MAKERERE UNIVERSITY

COMPARATIVE ANALYSIS OF GENOTYPIC DIVERSITY AMONG *XANTHOMONAS CAMPESTRIS* pv. *MUSACEARUM* AND *XANTHOMONAS VASICOLA* pv. *VASCULORUM* STRAINS

BY

WASUKIRA ARTHUR

BSc. Agric. (Mak), MSc. Crop Science (Mak)

Reg No: 2008/HD02/14609U

A THESIS SUBMITTED TO THE DIRECTORATE OF RESEARCH AND GRADUATE TRAINING FOR THE AWARD OF DOCTOR OF PHILOSOPHY OF MAKERERE UNIVERSITY

SEPTEMBER 2019

DECLARATION

This study is original and has not been submitted for any other degree award to any other university before

Signed: ……. Date……

ARTHUR WASUKIRA

School of Agricultural Sciences

Makerere University

This dessertation has been submitted for examination with the approval of the following Supervisors:

Signed…….. Date……

DR. GEOFREY TUSIIME

School of Agricultural Sciences

College of Agricultural and Environmental Sciences

Makerere University

Signed…… Date…….

DR. JEROME KUBIRIBA

National Banana Program

National Agricultural Research Organisation

COPYRIGHT

DEDICATION

To my dear family Stella, Patrice, Alvita, parents Mr. & Mrs. Patrick Stephen Wasukira Wanambwa and my Kuuku Lakeri Nakobe Nagimesi your consistent and persistent love and encouragement was the engine that has seen me thus far besides God's grace

ACKNOWLEDGEMENT

Firstly, I would like to express my sincere gratitude to my advisor Dr. J. Kubiriba and Dr. G. Tusiime for the continuous support of my Ph.D study for their patience, motivation, and immense knowledge. Your guidance helped me in all the time of research and writing of this thesis. I could not have imagined having better advisors and mentors for my Ph.D study.

Besides my advisors, I would like to thank the rest of my thesis committee: Prof. G. Murray, Dr. D. J. Studholme, and Dr.W. W Wagoire, for their insightful comments and encouragement, but also for the hard question which incented me to widen my research from various perspectives.

My sincere thanks also go to Dr. A. Agona, Dr. P. Musoli, Dr. G. Lubadde and Dr. L. Owere, who provided advice, an opportunity to join various teams, and who gave access to the laboratory and research facilities. I thank my fellow labmates for the stimulating discussions, working together before deadlines, and for all the fun we have had in the last many years. Also I thank my friends in the University of Exeter and Molecular Biology laboratory, NARL. In particular, I am grateful to Dr. Marta and the entire Molecular Lab team for enlightening me the first glance of advanced research.

Last but not the least, I thank my family: my parents and to my colleagues for supporting me spiritually and morally throughout writing this dessertation and my my life in general.

Philippians 4:13 I can do all things through him who strengthens me.

TABLE OF CONTENTS

List of Tables

List of Figures

Executive Summary

Banana bacterial wilt caused by *Xanthomonas campestris pv. musacearum* has caused significant decline in livelihoods dependant on the banana/musa enterprise in the great lakes region of Eastern Africa. First line control methods using cultural packages are not sustainable for long term application and yet resistant material is yet to be availed to the producers. The pathogen is also largely unknown and thus the need to understand host pathogen interaction at the molecular level. Recently, affordable genome sequencing has eased the application of genomic, post-genomic, and functional genomic approaches to problems of direct relevance to parasitic diseases in plants, humans and animals. Comparative genomic analysis of xanthomonas species pathogenic on musa was conducted as a means of unraffling the virulence factors associated with these strains. It refers to discovering how have Xanthomonas spp. are adapted for host- and tissue-specific pathogenesis, understanding bacterial plant pathogenesis and host- driven pathogen adaptation and potential to develop practical applications for crop protection through better disease control and prevention. Next generation sequencing of *Xanthomonas campestris pv. musacearum* (Xcm), *Xanthomonas vasicola vasculorum* (Xvv) and other musa associated xanthomonas was used to discover evolutionary relationships and composition of genetic factors in the strains. First the research sought to identify potential molecular markers of Xcm which information can be used in proposing a set of genetic material (primers) that are useful for diagnosis between banana infecting Xcm strains. Previous studies in other bacteria in this genus suggest that advanced draft genome sequences are valuable resources for molecular studies on their interaction with plants and could provide valuable tools for diagnostics and detection. Molecular markers are widely used in plant genetic research and breeding. Single-Nucleotide Polymorphisms (SNPs) are currently the marker of choice due to their large numbers in virtually all populations of individuals.

Secondly this study demonstrated the use genome-wide sequencing to systematically search genetic differences among Xvv and Xcm isolates. Comparative genome sequencing of bacteria is a powerful means of detecting sequence diversity among closely related, but distinct populations. Comparative whole-genome information about strain specific DNA variation has important implications for the development of new molecular markers for detection, pathovar classification, disease epidemiology and understanding evolutionary relationships. In nature, pathogen

populations with high genetic diversity have high evolutionary potential and thus are more likely to overcome host genetic resistance than pathogen populations with low genetic diversity. The resulting changes in population structure or virulence can lead to resistance breakdown. This is particularly true in agricultural production systems in which mono-culture is a common practice such as banana, maize and sugar cane production. Under these conditions, the frequency of pathogen genotypes with increased virulence may increase and ultimately lead to resistance breakdown and increased disease incidence. Therefore, availability of such genomic information on coding and noncoding- polymorphic loci will help in linking variability in pathogenicity of different strains to differences in their genetic backgrounds and monitoring changes in their genetic diversity. This study was undertaken to identify and characterize the genomes of six Xvv and two Xcm strains which identified variations are directly applicable for the development of molecular tools for pathogen detection and strain characterization, in understanding disease epidemiology and pathogen biology, and the development of novel disease management strategies.

Restriction endonuclease analysis (REA) and Restriction fragment length polymorphism (RFLP) analysis of genomic DNA are approaches to classify and differentiate strains. Some studies have developed fast molecular-based genus specific PCR diagnostic kits that are also independent of bacterial isolation and cultivation. Internal controls however are required by multiplexing with primers targeting specific genes which can further differentiate strains. Therefore genotypic characterization based on the analysis of restriction fragment length polymorphism of virulence factor fragment PCR products was conducted on selected members of the Xvm and Xvv from varying geographical locations. This study showed that indeed there can be differences in the effectors across bacterial strains and within strains of the same species and other clusters conserved in gram-negative bacteria. This variation indicates inherent capacity for host specialization using inbuilt resistance “R”genes of musa genomes, a response that can be explored to further understand the interaction of Xvm on susceptible and resistant hosts.

Finally it sought to genetically understand the interaction of selected banana-associated xanthomonas bacteria that are non-hypersensitive response and pathogenicity (Hrp). *Xanthomonas albilineans* is the causal agent of leaf scald, a major disease of sugarcane (*Saccharum* spp.). *Xanthomonas albilineans* closely related to Xanthomonas species isolated from banana in Samoa

and *Xanthomonas sacchari* belong to hyacinthi group together with X. *theicola, X. hyacinthi* and *X. translucens*. Comparative genomic analyses showed that most of the known pathogenicity factors from other *Xanthomonas* species are conserved in *X. albilineans*, with the notable absence of two major determinants of the "artillery" of other plant pathogenic species of Xanthomonas: the xanthan gum biosynthesis gene cluster, and the type III secretion system Hrp (hypersensitive response and pathogenicity). Six major clusters of xanthomanas based on Intergenic spacer sequences (ITS) as the phylogenetic tree with cluster I comprising *X. vasicola* and *X. campestris* whereas *X. albilineans* and *X. sacchari* are in Cluster II-VI. These are known to use a different infection mechanism from the Type Three Secretion System (TTSS).

Chapter One

General Introduction

1.1 Background

Bananas and plantains belong to the Musaceae family. There are 25-80 species in the genus Musa recorded. Musaceae contains only two genera, Musa and Ensete (Langhe *et al*., 2009). The wild form of *Ensete ventricosum* is widespread in tropical Africa from Kenya, Uganda and Tanzania south to Mozambique and South Africa (Transvaal), and west to the Democratic Republic of Congo. Enset is cultivated only in Ethiopia, where it was first domesticated possibly about 8000 years ago (Tsegaye and Struik, 2000). All banana and plantain cultivars were derived from two main species of Musa: *Musa acuminata* Colla (syn.s *M. cavendish* Lamb. ex Paxt., *M. chinensis* Sweet, *M. nana*, *M. zebrina* Van Houtee ex Planch.) and *Musa balbisiana* Colla. Hybrids of *M. acuminata* and *M. balbisiana* are sometimes given the names *Musa X paradisiaca* L., *Musa X sapientum* L., or perhaps most accurately, *M. acuminata* X *M. balbisiana* Colla. Genetic studies of the genus Musa which indicated that there are at least four wild species, M. acuminata (donor of the A genome), M. balbisiana (donor of the B genome), M. schizocarpa (donor of the S genome) and M. textilis (donor of the T genome), have contributed to the gene pools of domesticated bananas (Azhar M, 2008). In general, the most important banana cultivars in the world are triploids AAA while plantains are mostly AAB, ABB, or BBB. Communities of the East African coast and the nearby islands cultivate banana cultivars of different genome groups, including edible AA, various AAA, AB and ABB. The AB and ABB are considered to be of recent introduction both in inland Africa and along the coast (Karamura, 1998). The African plantains (AAB) are found in the humid forest lowlands of tropical central and western Africa. Others are cooking and beer type cultivars different from the *M. Acuminata* triploids which are mainly dessert types found along the East African coast.

The East African Highland Bananas (EAHB) are extensively pigmented black or brown with glossy Pseudo stems and robust dirty green leaves deeply split along the veins. The EAHB and the African plantains have a traditional dominance in the agricultural systems of tribes who cultivate them and they are staple foods in their respective regions. Furthermore, over a thousand

different types of cultivars of bananas and plantains are recognized today. The different varieties are characterized by wide differences in peel color and thickness, flavor, fruit size, and resistance to disease.

1.2 Economic importance of banana

Bananas are the world's most popular fruit and one of the world's most important staple foods, along with rice, wheat, and maize. World banana production is estimated at 107 million metric tons from more than 130 countries on 0.1 percent of the world's agricultural area, with a total trade value of US$9 billion (FAO, 2010) and with a retail value of approximately US$25 billion. Per capita annual consumption of bananas in Uganda is the highest in the world at 0.70 kg daily per person. Consumption in Rwanda and Burundi is less at 250 to 400 kg per person annually (about 3 to 11 bananas each day). Uganda itself is the second-largest producer of bananas in the world (Kilimo Trust, 2012). It is, however, one of the smallest exporters, the banana crop is mostly consumed by local markets. Less than 15% of the 144 million metric tonnes of banana produced in Uganda is exported. Africa exports 649,000 tonnes representing 3.9% of world traded banana. Uganda currently exports banana under the Fresh vacuum-sealed matooke (FREVASEMA) project to the United States and United Kingdom. A growing market segment is bananas marketed under Fair-trade and Global Gap (good agricultural practices) labels, but also organic bananas. The local and regional markets in Africa and Central Pacific countries are also regarded as markets of tomorrow. Uganda officially exports 1,200MT annually of matooke to Kenya and 20,000MT to Rwanda. Dessert bananas are in high and increasing demand in urban areas in Uganda and across the region. Uganda officially exports 5,000MT of dessert bananas to Kenya annually. Major regional outlets for Uganda include Kenya, Tanzania, South Sudan, Rwanda, and DR Congo. A recent comparative study of major staples in the EAC (Kilimo Trust, 2012) indicated that bananas are among the least traded. Although there is a growing cross-border trade, only modest amounts are exported internationally mainly to United Kingdom, Netherlands, Germany, and Portugal and USA (Kilimo Trust, 2012). According to three Afro-Carribean-pacific (ACP) countries (Uganda, Rwanda, and Burundi) are among the top 10 global banana producers, however the trade-in dessert bananas is dominated by five main exporting countries (accounting for 83% of the recorded global trade) of which none are ACP states (Agritrade, 2013).

The EAHBs are produced on 1.3 million ha in Uganda representing 38% of the arable land. Banana is 90% produced by smallholders on 0.15 – 10 ha of land (Bagamba et al., 2007) and is the main staple for over 20 million people. Of the bananas produced in Uganda, 90% is used for cooking while 10% is processed into beer or used as dessert. The volume of local processed banana beer produced in Uganda is 1.4 million liters whereas spirits amount to 2 million liters per year (Rietveld et al., 2013).

Banana is an important sustainable source of food since it fruits all year round, it is considered the cheapest source of potassium, Vit A & D and an income security crop. Banana also contributes to environmental conservation since it is perennial with good water-holding properties. Its broad leaves help cool and maintain soil structure by providing a protective soil cover throughout the year.

An analysis by Kilimo trust (2012) showed that banana is a better option than rice for food and income security. Uganda is the second-largest producer of banana, but the crop does not feature among the major traded commodities in the region. Regional demand for banana is high and mostly unsatisfied. It thus has a high potential for business and employment along the value chain with the growing demand for secondary processed products among higher-income groups. There are several opportunities for banana value chain development due to increasing urbanization, increased population growth, emerging new markets in South Sudan and Democratic Republic of Congo (DRC), improvements in road infrastructure and the increasing demand for industrial products namely food products/beverages/industrial spirits, crafts and animal feed.

1.3 Challenges in banana production

Constraints to banana production in Uganda are biotic, abiotic and socio-economic. In 2009 it was reported that banana production in Uganda had dropped by 49% mainly due to biotic stresses (pests and diseases). The biotic constraints and their corresponding yield losses according to the Uganda National banana program include banana weevils (60%), banana bacterial wilt (60 to 100%), nematodes (51%); black sigatoka (30-50%), banana streak virus (40%) and fusarium wilt (30%) (www.banana.go.ug). The banana suffers from several devastating diseases (Tushemereirwe *et al*., 2004). This is made worse a by new diseases and new virulent disease

strains. Among these are Black sigatoka which affects all East African Highland (EAHB) banana cultivars; Fusarium wilt which affects several introduced genotypes, banana streak virus that affects types of banana cultivars; and most recently Banana bacterial/ Xanthomonas wilt (BBW/BXW).

Xanthomonas campestris pv *musacearum* (Xcm) is the bacterial pathogen responsible for causing BXW. The spread of BXW is mainly via movement of contaminated farming and processing tools and insects feeding on floral parts of the plant (Karamura *et al*., 2006; Mwangi et al., 2006). Plants may be affected at any growth stage, including full maturity. Initial symptoms are a dull green color of the lamina which gradually assumes a scalded appearance and wilting back on its midrib. A cross-section of diseased petioles or pseudostems produces a yellowish ooze. There is uneven and premature ripening of the fruit and when fruit are cut, the sections show yellowish blotches in the flesh fingers and dark brown placental scars. Eventually the whole plant wilts and dies out leading to absolute loss of yield. The most effective control option, which is the use of resistant germplasm is hindered by lack of resistant varieties amongst the current gene pool unfortunately. With suitable land becoming scarcer, new approaches need to be considered to address the decline in food production in the sub-region.

Wairegi *et al*. (2010) reported that low soil fertility, low soil moisture, and sub-optimal crop management are the major abiotic constraints. Bananas take up a lot of nutrients from the soil; on average bananas remove 250 kg N, 26 kg P, 830 kg K, 105 kg Mg and 15 kg S per hectare to yield 40 tonnes/ ha (Ndabamenye *et al*., 2013). (Henao & Baanante, 2006) indicated that nutrient loss in Uganda stood at 66 Kg/ha/yr placing it among the highest in the region. Unfortunately fertilizers are often expensive or not available and yet blanket recommendations may not fit site-specific requirements. Lack of information on technologies with high cost of inputs and lack of water harvesting technologies are other compounding factors. Lack of supportive policies for the whole value chain especially the post-harvest sector, lack of varieties for post-harvest utilization mechanisms lead to reduced benefits of banana production and trade.

Figure 1.1: Typical symtoms of Xanthomonas pv. musacearum infected banana: a). Wilting of sheeths on male bud, b). Extensive wilting of leaves c). Discolored inner pulp of fingers on infected bunch d) premature ripening of infected bunch e) Deep yellow bacterial ooze on pseudostem (Photo credit: Bioversity international)

In Uganda, BXW was first reported in 2001, on bananas in the Central districts of Mukono and Kayunga. It has since been reported and confirmed present in many other districts of the country. BXW has also been reported in other countries of the great lakes region particularly Rwanda (2007), Burundi (2011), DR. Congo (2004), Tanzania (2006) and Kenya (2008) mainly in the areas bordering Uganda (Carter et al., 2010; Hakizima & Gallagher, 2008; Ndungo et al., 2008; Ndungo et al., 2006).

1.4 Banana bacterial wilt control

There has been reduction in banana production over the last decade. This coincided with the first occurrence and rapid spread of BXW that massively decimated the crop across production systems in East and Central Africa (Smith et al., 2008). BXW is generally controlled by cultural practices. It however involves investment in terms of financial, human and time resources consequently resurgence of BXW sometimes occurs. Research institutions in Uganda in collaboration with International and regional partners have developed a package of cultural practices for controlling BXW (Kubiriba et al., 2014). In Uganda, removal of male flower buds, disinfecting farm tools and hands after use, destruction of affected banana stools and burying the whole plant reduced the inoculum (Escalant et al., 2004; Kubiriba & Tushemereirwe, 2014; Smith et al., 2008; Tinzaara et al., 2013; Tripathi et al., 2009). Restriction of movement of banana plant material from infected areas to other areas has also reduced the spread of BXW. The package is usually abbreviated ABCC, which stands for avoidance (avoiding infected planting material by using clean planting material); Breaking off male buds (timely removal of male buds with forked stick); Cut out diseased plants and Cleaning cutting tools (disinfecting farm tools) (Kalyebara et al., 2006). This control package if fully (as a package instead of separate components) and correctly applied is effective in reducing the disease incidence and can eventually eradicate the disease on affected farms (Kubiriba & Tushemereirwe, 2014; Tripathi et al., 2009). Banana farmers find it difficult to employ the above package because they consider it slow and labor intensive. Accordingly the disease has persisted in several areas (Bioversity, 2011), Kubiriba et al. (2014) also affirmed that the continued spread of BXW was accelerated by the user unfriendliness of some of the control recommendations. Banana producers in the region previously abandoned BXW disease control options due to being overwhelmed by the rapid disease spread and slow response of plantations. It has been suggested that affected or threatened communities be involved in formulating and enforcing by-laws rather than government top-down enforced laws to effectively control BXW.

An effort to improve landraces for BXW resistant material through classical breeding has not been successful because no source of BXW resistance has been identified in banana germplasm (Ssekiwoko *et al*., 2006a). However recent screening of different *Musa* spp. has revealed

seemingly ‘resistant’ or tolerance in *M. balbisiana*, *M. zebrina* and *M. ornata* which can be used as inherent sources of resistance genes if the banana xanthomonas virulence factors are identified and functionally characterized (Ssekiwoko *et al*., 2006). Biotechnology approaches using genes from sweet pepper have been attempted to confer resistance to BXW with some successes (Tripathi *et al*., 2004; Karamura & Tinzaara, 2007). Tripathi *et al.* (2010) used the *Hrap* gene from sweet pepper while Tripathi *et al.*, 2014 used a Xa21 recognition receptor from rice and were able to show 100% resistance against BXW under confined field conditions. The available knowledge on BXW symptom identification, spread and severity, combined with improved coordination of disease management at regional and country level can reduce the losses significantly but is not sustainable.

1.5 Problem Statement

Xanthomonas is a complex group of plant-pathogenic bacteria that infect diverse crops in a host-specific manner (Pfeilmeier et al., 2016). Over 20 Xanthomonas species are divided into two main phylogenetic groups based on 16S rDNA and gyrB sequence analysis (Parkinson et al., 2007) and subdivided into pathovars loosely corresponding to host specificity. The BXW epidemic has spread fast and wide from ensete to banana in Eastern Africa region and yet origin of infection is attributed to Uganda although it was first reported in Ethiopia. There is need to estimate the emergence of establishment of Xcm lineages in Eastern Africa. The extent of isolation between pathogen populations is conditioned by the the degree of adaptation to hosts. Gene flow remains possible for populations that have diverged for a long time. Specialization on hosts greatly contributes to reproductive isolation and limits gene flow (Mhedbi-Hajri et al., 2013). The study of the relative importance of gene flow and divergence times between populations occupying different hosts is crucial for understanding evolutionary histories and emergences of pathogens. These populations exhibit detectable amounts of de novo evolutionary change among genetic sequences sampled at different time points (Biek et al., 2015). Host specialization is one of the characteristic features of highly evolved pathogens such as the Xanthomonas group of phytopathogenic bacteria. Since the hosts involve staple crops, detailed understanding of the diversity and evolution of such strains infecting diverse plants is important for quarantine purposes.Whole-genome information potentially enables the understanding of the variation and

evolution of pathogens at an unprecedented level (Midha and Patil, 2014). Molecular dating is used in the biological sciences to estimate the age of evolutionary events (Duchene and Bromham, 2013). Changes to DNA and amino acid sequences accumulate continuously in the genome over time, so comparing DNA sequences between lineages allows estimate of the time since they last shared a common ancestor. Molecular dating methods begin by estimating the amount of genetic difference that has occurred between sequences. Then, assumptions about the rate of change are used to infer the amount of time needed to explain that amount of genetic difference. Further, whole-genome sequences are a rich resource for developing markers for quarantine purposes and epidemiological surveys. Apart from resolving the relationship, there is also a need to study evolution of gene(s) that are known to be important for virulence, pathogenicity and fitness. There is need to gain insight into the evolution of Xanthomonas strains pathogenic to banana, maize and sugar cane.

Xanthomonas campestris is the most dominant xanthomonas species with at least 141 pathovars, which infect a wide range of plants including many of agricultural interest such as banana, tomato, pepper and cotton (Leyns et al., 1984). Several studies have been conducetd on the pathogenicity of Xcm on banana and other associated plants (Ssekiwoko et al., 2006, Chala et al., 2016, Tripathi et al., 2007). The response of several species to infection by Xcm has been largely studied but without indepth description of the pathogen infection mechanism. Alternative host plants are known to be important in the survival and perpetuation of several crop pathogens and were suspected to play a role in the survival of Xcm and perpetuation of Banana Xanthomonas wilt (BXW) disease of banana and enset (Ocimati et al., 2018). That study determined the potential risk posed by two weeds (Canna spp. and wild sorghum) and common banana intercrops (maize, millet, sorghum, taro, and sugarcane) as alternative hosts to Xcm. Both pathogen factors and host factors govern these interactions, and which of these two is dominant likely depends on the specific interaction. There is a growing body of data that clearly demonstrates that pathogens have evolved compatibility factors that facilitate the disease process in a host-specific manner (Alfano and Collmer, 2004). Plant cells recognize various microbial signature molecules such as flagellin, lipopolysaccharide, that act as Microbial/Pathogen Associated Molecular Patterns (MAMP/PAMP) (Boller and He, 2009). To circumvent this recognition and hence trigger immunity, bacteria release specialized protein molecules directly into the host cell known as type

III effectors using its well-evolved type III secretion system (White el al., 2009). In turn, plants have also evolved specialized Resistance (R) genes that act in response to effector molecules (Gu et al., 2005). There are limited studies into the characterization of virulence factors that play a general role in common plant disease epidemics caused by xanthomonas that act in a host-specific manner. The biology of the Xanthomonas-banana pathosystem remains largely unexplored. It is also not known if the factors that are responsible for pathogen compatibility on different host species are fundamentally different from the factors that determine compatibility among different cultivars of the same host species. For example, host specificity factors may play important roles during initial colonization, during pathogen migration to the appropriate tissue or cell type, during the initiation of cellular interactions, during the maintenance of these interactions, or even at the point where the pathogen disperses to a new host (Levin 1996). To address these questions, we first must have a way of identifying host-specific virulence factors. A reasonable hypothesis is that strains that are able to infect and cause disease on common host will share a common set of host-specificity factors. This can be answered by sequencing the genomes of several geographically distinct Xanthomonas strains. This study therefore aimed at identifying the virulence factors deployed by Xcm and some other banana infecting xanthomonas which information contributes to future development of disease management strategies against BXW.

