCPPB4379 strain isolated from banana (Fig 3.5C). The genes identified in these unique clusters encode for extracellular metalloproteases, methyl-accepting chemotaxis proteins, and the T4SS. Proteases play a role in the pathogen's protection against the host's extracellular defences, the break down of the host's cell walls, and virulence (Figaj *et al.*, 2019). Bacteria use chemotaxis to move according to a chemical gradient (Matilla & Krell, 2017).


BCC42
BCC931
BCC932
J5
J12
SAM119
Cluster count
Protein count
81 556
54 295
24 138
16 101
88 450
24 139
78 393
60 256
58 234
7 31
125 506
58 248
66 264
61 250
30 122
111 456
10 41
193 774
104 425
4 17
BCC42
BCC93
BCC93
J5
J12
SAM119
A


BCC42
BCC931
BCC932
J5
J12
NCPPB1060
Cluster count
Protein count
59 386
37 201
26 137
13 72
68 342
19 103
108 550
44 185
67 271
7 29
80 321
41 170
85 342
48 197
23 93
81 332
17 69
214 859
71 287
9 39
BCC42
BCC931
BCC932
J5
J12
NCPPB1060
B

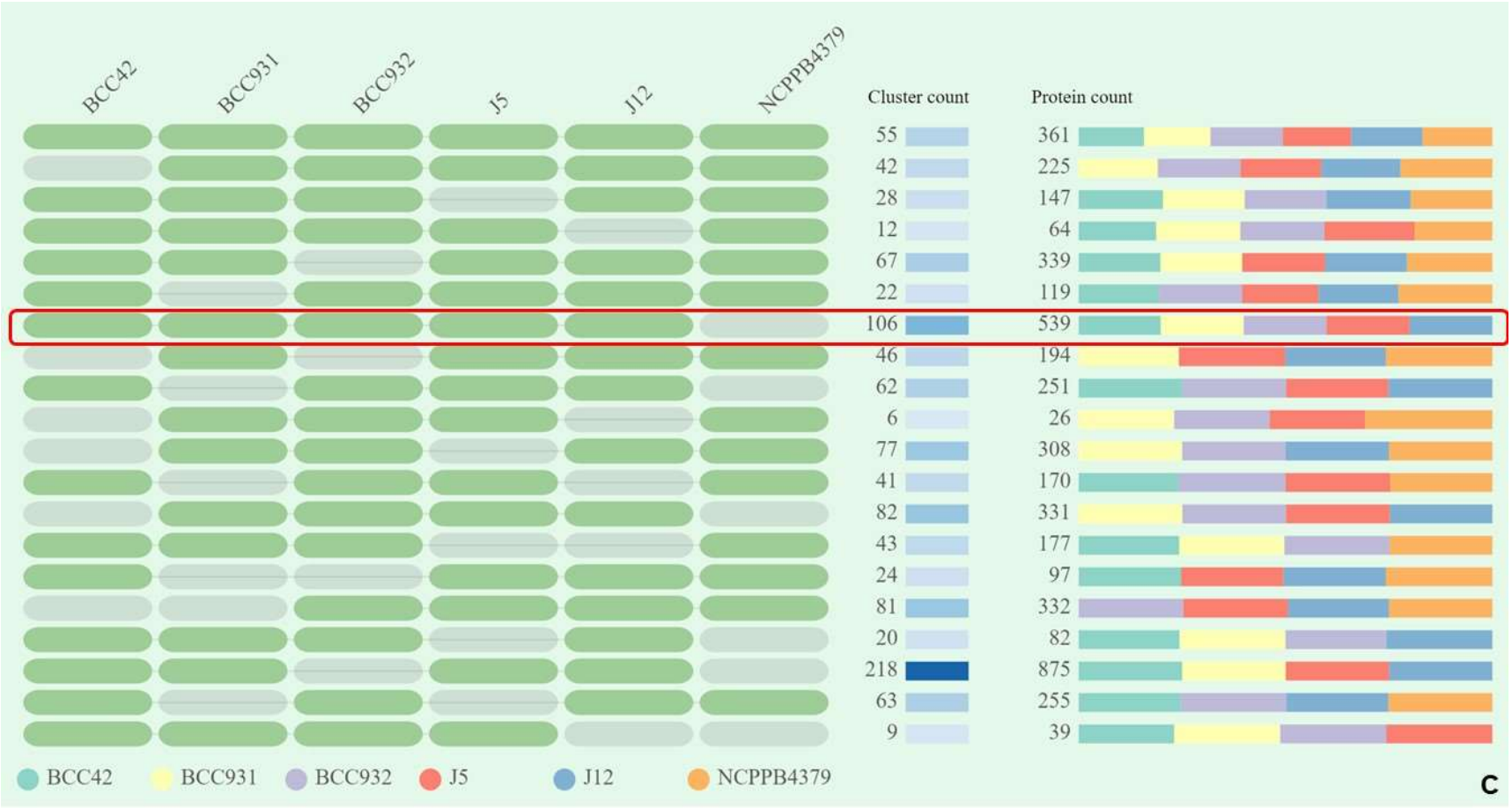


Figure 3.5: The occurrence pattern of the orthologous clusters which are shared between the *X. vasicola* from different hosts. The red block shows the orthologous clusters unique to the five *Xvv* strains isolated from *Eucalyptus*. A) The analysis between the *Xvv* SAM119 from maize and the *Xvv* strains from *Eucalyptus* identified 78 unique orthologous clusters. B) The analysis between the *Xvh* NCPPB1060 