Successful transformation of banana for resistance to BXW require understanding of molecular mechanisms underlying host-pathogen interactions. Bacterial blight caused by *Xanthomonas oryzae pv. oryzae* (Xoo) leads to a substantial yield reduction of up to 50% in most rice-growing regions. Host plant resistance is an effective control method, and more than 30 resistance genes have been identified in rice genotypes (Dossa et al., 2014). Susceptibility genes have the potential to be used in resistance breeding; however, because susceptibility genes have a function other than being a compatibility factor for the pathogen, the side effects caused by their mutation demands a one-by-one assessment of their usefulness for application. Global functional analyses will elucidate the complex regulatory networks and the diversity of proteins involved in resistance and susceptibility.Justification of the study

This study contributes to the knowledge of *Xanthomonas campestris pv. musaxearum* (Xcm) strains evolution in great lakes and the diversity of Xcm and *Xanthomonas vasicola pv.*

vasculorum (Xvv) virulence factors. According to Weller and Martin (2015), molecular evolutionary rate varies significantly among species and a strict global molecular clock has been rejected across the tree of life. Generation time is one primary life history trait that influences the molecular evolutionary rate. Theory predicts that organisms with shorter generation times evolve faster because of the accumulation of more DNA replication errors per unit time. Although the generation-time effect has been demonstrated consistently in plants and animals, the evidence of its existence in bacteria is lacking. Evolution genes generally infuence the kind and the frequency of genetic variation occurring in a population of organisms and sometimes they may ensure a certain degree of genetic isolation. Evolution genes have themselves undergone a long evolution and are now fine-tuned for their present functions. The theory of molecular evolution postulates that the products of evolution genes work hand in hand with other factors such as structural and stability features of biological macromolecules, the action of chemical and physical mutagens, as well as the chance of random encounters of interactive components (Arber, 2000). On the basis of established knowledge of microbial genetics there are three major natural strategies in the spontaneous generation of genetic variations in bacteria. These strategies include small local changes in the nucleotide sequence of the genome, intragenomic reshuffling of segments of genomic sequences and the acquisition of DNA sequences from another organism. The three general strategies differ in the quality of their contribution to microbial evolution.

In order to be more effective in limiting spread of Bxw, the availability of easy to use disease identification kits speed up the awareness of of the pathogen strain and thus developmentof sustainable disease control packages. Besides, good detection methods will facilitate regional trade because regulatory institutions need effective tools for pathogen detection. Significant progress has been made towards developing polymerase chain reaction (PCR) based diagnostic kits for general xanthomonas groups (Adriko et al., 2012, 2011). PCR based approach can reliably and effectively differentiate xanthomonads from other gram negative bacteria as well as identifying the isolates belonging to the two described groups of the genus Xanthomonas.

All strategies of plant disease management are founded on a fundamental understanding of the ecology and biology of plant pathogens. Crop protection methods include surveillance and quarantine, deployment of host resistance, use of biocides and biocontrol agents, agricultural

practices to limit the contact of crops with pathogens and means to forecast disease. All of these methods and the estimations of their durability are based on a knowledge of the genetic and phenotypic diversity of plant pathogens, their means and extent of dissemination, the extent of their host range and of habitats in which reservoirs of inoculum are found, their mechanisms of virulence to plants and how these processes are regulated by the biotic and abiotic environment (Morris et al., 2017). Xanthomonads population infecting plants of different species and families have high genetic diversity including variable patterns of avirulence genes. Proteins secreted via the Type 3 Secretion System (TTSS) into host cell are one of major pathogenicity mechanisms of gram negative bacteria. Proteins, delivered by the TTSS are called TTSS effectors. The majority of effector have unknown function yet. Some known effector proteins can block signals of innate immunity pathway. Entering the host cell they prevent development of plant reaction on non specific elicitors MAMPs (microbe associated metabolicprofiles). Effector proteins are essential for virulence and host specificity of all bacteria species with TTSS. Over 20 effector genes were described for xanthomonads in previous research. Homologous effectors were found in bacteria of different species and genera, and horizontal transfer of them was postulated. Publication of several complete genomes of plant pathogenic bacteria accelerated identification of new effector genes by identification of conservative promoters, specific signals and domens (Mokryakova et al., 2010). Many studies have identified virulence factors in model organism *Xanthomonas oryzae pv. oryzae* (Xoo) using the molecular approach while little is known of their biochemical properties (Dossa et al., 2014). Because of the essential biological functions of TTSS during bacterial association with eukaryotic hosts, a lot of research has been conducted to identify effector proteins. A major challenge for the identification of TTSS is that there is no defined signal peptide or motif discovered from the amino acid sequences of known effectors. The plant pathogen *Pseudomonas syringae* has been a model for the research of type III effectors. Thus far, over two hundred TTSS have been identified and confirmed in *P. syringae* (Schechter et al., 2006). Bioinformatic approaches for predicting TTSS based on genomic data have attracted a great deal of research interests. There are three main computational methods: promoter-based, secretion/translocation signal-based and homology-based, which could be used for the prediction of virulence factors (Arnold, Brandmaier, Kleine, Tischler, & Heinz, 2009). In a bacterium, genes encoding the TTSS apparatus and effectors usually have a conserved regulatory motif in their

promoters (Yang et al., 2010). Aligning candidates with known effectors for homology search would be the most straightforward way to identify TTSS effectors. However, TTSS effectors have great sequence diversity through fast evolution and many TTSS effectors have no homology with any protein in the public databases. According to Michelmore, (2000) genomic approaches can revolutionalize our understanding of plant disease resistance. Diseases arise from compatible interaction between plant and pathogen thus, altering a plant gene that critically facilitates compatibility could provide a more broad-spectrum and durable type of resistance (Birch & Kamoun, 2000). The potential sources of resistance genes may be identified through an assessment of genomes with resistance to related Xanthomonas sp. bacterial diseases including a comparison with the pathogen genomes (Stall et al., 2009). Tripathi *et al*. (2010) showed that constitutive expression of the sweet pepper hypersensitive response associated protein (*Hrap)* gene conferred resistance to BXW. The HRAP protein acts on harpin proteins, which are produced by many pathogens and sensed by plants triggering defense responses. HRAP dis-sociates harpin multimeric protein forms into monomers and dimersand these forms elevate HR and prevent propagation of a bacterial pathogen like Xcm. The Hrap is one of the important hypersensitive cell death associated genes that could be utilized for protecting plants from bacterial pathogen attack. Namukwaya *et al*. (2012) also indicated that *Sap 1* gene was promising against the wilt pathogen. Sweet pepper amphipathic protein (*sap*1), previously isolated and cloned from sweet pepper (*Capsicum annuum*), is reported to significantly enhance harpin-mediated hypersensitive response (Dayakar et al., 2003).

The above approaches will provide detailed information on pathogen infection processes. Genomic and post-genomic information give opportunities for comparison between different pathogens, or related non-pathogens. This ‘triangulation’ of information can reveal processes that are conserved, or which differ between pathogens and their hosts and hence highlight possible areas to target in disease control strategies (Raskin et al., 2006). Such information is limited for Xcm, unlike other Xanthomonads. There is need therefore to understand the genomic adaptations of Xcm and other musa associated xanthomonads that make it a successful pathogen in *Musa* or other hosts. Management strategies for plant bacterial diseases require a thorough knowledge of the pathosystem in order to identify the most appropriate timing for the targeting of pathogen populations and to determine when the critical host tissues are vulnerable to infection. This study

therefore aimed at determining the virulence factors associated with selected isolates of *Xanthomanas* on musa, sugar cane and maize and tracing the origin of Bxw in Eastern Africa.

1.6 General and specific objectives

The general objective of this study was to generate knowledge for understanding the molecular basis for X*anthomonas campestris* pv. *musacearum* pathogenesis on banana host.

Specific objectives:-

i) To validate the predicted virulence factors in two sequenced isolates of *Xanthomonas campestris* pv. *musacearum* and *Xanthomonas vasicola* pv. *vasculourum*

ii) To determine the evolution and variation of *Xanthomonas campestris pv. musacearum and Xanthomonas vasicola* pv. *vasculourum* isolates in the great lakes region

iii) To genetically characterize pathogenicity virulence factors in isolates of *Xcm, Xvv and* non-hrp banana associated xanthomonads

1.7 Research questions

1. Are all predicted virulence factors of Xcm4381 and Xvv702 present in other strains of Xcm and Xvv?
2. What is the evolution of supposedly monomorphic pathogen Xcm and other Xcm strains in the great lakes region?
3. How do the virulence factors of Xvv and Xcm isolates occur?
4. Do all Musa associated xanthomonas have the type three secretion system?

1.8 Hypotheses

1. All Xcm infections in East Africa are linked to the Ugandan isolate originating from Ethiopia
2. There is no genetic variation between Xanthomonas infecting banana, maize and sugar cane
3. All banana infecting xanthomonas isolates use the type three infection system for successful host colonisation

Chapter Two

Literature Review

2.1 Banana Xanthomonas Wilt

The genus *Xanthomonas* is formed by a complex group of yellow-pigmented gram-negative bacterial species with diverse physiological.traits. Pathogenic species and pathovars (pathogenic varieties, pv.) within species show a high degree of host plant specificity and combined are known to cause diseases on nearly 400 plant hosts, including both eudicots and monocots (H. Lu et al., 2008). This group of phytopathogens infects many economically important plants such as cotton, beans, rice, crucifers, banana, sugar cane and maize. Many exhibit tissue specificity, invading either host xylem vessels or the interveinal mesophyll apoplast. Some of the most important diseases caused by Xanthomonas species include citrus canker, caused by *Xanthomonas axonopodis* pv. *citri* and black rot, caused by *Xanthomonas campestris* pv. *campestris*. Additionally, Xanthomonas is also important in biotechnology due to the production of a polysaccharide known as xanthan gum, which is used as a stabilizer and emulsifier in the food and cosmetic industry, among other applications (da Silva et al., 2002).

The causal agent of BXW is *Xanthonomas campestris* pv. *musacearum* (*Xcm*). Various study comparisons have shown that isolates of *Xcm* are similar to *X. vasicola* (Karamura et al., 2015; Rodriguez-r et al., 2012;Aritua et al., 2008). This comprises a sorghum pathogen, *X. vasicola* pv. *holcicola,* and sugarcane and maize pathogen, *X. vasicola* pv. *vasculorum*. Key virulence factors of phytopathogenic bacteria are plant cell wall degrading enzymes, phytotoxins, extracellular polysaccharides and phytohormones, which are central for the pathogenesis of necrotrophic bacteria (Prasannath, 2013). The wilting symptoms in infected banana are as a result of intensive multiplication of the bacteria within the xylem vessels. The bacteria produce large quantities of exopolysaccharides (xanthan gum) that block off water flow in the vascular system resulting into wilting of the banana host (Mwangi, 2006; Ssekiwoko et al., 2006b; Tripathi et al., 2009). Molecular and genetic approaches have identified a wide range of genes and functions that play a role in the infectionand multiplication of a bacterium within plant tissues leading to development

of a disease. Todar, (2014) have also demonstrated that such genes are organized in complex regulatory networks. The genes, commonly termed effectors, are delivered into different cellular compartments of the plant hosts to modulate plants' defense circuitry and enable parasitic colonization (Mudgett, 2005; White et al., 2009). The avirulence (avr) and/or virulence (vir) genes in most gram-negative plant pathogenic bacteria are associated with host colonisation. The host specificity is determined via gene-for-gene interactions, and the hypersensitivity response and pathogenicity (hrp) genes. The genes are usually clustered in a chromosomal region that spans 20 to 30 kb and are involved in induction of Hypersensitive response (HR) in resistant host and nonhost plants and pathogenicity in susceptible host plant (Alfano & Collmer, 2004; Collmer et al., 2009). The hrp genes, on the other hand, encode a type III-secretion-systems (TTSS) through which the avr genes are injected into host cells. TTSS have been isolated in species of several gram-negative bacteria (*Salmonella, Yersinia, Shigella, Escherichia, Pseudomonas,* and *Xanthomonas*). They are known to consist of over 20 different subunits which enable these bacteria to translocate the effectors directly into the cytoplasm of the host cell to exert a broad range of virulence functions (Denancé et al., 2016; Lahaye & Bonas, 2001; Tampakaki et al., 2004).

2.2 Types of bacterial secretion systems

The dual membrane envelope of Gram-negative bacteria presents a double barrier between the cytoplasm and the external environment. Three secretion systems, type I, III and IV have evolved to transport proteins over both membranes in a single unified process (Preston et al., 2005). According to Tseng et al., (2009) bacteria form a very wide diversity of biotic associations, ranging from biofilms to mutualistic or pathogenic associations with larger host organisms. Protein secretion plays a central role in modulating all of these interactions. In gram-negative bacteria, where secretion involves translocation across inner and outer membranes, there are now known six general classes of protein secretion systems, each of which shows considerable diversity. The six types of secretory systems (SS), are critical for exporting proteins to the external environment or for directly contacting the host (Ren, 2009;Tseng et al., 2009). Secretion systems can be classified into two groups: (I) one-step secretion systems, wherein secreted proteins are exported across the inner and outer membranes in a single step, include the type I (T1SS), type

III (TTSS), Type IV (T4SS), and type VI (T6SS) secretion systems (fig. 2.1); or (II) two-step secretion systems, wherein proteins are first exported into the periplasmic space via the general secretion (Sec) or two-arginine (Tat) pathways, and then translocated across the outer membrane via the type II (T2SS), type V (T5SS), or less commonly, the T1 or T4SS systems (Saier, 2006) .

Bacteria have evolved a wide variety of highly specialized macromolecular nanomachines that secrete a wide range of substrates, including small molecules, proteins and DNA. These substrates have key roles in the response of a bacterium to its environment and also in several physiological processes such as adhesion, pathogenicity, adaptation and survival (Costa et al., 2015). Bacterial survival and multiplication depends on its ability to produce effector molecules to obtain nutrients from its host plant and cultivate the right environment in which to establish infection (Büttner & Bonas, 2002; Büttner & Bonas, 2009; Kay & Bonas, 2009). Depending on the secretion system, the secreted substrates have three possible fates: they remain associated with the bacterial outer membrane (OM), they are released into the extracellular space, or they are injected into a target cell (either a eukaryotic or bacterial cell).

This is the simplest pathway it allows direct secretion of effectors from the bacteria cytosol to the external environment. Secretion is achieved in a single step directly from the cytosol to the extracellular space. The translocation machinery is composed of three indispensable membrane proteins, two in the inner membrane, and the third in the outer membrane. The inner membrane proteins belong to the ABC transporter and membrane fusion protein families (MFPs), respectively, while the outer membrane component is a porin-like protein (Ib et al., 2015)

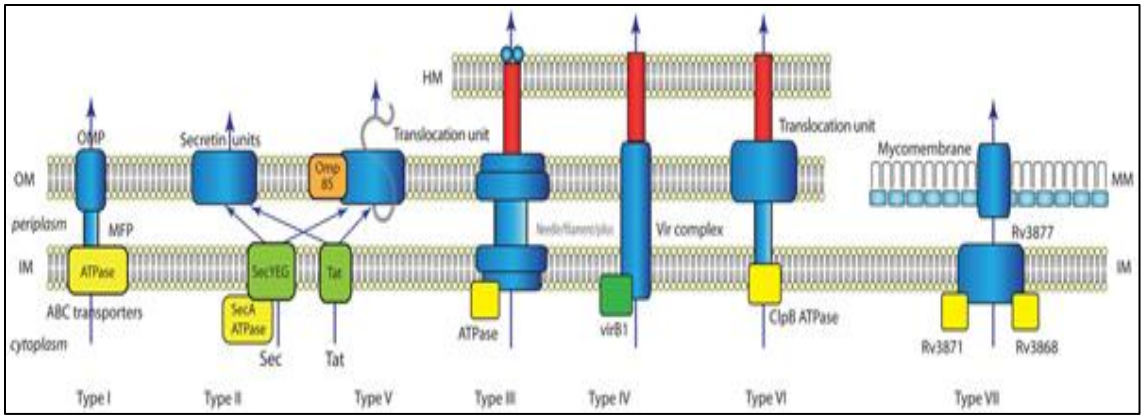

Figure 2.1: Summary of known bacterial secretion systems. In this simplified view only the basics of each secretion system are sketched. HM: Host membrane; OM: outer membrane; IM: inner membrane; MM: mycomembrane; OMP: outer membrane protein; MFP: membrane fusion protein. ATPases and chaperones are shown in yellow. (Credit: Tseng et al., 2009).

2.2.1 Type II secretion system (T2SS)

The type II secretion pathway is conserved in gram-negative bacteria with prevalence in bacterial pathogens of plants *Pseudomonas fluorescens*, Erwinia or Xanthomonas species (Cianciotto, 2005). It involves a set of 12 to 16 different proteins. Components of the T2SS are located in both the inner and outer membranes where they assemble into a supermolecular complex spanning the bacterial envelope, also called the secreton. The T2SS substrates transiently go through the periplasm before they are translocated across the outer membrane and exposed to the extracellular milieu. The T2SS is unique in its ability to promote secretion of large and sometimes multimeric proteins that are folded in the periplasm (Douzi et al., 2012). This pathway is composed of a more complex secretion structure, and requires transport of proteins to the periplasm and secretions across the outer membrane are required for secretion of an effector. Transport of virulence factors into periplasm requires an N-terminal signal sequence, and during transfer to the periplasm, the protein is processed by a signal peptidase. The intermediate location in the periplasm allows proper folding of the effector before it is secreted.

2.2.2 Type IV Secretion System (T4SS)

Type IV Secretion System (T4SS) uses complex structures similar to flagella and conjugation structures to interact with the eukaryotic host cells and deliver their effectors. The type IV pathway is best known from studies on *Agrobacterium tumefaciens*. This pathway is the only secretion pathway described to translocate both proteins and nucleic acids. Type IV pili (Tfp) are widespread among diverse members of the β-, γ-, and δ-Proteobacteria, Cyanobacteria and Firmicutes. T4P are therefore found in both Gram-negative and Gram-positive bacteria, supporting an ancient origin for this kind of pili. Tfp are proteinaceous, flexible filaments with a diameter of 5–8 nm, which extend up to several micrometers in length. They are generally located at one or both poles of a cell, where they mediate twitching motility, an efficient and versatile flagellar-independent form of bacterial surface motility. Tfp have also been shown to be involved in a variety of other bacterial activities and features, including surface adhesion and colonization, biofilm formation, genetic material uptake and virulence (Burdman, et al., 2011).

2.2.3 Type V Secretion System (T5SS)

Among all the types of secretory systems, Type V Secretion System (T5SS) is the simplest one. A characteristic of the secretion of these proteins is that it acquires an active confirmation as its extracellular transport proceeds. It is released as a proform from the membrane-bound helper by autoproteolysis. This precursor protein directs the transport of protease from the periplasm to the extracellular milieu, without the involvement of any other accessory protein. Therefore, the proteins secreted via T5SSs are also called “autotransporters” (Henderson & Navarro-garcia, 2004).

2.2.4 Type VI Secretion System (T6SS)

The Type VI Secretion System is the most recently discovered of the six mechanisms for effector secretion elucidated in gram-negative bacteria, particularly notable for its significant structural homology with T4 bacteriophage. This protein secretion system shares similarity with the puncturing device of bacteriophages in structure. The basic structure consists of thirteen proteins necessary for function. The compounds delivered by the T6SS have not been characterized in most bacterial species (Green et al., 2010; Ren, 2009; Shrivastava & Mande, 2008; Silverman et

al., 2012).

2.2.5 Type III secretion system (TTSS)

The type III secretion system (TTSS) is found in gram-negative bacteria that interact with both plant and animal hosts, either as pathogens or mutualists (Grant et al., 2006). The machinery of the TTSS, termed the injectisome, appears to have a common evolutionary origin with the flagellum (Buttner & He, 2009; He, 1998). TTSS play a central role in the host interactions of many Gram-negative plant pathogens including *P. syringae, P. aeruginosa, R. solanacearum, Xanthomonas sp.* and *E. carotovora*. The flagellar biosynthetic mechanism also represents a form of type III secretion. The principal known function of the injectisome is to deliver effector proteins across the bacterial and host membranes into the cytosol of host cells, where they may modulate a large variety of host cell functions, including immune and defense responses. In some cases however, effector proteins are simply secreted out of the cell (Büttner & Bonas, 2003; Cornelis & Gijsegem, 2000). The TTSS structure is characterized by a needle-like protruding structure (Fig 2.2) with a channel along which proteins travel thus resembling bacteria flagella, both at the structural and functional level. It is a syringe needle-like translocation apparatus, consisting of inner and outer membrane rings and a protruding filament called pilus. The genes encoding the TTSS are called the hrp (hypersensitive response and pathogenicity) genes in phytopathogenic bacteria. The hrp genes are usually arranged in clusters and are located in pathogenicity islands (PAIs), which are discrete regions that vary in G+C content from the overall genome and are flanked by insertion sequences, bacteriophage genes, and transposable elements (Galán & Collmer, 2007). PAI are defined as regions of DNA that contain virulence genes and are found in the genome of pathogens, but absent from or only rarely present in non-pathogenic variants of the same or related strains (Prasannath, 2013) Other characters include distinct boundaries from the rest of the genome and the presence of genes related to mobile elements such as transposases, integrases and insertion sequences. The role of the mobile elements associated with PAIs is still unclear, they are assumed to promote the spreading of PAIs to other bacterial strains. These features suggest that PAIs originated from other species, and that they were acquired by horizontal gene transfer. The newly acquired genetic material may confer new pathogenic and fitness traits to the bacteria. The hrp genes encode proteins that either regulate synthesis or assembly of the

TTSS, are structural components of the TTSS, or are extracellular proteins (e.g., harpins) secreted by the TTSS.

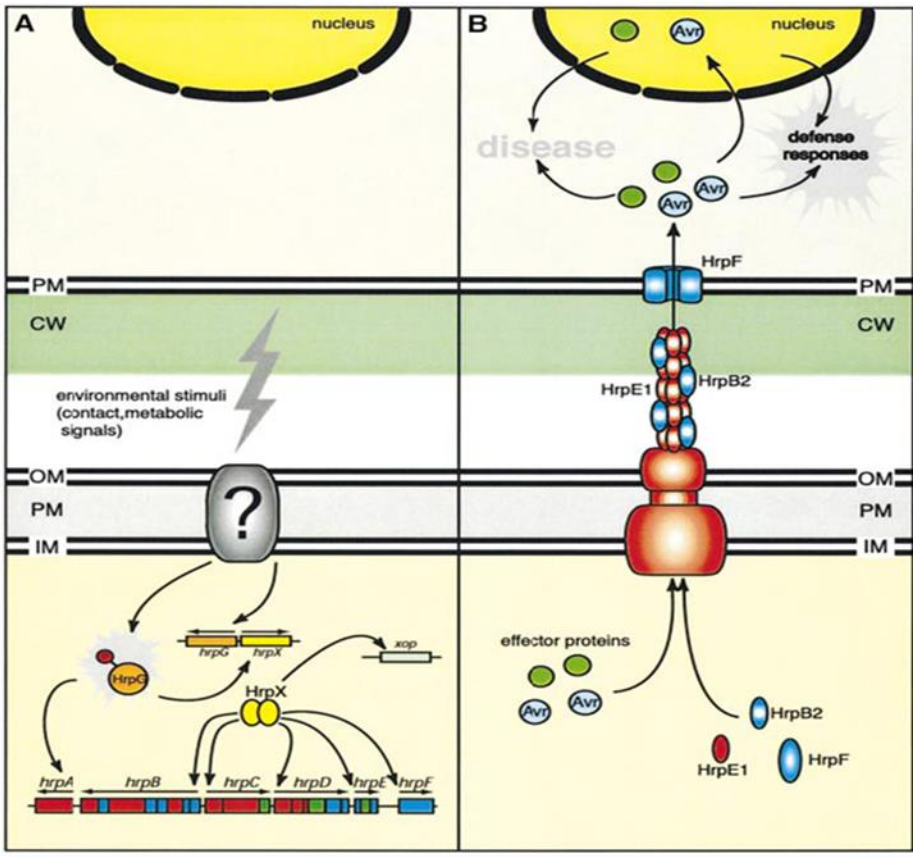

Figure 2.2: Schematic representation of the Type Three Secretion System Infection process: A) Composition of virulence factors in the bacterial cell before stimulus from the environment; B) Assembly of pilus and traslocation of virulence factors across the cell membrane (Credit: Buttner & He, 2009)

2.3 TTSS- Host plants interactions

Many bacterial pathogens use a specialized protein ‘‘injection needle’’ called a type III secretion system to help colonize cells of higher organisms. The type III secretion needle attaches to a host cell and is the delivery conduit for a variety of bacterial proteins that act inside of the host cell. These proteins are called type III effectors. They manipulate host cell biology in order to help the bacterial pathogen colonize the host. The TTSS secretes helper and accessory proteins into the extracellular environment to support the injection of effector proteins into the host cell (Cornelis & Gijsegem, 2000). Because of their shape and their ability to translocate proteins in a cell

contact-dependent manner, the TTSS are also referred to as 'injectisomes' or 'molecular needles'. The TTSS machinery in the phytopathogens is designated as hrp (hypersensitive response and pathogenicity) system (Gurlebeck et al., 2006; Lindgren, 1997). The injectisome consists of two parts, an envelope embedded multi-ring base and a long protruding surface appendage, called the hrp pilus. Hrp pili, have been described for *P. syringae, R. solanacearum, Erwinia amylovora* and *Xanthomonas campestris pv. vesicatoria*, elongate distally with the addition of their major component, Hrp pilin subunits, whereas TTSS effectors are secreted from the hrp pilus tip.

Stebbins (2005) determined that most TTSS effector molecules are dependent on chaperones, which keep the effectors in a partially unfolded state in the bacterial cytosol. Even though the pilus proteins, HrpA (*P. syringae* and *E. amylovora*), HrpY (*R. solanacearum*) and HprE (*X. campestris* pv. *vesicatoria*) do not share any significant homologous sequence, they exhibit a number of physico-chemical features in common. In the plant, the effectors presumably interfere with cellular processes to the benefit of the pathogen or have an avirulence activity that betrays the bacterium (fig. 2.3) to the plant surveillance system (Gürlebeck et al., 2006).

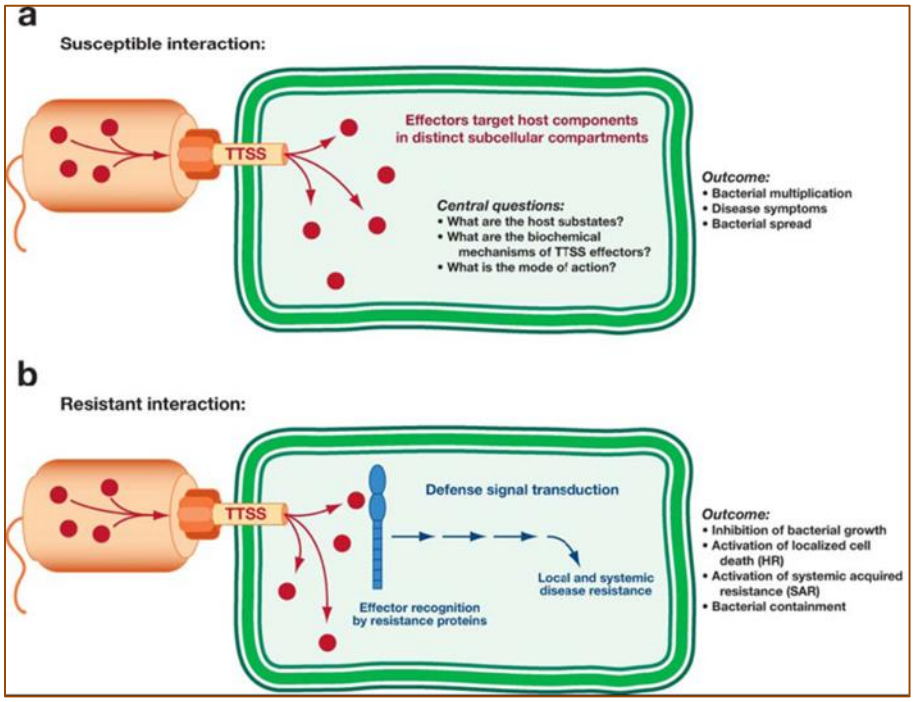

Figure 2.3: Schematic representation of susceptible (a) and resistant (b) reaction during infection by TTSS bacteria such as Xanthomonas (Cornelis & Gijsegem, 2000)

The type III pilus functions as a conduit to guide the translocation of the effectors to the interior of the host cells (Tang et al., 2006). Some hrp genes are highly conserved between diverse plant and animal pathogens and thus are renamed as hrc (hypersensitive response and conserved) genes. Most of hrp/hrc genes are clustered on a 20 to 30 kb chromosomal region consisting of more than 20 genes that are organized into multiple transcriptional units (Alfano & Collmer, 2004).

2.4 Translocation of TTSS effectors into host

Delivery of effector genes from the cytoplasm of gram-negative bacteria to the plant cell interior requires the TTSS to transport genes across multiple physical barriers. The TTSS assemble filamentous supermolecular structures that can direct the translocation of proteins across the bacterial membranes to the extracelluar space or into host cells (Büttner & He, 2009). Proteins that translocate through the TTSS are called Hop (Hrp-dependent outer proteins) proteins. Another subset of Hops is secreted into and function in the host cells, and these Hops are designated as effectors. Hrp pili have been found and characterized in most major plant pathogens (Cornelis & Gijsegem, 2000; Weber et al., 2005). The essential contribution of Hrp pili to TTSS suggest that Hrp pili like the TTSS needles provides a protein transport channel for effector proteins to the host-pathogen interface. The TTSS from plant pathogenic bacteria is connected to an extracellular pilus that spans the plant cell wall (Li et al., 2002; Wei et al., 2000).

The pilus may connect to additional TTSS associated protein complexes in the eukaryotic cell wall to provide a continuous conduit for the delivery of effector proteins into the eukaryotic cell. TTSS secretes several translocator proteins whose function is to facilitate the translocation of effector proteins across the eukaryotic cell. Several putative translocator proteins have been identified HrpF (*X. campestris* pv *vesicatoria*), PopF1 and PopF2 (*R. solanacearum*) and HrpK (*P. syringae*) (Büttner & He, 2009). TTSS is controlled by other accessory proteins that act in the bacterial cytoplasm or are secreted by the TTSS. HrpJ proteins contribute to the secretion of harpin proteins and thus may indirectly affect effector protein translocation. Translocation of effector proteins is differentially regulated by the global TTSS chaperone HpaB, which specifically promotes the translocation of a certain class of effector proteins (Büttner & Bonas, 2006). The activity of HpaB is controlled by regulator HpaA that binds HpaB in the bacterial cytoplasm and allows secretion of extracellular components of the TTSS. After assembly of the

TTSS, secretion of HpaA liberates HpaB and thus activates secretion and translocation of effector proteins (Lorenz et al., 2008).

2.5 Functions of TTSS effectors

The establishment of bacterial pathogens in their host and their ability to cause disease expand beyond the ability to form the secretion system and transfer effectors through it. The infection process is a highly coordinated activity of a myriad of bacterial genes whose identity and mode of action are largely unknown. Some effectors appear to be widely distributed among different strains, whereas others are shared by only a few strains (Alfano & Collmer, 2004). A number of TTSS effector genes are designated as avr genes because they were initially detected as proteins that induced resistance or avirulent reactions in the otherwise susceptible host plants.

Functional classification of virulence factors: Most virulence factors fall into one of a few categories when classified according to function. These include the following:

i) Toxins, which are secreted bacterial molecules that poison host cells by a variety of mechanisms following uptake by the host cell. They often have essential roles in altering and killing host cells, usually by an enzymatic process (Schiavo & van der Goot, 2001). Exotoxins are usually proteins that are secreted by bacteria out into the supernatant. Endotoxin is lipolysaccharide (LPS), which is principally carbohydrate, not protein, non-enzymatic, and not normally secreted. Because LPS is a critical component of all gram-negative bacteria, generally it is not considered a bacterial virulence factor, although it stimulates toxic effects in the host by activating immune responses and cytokine production. Exotoxins come in several types and specificities. They usually have two general functions, which can often be uncoupled: the ability to bind to a host cell receptor, and enzymatic activity.

ii) Adhesins, which are bacterial surface structures that allow pathogens to adhere to host cells. The ability to adhere to the host cell is usually central to a pathogen's ability to cause a disease. Like toxins, there are many types of adhesins (Hultgren et al., 1993). Because of the necessity to be on the bacterial surface to access host cells, many adhesins are at the tip of hair like projections called pili or fimbriae. This presumably allows surface exposure, as well as flexibility upon initial contact, much like a harpoon on the end of a line. Like flagella and other

surface organelles, adhesins have a complex base structure embedded in the outer membrane, plus several other chaperones and secretins that assist in pilus assembly (Kline et al., 2009; Soto & Hultgren, 1999).

iii) Invasins, which are bacterial surface molecules that trigger host cells to enable bacterial uptake into them. Most invasins are proteins (enzymes) that act locally to damage host cells and/or have the immediate effect of facilitating the growth and spread of the pathogen. The damage to the host as a result of this invasive activity may become part of the pathology of an infectious disease (Pizarro-Cerd & Cossart, 2006).

iv) Virulence factors, which mediate intracellular survival within host cells. Virulence factors in bacteria may be encoded on chromosomal DNA, bacteriophage DNA, plasmids, or transposons in either plasmids or the bacterial chromosome (Chen et al., 2012). Recent studies have shown that a growing list of pathogen-encoded effectors functions as proteases that are secreted into plant cells to modify host proteins. In addition, several plant proteases have been found to function in activation of the defence mechanism. These findings reveal that post-translational modification of host proteins through proteolytic processing is a widely used mechanism in regulating the plant defence response (Xia, 2004) . Effectors are often pathogen proteins that probably evolved to subvert various host processes for promotion of the pathogen life cycle. Five classes of effector-specific R proteins are known, and their sequences suggest roles in both effector recognition and signal transduction. Although some R proteins may act as primary receptors of pathogen effector proteins, most appear to play indirect roles in this process. The functions of various R-proteins require phosphorylation, protein degradation, or specific localization within the host cell. Some signaling components are shared by many R gene pathways whereas others appear to be pathway specific (Martin et al., 2003).

TTSS effectors presumably promote bacterial growth in the apoplast by defeating host defenses and releasing nutrients from plant cells. Many of the Hrp effector genes are induced in *Xanthomonas* and *Ralstonia* by an AraC-type transcriptional activator that is activated by a two-component regulator in Pseudomonas (Narayanasamy, 2008). Genes controlled by these factors have been globally identified by cDNA-AFLP, reporter transposons, or as part of in vitro expression technology (IVET) studies. Alternatively, candidate members of Hrp regulons have

been identified on the basis of predicted promoter sequences in the complete genomes of *R. solanacearum* GMI1000, *X. campestris* pv. *campestris*, *X. axonopodis* pv. *citri*, and *P. syringae* pv. *tomato* DC3000. These studies have revealed many genes in each Hrp regulon (in addition to the TTSS machinery genes), but not all of these genes encode TTSS effectors, and some TTSS effector genes are not regulated by the Hrp system (Alfano & Collmer, 2004). Currently more candidate TTSS effectors are being discovered through bioinformatic analyses of bacterial genome sequences. These screens make use of conserved sequences, such as conserved regulatory domains or the signal sequence, to target these proteins through the TTSS that reside in the extreme N terminus of the protein and are rich in polar amino acids (Büttner, et al., 2003; Cunnac et al., 2004; Dutilh et al., 2013; Samudrala et al., 2009; Vencato et al., 2006).

According to Ponciano et al., (2003) TTSS effector proteins are predicted to stimulate an increase in the pH and nutrient content of the plant apoplast. This makes the apoplastic fluids more favourable for bacterial multiplication, activate the host defense responses through recognition by a corresponding host R protein (avirulence function). It also inhibits the activation of host defense responses that are signaled by other TTSS avirulence effectors thus inhibiting the basal plant resistance mechanisms. The contributions of effector genes to pathogen avirulence and pathogenicity are determined by using gene specific knockout mutants in the effector genes or by comparing naturally occurring variants of the genes isolated from field populations. The molecular basis for how TTSS effector genes contribute to disease or resistance may vary, but there are commonalities such as described for bean, Arabidopsis, tomato, and tobacco. Effector genes may contribute to different attributes of pathogenicity like control of intercellular multiplication, whereas in other cases, the effectors control the ability of the bacteria to exit the intercellular spaces to the leaf surface(Bent & Mackey, 2007; Jones & Dangl, 2006)

Not all effectors contribute measurably to pathogenicity, and those that do may have quantifiably different contributions. Redundancy of function may make it difficult to measure a pathogenicity function for a given gene. In some xanthomonads, there are many copies of members of the avrBs3/pthA gene family, and some of these contribute additively and redundantly to pathogenicity. In other cases, although the genes are redundant, the functions of those genes are not redundant, i.e., the pathogenicity function of the effector gene may be gene specific.

Prediction of the importance of effectors in pathogenicity and fitness from sequence structures has not been fruitful, mainly because few show similarities to known genes in databases. In a few cases, where effectors are related across pathovars, species, and even genera, and their roles in pathogenicity have been confirmed, their importance to pathogenic fitness has been speculated. As more commonalities in structure and function among effectors are observed and tested, it may be possible to predict the impact of loss of effector function on pathogen fitness. The role of a particular effector gene in pathogenicity may vary in different bacterial genetic backgrounds (Greenberg & Vinatzer, 2003; Merrell & Camilli, 2000; Rediers et al., 2005).

2.6 Evolutionary change and comparative genomics

Determining the relative part played by geography and ecological specialization in the reproductive isolation between populations of pathogens is still a great challenge. The extent of isolation between pathogen populations is conditioned by the divergence accumulated in allopatry or by the degree of adaptation to hosts (Mhedbi-Hajri et al., 2011). Kohn, (2004) observed that phylogenetic or genealogical interpretation of DNA sequence data from multiple genomic regions was the gold standard for species delimitation and population genetics. Precise species concepts could inform quarantine decisions but were likely to reflect evolutionary events too far in the past to impact disease management. The field of phylogeography has long since realized the need and utility of incorporating nuclear DNA sequences into analyses. However, the use of nuclear DNA sequence data, at the population level, has been hindered by technical laboratory difficulty, sequencing costs, and problematic analytical methods dealing with genotypic sequence data, especially in non-model organisms Multilocus approaches at the population level could identify patterns of endemism or migration directly associated with episodes of disease, including host shifts and associated changes in determinants of pathogenicity and avirulence (Puritz et al., 2012). Bacteria must retain their genetic information from one generation to the next. Bacteria need to evolve strategies allowing the generation of new genetic diversity in order to survive and adapt to the continually changing environmental conditions and niches in which they live. Therefore, genetic plasticity mirrors different bacterial lifestyles and physiological versatilities.

Plant pathogenic bacteria have evolved several unique virulence strategies to successfully infect their hosts. One current area of intense research in the field of plant-pathogen interactions is the

identification and characterization of pathogen virulence factors and the elucidation of their mode of action within the host (Pfeilmeier et al., 2016). Three main forces are shaping bacterial genomes: gene gain, gene loss and gene change. The evolution of these complex mechanisms was shaped by the forces of random change and natural selection. Random genomic change can generate new functions or disrupt existing ones, and natural selection favors and keeps the fittest combinations (Van Baarlen et al., 2007). TTSS effectors are evolutionarily diverse and highly variable in their distribution, both within and among species. The genotypic differences accumulated at the DNA level lead to observed phenotypic differences between individuals of a population. Genomic changes can be as subtle as the mutation, insertion or deletion of individual nucleotides, and as drastic as the duplication or loss of chromosomal segments, entire chromosomes, or complete genomes. Changes in a protein-coding gene can lead to multiple co-existing variants, or *alleles*, of that gene within a population, that differ in specific residues and perform the same function with slight differences. As the result of mating, the progeny will inherit a combination of paternal and maternal alleles for different genes. The random mating of individuals within a populations and the random segregation of chromosomal segments in gamete formation creates new allelic combinations at each generation. The frequency of these allelic combinations will vary through evolutionary time, either by selection for their evolutionary fitness or by random genetic drift. As populations segregate and adapt to their environment, different combinations of alleles dominate in each population. The resulting differences in behavior or chromosomal organization can lead to loss of reproductive ability across sub-populations and the emergence of new *species*. The emergence of new functions in these changing species allowed adaptation to all niches on land, in the air, underground, or in the deepest oceans, in species as diverse as dinosaurs and amoebae (Wilmes et al., 2009).

The genomes of related species exhibit similarities in functional elements that have undergone little change since the species' common ancestor. Deleterious mutations in these functional regions have certainly occurred, but the individuals carrying them have been at a disadvantage and eventually eliminated by natural selection. Mutations in non-functional regions have no effect to an organism's reproductive fitness, and will accumulate over evolutionary time. Hence, the combined effects of random mutation and natural selection allow comparative approaches to separate conserved functional regions from diverged non-functional regions (Desai and Fisher,

2007). According to Wei et al., (2002) a genome taken in isolation from a single organism does not reveal much by itself. Genomes, and genes, need to be studied in comparison with other species (or subspecies, or strains), in the phylogenetic context of the evolutionary process. Comparative genomics is a large-scale, holistic approach that compares two or more genomes to discover the similarities and differences between the genomes and to study the biology of the individual genomes. Comparative genome analysis allows identification of functional elements without prior knowledge of function because of greater degree of conservation across the genomes of related species (Hardison, 2003). When selecting species for a pairwise comparative analysis, we face a tradeoff between closely related species (with many common functional elements but additional spuriously conserved non-functional regions), and distantly related species (with mostly diverged non-functional regions but fewer common functional elements). The use of multiple closely related species may present an attractive alternative, exhibiting an accumulation of independent mutations in non-functional regions, while having most biological functions in common (Ali et al., 2013). Not only does comparative genomics aim to discriminate conserved from divergent and functional from nonfunctional DNA, this approach is also contributing to identifying the general functional class of certain DNA segments, such as coding exons, noncoding RNAs, and some gene regulatory regions (Wray et al., 2002).

Cesbron et al., (2015) indicate that cross-species conservation has previously been used to identify putative genes or regulatory elements in small genomic regions. Light sampling of whole-genome sequence has been studied as a way to improve genome annotation (Land et al., 2015). Complete bacterial genomes have been compared to identify pathogenic and other genes. Besides the existence or absence of gene(s), discovering the evolutionary pressure of common genes is another contribution of comparative genomics. Over 250 bacterial whole genome sequences are currently available in public databases, representing hundreds of species as well as multiple strains of the same species (Smith, 1987). The study of these genomes by both computational and experimental approaches has significantly advanced the understanding of the physiology and pathogenicity of many microbes and provided insights into the mechanisms and history of genome evolution (Soderlund, 2009). In summary comparative genomic analysis are powerful tools for identifying the orthologous genes in species, presence and absence of specific genes, evolutionary signals, and candidate regions associated with pathogenicity (Sivashankari &

Shanmughavel, 2007; Wassenaar et al., 2009). These approaches and principals will be used in this study to unravel the pathogenicity factors within the different isolates of *Xanthomonas campestris pv. musacearum* and *Xanthomonas campestris pv. vasculorum.*

2.7 Genome sequencing methods

Nucleotide sequencing refers to the identification of the nucleotides in a polymer of nucleic acids, whether DNA or RNA.Adams, (2008) noted that scientists had developed a number of sequencing techniques that emphasized speed without loss of accuracy. DNA sequence represents a single format onto which a broad range of biological phenomena can be projected for highthroughput data collection. Initially, these techniques expanded upon the so-called Sanger process that was first developed in the 1970s, gradually automating this process and increasing the number of samples that could be sequenced at one time. Sequencing thus brought about the field of genomics and increased our understanding of the organization and composition of several genomes. Tremendous improvements in sequencing have led to the generation of large amounts of DNA information in a very short period of time (Shendure & Ji, 2008). Sanger sequencing was based on capillary electrophoresis which improved the ability to mine genetic information from any biological system. However this was hampered by limitations in throughput, scalability, speed and resolution that are required by scientists to acquire essential information from research.

According to Llaca, (2012), most sequencing applications can be divided into 2 categories: 1) de novo sequencing, and 2) resequencing. In the case of de novo sequencing, reads are obtained from an unknown sequence and either assembled to reconstruct this sequence or compared directly to reads from other unknown sequences. In the case of resequencing, reads are mapped or aligned to a known reference sequence. De novo applications are usually slower and more computer intensive than resequencing, but are needed to reconstruct genomes and transcriptomes in species with unknown genomes. Major resequencing applications include polymorphism discovery and transcription profiling.

Egan et al., (2012) describes Next Generation Sequencing (NGS) as a fundamentally new approach that has triggered numerous grounds breaking discoveries and revolutionalized genomic science. The concept is similar to capillary electrophoresis (CE) of Sanger with the bases of a

small fragment of DNA sequentially identified from signals emitted as each fragment is re-synthesized from a DNA template strand, extended across millions of reactions. This has enabled rapid sequencing of large stretches of DNA base pairs spanning entire genomes. NGS is highly scalable with use in lower output volumes for targeted studies/smaller genomes or large number of samples by multiplexing on a high throughput instrument. With largely automated easy to use protocols, researchers can go from idea to experiment to data to publication faster and easier than with Sanger sequencing. There several NGS methods but the key ones are three: sequencing by synthesis, sequencing by ligation, and single-molecule sequencing (Abbasi, 2010; Ansorge, 2009; Illumina, 2013; Kumar, Banks, & Cloutier, 2012; Metzker, 2008; Morozova, Hirst, & Marra, 2009; Pareek & Smoczynski, 2011; Quail et al., 2012; Srinivasan & Batra, 2014; Varshney, Nayak, May, & Jackson, 2009; H. Yang et al., 2012) .

2.7.1 Sequencing by synthesis

In sequencing by synthesis, DNA is fragmented to the appropriate size, ligated to adaptor sequences, and then clonally amplified to enhance the fluorescent or chemical signal. Templates are then separated and immobilized in preparation for flow cell cycles. Although there are different methods for separating and immobilizing templates, Illumina is the most widely used method of the sequencing by synthesis technique. Illumina initially developed by Solexa, the Illumina Genome Analyzer uses solid-phase bridge amplification in which 5 ′ and 3 ′ adapters are ligated to each end of a DNA template (Illumina, 2012). One end of the fragment is then attached to the substrate. The adapters hybridize to immobilize forward or reverse primers, creating a bridge that facilitates amplification, generating amplicons that remain attached to the substrate, thus forming clusters of identical templates, which enhances chemiluminescent detection. Millions of such clusters are formed within each channel of the flow cell. Following amplification, the DNA amplicons are denatured and primed. Elongation is conducted through a series of cyclical washes, the first being the addition of a mixture of all four nucleotides, each labeled with a different fluorophore and modified as 3 ′ - *O* - azidomethyl reversible terminators. Following image capture, elongation continues after the fluorescent dye moiety is cleaved and the 3 ′ -OH group is restored through reaction with tris (2-carboxyethyl) phosphine. This cycle is repeated

until the DNA fragment has been synthesized to its target length. This method is currently the most widely used NGS platform and will be used in this study.

2.7.2 Next generation sequencing and its applications

Applications of Next Generation Sequencing (NGS) specifically using Illumina sequencing include DNA sequencing, gene regulation analysis, qualitative and quantitative sequencing based transcriptome analysis, cytogenetic analysis, DNA-protein interaction analysis (ChIP-seq), sequencing -based methylation analysis, small RNA discovery and analysis, de novo metagenomics and metatranscriptomics. “Restriction site associated DNA (RAD)” is based on microarray platform. The RAD has been adapted on the massively-parallel NGS platform to efficiently detect DNA polymorphisms without the requirement of any prior molecular knowledge for the species under investigation. RAD sequencing produces two types of DNA markers: one type of markers is from DNA variations within the restriction sites which are dominant markers; the other is from sequence variation adjacent to the restriction sites which are codominant markers. RAD markers have been employed in genetic mapping on bacteria, fungi, fish, insects, and more recently on plants. Fingerprints in RAD sequencing are presented as DNA sequence reads. SNP markers developed from RAD sequencing are suitable for high throughput multiplex implementation in molecular plant breeding on modern SNP genotyping platforms. Yang *et al.*, 2012 demonstrated that the NGS-based RAD sequencing technology can be cohesively integrated into the marker development protocol for molecular plant breeding. The sequencing reads generated from the RAD sequencing have the same function and effects as the DNA fingerprints produced by traditional DNA fingerprinting methods for marker development in molecular plant breeding. Abbasi, (2010) summed up that the key areas of research and development that will benefit from NGS include food security (e.g. plant and animal breeding, disease diagnostics and management), industrial biotechnology (e.g. drug discovery, biopharmaceutical developments) and medicine (personalised medicine, disease diagnostics and management). The technology has both economic and societal impact and the use of data arising from some potential applications (e.g. personalised medicine) raises significant ethical and societal issues. The molecular composition of Xcm and Xcv is not well documented thus use of Illumina sequencing was ideal for completing a large part of this study.