from sorghum and the *Xvv* strains from *Eucalyptus* identified 108 unique orthologous clusters. C) The analysis between the *Xvm* NCPPB4379 from banana and the *Xvv* strains from *Eucalyptus* identified 106 unique orthologous clusters.

3.4.6 Presence/absence of type III secretion system effectors and type IV pilus in the unique gene set

The *Xvv* strains from *Eucalyptus* have unique genes when compared to X*. vasicola* from other hosts (Fig 3.6). The T3SS effector homologs HopAF1 and *HolPtoQ* were present only in the genomes of *Xvv* isolated from *Eucalyptus*, *AvrRxv* is unique to *Xvm* NCPPB4379 from banana, and *HopH1* is unique to *Xvh* NCPB1060 from sorghum. The type IV pilus (T4P) genes are unique to the *Xvv* strains from *Eucalyptus*, except for *pilU,* which is also present in the other *X. vasicola* genomes. The T3SS effectors play a role in the virulence of the strains (Deng *et al.*, 2017), and the T4P play a role in the sensing of the environment by the pathogen and adhesion to the plant surface (Craig *et al.*, 2019).

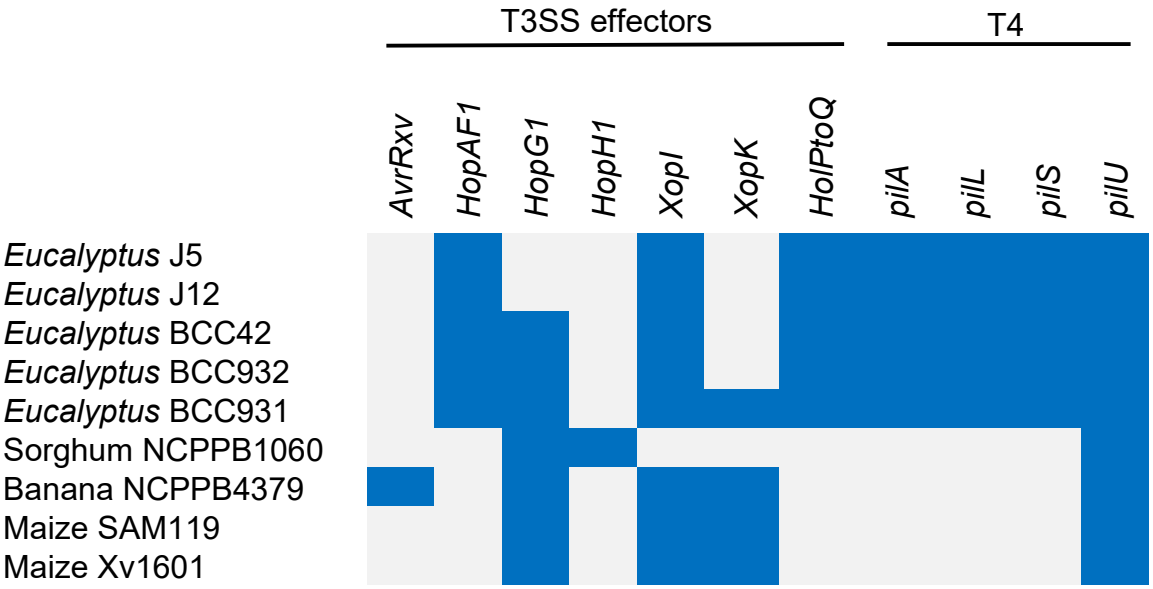


Figure 3.6: The heatmap showing the presence (blue) and absence (grey) of T3SS effectors which play a role in virulence and type IV pilus, which plays a role in adhesion. The T3SS effectors *HopAF1* and *HolPtoQ*, and the T4SS pili genes are unique to the five *Xvv* strains from *Eucalyptus*.