The application of next-generation sequencing (NGS) approaches has reformed the plant breeding technology especially in the area of molecular marker assisted breeding. The availability of genomic information aids in the deeper understanding of severalmolecularmechanisms behind the vital physiological processes. In addition, the NGS methods facilitate the genome-wide discovery of DNA based markers linked to key genes involved in important biological phenomenon. Among the molecular markers, single nucleotide polymorphism (SNP) indulges various benefits in comparison with other existing DNA based markers (Manivannan et al., 2018). The use of genome survey sequence (GSS) and partial targeted de novo assembly strategies can be useful in gene discovery research projects involving non-model plants species, or species with significant sequence diversity and high structural polymorphism (Lu et al., 2016). Partial de novo assembly is also useful for gene discovery in large genomes or any non-model species where a region of interest can be genetically mapped and a partial physical map can be derived from the genetic map (Vandroemme et al., 2013). Aritua et al., (2015) sequenced and analyzed the genomes of 26 strains of *Xanthomonas axonopodis* pv. *phaseoli* and *X. fuscans* subsp. *fuscans*, the causative agents of bean common bacterial blight disease, collected over four decades and six continents. This revealed considerable genetic variation within both taxa, encompassing both single-nucleotide variants and differences in gene content that could be exploited for tracking pathogen spread.

Chapter Three

Identification of predicted Virulence factors in isolates of *Xanthomonas campestris* pv. *musacearum* and *Xanthomonas vasicola* pv. *vasculourum*

Abstract:

The identification and understanding of the genetic determinants of bacterial virulence are essential to be able to design efficient protection strategies for infected plants. In this study, we exploited the genetic diversity of *Xanthomonas campestris pv. musacearum* and *vasculorum* to mine for predicted pathogenicity determinants. This work is a foundation to understanding the contribution of known and unknown loci to pathogenicity relevant at the pathovar level. The identified virulence determinants can be prime targets for breeding resistance to *X. campestris* pv. *musacearum* in banana. Genotypic characterization using polymerase chain reaction of prediceted effectors was performed on thirteen (13) isolates of the *Xanthomonas vasicola pv. musacearum* (Xvm) and *Xanthomonas vasicola* pv. *vasculorum* (Xvv) from Ethiopia, Kenya, Burundi, Rwanda, Uganda and DRC. Bacterial isolates of similar species can have unique Type four pili (Tfp) and Tfp pilus assembly protein PilF a fimbrial biogenesis protein was hybridized in all Xanthomonas isolates except Xsp1131 only. Type III effector protein RipT was confirmed to be present in all isolates of Xcm and Xvv but not Xsp1131 and Xsp1132. All the Xvm and Xvv isolates under test showed presence of type III effector HopAF1 except Xvv206, Xsp1131 and Xsp1132. YopJ type III secretion system effector protein hybridizes in DNA of all Xcm isolates tested but not in Xsp1131 nor Xsp1132. This association study performed using the presence/absence of predicted virulence genes gives a basis for further comparison of the specific isolates and tracking of the inherent diversity.

Keywords: Banana, Effectors, Pathogen-host, Xanthomonas wilt

3.1 Introduction

Pathogenic species and pathovar within species show a high degree of host plant specificity and exhibit tissue specificity, invading either the vascular system or the mesophyll tissue of the host (Ryan et al., 2011). Plant pathogenic Xanthomonas pathovar require a Type Three Secretion

System (TTSS) to secrete and translocate effector proteins in order to cause disease. The repertoire of effectors can vary between species and isolates within species and is believed to be a key determinant in the host range of a given pathogen (Lemaire et al., 2009). Genome-wide sequencing of draft genomes of Xcm4381 and Xvv702 by Studholme et al., (2010) shows that both draft genomes encode homologues of the candidate TTSS effectors including AvrBs2, AvrGf1, HopW1; HopAF1; PilvD; RipT; YopT-like cysteine protease and a variety of homologues of XopF, XopK, XopL, XopN, XopP, XopQ, XopR, XopX, XopZ, XopA, XopB, XopG, XopH, XopI, XopY, XopAA, XopAD, XopAE and XopAK, which are conserved in a subset of Xanthomonas genomes. They established that Xcm4381 encodes two predicted YopJ-like C55 cysteine proteases. These are absent from Xvv702, whereas Xvv702 encodes a protein XopAF (also known as AvrXv3) which is absent from Xcm4381. It also shares 35% identity with the HopAF1-like genes. Such differences in effector repertoires are said to be significant for host adaptation.

The lipopolysaccharides are important virulence factors in plant pathogens. Lipopolysaccharide (LPS), which is produced by gram-negative bacteria, is a powerful activator of inherent immune responses (Erbs & Newman, 2003). The lipopolysaccharide locus in Xvv702 showed no significant sequence similarity to that of Xcm4381.

Type IV pili are an efficient and versatile device for bacterial surface motility. They are widespread among the ß-, α-, and ∞-proteo-bacteria and the cyano-bacteria. Within that diversity, there is a core of conserved proteins that includes the pilin (PilA), the motors PilB and PilT, and various components of pilus biogenesis and assembly, PilC, PilD, PilM, PilN, PilO, PilP, and PilQ (Kager and Basel, 2004). The respective Type IV pilus (Tfp) cluster showed little sequence similarity between proteins, respectively encoded on the Xvv702 and the Xcm4381 Tfp clusters. An 8-kb gene cluster in Xcm4381 is reported to encode Tfp components FimT, PilV, PilW, PilX, PilY1 and PilE. A different gene cluster in Xvv 702 encodes homologues of Tfp components FimT, PilE, PilY1, PilW and PilV. These differences might be adaptive for attachment to and motility on different plant surfaces. Type IV pili are multifunctional complexes that can act as bacterial virulence factors because pilus-based motility is used to spread pathogens over the surface of a tissue, or to build multicellular structures such as biofilms and fruiting bodies

(Burdman et al., 2011).

The draft genome analysis of the two bacteria species revealed several genetic differences between the two isolates that might be important for host specificity, virulence and epiphytic fitness. They included differences in the repertoires of secreted and translocated effector proteins, Tfp and enzymes for lipopolysaccharide biosynthesis. (Hajri et al., 2012; Ho Sui et al., 2009) proposed two hypotheses can be formulated to explain host specificity: either it can be explained by the phylogenetic position of the strains, or by the association of virulence genes enabling a pathological convergence of phylogenically distant strains. In this latter hypothesis, host specificity would result from the interaction between repertoires of bacterial virulence genes and repertoires of genes involved in host defenses (Lemaire et al., 2009). There is urgent need to clearly identify and understand the infection system applied by Xvm and also propose pathogen protein targets for suppressing the motility, mode of host plant surface identification, vascular rapid multiplication and evasion of host plant resistance mechanism. This study therefore aimed at identifying the predicted virulence factors in different isolates of Xvm, Xsp. and Xvv.

3.2 Materials and Methods

3.2.1 Bacterial culture conditions

Bacterial Cultures of isolates in Table 3.1 were obtained from the National Collection of Plant Pathogenic Bacteria (NCPPB) at the FERA, York, UK. The samples on filter paper in glass vials were recovered through streaking on solid Kings Broth (KB) for 12hours. The KB was prepared using 10g Peptone meat, 10g N-Z casein; 1.5g $MgSO_4$-$7H_2O$; 1.5g K_2HPO4; 12.6g glycerol to a litre of dH_2O and stabilized at pH 7.2. Cultures were further purified by re-streaking onto new KB plates overnight. A loop of single colony from each isolate was inoculated onto 10mls liquid KB and grown at 28ºC in a shaking incubator at 150 rpm overnight. All of the isolates used in this study were then put in glycerol stock and kept at -20 ºC for further use during the experiment and -80ºC for long term storage.

Table 3.1: Bacterial isolates provided by NCPPB and used for the study

Code	Species	Host	Origin Country	Sampling date	Donor Reference
Xcm4383	*Xanthomonas vasicola pv. musacearum*	*Musa*-Banana	Uganda	2007	071/wsk/05
Xcm4387	*Xanthomonas vasicola pv. musacearum*	*Musa*-Banana	Democratic Republic of Congo	2007	IMI392223
Xcm4389	*Xanthomonas vasicola pv. musacearum*	*Musa*-Banana	Rwanda	2007	IMI393640
Xcm4433	*Xanthomonas vasicola pv. Musacearum*	*Musa*-Banana	Burundi	2008	20709953
Xcm4434	*Xanthomonas vasicola pv. Musacearum*	*Musa*-Banana	Kenya	2008	20702295
Xcm2251	*Xanthomonas vasicola pv. musacearum*	*Musa sp.*	Ethiopia	1969	B3645
Xcm2005	*Xanthomonas vasicola pv. musacearum*	*Ensette ventricosum*	Ethiopia	1967	B3100

Xcm4392	*Xanthomonas vasicola pv. musacearum*	*Musa-Banana*	Tanzania	2007	394051
Xvv206	*X. vasicola pv. vasculorum*	*Zea mays*	South Africa	1948	No Donor reference
Xvv1326	*X. vasicola pv. vasculorum*	*Saccharum offinarum*	Zimbabwe	1962	No Donor reference

3.2.2 Total genomic DNA Extraction from cultures

Genomic DNA from each of the isolates was extracted using modified protocol from Gomes et al., (1988). Cultures were grown at 24ºC for 12hrs and then centrifuged at 4000rpm 4ºC for 10 min to obtain a cell pellet, and supernatant was poured off and the pellet dried. Cells were suspended and dissolved in 2ml of TE buffer (25mM Tris-Hcl pH 8.0, 10mM EDTA). lysis of 300µl cells done by adding 12 µl Lysozyme 20mg/ µl. 1.5 µl RNase 10mg/ µl were then added and mixed by inverting 10 times and incubated at 25^0C for 10min. SDS (17 µl 10%) were then added and mixed by inverting 10 times then incubated on ice for 5 min. Proteins were pelleted by adding 170 µl 8M ammonium acetate, mixed by inverting 10 times, vortexed 20 secs and spinned at 4ºC, maximum speed for 15min. The supernatant was carefully pippeted off into clean tubes avoiding the pellet and membrane formed on the surface. DNA was pelleted by adding 0.75% Vol Isopropanol, mixed by inverting until a ball of DNA was visible; then centrifuged at 4ºC, at 13500rpm for 10 mins. The supernatant was pippeted off and air dried for 5 mins. The pellet was washed with 100µl 70% ethanol; centrifuged for 1 min maximum speed, the ethanol pipetted off, spinned short burst and pipette off excess ethanol; air dry pellet 5 mins. The pellet was then dissolved in 200 µl TE, incubated in 65ºC water bath for 30min to 1 hr. The quality and quantity of DNA was then quantified using a NanoDrop® ND-1000 (NanoDrop Technologies USA,

www.nanodrop.com). Agarose gel electrophoresis of each sample DNA was run on 0.8% agarose gel for 45min at 100V.

3.2.3 Primer design and validation of putative virulence factors

The primers for 26 virulence factors were designed using program Primer3 from the EMBOSS package (Table 3.2). The Primer-Search program was used to confirm the specificity of the primers (cut-off of 20% mismatch) against the sequenced Xcm4381 and Xvv702. Each 25µl of PCR mixture contained 1.5 mM MgCl2, 0.2µM forward and reverse primers, PCR buffer (Invitrogen, Carlsbad, CA), 0.2 mM each deoxy nucleoside triphosphate (dNTP), 0.5 units Taq polymerase (Invitrogen, Carlsbad, CA), and 10 ng DNA. After an initial denaturing step, PCR was conducted for 28 cycles of 30 s at 94°C, 30 s at 55°C, and 70 s at 72°C. The primers were tested on DNA of Xcm4381 and Xvv702 and those validated used to amplify bacterial isolates indicated in Table 3.1.

3.3 Results

Virulence factors were grouped into predicted lipopolysaccharides (LPS), type iv pilus proteins (Tfp) and Type Three Secretion System (TTSS) proteins for analysis across 8 isolates of Xcm, 2 isolates of Xsp and 3 Xvv isolates. The amplification of DNA was scored as present or absent for the different isolates in the study and were subjected to cluster analysis using PAST (version 2.17c) with dissimilarity index using Ward method. This study has clustered the genes in 3 groups and the Xcm isolates into 3 groups.

Table 3.2: Sequences of primers for 25 proteins used in the study and the expected fragment band sizes

Code	Genes	Expected product size (Kb)		Forward primer	Reverse primer
		Xcm	Xvv		
LPS-A	ABC transporter permease	783	NP	ATGCAGAGCCATGCAAATCTG	TCATAGGACGTCCGCAAAACC
LPS-B	SAM-dependent methyltransferases-WsaE	1656	NP	ATGGGCGTTGATCTCCAAGC	TCAATTGGCCAATAAGCGGC
LPS-C	Truncated O-antigen biosynthesis protein	1701	NP	ATGGCGGCATATGCTTCTGG	TCAAGCATCTGCCTGCGC
LPS-D	Hypothetical protein-ZP_06489485	1269	NP	ATGCTTGATCGTTTGCTACGGC	CTACACCCGGTCAACCTGG
LPS-E	Predicted membrane protein	399	NP	TCATCGATGCCCCCGAAAAGTG	GTGATCGATCAGAAGTTCTTC
LPS-F	NAD dependent epimerase	945	NP	TCAGCGCTTGAGCACGATGTGGC	ATGGCACAGAAGAATGACAAG
LPS-G	Indolepyruvate ferredoxin oxidoreductase	1302	1302	CTACTCCGTGACCCGGCG	ATGGCGGCGGGGCAGTCC
LPS-H	Indolepyruvate ferredoxin oxidoreductase	1302	1302	ATGGCGGCGGGGCAGTC	CTACTCCGTGACCCGGCG
LPS-K	Lipopolysaccharide biosynthesis protein	2838	2838	TCATCCGGCTTGAATCAATCCG	ATGGCGCACCGCGTGACTG
LPS-M	Short chain dehydrogenase	NP	729	ATGCAACGCGTTCTTATTATCG	TCAGAGCTTGATCCTACGAAAAACG

LPS-N	Putative transmembrane GtrA	NP	405	ATGATTAGTCGACAGTTCATCG	TCATTGTGGTGTGGCGATCTTG
LPS-P	GDP-mannose 4,6-dehydratase	NP	969	TCAGATTGAAAGTCGACGCATATC	ATGAGTAAGAATGCCCTGATC
Tfp-A	Tfp pilus assembly protein PilE	426	NP	TCAAAAACATTGTGACGCGCCATC	ATGTACAGAATTGCTCGCTCG
Tfp-B	Tfp pilus assembly protein PilE	NP	381	TTGATCGAGTTGATGATCGTCG	CTACCAACAGCCCGGAGTG
Tfp-C	Tfp pilus assembly protein FimT	NP	474	ATGATCACGATCGTCGTGC	TCATTGGCAGGAGCTCTTCTG
Tfp-D	Tfp_fimbrial biogenesis protein	534	NP	ATGCGATACGCTCCTCTGG	TCACGAAGGGCAGGGATG
Tfp-F	Tfp pilus assembly protein PilV	411	NP	TCATAGACGGGTCTTGATAG	ATGTCTGGTATCGGTTTGATCG
Tfp-G	Tfp pilus assembly protein PilW	975	NP	TCATGCATTGCGATTCCGAAG	ATGCGCGGCGTGACGCTG
Tfp-H	Tfp pilus assembly protein PilX	480	NP	TTACTGCTTGATGTAGTTAGTCTG	GTCTCGCTCATTGTTGTGCTG
Tfp-K	Tfp pilus assembly protein PilY1	2151	NP	TTAGTTCCTGATGATCTCGCG	TTGGCCGACGTGGCCATG
RipT	Type III effector protein RipT	720	NP	GTGTTCAATGGTGATGAGATAG	CTAAGGCACGAGACGAGC
HopA	Type III effector HopAF1	861	860	ATGCTTGTCATGATGCCTGTTACC	CTATTCCGCAGCGACGAG
HopW	Type III effector HopW1	1503	1503	TCAGGAGGTAATCCCCTTTTTGG	TTGCAAAAACAAGCGGGCCTC
	Type III secretion system effector protein XopAF	NP	657	CTATTTAACAAGATCTGTTACAAATC	ATGACTGATGGTTTAGATCTTTGCG

Yop1	Virulence factor yopJ-like 1	1077	NP	TTGCAGGACTTCATTGACGC	TTAGCTCCCATACCCGGAGTCG
Yop2	Virulence factor yopJ-like 2	1068	NP	ATGGACATTGAAAATCTCCCC	TCAGGATTCTAAGGCGTGACG

NP: No amplification product/fragment expected

Table 3.3: Summary observations of amplified virulence factors in genomic DNA of 13 selected Xanthomonas isolates

Code	Bacterial isolates	Xcm (kb)	Xvv (kb)	Xcm 4383	Xcm 4387	Xcm 4389	Xcm 4433	Xcm 4434	Xcm 2251	Xcm 2005	Xcm 6267	Xsp 1131	Xsp 1132	Xvv 206	Xvv 1326	Xvv 1381
	Genes		Host	Musa	Musa	Musa	Musa	Musa	Musa	Enset	Musa	Musa	Musa	Maize	S/cane	S/cane
LPS-M	Short chain dehydrogenase	NP	729	-	+	-	-	-	-	+	+	-	-	+	+	+
LPS-N	Putative transmembrane GtrA	NP	405	-	-	-	-	-	-	+	-	-	-	+	+	+
LPS-P	GDP-mannose 4,6-dehydratase	NP	969	-	-	-	-	-	-	+	-	-	-	+	+	+
LPS-A	ABC transporter permease	783	NP	+	+	+	+	+	+	+	+	-	-	-	-	-

LPS-B	SAM-dependent methyltransferases-WsaE	1656	NP	+	+	+	+	+	+	+	+	-	-	-	-	-
LPS-C	Truncated O-antigen biosynthesis protein	1701	NP	+	+	+	+	+	+	+	+	-	-	-	-	-
LPS-D	Hypothetical protein-ZP_06489485	1269	NP	+	+	+	+	+	+	+	+	-	-	-	-	-
LPS-E	Predicted membrane protein	399	NP	+	+	+	+	+	+	+	+	-	+	-	+	+
LPS-F	NAD dependent epimerase	945	NP	-	+	+	+	+	+	+	+	+	+	*	*	*
LPS-K	Lipopolysaccharide biosynthesis protein	2838	2838	-	+	+	+	+	+	-	-	-	-	-	-	-
LPS-G	Indolepyruvate ferredoxin oxido	1302	1302	+	+	+	+	+	+	+	-	-	-	-	-	-

	reductase																
LPS-H	Indolepyruvate ferredoxin oxido reductase	1302	1302	+	+	+	+	+	+	+	-	-	-	+	+	-
Tfp-A	Tfp pilus assembly protein PilE	426	NP	+	+	+	+	+	+	+	+	-	-	+	+	-
Tfp-B	Tfp pilus assembly protein PilE	NP	381	-	-	-	-	-	-	+	-	-	-	-	+	+
Tfp-C	Tfp pilus assembly protein FimT	NP	474	-	-	-	-	-	-	-	-	-	-	-	+	+
Tfp-D	Tfp_fimbrial biogenesis protein	534	NP	+	+	+	+	+	+	+	+	-	+	+	+	+
Tfp-F	Tfp pilus assembly protein PilV	411	NP	+	+	+	+	+	+	+	+	-	-	+	-	-

Tfp-G	Tfp pilus assembly protein PilW	975	NP	+	+	+	+	+	+	+	+	-	-	+	-	-
Tfp-H	Tfp pilus assembly protein PilX	480	NP	-	+	+	+	+	+	+	+	-	-	+	+	+
Tfp-K	Tfp pilus assembly protein PilY1	2151	NP	-	+	-	+	-	+	+	+	-	-	+	-	-
HopA	Type III effector HopAF1	861	860	+	+	+	+	+	+	+	+	-	-	-	+	+
HopW	Type III effector HopW1	1503	1503	-	+	+	+	+	+	+	+	-	-	+	+	+
XopAF	Type III secretion system effector protein	NP	657	-	-	-	-	-	-	-	-	-	-	-	+	+
Ript	Type III effector protein Ript	720	NP	+	+	+	+	+	+	+	+	-	-	+	+	+
Yop1	Virulence	1077	NP	+	+	+	+	+	+	+	+	-	-	-	-	-

	factor yopJ-like 1																	
Yop2	Virulence factor yopJ-like 2	1068	NP	+	+	+	+	+	+	+	+	-	-	-	-	-

NP: No amplification product expected; +: amplification product; -: No amplification product

3.3.1 LPS variation in selected Xcm and Xvv isolates

All Xcm isolates had lipopolysaccharides: ABC transporter permease/ATP-binding protein (permease); S-adenosyl-L-methionine (SAM) dependent methyltransferase, truncated O-antigen biosynthesis protein and hypothetical protein ZP_06489485 were detected in other isolates but not in Xsp1131, Xsp1132, Xvv206, Xvv1326 and Xvv1381 isolates (fig. 3.1). Lipopolysaccharides (LPSs) are essential and distinctive structures of Gram negative bacteria being a major component of the bacterial cell surface. In general, LPS molecules consist of a hydrophilic hetero-polysaccharide formed by three major substructures, the O-specific polysaccharide (O-antigen), composed of a repetitive sugar subunit; the core oligosaccharide region that is covalently linked to the glycolipid moiety lipid A; and the lipid A anchored to the outer side of the plasmatic external membrane. A broad range of pathogens are recognized by plants through molecular-associated molecular patterns (MAMPs), which are highly conserved fragments of pathogenic molecules. In plant pathogenic bacteria, lipopolysaccharide (LPS) is considered a virulence factor and it is being recognized as a MAMP (Mackey & McFall, 2006). The bacterial LPS molecule confers protection against different environmental stresses, including the hostile medium found inside plant tissues. One of the most widely studied effects of LPSs on plant cells is their ability to prevent the hypersensitive response (HR) induced in plants by avirulent bacteria. *Xanthomonas* sp. isolates mutant in LPS biosynthesis frequently show reduced virulence with a rapid decline in viable bacterial numbers inside plant tissues. Furthermore, since defective LPSs can no longer protect the cell against aggressive environments, such mutants are often more sensitive to Reactive Oxygen Synthesis (ROS), antibiotics, detergents and antimicrobial peptides.

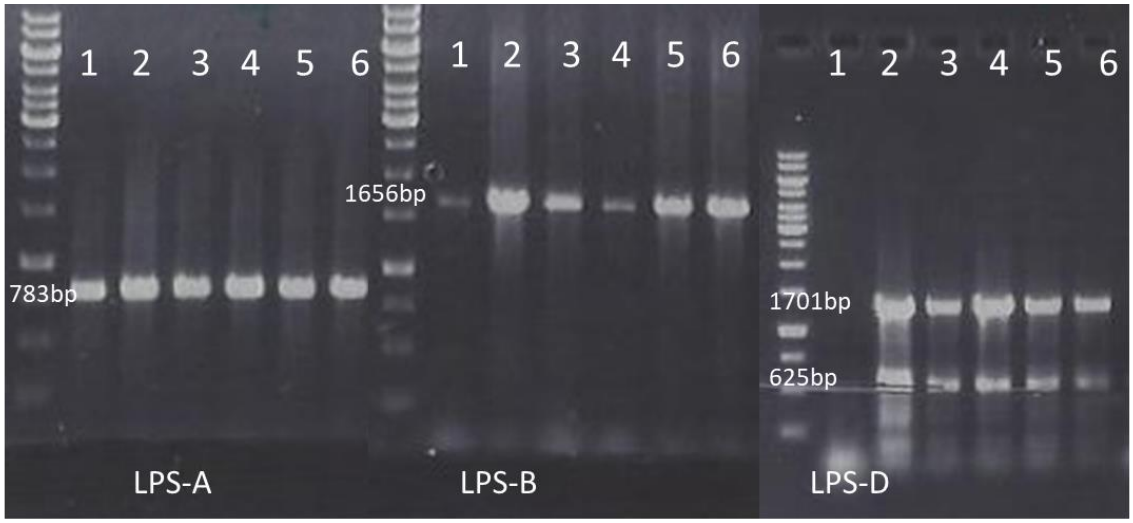

Figure 3.1: Amplification of fragments from ABC transporter permease (LPS-A), SAM-dependent methyltransferases (LPS-B) and hypothetical protein (LPS-D) in Xcm 4383 (lane 1), Xcm 4387 (lane 2), Xcm 4389 (Lane 3), Xcm 4433 (Lane 4), Xcm 4434 (lane 5) and Xcm 2251 (Lane 6)

Primers for predicted membrane protein (YP_007652590) are present in all Xcm isolates (LPS-E) while primers for protein short chain dehydrogenase/reductase (M) was amplified in Xcm4387, Xcm2005 and Xcm6267 just like was expected in Xvv206, Xvv1326 and Xvv1381. Fragments of Nicotinamide adenine dinucleotide (NAD) dependent epimerase were observed in Xcm 4387, Xcm 4389, Xcm 4433, Xcm 4434, Xcm 2251, Xcm 2005, Xcm 6267, Xsp 1131, Xsp 1132 and in Xvv 206, Xvv 1326 and Xvv 1381 only (LPS-F) (Fig 3.2).

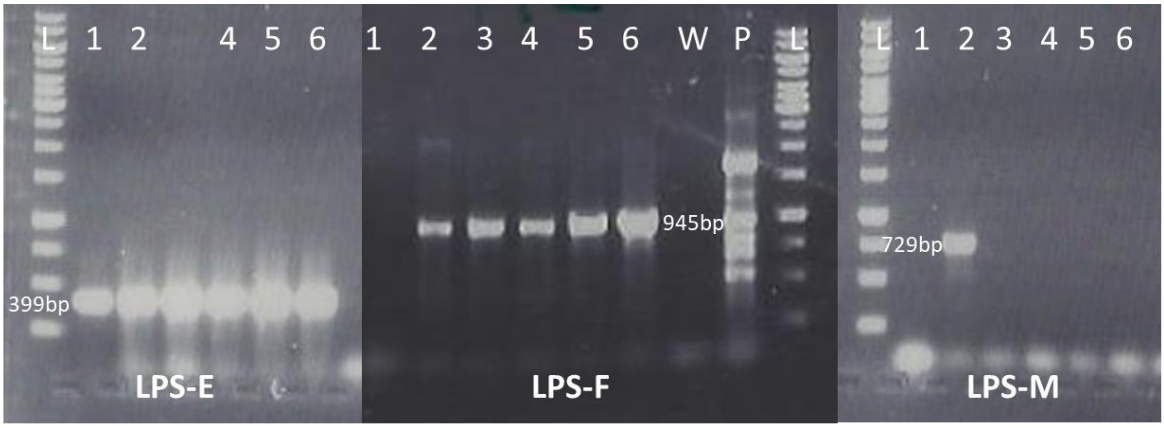

Figure 3.2: Amplified products of isolates for primers predicted membrane protein (LPS-E); NAD dependent epimerase (LPS-F) and short chain dehydrogenase/reductase (LPS-M) in Xcm4383 (1), Xcm4387 (2), Xcm4389 (3), Xcm4433 (4), Xcm4434 (5), Xcm2251 (6) against 1kb ladder (L).

3.3.2 Type IV pili of Xcm, Xvv and Xsp isolates

The multiplex pcr showed that bacterial isolates of the same species can have different Tfp. Xcm4381 based Tfp primers specific to, amplified a 1500kb fragment for Xcm2251. Xcm4383, Xcm4387, Xcm4389, Xcm4433, Xcm4434 isolates had same size fragment and a different fragment in Xsp1131, Xsp1132 (Tfp-A). Tfp pilus assembly protein PilE specific to Xvv702 hybridized a fragment in Xcm2005, Xvv1326 and Xvv1381 isolates. No fragment was observed in the other Xcm isolates (Tfp-B). Tfp pilus assembly protein FimT amplified fragments in isolates of Xvv1326 and Xvv1381 only (Tfp-C) as shown in Table 3.3.

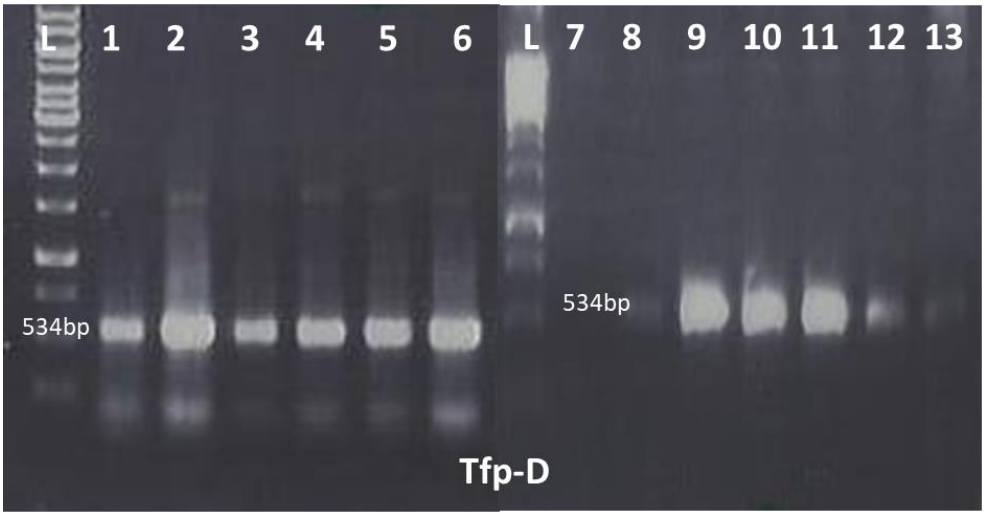

Figure 3.3: Amplification of Xcm-Tfp fimbrial biogenesis protein (D) and Xvv-Tfp fimbrial biogenesis protein (E) in Xcm4383 (1), Xcm4387 (2), Xcm4389 (3), Xcm4433 (4), Xcm4434 (5), Xcm2251 (6), Xsp.1131 (7), Xsp.1132 (8), Xcm2005 (9), Xcm(P)6267 (10), Xvv206 (11), Xvv1326 (12) and Xvv1381 (13)against 1kb ladder (L)

Tfp pilus assembly protein PilF a fimbrial biogenesis protein was hybridized in all isolates but not Xsp 1131anthomonas isolates. Tfp pilus assembly protein, PilV (Tfp-F) and PilW (Tfp-G) were not hybridized in Xsp1131/1132, Xvv1326 and Xvv1381 but hybridized for the rest of the isolates under test (fig. 3.4).

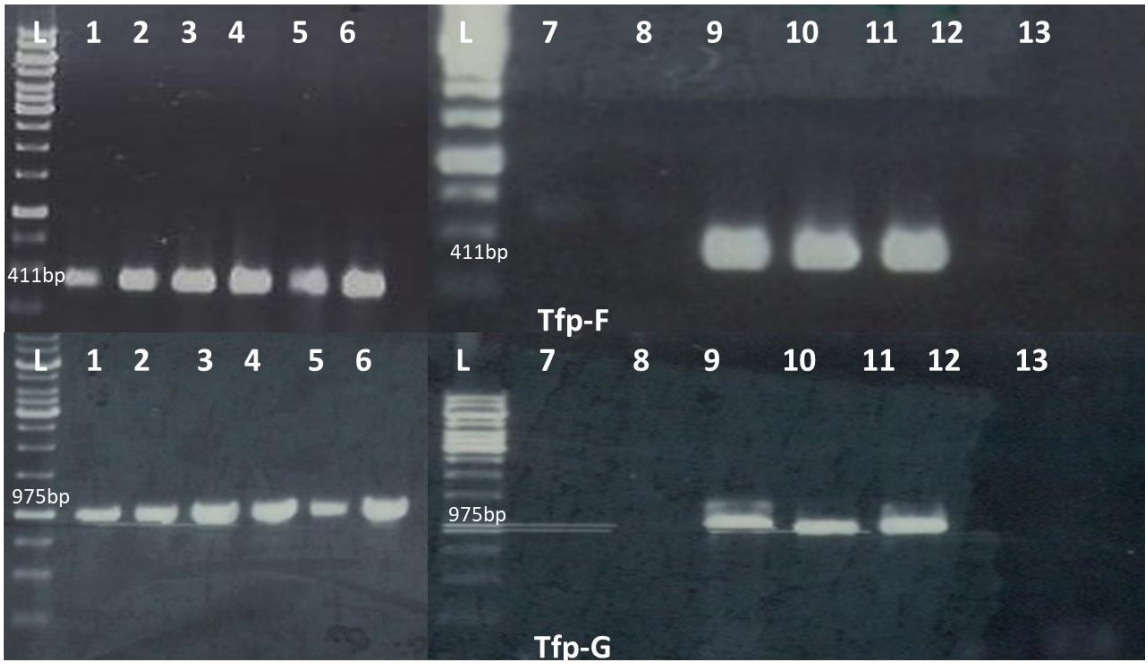

Figure 3.4: Amplification of Tfp pilus assembly protein PilV (F) and Tfp pilus assembly protein PilW (G) in Xcm 4383 (1), Xcm 4387 (2), Xcm 4389 (3), Xcm 4433 (4), Xcm 4434 (5), Xcm 2251 (6), Xsp.1131 (7), Xsp.1132 (8), Xcm2005 (9), Xcm (P)6267 (10), Xvv 206 (11), Xvv 1326 (12) and Xvv 1381 (13) against 1kb ladder (L)

All DNA of isolates hybridized with Tfp pilus assembly protein PilX primers except Xcm 4383, Xsp 1131 and Xsp1132. PilY1 a Tfp pilus assembly protein (tip-associated adhesion) was hybridized in DNA of Xcm 4387, Xcm 4433, Xcm 2251, Xcm 2005, Xcm 6267, and Xvv 206. No fragment was observed for isolates Xcm 4383, Xcm 4389, Xcm 4434, Xsp 1131, Xsp 1132, Xvv 1326, and Xvv 1381 (fig. 3.5).

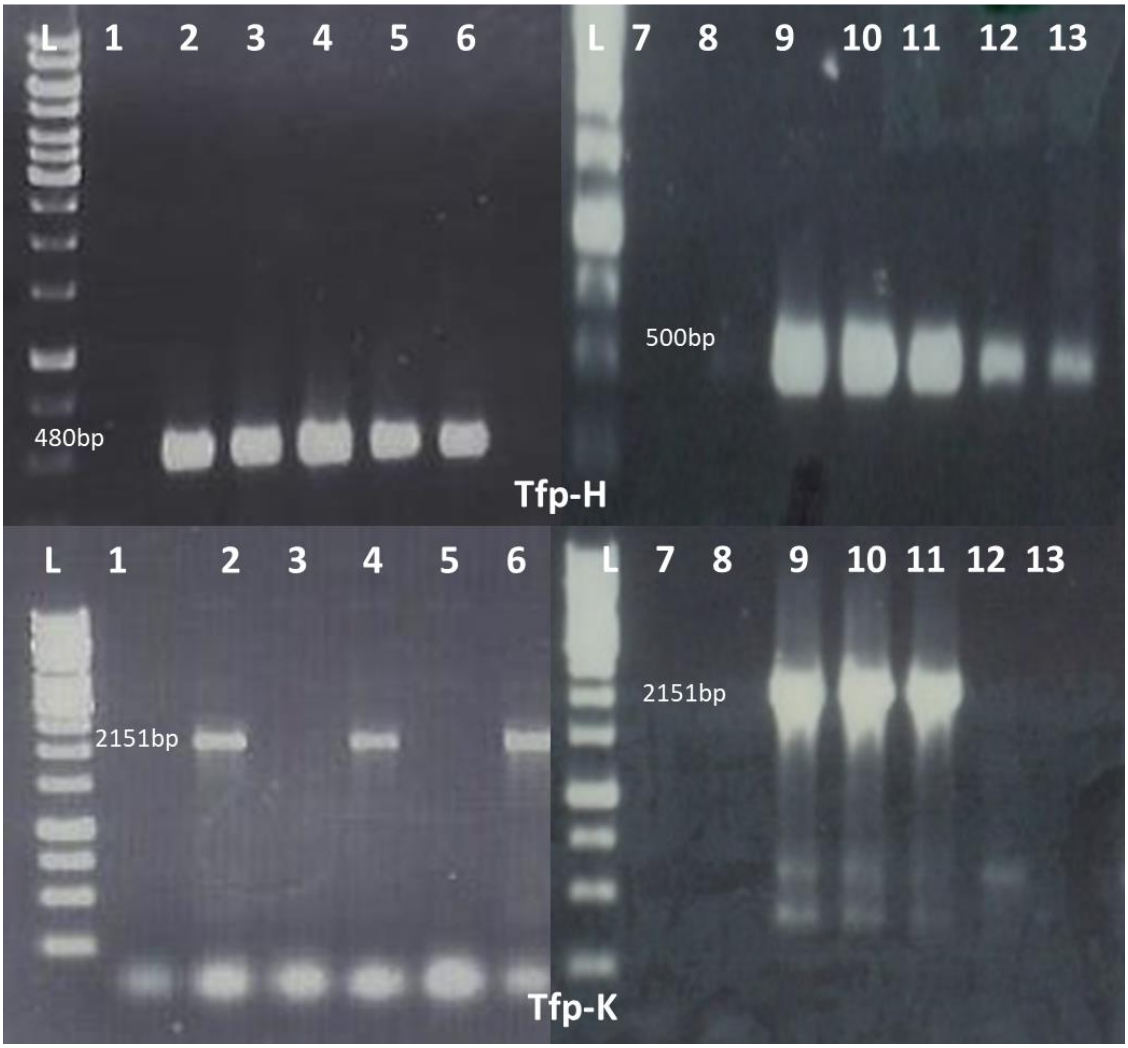

Figure 3.5: Amplification of Tfp assembly protein PilY1 (Tfp-K) in Xcm4383 (1), Xcm 4387 (2), Xcm 4389 (3), Xcm 4433 (4), Xcm 4434 (4), Xcm 2251 (6), Xsp. 1131 (7), Xsp. 1132 (8), Xcm2005 (9), Xcm (P)6267 (10), Xvv206 (11), Xvv1326 (12) and Xvv 1381 (13)) against 1kb ladder (L)

3.3.3 Type Three Secretion System (TTSE) effectors

Type III secretion systems (TTSSs) are central to the pathogenesis and specifically deliver their secreted substrates (type III secreted effectors, TTSEs) into host cells. Since TTSEs play a crucial role in pathogen-host interactions, identifying them is crucial to the understanding of the pathogenic mechanisms of TTSSs. In this study we show that indeed there can be differences in the effectors across bacterial isolates and within isolates of the same species and other clusters

conserved in gram-negative bacteria. Type III effector protein RipT was confirmed to be present in all isolates of Xcm and Xvv but not Xsp1131 and Xsp1132 (RipT). All the Xcm and Xvv isolates under test yielded bands of type III effector HopAF1 except Xvv206, Xsp1131 and Xsp1132 (fig. 3.6).

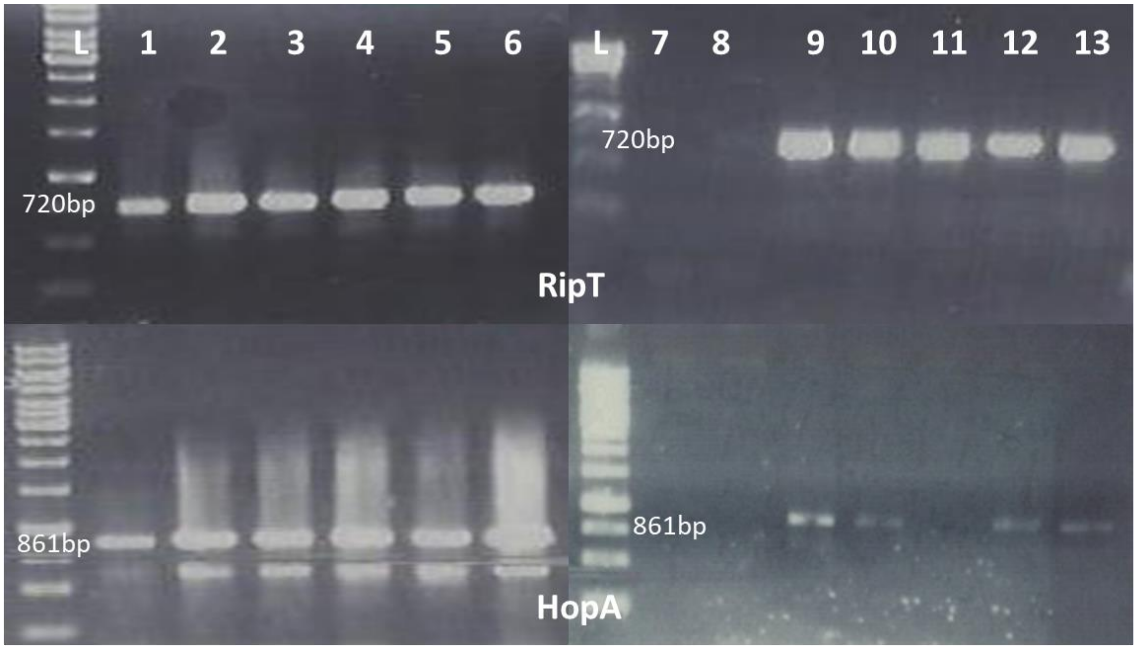

Figure 3.6: Amplification of Type three effector protein RipT and Type three effector protein HopAF from DNA isolates of Xcm 4383 (1), Xcm 4387 (2), Xcm 4389 (3), Xcm 4433 (4), Xcm 4434 (5), Xcm 2251 (6), Xsp.1131 (7), Xsp.1132 (8), Xcm 2005 (9), Xcm (P)6267 (10), Xvv 206 (11), Xvv 1326 (12) and Xvv 1381 (13) against 1kb ladder (L)

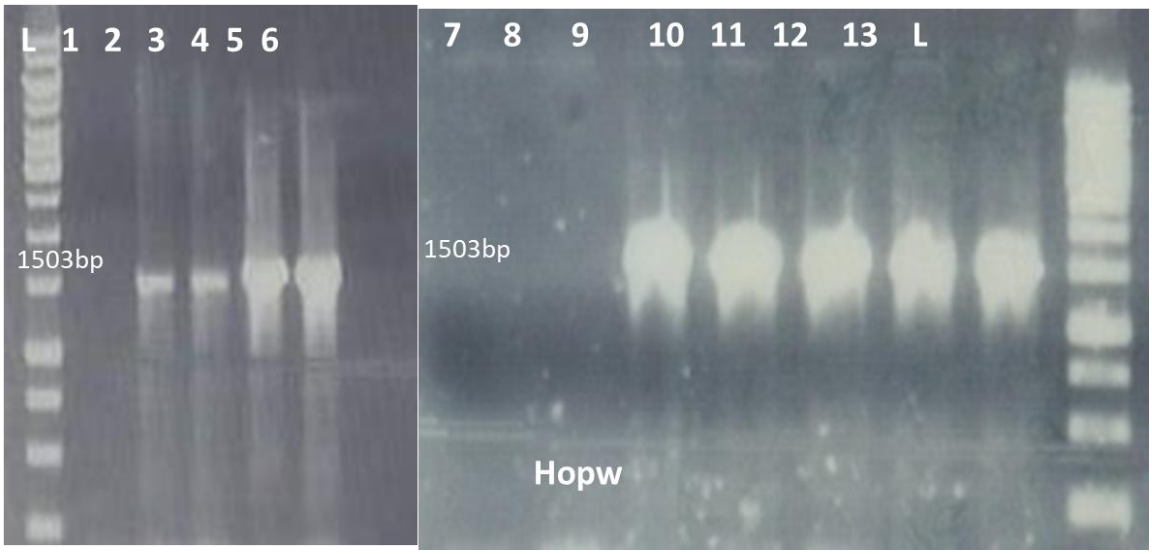

Figure 3.7: Amplification of Type three effector protein HopW1 in Xcm 4383 (1), Xcm 4387 (2), Xcm 4389 (3), Xcm 4433 (4), Xcm 4434 (4), Xcm 2251 (6), Xsp.1131 (7), Xsp.1132 (8), Xcm 2005 (9), Xcm (P)6267 (10), Xvv 206 (11), Xvv 1326 (12) and Xvv 1381 (13)

Primers for type III effector HopW1 hybridized fragments in all Xcm and Xvv isolates tested except Xcm4383, Xsp1131 and Xsp1132 (fig. 3.7). YopJ type III secretion system effector protein hybridizes products in DNA of all Xcm isolates tested but not in Xsp1131, Xsp1132. Xvv702, Xvv1326 and Xvv1381 isolates yielded the only products for type III effector protein XopAF, with no hybridization in any of the other Xanthomonas isolates under study (fig. 3.8).

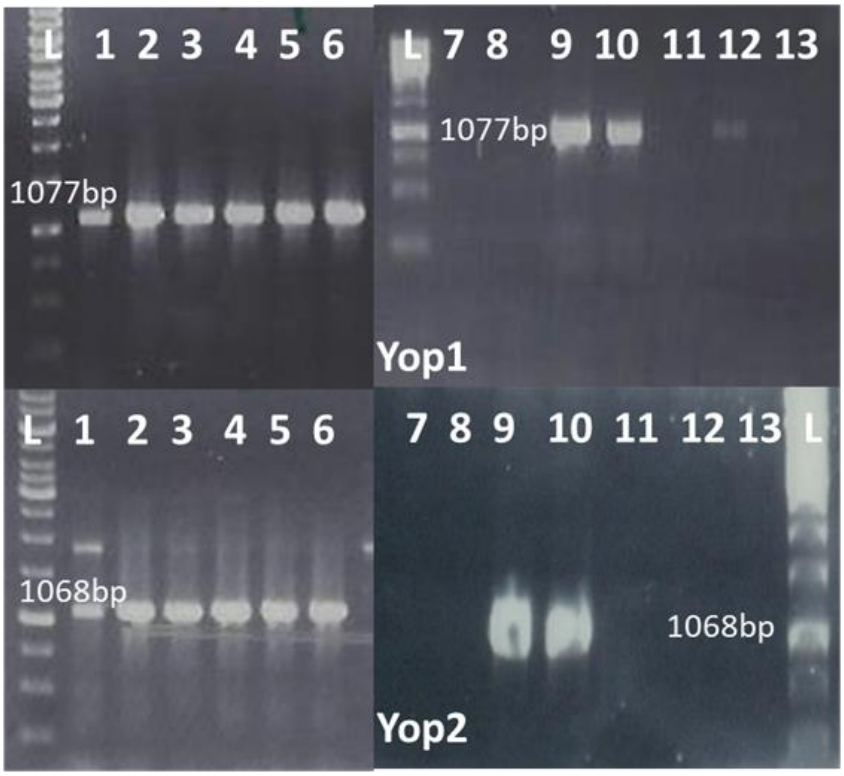

Figure 3.8: Amplification of TTSS effector protein YopJ-like1 and YopJ-like2 in Xcm 4383 (lane1), Xcm 4387 (lane2), Xcm 4389 (Lane3), Xcm 4433 (Lane4), Xcm 4434 (Lane4), Xcm 2251 (Lane6), Xsp. 1131 (lane7), Xsp. 1132(lane8), Xcm 2005 (Lane9), Xcm (P)6267 (Lane10), Xvv206 (Lane11), Xvv1326 (Lane12) and Xvv 1381 (Lane 13) against 1kb ladder (L)

3.3.4 Cluster analysis of virulence factors

Even though genes are continuously acquired from outside or lost over time, orthologous genes can be shared among multiple genomes. If two genomes are closely related or share recent common ancestor, they will share more genes. In contrast, if they are distantly related or share distant common ancestor, the number of shared genes would be less (as they are lost along the way). This is a basis for inferring phylogeny from the gene content information (as presence/absence). Cluster analysis of the presence or absence of virulence factors clusters into 4 separate clades (fig. 3.9). This could indicate groups of virulence factors with significant effect on the pathogenicity of the different isolates on its host. Analysis of similarity (ANOSIM) indicated that the cluster 3 was most significantly different from cluster 1 and 2. NAD-dependent epimerase/dehydratase and short chain dehydrogenase proteins are shown to be outliers and thus a subject of interest for further analysis.