3.5 Discussion

The aim of this study was to determine the changes which occurred in the genomes of the five *X. vasicola* strains isolated from *Eucalyptus,* which may have enabled them to jump hosts to sugarcane. The results showed that the five *Xvv* strains from *Eucalyptus* carry plasmids, and differences were observed in genes responsible for virulence, motility and attachment. The acquisition of these genetic factors may have played a role in the adaptation if the *X. vasicola* strains to *Eucalyptus*.

The five strains of *X. vasicola* from *Eucalyptus* were confirmed as *Xvv*. The sizes of the genomes were ≈4.8Mbp, had a G+C content of 63,5% ≥ 63,2%, and have a protein count of ≈4500. These characteristics are typical of genomes of *X. vasicola* species (Saddler & Bradbury, 2015, Sanko et al., 2018). The strains also share a high orthologous ANI with *Xvv* SAM119 and Xv1601 (>99%), *Xvh* NCPPB1060 and *Xvm* NCPPB4379 (>98%). This high ANI supports the identification of these strains as *Xvv* (Studholme *et al.*, 2020).

The five *Xvv* strains from *Eucalyptus* each harbour a bacteriophage and they all share a plasmid from *X. vesicatoria,* which infects the dicotyledonous plant tomato (Roach *et al.*, 2018). A previous study found evidence of plasmids from *X. axonopodis* and *X. citri* by aligning raw reads of *Xvv* strains against the RefSeq plasmid database (Wasukira *et al.*, 2014). The strains share a 47kbp plasmid pLMG911.2 CP01872 which carries genes that encode for the T4SS and tannase. The T4SS is a versatile secretion system, and it has three broad functions, namely conjugation, natural transformation and transfer of proteins. This secretion system is essential in the virulence of pathogenic bacteria as they use it for the transfer of toxic effector proteins, and it plays a role in the interaction between the pathogen and the host (Wallden *et al.*, 2010, Christie *et al.*, 2014). Tannins are polyphenolic compounds used by plants to defend themselves against microorganisms, and these are present in the leaves of woody plants but not in the leaves of herbaceous plants (Aguilar-Zarate *et al.*, 2014, Constabel *et al.*, 2014). Tannase is an enzyme that hydrolyses tannins and pathogens

encoding this enzyme use it overcome the inhibitory effects of tannins (da Silva *et al.*, 2002, Aguilar-Zarate *et al.*, 2014). Therefore, the acquired genes provide the *Xvv* strains with an adaptive advantage by being able to protect themselves against the effects of the tannins present in the woody-plant leaves and having additional virulence factors to overcome the dicotyledon defence repertoire.

The maximum likelihood phylogenetic tree resolved the evolutionary relationship of the *X. vasicola*. The strains clustered according to their host and the tree topology is consistent with what has been found by Studholme *et al*., (2020). The novel finding is that the five *Xvv* strains from *Eucalyptus* form a monophyletic group with *Xvv* from sugarcane, and this group has high bootstrap support. This evolutionary relatedness between the *Xvv* from the two different hosts support a suggestion made by Coutinho *et al*., (2015) and the pathogenicity experiments in Chapter 2 that the outbreak of bacterial blight and dieback of *E. grandis* in KwaZulu-Natal in the mid-2000s was as a result of a host jump event from sugarcane.

The orthologous genome clustering identified genes encoding metalloproteases, chemotaxis proteins and the T4SS. Metalloproteases are involved in the breakdown of the plant host's cell walls, allowing the pathogen entry into and dissemination within the host (Figaj *et al.*, 2019). Chemotaxis is the movement towards or away from environmental cues. It has been suggested that this, together with motility, is important in the penetration into the host by xanthomonads (Matilla & Krell, 2017). Therefore, these unique virulence proteins may play a role in enabling *Xvv* to overcome the more complex *E. grandis* physical leaf defences, and cause disease.