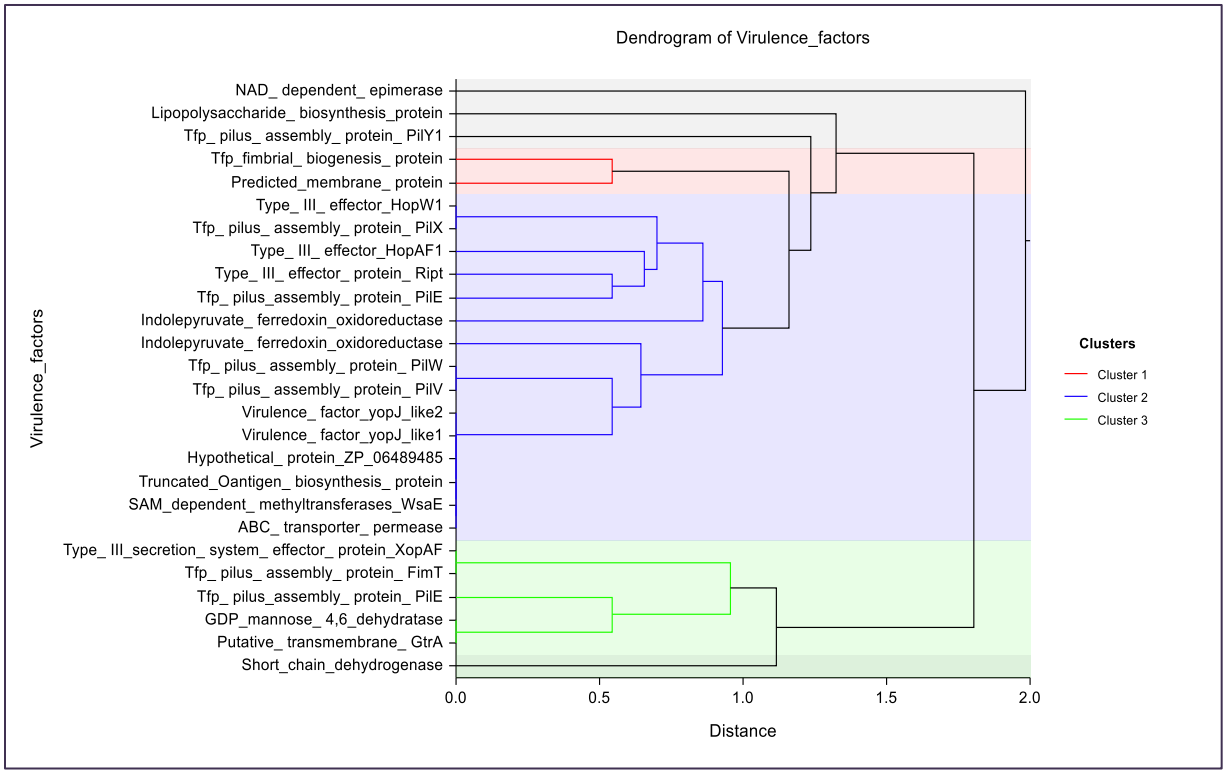

Figure 3.9: Hierarchical clustering dendogram showing clusters of virulence factors from amplification of 25 virulence factors against 13 Xanthomonas isolates. R=0.749

Further cluster analysis focusing on the *Xcm* and *Xvv* isolates including *Xsp* is shown in fig. 3.10. The major clustering shows the strains grouped into *Xanthomonas* sp. (*X*sp. 1131 and *X*sp. 1132) and the *Xcm*/*Xvv* clades. The Xvv 206, Xvv 1326 and Xvv 1381 forms a separate cluster from the Xcm in another major grouping. Within the Xcm cluster, another sub cluster is observed which had Xcm 4387, Xcm 4433, Xcm2251, Xcm 4389 and Xcm 4434 isolated from DRC, Burundi, Ethiopia, Rwanda and Kenya respectively. The other sub cluster included Xcm 6267, Xcm 2005 and Xcm 4383 from Tanzania, Ethiopia and Uganda respectively.

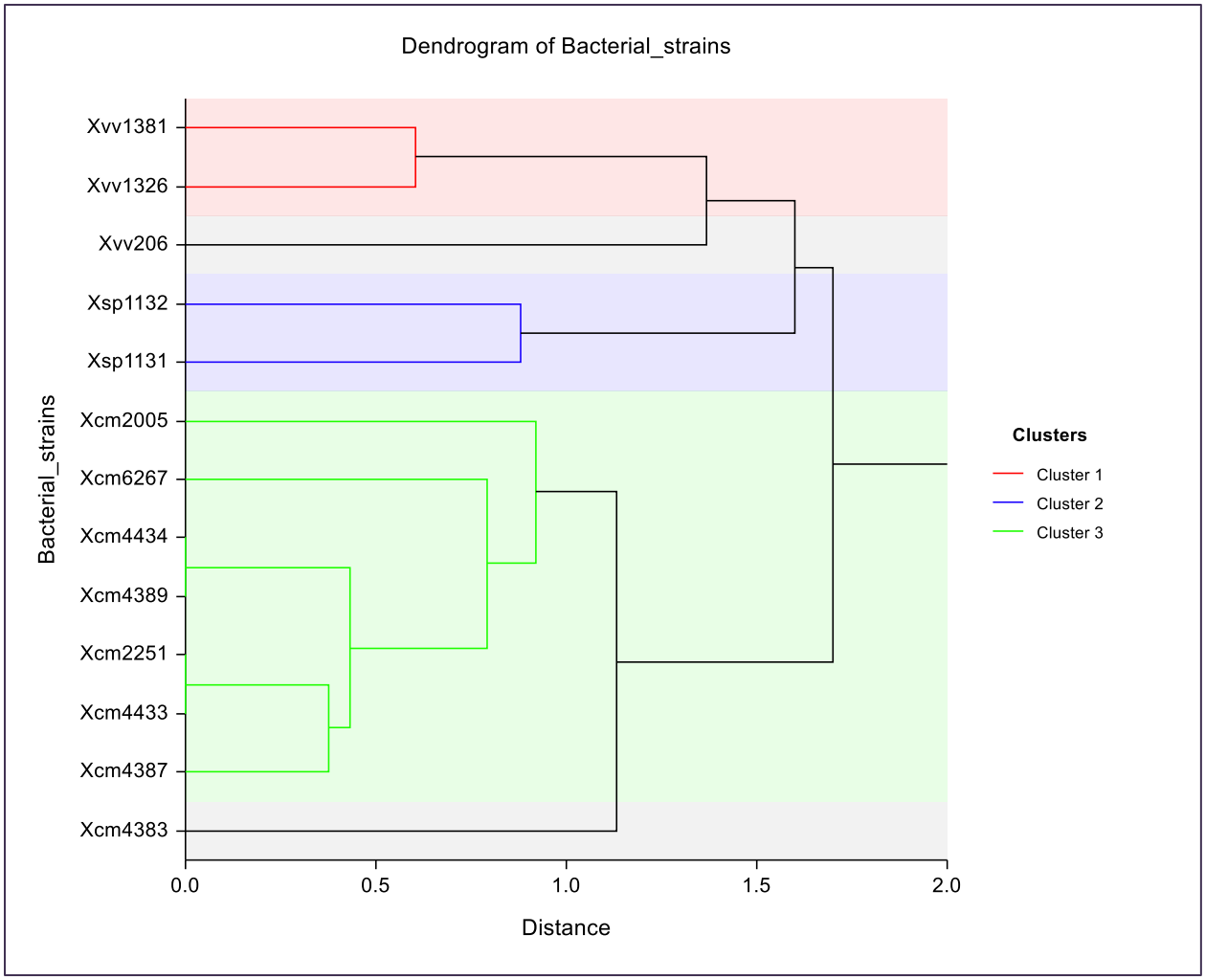

Figure 3.10: Hierarchical clustering dendogram showing clustering of Xanthomonas isolates based on presence or absence of selected virulence factors.

3.4 Discussion

3.4.1 Presence of Lipopolysaccharides in Xcm and Xvv isolates

It has been shown that Xcm isolates tested contain some lipopolysaccharides that are not detected in Xvv isolates tested. Studholme et al., (2010) had predicted that lipopolysaccharide locus in Xvv702 and Xcm4381 were not significantly similar which has been confirmed by this study. However a predicted membrane protein was detected in isolates of Xvv1326 and Xvv1381 but not Xvv206. This seems to indicate that the successful pathogenicity of Xcm on banana is due to major

components of the outer membrane. These could be involved in the protection of bacterial cells, contributing to reduce the membrane permeability and thus allowing growth of bacteria in the unfavorable conditions of the plant environment.

The study has shown that O-antigen is only present in the *Xcm* isolates but not *Xvv* nor *Xsp*. One of the most widely studied effects of LPSs on plant cells is their ability to prevent the hypersensitive response (HR) induced in plants by avirulent bacteria. *X. axonopodis pv. citri* 306 shares 93% nucleotide sequence similarity with Xcm4381 Studholme et al., (2010) and Casabuono et al., (2011) suggested that the O-antigen region of *X. axonopodis pv. citri* 306 LPS could be involved in the innate immunity of citrus. Horizontal gene transfer events (HGT) are frequently observed in genomic regions that encode functions involved in biosynthesis of the outer membrane located lipopolysaccharide (LPS).

Although LPS has been suggested as a potentiator of plant defense responses, inter-strain variation at LPS biosynthetic gene clusters has not been reported for any plant pathogenic bacterium. Aritua et al., (2008), found that although banana and sugarcane strains were of similar phylogenic group, their host ranges were different, since only the *Xcm* pathovars could cause disease in banana. This is consistent with the observation that lipopolysaccharide (LPS) O-antigen is pathovar/strain specific and may be involved in host range selection and pathogenicity by acting as a barrier against plant toxins (da Silva et al., 2002). Also different strains of the same pathogen can have substantially different LPS biosynthetic gene clusters as has been shown for NAD dependent epimerase and Lipopolysaccharide biosynthesis protein in our study. This can be attributed to the advantage in evading the host immune system since LPS is highly antigenic.

3.4.2 Role of Tfp in Xcm and Xvv virulence

The ability of pathogenic bacteria to infect plants involves more than their ability to form the Hrp TTSS and transfer virulence effectors. Successful establishment of a pathogenic bacterium in the host tissue requires the coordinated activity of a plethora of genes, the identity and mode of action of many of which are still not fully defined especially for *X. vasicola pv. musacearum*. Qian et al., (2005) showed that infection is initiated with bacterial attachments to and colonization of host tissues via surface structures and appendages. Attachment of plant pathogenic bacteria to host-

specific surfaces/cells is necessary for colonization of host tissue and is mediated by surface-exposed adhesins, which generally behave as lectins, recognizing oligosaccharide residues of glycoprotein or glycolipid receptors on the host cell (Burdman et al., 2011; Kline et al., 2009)

In this study it has been showed that Xcm and Xvv isolates differ in their Tfp composition as was earlier predicted by Studholme et al., (2010). Also showed that the structural subunit PilE was existent in all the isolates of *Xanthomonas* under test except X.sp1131 and X.sp1132. Interestingly the primers for PilE were able to amplify a product in Xcm2005 (Enset), Xvv1326, Xvv1381(sugar cane) and yet no product was observed for Xvv206(maize). (Moreira et al., (2010) and Moreira et al., (2005) report of the presence and divergence of copies of PilE-FimT cluster in ancestors of xanthomonadaceae is consistent with the findings of this study where FimT was hybridized in only the Xvv1326 and Xvv1381 strain from sugarcane. PilV is reported to be essential for pilus mediated cell adherence. The involvement of minor pilins (i.e. PilE, PilV, PilW, PilX, FimT) in pilus assembly may play roles in priming of pilus extension or prevention of pilus retraction; in control of pilus length; or in pilus-specific functions including adherence, transformation competence or motility (Giltner et al.,2012). The putative type IV pilus protein PilY1 is likely important for attachment to surfaces. In Xanthomonas, type IV pili may be associated with the establishment of an aggregated bacterial population necessary to counteract the turbulent environment of the xylem, facilitating its adherence to the vessels in conjunction with other components, such as exopolysaccharides (Sluys et al., 2002). Moreira et al., (2004, 2005) reported a cluster *pilE-pilY1-pilX-pilW-pilV-fimT* which was common to both *Xylella fastidiosa* and *X. axonopodis pv. citri*. In our study we show that isolates of the same species of plant pathogenic bacteria may vary in clusters of Tfp. The variation in Tfp particularly between Xcm and Xvv remains to be investigated in its role in the successful colonization of the host and nonhost. However Studholme et al., (2010) indicated that the Tfp cluster which is adjuscent to tRNA-Asn gene can be completely absent/partially deleted. The cluster also is dispensable in some xanthomonas species, functionally redundant with very little sequence similarity.

3.4.3 Type III secretion system effectors (TTSS)

In this study we were able to confirm that proteins RipT, HopAF1 and YopJ-like1/2 amplified with Xcm isolates under test. It was also observed that Xvv206, Xvv1326 and Xvv1381 did not yield

any product for YopJ-like proteins. These YopJ family members act as cysteine proteases which is required for inhibition of the mitogen-activated protein kinase (MAPK) and nuclear factor kB (NF-kB) signaling for induction of localized cell death in plants (Orth, 2007; Orth et al., 2000). This seems to imply that host colonization of Xcm is linked to the presence of *YopJ*-like genes and therefore silencing of this gene can offer resistance.

Although effector protein HopW1 and XopAF (AvrXv3) was amplified in Xvv1326 and Xvv1381, it was not detected in Xvv206. HopW1 is also encoded in *Pseudomonas syringae* where it forms complexes with actin, disrupts the actin cytoskeleton, inhibits endocytosis and trafficked certain proteins to the vacuoles (Kang et al., 2014). In plants, the actin cytoskeleton is important for various cell biological processes that may be important for immune signaling. The C-terminal domain of HopW1 is critical for triggering defenses and interacting with three HopW1-interacting proteins (WIN1, WIN2 and WIN3). Jalan et al., (2013) in characterizing pathotypes of *X. citri* subsp. *Citri* showed that XopAF was present in Xcaw12879 but not XccA306 thus contributing to virulence in Xcaw.

3.5 Conclusion

The identification of factors involved in the host range definition and evolution is a pivotal challenge in the goal to predict and prevent the emergence of plant bacterial disease. The LPS NAD dependent epimerase and O-antigen are present in all the Xcm isolates but not Xvv and thus can be used for further genomic studies on species differentiation. The structural subunit *PilE* was confirmed all Xcm except NCPPB206, NCPPB1131 and NCPPB1132. In order to comprehensively characterize the pathogenicity repertoire of plant pathogens such as Xcm and Xvv it is necessary to move beyond single gene approaches and to apply genomics tools. Further, genomics approaches reveal origins of pathogenicity and virulence factors and thus contribute to our understanding of how microbial pathogenesis evolves. The importance of horizontal gene transfer (HGT) events that lead to the acquisition of foreign DNA sequences has been well documented in some *Xanthomonas* bacteria. Efforts to define the impact of these events in the genomic structure and variations between closely related species have to be made for Xcm and Xvv.

Chapter Four

Diversity and spread of the Banana Pathogen *Xanthomonas Campestris* pv. *musacearum isolates* in Great Lakes region of Africa

Abstract:

Many gram-negative pathogenic bacteria directly translocate effector proteins into eukaryotic host cells via type III delivery systems. Type III effector proteins are determinants of virulence on susceptible plant hosts; they are also the proteins that trigger specific disease reaction in susceptible/resistant plant hosts. Evolution of type III effectors is dominated by competing forces: the likely requirement for conservation of virulence function, the avoidance of host defenses, and possible adaptation to new hosts. Here genome-wide sequencing was used to discover a set of SNPs and also to trace the relationships among East African isolates of *Xanthomonas*. These SNPs have potential as molecular markers for phylogeographic studies of the epidemiology and spread of the pathogen. East African isolates of *Xanthomonas campestris pv. musacearum* (Xcm) from banana and ensete comprise a monophyletic clade closely related to Xanthomonas vasicola. This analysis reveals two major sub-lineages of the pathogen Xcm, suggesting that the current outbreaks of BXW on *Musa* species in the region may have more than one introductory event. Isolates of Xcm from Burundi, Kenya, Tanzania and Uganda comprise a separate sub-lineage that is distinct from isolates from and D. R. Congo, Ethiopia and Rwanda. Also, based on comparisons of genome-wide sequence data from multiple isolates of *Xcm* and multiple isolates of *X. vasicola* pathovar *vasculorum*, genes specific to *Xcm* that could be used to specifically detect *Xcm* by PCR based methods were identified.

Keywords: banana *Xanthomonas* wilt; whole-genome sequencing; SNP; molecular markers; *Ensete ventricosum*; *Xanthomonas vasicola*

4.1 Introduction

The bacterium *Xanthomonas campestris* pathovar *musacearum* (*Xcm*) is the causal agent of BXW. This disease has devastated economies based on banana and plantain crops (*Musa* species) in East Africa (Biruma et al., 2007). *Xcm* was first described as a wilt-causing pathogen on ensete (*Ensente ventricosum*), a plant closely related to banana that is a staple crop in the highlands of Ethiopia (Yirgou and Bradbury, 1968). In 1974 Yirgou and Bradbury (1974) wrote that "*Great care should be taken to see that enset wilt does not escape and establish itself on banana in other parts of the world where it could pose a serious problem on this crop*". Ominously, one and a half decades later, a major epidemic of this disease was reported in Uganda (Tushemereirwe et al, 2004). Subsequently it has spread into many banana-growing regions around the Great Lakes in Uganda, Kenya, Tanzania, Democratic Republic of Congo, Rwanda and Burundi (Eden-Green *et al.*, 2006, Reeder *et al.*, 2007, Carter *et al.*, 2010). Efforts are underway to tackle this pathogen by a number of different measures including cultural practices (Kubiriba et al., 2012) and genetic modification of the crop (Namukwaya et al., 2012; Tripathi et al., 2010).

Over 20 Xanthomonas species are divided into two main phylogenetic groups based on 16S rDNA and gyrB sequence analysis (Parkinson et al., 2007; Rodriguez-r et al., 2012; Simões et al., 2007) and subdivided into pathovars loosely corresponding to host specificity. Group 1, also known as the early branching group, comprises highly diverse Xanthomonas species including important sugarcane and cereal pathogens (e.g., *Xanthomonas sacchari, Xanthomonas albilineans and Xanthomonas translucens*, (Hauben et al., 1997; Parkinson et al., 2007). Group 2, the largest and best-described group, includes species such as *Xanthomonas oryzae, Xanthomonas citri, Xanthomonas vasicola, Xanthomonas euvesicatoria, Xanthomonas axonopodis* and *Xanthomonas campestris* (Hauben et al., 1997; Parkinson et al., 2007). This diverse genus of bacteria infects and associates with many plant hosts, but individual strains typically possess very restricted host ranges limited to a single genus.

Although it is classified as a member of the species *Xanthomonas campestris*, it was shown that *Xcm* is more likely to be a strain of the species *Xanthomonas vasicola* (V Aritua et al., 2008). Studholme et al., (2010) previously generated complete genome sequences for a single isolate of

Xcm from banana in Uganda (NCPPB4381) and for a single isolate of *X. vasicola* pathovar *vasculorum* (*Xvv*) that is non-pathogenic on banana, isolated from sugarcane in Zimbabwe (NCPPB702). These two isolates share identical gyraseB DNA sequences, consistent with their very close evolutionary relationship (Aritua et al., 2008).

Differences between these two genome sequences revealed several candidate genes that might play a role in adaptation to the banana host. These may also be useful tools in identifying genes for deployment of disease resistance. Specifically, these included homologues of effectors secreted and translocated by the type III secretion system (TTSS). TTSS effectors have previously been shown to contribute to host specificity acting as virulence and/or avirulence factors (Boucher & Roby, 2009; Leach & White, 1996). In common with most previously sequenced *Xanthomonas* genomes, *Xcm* encodes homologues of the effectors AvrBs2, AvrGf1, XopF, XopK, XopL, XopN, XopP, XopQ, XopR, XopX and XopZ as well as homologues of XopA, XopB, XopG, XopH, XopI, XopY, XopAA, XopAD, XopAE and XopAK, which are found in some other *Xanthomonas* species (Studholme et al., 2010). *Xcm* also encodes homologues of *P. syringae* effectors HopW1 and HopAF1 and *Ralstonia solanacearum* putative effector RipT. *Xcm* encodes two predicted YopJ-like C55 cysteine proteases (RefSeq accessions ZP_06491730 and ZP_06492219) (Zheng et al., 2012) that are absent from *Xvv* 702. On the other hand, *Xvv* 702 encodes a close homologue (ZP_06483517) of XopAF (also known as AvrXv3), which is absent from *Xcm* (Studholme et al., 2010).

Previous work showed that *Xcm* is a highly monomorphic pathogen and no specific genetic differences have yet been detected among different isolates using traditional typing and diagnostic methods (Aritua et al., 2008). Affordable complete genome sequencing now makes it feasible to identify cryptic genetic diversity among isolates of a genetically monomorphic pathogen (Cai et al., 2011; Jones & Dangl, 2006; Lefeuvre et al., 2013) .

This study compared genomes of multiple isolates of the *Xanthomonas campestris* pv. *musacearum* from Ethiopia, Democratic Republic of Congo (DRC), Rwanda, Burundi, Tanzania, Kenya and Uganda and also against Xvv isolated from sugarcane and maize. It also aimed at performing a Single-Nucleotide Polymorphism (SNP) analysis of a large collection of isolates, to unravel the microevolution and regional spread of a banana bacterial plant pathogen. These SNPs have potential

as molecular markers for phylogeographic studies of the epidemiology and spread of the pathogen.

4.2 Materials and Methods

Bacterial isolates were obtained from the National Collection of Plant Pathogenic Bacteria (NCPPB) at Food and Environmental Research Agency (FERA). All Xcm isolates were originally collected from diseased banana plants except for NCPPB2005, which was isolated from *Ensete ventricosum* from Ethiopia, DRC, Rwanda, Burundi, Tanzania, Kenya and Uganda.. All Xvv isolates were originally collected from sugarcane, except for NCPPB206, which was isolated from maize. DNA library preparation and genome sequencing using the Illumina GA2x were performed using standard Illumina protocols as previously described (Studholme et al., 2010). The DNA preparation, purification and amplification were done as previously described in Chapter three.

The Illumina GA2x sequencing platform was used to generate genome-wide sequence data for 13 isolates of Xcm available from the NCPPB. A further three African isolates of Xvv, a pathovar that are very closely related to Xcm but non-pathogenic on banana were also sequenced. Genome sequence data from Xcm NCPPB4381 and Xvv NCPPB702 was also included in the analyses (Studholme et al., 2010). Table 4.1 lists the sequenced isolates and the depth to which each was sequenced. All raw sequence data was submitted to the Sequence Read Archive (Kodama et al., 2012; Leinonen et al., 2011).

The raw unassembled sequence reads were aligned against the previously published Xcm NCPPB4381 reference genome sequence (Studholme et al., 2010) using the Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2009) against a reference genome sequence and used IGV (Robinson *et al.*, 2011) to visualize the alignments (fig. 4.1). Burrows-Wheeler Aligner BWA (Li and Durbin, 2009) approach has the advantage of avoiding assembly artifacts and problems arising from incomplete assemblies. In the BWA alignments of raw Illumina sequence reads versus the reference genome sequence, the full length of each gene was covered by reads (depth of one or greater) from all our Xcm Illumina sequence datasets. *De novo* assembly of Illumina sequence reads was performed using Velvet 1.1.04 (Zerbino & Birney, 2008). Any sequence reads that contained one or more 'N' prior to assembly were discarded.

A very conservative approach to infer SNPs from alignments of Illumina reads against the previously published *Xcm* NCPPB4381 reference draft genome assembly was applied. To avoid false positives and false negatives, only those regions of the *Xcm* genome with coverage depth between 5-95% for every sequenced *Xcm* genome were analyzed. Over 60% of the length (2,908,042 out of 4,782,144 nt) of the *Xcm* NCPPB4381 genome fulfilled these two criteria, where data was in sufficient quantity and consistent in all of the fourteen isolates. For the remaining 40% of the genome, there was some degree of ambiguity in the data for at least one of the isolates and was removed from the analyses.

Comparative analysis of molecular sequence data is essential for reconstructing the evolutionary histories of species and inferring the nature and extent of selective forces shaping the evolution of genes and species (Gilbert & Webb, 2007). Phylogenetic relationships were inferred using a maximum likelihood method based on the Tamura- Nei model conducted using the MEGA5 software package (Hall, 2013; Koichiro Tamura et al., 2011). Bootstrap consensus trees inferred from 500 replicates were taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in fewer than 50% bootstrap replicates were collapsed. Initial tree(s) for the heuristic search were obtained automatically as follows. When the number of common sites was <100 or less than one quarter of the total number of sites, the maximum parsimony method was used; otherwise the BIONJ method with MCL distance matrix was used. The trees are drawn to scale, with branch lengths measured in the number of substitutions per site. All positions containing gaps and missing data were eliminated.

Amplified DNA fragments were digested with restriction endonucleases (AluI, Fok or NdeI). The restriction analysis was performed with 2.5 U of the endonuclease using the buffer and temperature recommended by the manufacturers (New England Biolabs). Restriction fragments were separated in a 8% (w/v) agarose gel with 100 bp ladder (Promega, G210A) at 100 V for 1 h in TAE buffer and visualized with UV light after staining in ethidium bromide (0.5 µg mL).

Digestion of each of the PCR products from each with a restriction enzyme AluI, EcoRI, FokI, NdeI or RsaIwas conducted following the manufacturers guidelines. Electrophoresis of the digested PCR products on a 2% agarose gel alongside a 100 bp ladder (Promega G210A) in which the brightest band, marked with a black arrow, indicates 500 bp (fig. 4.4).

4.3 Results and Discussion

4.3.1 Genome Sequencing

Genome assemblies *de novo* for NCPPB2005, NCPPB4379, NCPPB4380, NCPPB4384, NCPPB4392, NCPPB4394, NCPPB1326, NCPPB1381 and NCPPB206 using Velvet 1.1.04 (Zerbino *et al.*, 2008) were generated. These have been submitted to GenBank (Benson *et al.*, 2012) with accession numbers AKBE00000000, AKBF00000000, AKBG00000000, AKBH00000000, AKBI00000000, AKBJ00000000, AKBK00000000, AKBL00000000 and AKBM00000000.

Table 4.1: Isolates of X. campestris pv. musacearum (Xcm) and X. vasicola pv. vasculorum (Xvv) subjected to genome-wide sequencing

Isolate	Source and Date of Isolation	Host	Genomic coverage	SRA Accession
Xcm NCPPB2005	Ethiopia 1967	Ensete	72×	SRR489154.7
Xcm NCPPB2251	Ethiopia 1969	Banana	13×	SRR494492.2
Xcm NCPPB4379	Uganda (Kayunga) 2007	Banana	102×	SRR494484.2
Xcm NCPPB4380	Uganda (Kiboga) 2007	Banana	113×	SRR494485.2
Xcm NCPPB4381	Uganda (Luwero) 2007	Banana	56×	SRR020203.3
Xcm NCPPB4383	Uganda (Wakiso) 2007	Banana	11×	SRR494493.2
Xcm NCPPB4384	Uganda (Nakasongola) 2007	Banana	55×	SRR494488.2
Xcm NCPPB4387	D. R. Congo (Kivu) 2007	Banana	13×	SRR494494.1
Xcm NCPPB4389	Rwanda (Gishenyi) 2007	Banana	16×	SRR494495.2
Xcm NCPPB4392	Tanzania (Maleba) 2007	Banana	72×	SRR494498.3
Xcm NCPPB4394	Tanzania (Maleba) 2007	Banana	92×	SRR494489.1

Xcm NCPPB4395	Tanzania (Maleba) 2007	Banana	117×	SRR494490.2
Xcm NCPPB4433	Burundi 2008	Banana	13×	SRR494496.1
Xcm NCPPB4434	Kenya 2008	Banana	15×	SRR494497.1
Xvv NCPPB206	South Africa 1948	Sugarcane	70×	SRR494500.3
Xvv NCPPB702	Zimbabwe 1959	Maize	35×	SRR020202.3
Xvv NCPPB1326	Zimbabwe 1962	Sugarcane	63×	SRR494491.5
Xvv NCPPB1381	Zimbabwe 1962	Sugarcane	66×	SRR494499.3

The most contiguous of these assemblies was for NCPPB4384. This consisted of 84 scaffolds, of which the 12 longest scaffolds were each at least 154 Kb long and accounted for more than 2.5 Mb; that is the N_{50} length was 154 Kb for NCPPB4384. The N_{50} lengths for the other *Xcm* assemblies were 56 Kb (NCPPB2005), 146 Kb (NCPPB4379), 147 Kb (NCPPB4380), 87 Kb (NCPPB4392) and 151 Kb (NCPPB4394).

4.3.2 Genomic differences between Xcm and Xvv

A set of genes that are conserved in all of the sequenced isolates of *Xcm* were identified. These were, however, absent from all the sequenced isolates of *Xvv* and are therefore candidates for use in an *Xcm*-specific PCR assay. Examples of these genes are listed in Table 4.2. Note that this list of genes was not generated by aligning assembled genome sequences. Each of these genes has no matches (*i.e.*, zero depth of) to any sequence reads in the *Xvv* Illumina datasets (as judged from the BWA alignments)

Table 4.2: Candidate genes for development of an Xcm-specific PCR based assay

RefSeq Locus tag	**Predicted gene product**
XcampmN_010100002667	hypothetical protein
XcampmN_010100009057	general secretion pathway protein D
XcampmN_010100016989	Transposase
XcampmN_010100016984	phage-related integrase
XcampmN_010100014552	hypothetical protein
XcampmN_010100013878	DNA-cytosine methyltransferases
XcampmN_010100013483	hypothetical protein
XcampmN_010100011643	conjugal transfer relaxosome component TraJ
XcampmN_010100011578	hypothetical protein
XcampmN_010100011573	Fis family transcriptional regulator
XcampmN_010100011558	hypothetical protein
XcampmN_010100011553	hypothetical protein
XcampmN_010100010854	hypothetical protein
XcampmN_010100010849	XRE family transcriptional regulator
XcampmN_010100006985	hypothetical protein
XcampmN_010100004971	exported protein
XcampmN_010100004961	virulence regulator

XcampmN_010100004956	hypothetical protein
XcampmN_010100004736	hypothetical protein
XcampmN_010100001342	ISXo2 putative transposase
XcampmN_010100001332	ABC-type antimicrobial peptide transport system ATPase component
XcampmN_010100001327	RND family efflux transporter MFP subunit
XcampmN_010100013888	ISxac1 transposase
XcampmN_010100011563	putative DNA methylase
XcampmN_010100004966	Integrase
XcampmN_010100001337	peptide ABC transporter permease
XcampmN_010100013883	restriction endonuclease-like protein
XcampmN_010100000225	putative secreted protein
XcampmN_010100000622	Fimbrillin
XcampmN_010100015677	Methyltransferases
XcampmN_010100016677	Putative acetylhydrolase

4.3.3 The Sequenced Xcm Isolates Comprise a Single Monophyletic Clade

Based on nucleotides found at 21,525 polymorphic positions from 19 taxa, the maximum likelihood phylogenetic tree shown in fig. 4.1 was generated. This clearly groups all of the sequenced *Xcm* isolates into a single distinct clade that is closely related to but distinct from the sequenced *Xvv* isolates.

The phylogenic tree was rooted with *X. oryzae* pv. *oryzae* MAFF 311018 ("*Xoo*") as the outgroup. Branch lengths are drawn to scale and measured in the number of substitutions per site. Bootstrap values are given as percentages from 500 bootstrap trials.

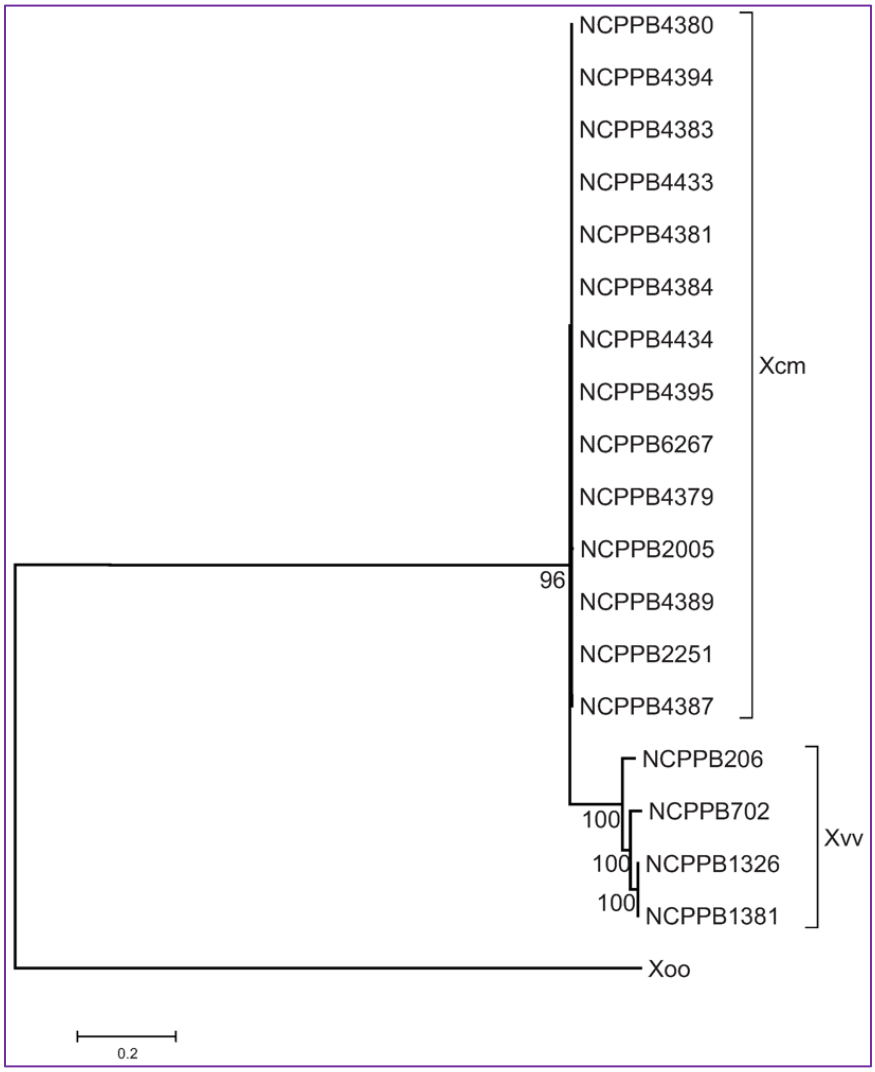

Figure 4.1: East African isolates of *X. campestris pv. musacearum* (Xcm) from banana and ensete comprise a monophyletic clade closely related to *Xanthomonas vasicola*

4.3.4 Xcm Isolates from Uganda, Kenya, Tanzania and Burundi are Genetically Distinct from Isolates from Ethiopia, DR Congo and Rwanda

Out of 2,908,042 nucleotides (nt) over which there was no ambiguity, 2,907,999 were invariant across all isolates; that is the *Xcm* genomes shared at least 99.9985% identity relative to *Xoo*. The two sub-lineages are distinguishable from each other by 86 polymorphic positions (SNPs). These are listed in full in the in Table 4.3.

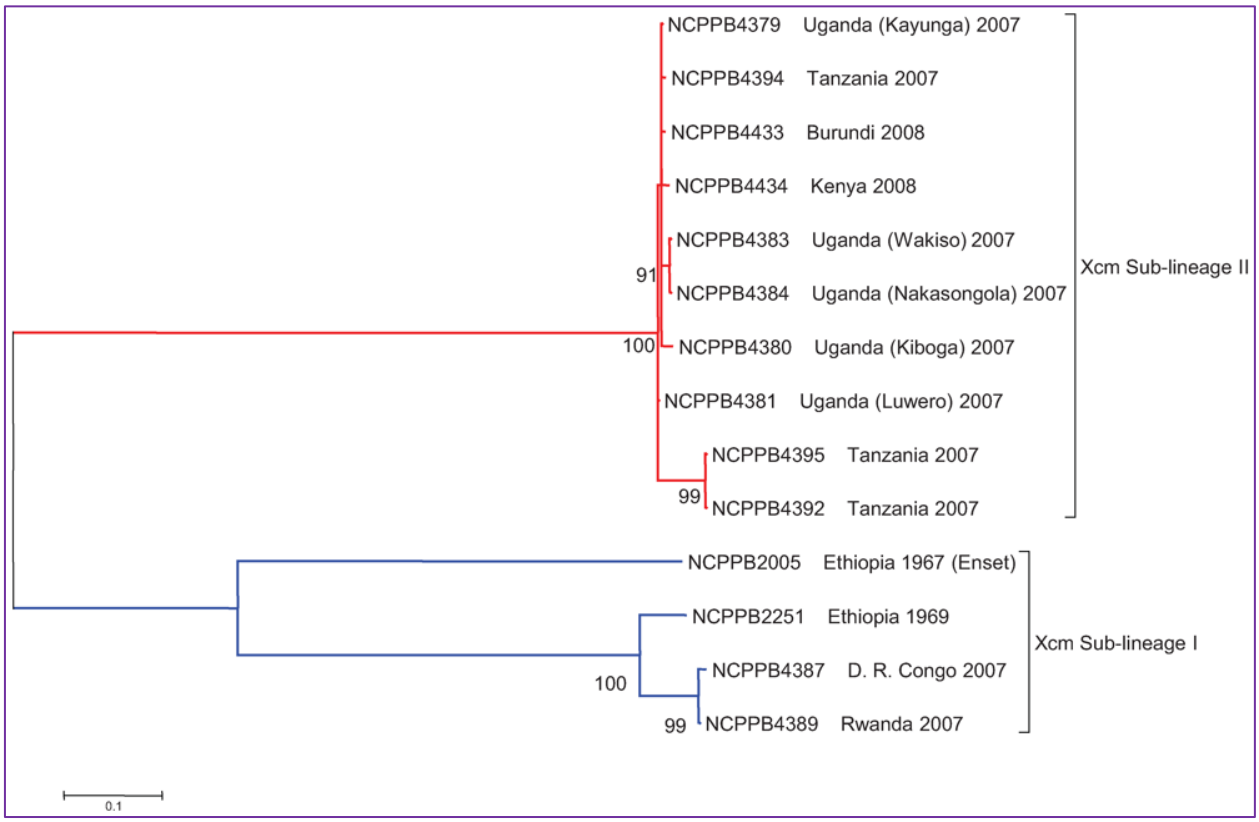

Figure 4.2: Isolates of Xcm from Burundi, Kenya, Tanzania and Uganda comprise a separate sub-lineage that is distinct from isolates from and D. R. Congo, Ethiopia and Rwanda

Table 4.3: Non-silent single-nucleotide polymorphisms that distinguish Xcm sub-lineages I and II

RefSeq Accession	Position	I	II	Locus Tag and Predicted Gene Product
NZ_ACHT01000013	861	g	C	XcampmN_010100000120 putative ISXo8 transposase
NZ_ACHT01000014	6000	a	G	XcampmN_010100000165 putative monovalent cation/H+ antiporter subunit A
NZ_ACHT01000034	8898	t	C	XcampmN_010100000807 putative integrase protein
NZ_ACHT01000045	1261	a	G	XcampmN_010100001162 bifunctional aspartate kinase/diaminopimelate decarboxylase protein
NZ_ACHT01000045	45,548	a	C	XcampmN_010100001377 chemotaxis protein
NZ_ACHT01000059	1907	c	T	XcampmN_010100001687 putative sugar transporter component
NZ_ACHT01000101	995	g	t	XcampmN_010100003517 soluble lytic murein transglycosylase
NZ_ACHT01000104	13,081	t	g	XcampmN_010100003612 GTP-dependent nucleic acid-binding protein EngD
NZ_ACHT01000113	10,410	t	g	XcampmN_010100004062 acetyltransferase (GNAT) family protein
NZ_ACHT01000236	10,652	t	c	XcampmN_010100007340 metallopeptidase
NZ_ACHT01000242	10,465	a	c	XcampmN_010100007585 dihydrolipoamide acetyltransferase
NZ_ACHT01000294	2184	g	t	XcampmN_010100009424 xanthan biosynthesis glucuronosyltransferase GumK
NZ_ACHT01000345	1576	t	c	XcampmN_010100010814 cytochrome C peroxidase
NZ_ACHT01000402	4858	t	c	XcampmN_010100012145 heavy metal transporter
NZ_ACHT01000404	632	g	a	XcampmN_010100012200 tryptophan halogenase
NZ_ACHT01000500	23,584	a	g	XcampmN_010100016057 putative polysaccharide deacetylase
NZ_ACHT01000520	5360	a	g	XcampmN_010100016692 5-methyltetrahydrofolate-homocysteine methyl transferase
NZ_ACHT01000549	7371	a	c	XcampmN_010100018271 two-component system sensor protein
NZ_ACHT01000560	4001	t	c	XcampmN_010100018673 exodeoxyribonuclease III
NZ_ACHT01000590	927	c	t	XcampmN_010100019303 RNA polymerase sigma factor
NZ_ACHT01000626	10,220	t	c	XcampmN_010100019733 putative glutathionylspermidine synthase
NZ_ACHT01000634	2345	t	c	XcampmN_010100019848 beta-mannosidase precursor
NZ_ACHT01000644	2590	g	a	XcampmN_010100020168 two-component system sensor protein
NZ_ACHT01000694	10,665	a	t	XcampmN_010100022153 peptide-acetyl-coenzyme A transporter family Protein
NZ_ACHT01000720	19,485	t	c	XcampmN_010100023003 drug: proton antiporter (19121–20371)

Xcm sub-lineages I (Ethiopia, DR Congo and Rwanda) and II (Uganda, Kenya, Tanzania, Burundi)

The geographic locations of the sequenced *Xcm* isolates are shown in fig. 4.3. It has been widely assumed that the outbreaks in Uganda, and subsequent outbreaks in neighboring countries, ultimately originated in Ethiopia, with the pathogen perhaps being inadvertently transmitted via international trade. Consistent with this model, the *Xcm* isolates from DR Congo and Rwanda do indeed show extremely high levels of genetic similarity to Ethiopian isolate NCPPB2251. However, the isolates from Uganda, Kenya, Tanzania and Burundi show a distinct genotype, characterized by the 86 consistent SNP differences. This is not consistent with a single introduction from Ethiopia into East Africa. These lineages could be a result of mutations acquired for survival of the isolates in specific host environment in DRC/Rwanda and Uganda/Kenya. These two have varying banana cultivation techniques and thus micro environments.

The data do not exclude the possibility that the current outbreaks can ultimately be traced back to Ethiopia; it is possible that both lineages I and II are endemic there and it is simply by chance that the two available isolates happen to belong to the DR Congo/Rwanda sub-lineage I. There is an urgent need to collect a range of isolates from Ethiopia and survey their genotypes to ascertain the level of genetic diversity in this pathogen's presumed center of origin. Genotyping new isolates should be possible and will be expedited by developing these newly discovered SNPs into PCR based molecular markers. Similarly, there is a pressing need to survey genotypes of a much larger collection of isolates from outbreaks in all the banana-growing areas to uncover the routes of geographical spread at a much higher degree of resolution. Ideally a survey of genotypes should be conducted on isolates for which precise details are available on the date and the geographic location at which they are collected.

Figure 4.3: Geographical distribution of the two major sub-lineages of Xcm. Blue rectangles indicate locations of isolates belonging to sub-lineage I and red ovals indicate those of sub-lineage II. The approximate geographical locations are indicated for each of the fully sequenced Xcm isolates from Ethiopia (NCPPB2005 and NCPPB2251), Uganda (NCPPB4379, NCPPB4380, NCPPB4381, NCPPB4383 and NCPPB4384), Kenya (NCPPB4434), Tanzania (NCPPB4392, NCPPB4394 and NCPPB4395), DR Congo (NCPPB4387), Rwanda (NCPPB4389) and Burundi (NCPPB4433).

It should also be noted that although we have categorized the isolates according to the country in which they were collected, paths of transmission may be more influenced by geographical boundaries rather than by political ones. For example, Rwanda shares two of its borders with Uganda and Tanzania its population regularly exchanges goods and plant materials including suckers.

All of the available isolates from Uganda were collected in 2007 from sites in the central region geographically close to Mukono where the disease was first reported and probably all represent

the same single outbreak. It would be interesting to compare these with isolates from outbreaks in Kabale (near Rwanda) or Kasese (near DR Congo).

All of the available isolates from Tanzania also belonged to sub-lineage II along with those from Uganda. The disease was reported in Tanzania shortly after it was discovered in Uganda and there has been unconfirmed speculation that it may have been inadvertently carried to Tanzania by banana alcohol traders from the Buganda region, close to where the sequenced Ugandan isolates were collected.

Although BXW was reported in DR Congo after it was reported in Uganda, the field pictures first sent to Uganda from DR Congo, showed greater devastation. It is not clear where banana *Xanthomonas* wilt occurred first: DR Congo or Uganda. There is a lot of movement of people and bananas from Congo to Rwanda and back, conflicts notwithstanding, and so it is perhaps not surprising that we observe isolates from these two countries belonging to the same sub-lineage I. However, the close relationship between an Ethiopian isolate and those in Rwanda and DRC is not so easily explained. The disease spread could be primarily determined by movement of local people and bananas between countries, then would instead expect isolates from Rwanda, Uganda and DR Congo to cluster together.

4.3.5 Comparison of Xcm Isolated from Ensete versus Xcm Isolated from Banana

Most of the available isolates of *Xcm* were originally isolated from banana. The exception is NCPPB2005, which was isolated from *Ensete ventricosum* in Ethiopia in 1967. This isolate clearly falls within *Xcm* sub-lineage I (fig. 4.3). This ensete-associated isolate differs from the banana-associated isolates NCPPB2251, NCPPB4387 and NCPPB4389 at 67 SNP positions. Some examples of these differences are listed in Table 4.4 and include non-silent polymorphisms (base substitutions that result into change of amino acid function) in several potential virulence genes (e.g., homologues of HrpF and a gene encoding a HopW1 TTSS effector). However, since *Xcm* is able to infect both banana and ensete it is not clear whether these differences have any neutral or functional significance. It would be interesting to survey a much larger sample of isolates from both banana and ensete to search for any significant associations between genotype and host species that might reveal adaptation. These isolates should be representative of all the

banana production zones in the region. Using the DNA sequences to date evolutionary events of bacteria by molecular dating is an important method to temporally determine diversification of the Xcm (Rutschmann, 2006).

Table 4.4: Non-silent single-nucleotide polymorphisms that distinguish NCPPB2251 from banana versus NCPPB2005 from ensete

Refseq Accession	Position	NCPPB 2005 (Ensete)	NCPPB 2251 (Banana)	NCPPB 4389 (Banana)	Locus Tag and Predicted Gene Product
NZ_ACHT01000041	15,615	C	T	T	XcampmN_010100000977 hemolysin III
NZ_ACHT01000072	4507	A	C	C	XcampmN_010100002109 VirB3 protein
NZ_ACHT01000140	1116	C	T	T	XcampmN_010100004536 LacI family transcription regulator
NZ_ACHT01000199	8012	G	T	G	XcampmN_010100006143 type III secreted effector HopW1
NZ_ACHT01000215	3229	C	T	C	XcampmN_010100006660 HrpF protein
NZ_ACHT01000236	9512	C	T	T	XcampmN_010100007340 metallopeptidase
NZ_ACHT01000294	31,553	G	A	A	XcampmN_010100009559 MFS Transporter
NZ_ACHT01000303	7530	A	C	C	XcampmN_010100009850 histidine kinase/response regulator hybrid protein
NZ_ACHT01000332	2191	A	G	G	XcampmN_010100010574 putative filamentous hemagglutinin-like protein
NZ_ACHT01000360	1961	A	G	G	XcampmN_010100011266 two-component system sensor protein
NZ_ACHT01000374	12,027	T	C	C	XcampmN_010100011573 Fis family transcriptional regulator

NZ_ACHT01000388	5277	T	G	G	XcampmN_010100011860 AraC family transcriptional regulator
NZ_ACHT01000396	3578	C	G	G	XcampmN_010100011920 catalase
NZ_ACHT01000439	5166	C	G	G	XcampmN_010100013743 ECF subfamily RNA polymerase sigma factor
NZ_ACHT01000532	743	C	T	T	XcampmN_010100017284 beta-glucosidase
NZ_ACHT01000560	2783	C	T	T	XcampmN_010100018663 molybdopterin biosynthesis
NZ_ACHT01000668	1036	C	A	A	XcampmN_010100021383 ABC transporter Permease
NZ_ACHT01000690	6284	T	G	G	XcampmN_010100022008 isocitrate Dehydrogenase

4.3.6 Loss of Phage-Associated Genes in Some Xcm Isolates

In addition to surveying SNPs, a search for loss or gain of genes was conducted. By aligning sequence reads against the previously published NCPPB4381 genome assembly and systematically comparing gene-coverage in each of the alignments, a genomic region (GenBank: GG699410.1) that showed differential coverage among different isolates of *Xcm* was identified (fig. 4 . 4). This region shows significant similarity at the amino acid and nucleotide sequence levels to two previously sequenced phage: *Xanthomonas* phage Cfc1 (RefSeq: NC_001396.1) and *Stenotrophomonas* phage phiSHP2 (GenBank: HM150760.1). Specifically, two Tanzanian isolates (NCPPB4392 and NCPPB4395) appear to have completely lost at least 17 genes from this region, while Ethiopian isolate NCPPB2005 has lost 11 of the same genes. Interestingly, another isolate from the same area of Tanzania (NCPPB4394) appears to have these genes intact as do all the Ugandan, Kenyan, Rwandan and Burundi isolates. Furthermore, there is a high concentration of SNPs in this genomic region. Therefore, it seems likely that this genomic region represents the relic of a phage or similar mobile element that is in the process of

degenerating, convergent, in some members of both sub-lineages.