The five *Xvv* strains from *Eucalyptus* had two unique effectors genes of the T3SS which plays a role in the manipulation of the host's cellular processes, and genes for the T4P which plays a role in adhesion. *Pseudomonas syringae* T3SS effector *HopAf1* suppresses ethylene production (Washington *et al.*, 2016), and sequences of this gene

were identified in draft genomes of *X. vasicola* from banana and sugarcane (Studholm *et al.*, 2010). The *P. syringae* T3SS effector gene *HolPtoQ* is induced by reactive oxygen species, and the homolog of this gene was identified in *X. campestris* pv. *vesicatoria* causing disease in pepper (Roden *et al.*, 2004). The T3SS effector genes identified in the five *Xvv* strains from *Eucalyptus* may have facilitated infection and population growth them to promote their growth in a woody plant environment which has different ethylene and reactive oxygen species to their grass hosts. This then may have allowed them to reach population numbers required to cause disease. The T4P are surface exposed fibres essential for the adhesion and motility of a bacterial pathogen among other functions (Craig *et al.*, 2019), and differences in the sequences of this pilus were observed by Wasukira *et al*., (2014). Therefore, the acquisition if these genes may have given an adaptive advantage to the five *Xvv* strains of *Eucalyptus*, allowing them to attach on a leaf surface different from that of the common monocotyledon hosts.

3.6 Conclusion

In this study, the aim was to determine the changes which occurred in the genomes of the *X. vasicola* isolated from the dicotyledonous plant *Eucalyptus,* which may have enabled them to move from the monocotyledonous plant sugarcane. The study identified a plasmid shared by the five *Xvv* stains of *Eucalyptus* which carried genes for the versatile T4SS which has a role in virulence, and tannase which breaks down tannins present in the woody plant's leaves. These strains also showed a close evolutionary relationship with an *Xvv* strain from sugarcane, supporting a previous suggestion that the pathogen jumped from sugarcane to a nearby *E. grandis* plantation, resulting in a disease outbreak. Orthologous clusters unique to these *Eucalyptus* strains were identified, and they contained genes which play a role in chemotaxis and the break-down of plant cell wall proteins. There were also homologs of T3SS effector proteins involved in the suppression of ethylene and protection against reactive oxygen species which were unique to these strains. Also, the strains have genes for the T4P, which plays a role in adhesion and pathogen-host interaction. These unique genetic elements were probably acquired through horizontal transfer due to multiple introduction events of this pathogen to the dicotyledonous host, resulting in a host jump. The findings of this study are novel, being the first to sequence whole genomes of *Xvv* strains isolated from *E. grandis* and identifying factors which may have played a role in their adaptation to a dicotyledonous plant species.

Future studies aiming to understand the factors which allow host shift/jump events should consider a comparative transcriptomic study to determine the gene expression profile of *X. vasicola* in different hosts, and at different time points. The role of the plasmids in the pathogenicity of the *Eucalyptus* strains should also be determined, and the functions of the candidate genes have to be validated.

3.7 References

Aguilar-Zarate P, Cruz-Hern´andez MA, Monta˜nez JC, Belmares-Cerda RE & Aguilar CN (2014) Bacterial tannases: Production, properties and applications. *Revista Mexicana de Ingeniería Química* **13**: 63-74.

Altschul SF, Gish W, Miller W, Meyers EW & Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.

Aziz RK, Bartels D, Best AA*, et al.* (2008) The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics* **9**: 1-15.

Bertels F, Silander OK, Pachkov M, Rainey PB & van Nimwegen E (2014) Automated Reconstruction of Whole-Genome Phylogenies from Short-Sequence Reads. *Molecular Biology and Evolution* **31**: 1077-1088.

Bonneaud C, Weinert LA & Kuijper B (2019) Understanding the emergence of bacterial pathogens in novel hosts. *Philosophical Transactions of the Royal Society B* **374**: 1-7.

Brettin T, Davis JJ, Disz T*, et al.* (2015) RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Nature: Scientific Reports* **5**: 1-6.

Brooks L, Kaze M & Sistroma M (2019) A Curated, Comprehensive Database of Plasmid Sequences. *Microbiology Resource Announcements* **8**: 1-2.

Christie PJ, Neal Whitaker N & González-Rivera C (2014) Mechanism and structure of the bacterial type IV secretion systems. *Biochimica et Biophysica Acta* **1843**: 1578–1591.

Cleenwerck I, Vandermeulebroecke K, Janssens D & Swings J (2002) Re-examination of genus *Acetobacter*, withdescriptions of *A. cerevisiae* sp. nov. and *A. malorum* sp. nov. *Internatinal Journal of Systemic and Evolutionary Microbiology* **52**: 1551-1558.

Constabel CP, Yoshida K & Walker V (2014) Diverse ecological roles of plant tannins: Plant defense and beyond. *Recent Advances in Polyphenol Research,* Vol. 4 (Romani A, Lattanzio V & Quideau S, eds.), p.^pp. 115-142. Wiley Blackwell, John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK.

Coutinho TA, van der Westhuizen L, Roux J, McFarlane SA & Venter SN (2015) Significant host jump of *Xanthomonas vasicola* from sugarcane to a *Eucalyptus grandis* clone in South Africa. *Plant Pathology* **64**: 576-581.

Craig L, Forest KT & Maier B (2019) Type IV pili: dynamics, biophysics and functional consequences. *Nature Reviews Microbiology* 429-440.

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

Deng W, Marshall NC, Rowland JL, McCoy JM, Worrall LJ, Santos AS, Strynadka NCJ & Finlay BB (2017) Assembly, structure, function and regulation of type III secretion systems. *Nature Reviews Microbiology* **15**: 323-337.

Figaj D, Ambroziak P, Przepiora T & Skorko-Glonek J (2019) The role of proteases in the virulence of plant pathogenic bacteria. *International Journal of Molecular Sciences* **20**: 1-30.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W & Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology* **59**: 307-321.

Jacques M-A, Arlat M, Boulanger A*, et al.* (2016) Using ecology, physiology, and genomics to understand host specificity in *Xanthomonas*. *Annual Review of Phytopathology* **54**: 163-187.

Jalali A, Alavi SM & Sangtarash MH (2017) Comparative genomic analysis of wide and narrow host range strains of *Xanthomonas citri* subsp. *citri*, showing differences in the genetic content of their pathogenicity and virulence factors. *Australasian Plant Pathology* **46**: 49-61.

Kirzinger MWB & Stavrinides J (2012) Host specificity determinants as a genetic continuum. *Trends in Microbiology* **20**: 88-93.

Kolmogorov M, Raney B, Paten B & Pham S (2014) Ragout-a reference-assisted assembly tool for bacterial genomes. *Bioinformatics* **30**: i302–i309.

Korus K, Lang JM, Adesemoye AO, Block CC, Pal N, Leach JE & Jackson-Ziems TA (2017) First report of *Xanthomonas vasicola* causing bacterial leaf streak on corn in the United States. *Plant Disease* **101**: 1030.

Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357-359.

Lee I, Kim YO, Park S-C & Chun J (2016) OrthoANI: An improved algorithm and software for calculating average nucleotide identity. *International Journal of Systematic and Evolutionary Microbiology* **66**: 1100–1103.

Madden T (2002) The BLAST sequence analysis tool. *The NCBI Handbook,*(McEntyre J & Ostell J, eds.), p.^pp. National Center for Biotechnology Information (US), Bethesda (MD).

Matilla MA & Krell T (2017) The effect of bacterial chemotaxis on host infection and pathogenicity. *FEMS Microbiology Reviews* **42**: 40–67.

Mhedbi-Hajri N, Darrasse A, Pigné S, Durand K, Fouteau S, Barbe V, Manceau C, Lemaire C & Jacques M-A (2011) Sensing and adhesion are adaptive functions in the plant pathogenic xanthomonads. *BMC Evolutionary Biology* **11**: 1-12.

NCBI (2008) BLAST® Command Line Applications User Manual. p.^pp. National Center for Biotechnology Information (US), Bethesda (MD).

O'Leary NA, Wright MW, Brister JR*, et al.* (2015) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research* **44**: D733–D745.

Overbeek R, Olson R, Pusch GD*, et al.* (2014) The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research* **42**: D206–D214.

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P & Tyson GW (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research* **25**: 1043–1055.

Roach R, Mann R, Gambley CG, Shivas RG & Rodoni B (2018) Identification of *Xanthomonas* species associated with bacterial leaf spot of tomato, capsicum and chilli crops in eastern Australia. *European Journal of Plant Pathology* **150**: 595-608.

Roden JA, Belt B, Ross JB, Tachibana T, Vargas J & Mudgett MB (2004) A genetic screen to isolate type III effectors translocated into pepper cells during *Xanthomonas* infection. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 16624-16629.

Saddler GS & Bradbury JF (2015) Xanthomonas. *Bergey's Manual of Systematics of Archaea and Bacteria*.

Sanko TJ, Kraemer AS, Niemann N, Gupta AK, Flett BC, Mienie C & Bezuidenhout CC (2018) Draft genome assemblages of 10 *Xanthomonas vasicola* pv. *Zeae* strains, pathogens causing leaf streak disease of maize in South Africa. *Genome Announcements* **6**: 1-2.

Schulze-Lefert P & Panstruga R (2011) A molecular evolutionary concept connecting nonhost resistance, pathogen host range, and pathogen speciation. *Trends in Plant Science* **16**: 117-125.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV & Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**: 3210-3212.

Studholm DJ, Kemen E, MacLean D, Schornack S, Aritua V, Thwaites R, Grant M, Smith J & Jones JDG (2010) Genome-wide sequencing data reveals virulence factors implicated in banana *Xanthomonas* wilt. *FEMS Microbiology Letters* **310**: 182-192.

Studholme DJ, Wicker E, Abrare SM*, et al.* (2020) Transfer of *Xanthomonas campestris* pv. *arecae*, and *Xanthomonas campestris* pv. *musacearum* to *Xanthomonas vasicola* (Vauterin) as *Xanthomonas vasicola* pv. *arecae* comb. nov., and *Xanthomonas vasicola* pv. *musacearum* comb. nov. and description of *Xanthomonas vasicola* pv. *vasculorum* pv. nov. *Phytopathology* 1-19.

Vicente JG & Holub EB (2013) *Xanthomonas campestris* pv. *campestris* (cause of black rot of crucifers) in the genomic era is still a worldwide threat to brassica crops. *Molecular Plant Pathology* **14**: 2-18.

Wallden K, Rivera-Calzada A & Waksman G (2010) Type IV secretion systems: versatility and diversity in function. *Cellular Microbiology* **12**: 1203–1212.

Washington EJ, Mukhtar MS, Finkel OM, Wan L, Banfield MJ, Kieber JJ & Dangl JL (2016) *Pseudomonas syringae* type III effector HopAF1 suppresses plant immunity by targeting methionine recycling to block ethylene induction. *Proceedings of the National Academy of Sciences* **113**: E3577-E3586.

Wasukira A, Coulter M, Al-Sowayeh N, Thwaites R, Paszkiewicz K, Kubiriba J, Smith J, Grant M & Studholme DJ (2014) Genome sequencing of *Xanthomonas vasicola* pathovar *vasculorum* reveals variation in plasmids and genes encoding lipopolysaccharide synthesis, type-IV pilus and type-III secretion effectors. *Pathogens* **3**: 211-237.

Wick RR, Judd LM, Gorrie CL & Holt KE (2017) Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Computational Biology* **13**: e1005595.

Xu L, Dong Z, Fang L*, et al.* (2019) OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Research* **47**: W52–W58.

Summary

A previous study suggested that an outbreak of bacterial blight and dieback of *Eucalyptus grandis* was as a result of a host jump event of *Xanthomonas vasicola* pv. *vasculorum* (*Xvv*) from sugarcane. The aim of this dissertation was to determine the host range of *X. vasicola* and to identify the genetic determinants which enabled this pathogen to switch hosts.

In **chapter 1**, a review of previous literature on the changes in the bacterial genome, which lead to its adaptation to a new host was conducted. The factors that lead to host adaptation include the modification of chromosomal genes by phase variation, and single sequence repeats contingency loci, the acquisition of new genes with novel functions by horizontal gene transfer, or the loss of redundant genes new environment.

In **chapter 2**, the pathogenicity of *X. vasicola* strains from different hosts was determined on the monocotyledonous plant species banana, maize, sorghum and sugarcane, and on the dicotyledon *E. grandis*. All the *X. vasicola* strains were pathogenic to maize, sugarcane, and *E. grandis*. Only the sorghum plants inoculated with *X. vasicola* from maize and sorghum displayed disease symptoms, and the banana plants remained asymptomatic throughout the experiment.

In **chapter 3**, the differences between the genomes of the five *Xvv* strains from *E. grandis* and genomes of *X. vasicola* from other hosts were identified. The five *Xvv* strains from *E. grandis* share a plasmid which has genes which encode for the type IV secretion system and tannase. The five strains are phylogenomically close to an *X. vasicola* strain from sugarcane, and these strains form a monophyletic group. Genes which encode for the type IV secretion system, metalloprotease, and chemotaxis proteins were identified in the orthologous clusters unique to the five *Xvv* strains from *Eucalyptus*. Type III effectors which protect the cells against ethylene and reactive oxygen species and type IV pilin genes were present only in the genomes of the five *Xvv* strains isolated from *Eucalyptus*.