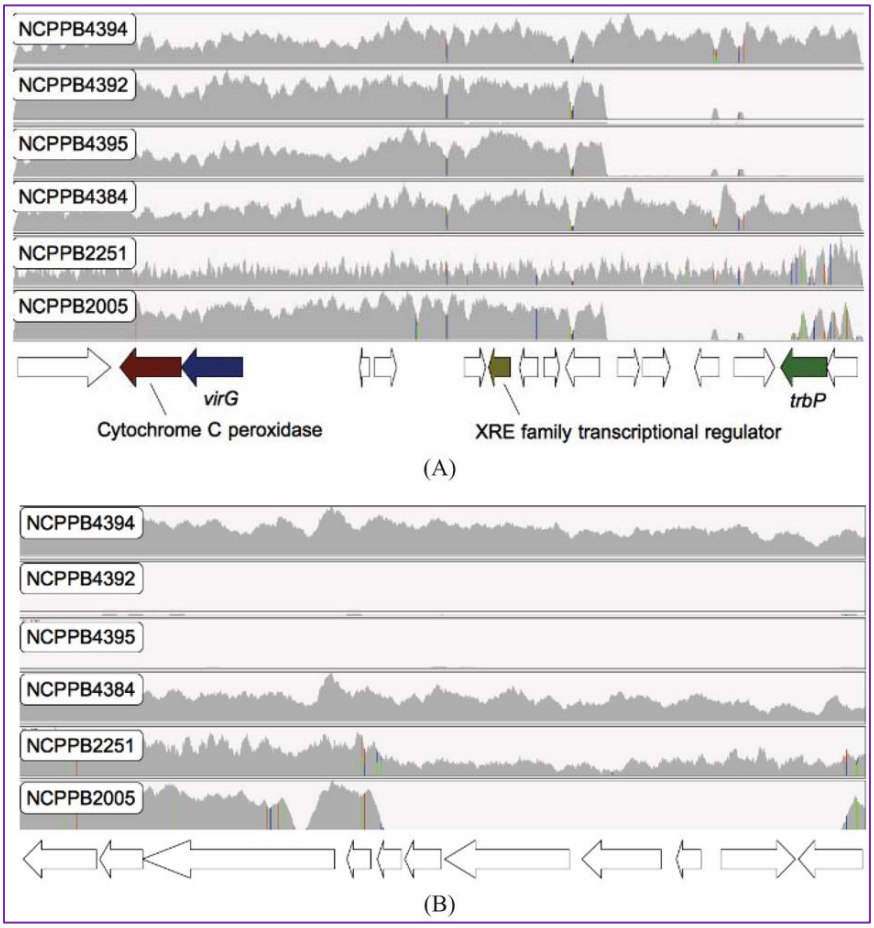

Figure 4.4: Loss of phage-related genes in two Tanzanian isolates and one Ethiopian isolate of Xcm. The figure shows alignments of genomic sequence reads from six Xcm isolates versus two contigs from the previously published NCPPB4381 genome assembly (Studholme et al., 2010) as viewed in IGV. (Panel A): contig_scf_7264_3425_27 (GenBank: ACHT01000345.1); (Panel B): contig_scf_7264_3425_29 (GenBank: ACHT01000346.1). Both contigs are contained within genomic scaffold scf_7264_3425 (GenBank: GG699410.1). The vertical axes are the depth of coverage by aligned sequence reads. Colored vertical bars indicate discrepancies with the NCPPB4381 reference sequence, including SNPs. The horizontal axis is the position on the contig. Positions of predicted genes are indicated below the horizontal axis. Hypothetical genes of unknown function are indicated by white arrows while homologues of characterized genes are indicated by colored arrows.

4.3.7 Experimental Validation of Genetic Polymorphisms

One of the major motivations for comparing these genome sequences is to provide a resource for generating molecular markers that can be used in epidemiological and biogeographical

studies on a much larger panel of isolates without the need for whole-genome sequencing using, for example, the polymerase chain reaction (PCR). Therefore, the results of these genome comparisons are to design pairs of PCR primers that can be used to distinguish between the two sub-lineages of *Xcm* were used (Table 4.5). A similar approach could also be taken to assay other classes of SNPs identified from the genome sequence data.

Many of the SNPs that were identified are predicted to fall within restriction sites. For example, position 6150 in RefSeq accession NZ_ACHT0100081 is a G that falls within an AluI restriction site (A**G**↓CT). However, in NCPPB2005 and the other members of sub-lineage I, this G is substituted for an A abolishing the recognition sequence for the AluI restriction enzyme .There are no other AluI sites in the vicinity of this SNP. We generated a pair of primers flanking approximately 250 bp either side of the SNP or amplified this 500 bp sequence by PCR. The two alleles could then be readily distinguished by digestion with AluI (fig. 4.5). The same approach was used to design assays for three other SNPs (Table 4.5 and fig. 4.5).

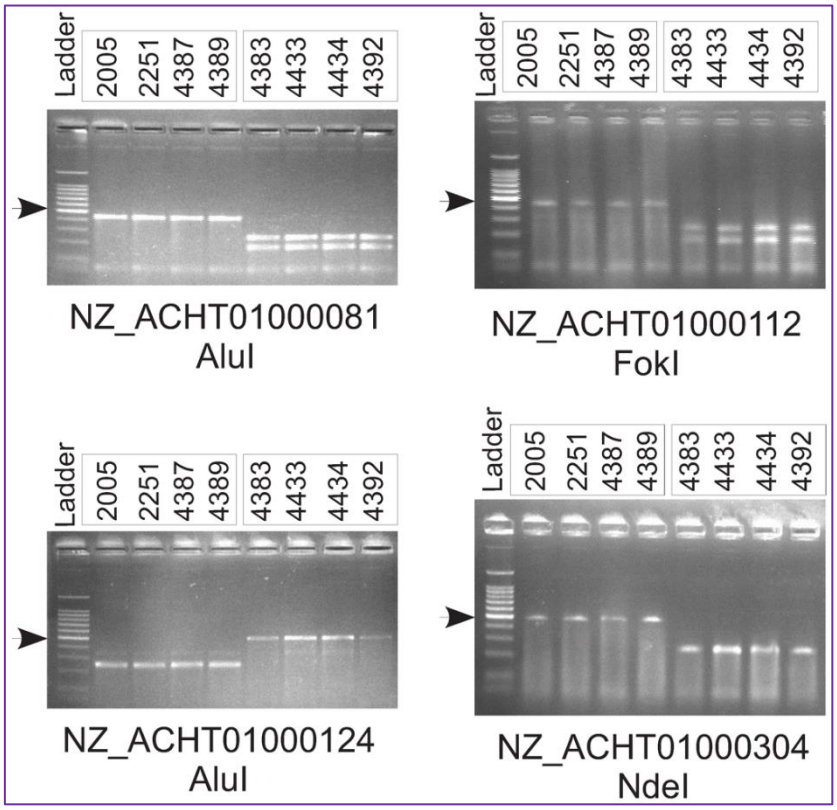

Figure 4.5: Experimental validation of single-nucleotide polymorphisms (SNPs) between the two sub-lineages of

Further analysis of the observed relationship by the alignment of NCPPB2005 to represent sub lineage I and NCPPB4392 to represent sub lineage II. Other products of the digestion using BWA and visualized using IGV for each of restriction enzymes were observed but not presented in this thesis.

Table 4.5: Polymerase chain reaction (PCR) primers used for distinguishing the two sub-lineages by restriction fragment length polymorphisms

Primer Sequences	**Target Sequence RefSeq Accession Number and Coordinates**	**Restriction Enzyme**
GAGCTCCTGCGCCGATGC GTGAGCGTAAAGGCGGCTATTCTA	NZ_ACHT01000081: 5900–6398	AluI
CGGCGTGGTTTTGCCTTTGC CGTACGGCCTGGCGGTGAT	NZ_ACHT01000112: 10863–11347	FokI
TCACCTGTTCGATGCGGCC GCTACTGGCTGTCGCGGC	NZ_ACHT01000124: 5385–5873	AluI
ATGTTTGCCGATACCTGGATGC GCATGCTTGCCGGTTTCGACGA	NZ_ACHT01000304: 10080–10567	NdeI

4.4 Conclusions

High throughput whole-genome sequencing was used to explore genetic diversity among isolates of *Xcm*, the bacterial pathogen responsible for BXW, which is devastating banana and plantain crops in East Africa and threatens the food security of millions. To understand the

evolution and geographical spread of this newly emerging pathogen, there is need for molecular markers such as SNPs. Given the high degree of genome sequence identity among isolates (99.9985 %), genome-wide sequencing is the only tractable way of discovering sequence polymorphisms and has been applied to bacterial phytopathogens (Cai et al., 2011; Kumar, et al., 2012).

The high degree of sequence similarity among isolates indicates a very recent origin of the pathogen. However it can only able to know for a fact that genes are homologous only if we could directly explore their common ancestor and all intermediate forms. Under the notion of homology, a sequence or structural alignment becomes a powerful tool for evolutionary and functional inferences. Homology lends legitimacy to the transfer of functional information from experimentally characterized proteins (or nucleic acids) to uncharacterized homologs, the single most common and practically important application of computational methods in molecular biology. The molecular markers discovered here enabled the reconstruction of the phylogenetic relationships between *Xcm* isolates from diverse geographical locations within the known range of the pathogen in Africa. Interestingly, the isolates fell into two major sub-lineages. This may indicate that there have been at least two separate introductions of *Xcm* into the banana-growing regions around Lake Victoria. This contrasts with the widely held working assumption that *Xcm* spread from Ethiopia to Uganda and thence subsequently into neighboring countries. This view is largely based on the fact that *Xcm* was reported in Ethiopia in the late 1960s, then Uganda in 2001 and only later in other African countries (2004: D. R. Congo, 2005: Rwanda and Tanzania, 2006: Kenya and Burundi). However, it is possible that the disease has existed for some time before being officially reported. An alternative hypothesis is that all outbreaks in the region can be traced back to a single introduction of inoculum that contained some genetic diversity and that some genetic diversity is maintained within the population. Under that scenario, the results could be explained by stochastic effects of sampling error given such a small number of isolates. No extensive biogeographical or divergence dating analytic approaches have been so far applied to available genomes. Therefore, there is a pressing need to collect and genotype many more isolates, including multiple isolates from within single outbreaks. On the other hand, where sequencing of multiple isolates from a single geographical area has been done (the five isolates from central Uganda and the three isolates from North Western Tanzania), only a single sub-

lineage was observed. This would be an unlikely outcome if both sub-lineages were approximately equally abundant at these sites. Therefore, currently the multiple-introduction model until further isolates and genetic data are available is more favored. This is possible due to the clear differences in the banana cultivars grown, management of gardens, combination of crops and overall ecosystem divergence in Eastern Africa. To clarify this further molecular dating of the *Xcm* and *Xvv* strains in the region will need to be conducted. One important aspect in comparative genomics is the set of genes associated with evolution of different lineages, so that their biological importance through (relative or absolute) time across clades can be inferred. Accurate inference of branch lengths is essential to molecular dating and is governed by the type and amount of data used, the phylogenetic technique chosen, and the selection of an appropriate model of molecular evolution.

Chapter Five

Genotypic diversity of sequenced *Xanthomonas campestris* pv. *musacearum* isolates and *Xanthomonas vasicola* pv *vasculorum* strains

Abstract:

Xanthomonas vasicola pathovar *vasculorum* (*Xvv*) is the bacterial agent causing gumming disease in sugarcane and maize. Molecular sequence analyses have recently confirmed that strains formerly classified as *X. campestris* pv. *vasculorum* fall within two major phylogenetic groupings that correspond to two distinct species: *X. axonopodis* and *X. vasicola*. Characterization studies of Xcm isolates through Fatty Acid Methyl Ester (FAME) analysis, RepPCR and gyraseB sequencing have revealed that Xcm is phylogenetically closely related to strains of *X. vasicola* pathovars. This study compared the complete genome sequences for five isolates of *Xvv* originating from sugarcane and one from maize. The sequence analyses presented here revealed large differences in gene content among isolates of Xvv, both among isolates from sugarcane and also between isolates from sugarcane and a single isolate from maize. Some of these differences in gene content are ascribed to the gain and loss of plasmids, though many of the differences are likely to be chromosomally located. The study also reports genetic differences implicated in important extracellular structures, such as lipopolysaccharide (LPS) synthesis, the TFP and candidate substrates (i.e., effectors) of the TTSS. Further analysis of previously published data revealed some hitherto undetected differences in gene content among Xcm isolates, including a homologue of TTSS effector XopL and a plasmid. This identified two distinct types of lipopolysaccharide synthesis gene clusters among *Xvv* isolates. Four of six *Xvv* isolates harbored sequences similar to the *Xac* plasmid, pXAC47, and showed a distinct Type IV pilus (TfP) sequence type, whereas the TFP locus of the other two isolates resembled that of the closely related banana pathogen, *Xanthomonas campestris* pathovar *musacearum* (*Xcm*). The *Xvv* isolate from maize has lost a gene encoding a homologue of the virulence effector, XopAF, which was present in all five of the sugarcane isolates, while *xopL* contained a premature stop codon in four out of six isolates.

Keywords: gumming disease; sugarcane; banana *Xanthomonas* wilt

5.1 Introduction

The bacterial genus, *Xanthomonas*, includes many economically important pathogens of crop plants (Meyer & Bogdanove, 1995; Parkinson et al., 2007; Rademaker et al., 2000; Schneider et al.,2011). Genome sequencing of species and pathovars of *Xanthomonas* has led to insights into the evolution and mechanisms of virulence and their ability to overcome host defenses (Chan, Sherman, & Bourke, 2006; Hajri et al., 2012; Midha & Patil, 2014; Pieretti et al., 2009; Simonson et al., 2005; Thieme et al., 2005; Wroblewski et al., 2009; F. Zhang et al., 2013). One pathovar whose study at the molecular level has been relatively limited is *Xanthomonas vasicola* pathovar *vasculorum* (*Xvv*).

Strains of *Xvv* are responsible for gumming disease in sugarcane and are also found naturally infecting some other monocotyledonous hosts, including maize and sorghum (Aritua et al., 2008; Champoiseau et al., 2006; Harrison & Studholme, 2014; Vandroemme et al., 2013). In addition to the importance of gumming disease and *Xvv per se*, strains of this pathovar share a recent common ancestor with the causal agent of banana *Xanthomonas* wilt, namely *Xanthomonas campestris* pathovar *musacearum* (*Xcm*), with which they share an identical *gyrB* DNA sequence (Aritua et al., 2008). Strains of *Xvv* do not cause disease in banana, but *Xcm* can infect maize and sugarcane.

The narrow genetic diversity in *Xcm* (Aritua et al., 2008; Odipio et al., 2009) suggests a population bottleneck, and *Xcm* may represent a single clonal lineage of the species, *X. vasicola*, a closely related species, that has recently emerged to colonize bananas in east Africa. Whole-genome sequence data can provide greater phylogenetic resolution than analysis of a single gene fragment. Although all isolates shared identical *gyrB* amplicon sequences, analyses of genome-wide sequence data revealed that *Xcm* and *Xvv* comprise two distinct clades, although they are closely related to each other (Wasukira et al., 2012). This suggests that *Xcm* did not arise from *Xvv*, but rather from some currently unknown close relative of *Xvv*, possibly another pathovar belonging to the same species (*X. vasicola*).

Given the close phylogenetic relationship between *Xcm* and *Xvv*, molecular comparison between *Xcm* and *Xvv* is of importance for understanding the evolution of *Xcm* and the adaptation to banana. To better understand the range of diversity within *Xvv*, analysis of available genome sequence data for four isolates of *Xvv* was conducted (described in previous publications (Studholme et al., 2010; Wasukira et al., 2012) and new sequence data from a further two isolates of *Xvv*.

There has been some confusion in the literature regarding the taxonomy of the *Xvv* strains, whose genome sequences are analyzed here. Ivey et al., (2010) listed strains NCPPB 702, NCPPB 1326 and NCPPB 206 as *X. axonopodis* pathovar *vasculorum.* In contrast (Qhobela & Claflin, 1988; Qhobela & Leach, 1991) classified strain NCPPB 1326 as *X. campestris* pathovar *vasculorum*. However, Vauterin et al., (1995) proposed that these strains be included in the species, *X. vasicola* (as pathovar *vasculorum*), on the basis of genomic DNA hybridization studies. A subsequent analysis of cellular fatty acid profiles also confirmed a close affinity between these strains and strains of the species, *X. vasicola*, and only rather distant similarity to species *X. axonopodis* and *X. campestris.* Rademaker et al., (2000) also list strain NCPPB 206 (synonymous with LMG 8284) as belonging to species *X. vasicola* (pathovar *vasculorum*, *i.e.*, *Xvv*). Overall, the available evidence overwhelmingly supports the inclusion of these isolates in species *X. vasicola*, and therefore, in this study, they are referred to as *Xvv* (rather than as *X. campestris* or *X. axonopodis*).

Studholme et al., (2010) previously generated draft genome sequences for one isolate of *Xcm* and one isolate of *Xvv.* They identified several differences between the two isolates that might contribute to their distinct host ranges. Differences were found between *Xcm* NCPPB 4381 and *Xvv* NCPPB 702 in their repertoires of Type III secretion system (TTSS) effectors, lipopolysaccharide (LPS) synthesis genes and Type IV pilus (TfP) genes. It should be noted that these differences were between just a pair of isolates, and so, it was unclear how generalizable these results were to other isolates of *Xcm* and *Xvv*. In a more recent publication, it was reported that sequencing the genomes of a further 13 isolates of *Xcm* and three isolates of *Xvv* (Wasukira et al., 2012). The main focus of the Wasukira et al., (2012) study was genetic variation among isolates of *Xcm*, revealing two main phylogenetic groups (or sub-lineages) among *Xcm*. That previous study reported a list of genes that were consistently conserved in *Xcm* and absent from the four *Xvv* genome sequences then available, and the study used data from the four *Xvv* genome sequences to generate a phylogenetic tree; however, no further analysis of the *Xvv* genome sequences was reported.

The current study extends the previous work by systematically searching for genetic differences among *Xvv* isolates, rather than focusing on variation between *Xvv* and *Xcm* or variation among isolates of *Xcm*. Furthermore, included in the current study are sequence data from two additional isolates of *Xvv*, bringing the total number of sequenced *Xvv* isolates up to six.

5.2 Materials and methods

5.2.1 Bacterial Isolates and Preparation of Genomic DNA

Bacterial isolates were obtained from the National Collection of Plant Pathogenic Bacteria (NCPPB) at The Food and Environment Research Agency, UK (FERA). DNA preparation, bacterial isolates were grown overnight at 28 °C in 10 mL King Broth shaken at 200 rpm. Bacterial cells were harvested by centrifugation and re-suspended in TE buffer (50 mM Tris-HCl, 40 mM EDTA, pH 8.0) containing 12 µL of 20 mg/mL lysozyme and 10 mg/mL RNase and incubated at 25 °C for 10 min with 17 µL 10% sodium dodecyl sulphate, then incubated on ice for 5 min. Proteins were dissolved with 170 µL of 8 M ammonium acetate, vortexed vigorously for 30 s centrifuged at 4 °C and for 15 min. DNA was precipitated with isopropanol and re-dissolved in 100 µL of 10 mM Tris, pH 8.0, and 1 mM Na2EDTA.

5.2.2 Genome Sequencing

DNA library preparation and genome sequencing using the Illumina GA2x were performed using standard Illumina protocols (Illumina, 2012). The Illumina GA2x platform was followed to sequence genomes of *Xvv* isolates NCPPB 895 and 890. This generated paired sequence reads of length 67 nucleotides.

5.2.3 Alignment of Sequence Reads Against Reference Genome Sequences

BWA was used to align GA2x sequence reads against a reference genome sequence (Li & Durbin, 2009). IGV (Robinson et al., 2011) was used to visualize the alignments and SAMtools (Li *et al.*, 2009) to manipulate the alignments and convert between formats.

5.2.4 SNP Calling and Phylogenetic Analysis

A very conservative approach to infer SNPs from the alignments of Illumina reads against the previously published *Xoo* reference draft genome assembly approach was used. To avoid false positives and false negatives, only used those regions of the *Xoo* genome with a coverage depth of 10 or more for every sequenced *Xcm* and *Xvv* genome and where there was at least 95% consensus among the sequence reads within each isolate. Just over 30% of the length (1,507,606 out of 4,940,217 nucleotides) of the *Xoo* genome fulfilled these two criteria. In other words, for 30% of the *Xoo* chromosome, there was sufficient quantity and consistency in the data to be almost certain of the

sequence in all of the eight isolates (six *Xvv* and two *Xcm*; see fig. 5.1). For the remaining 70% of the *Xoo* genome, there was some degree of ambiguity in the data for one or more of the isolates. The phylogenetic tree was inferred using the Maximum Parsimony method implemented in MEGA6 based on 39,665 single-nucleotide variants with respect to the chromosome of *Xoo* MAFF 311018. Bootstrap values were calculated as percentages of 500 trials (Tamura et al., 2013). The aim was to acquire basic evolutionary understanding of the *Xvv* isolates so as to mine for variation with the sequenced *Xcm* isolates. This method is best suited for sequences that are quite similar and is limited to small numbers of sequences.

5.2.5 Genome Assembly

De novo assembly of Illumina sequence reads was performed using Velvet 1.1.04 (Zerbino & Birney, 2008). Any sequence reads that contained one or more "N" prior to assembly were discarded. It is difficult or impossible to predict the optimal parameter values for Velvet assembly. Therefore, assemblies were generated using a range of combinations of hash length and coverage cut-off and chose the assemblies giving the largest N_{50} values. The N_{50} is defined as the minimum contig length needed to cover 50% of the genome. It means, half of the genome sequence is in contigs larger than or equal the N_{50} contig size. Or, that the sum of the lengths of all contigs of size N_{50} or longer contain at least 50 percent of the total genome sequence. For the newly presented assemblies (*i.e.,* for *Xvv* 890 and 895), the parameter values were for *Xvv* 895: hash length = 25 and coverage cut-off = 2 and for *Xvv* 890: hash length = 29 and coverage cut-off = 3.

5.2.6 Identification of Presence and Absence of Genes

BEDtools (Quinlan & Hall, 2010) was used to infer the breadths of coverage for genomic features based on Binary Alignment Map (BAM) files from the Burrows-Wheeler Aligner (BWA) and General Feature Format (GFF) files from Rapid Annotation using Subsystem Technology (RAST)(Aziz et al., 2008). The Heatmap package in R was used to generate Heatmaps (Kolde, 2015). However, it should be noted that incompleteness of *de novo* assemblies do not invalidate these, since our inferences are based upon comparisons of alignments of raw reads rather than comparisons between assemblies; these alignments consist of unassembled read-pairs aligned with BWA against various reference genome sequences. The single exception is fig. 5.2, in which the comparison consists of alignments between *de novo* assemblies; it is possible that some gaps in the alignments in fig. 5.2 could arise through the

incompleteness of the *de novo* assemblies. The figs 5.4 and 5.5 are based on BWA alignments between de novo assemblies, but the findings were also validated by the inspection of alignments of raw (unassembled) sequence reads.

5.2.7 Visualization of Genome-Wide Patterns of Sequence Conservation

BLASTN (Altschul et al., 1990; Metzker, 2008) was used to align assembled sequences and visualized the alignments using the Artemis Comparison Tool (Carver et al., 2005) and BLAST Ring Image Generator (BRIG) (Alikhan et al., 2011), which is a wrapper for Circular Genome Viewer (CGView) (Grant & Stothard, 2008; Stothard & Wishart, 2005)

5.2.8 Identification of Potential plant-inducible promoter (PIP) Boxes

A profile hidden Markov model (HMM) based on a multiple sequence alignment of 22 known PIP boxes from *X. vesicatoria* (from Table 3 in (Weber & Koebnik, 2006) using hmmb from the HMMER 1.8.5 package (Eddy, 2003). The DNA sequence was scanned against this profile-HMM using hmmls from HMMER 1.8.5 with a bit-score cut-off of 10.0.

5.3 Results and Discussion

By comparison of genome sequence data, several likely important events in the evolutionary history of *Xvv* and *Xcm* were identified; these are summarized in fig. 5.1 and include the acquisitions of plasmids and the exchange of genes encoding LPS biosynthesis and TFP, as well as the loss and gain of candidate TTSS effector genes. The evidence supporting the proposal of each of these events is presented figs. 5.2–5.6. The main findings were that: a plasmid similar to *pXAC*47 is found in a clade of four *Xvv* isolates from sugarcane ; the most ancestrally branching *Xvv* sugarcane isolate (NCPPB 895) may contain a plasmid similar to that from a cassava pathogen ; some (but not all) *Xcm* isolates harbor sequences similar to that from a plasmid in a cotton pathogen; there is considerable variation in TFP genes among *Xvv* isolates ; there are two distinct sequence types of the LPS biosynthesis gene cluster among *Xvv* isolates; and *Xvv* isolates vary with respect to their repertoires of putative TTSS effector genes.

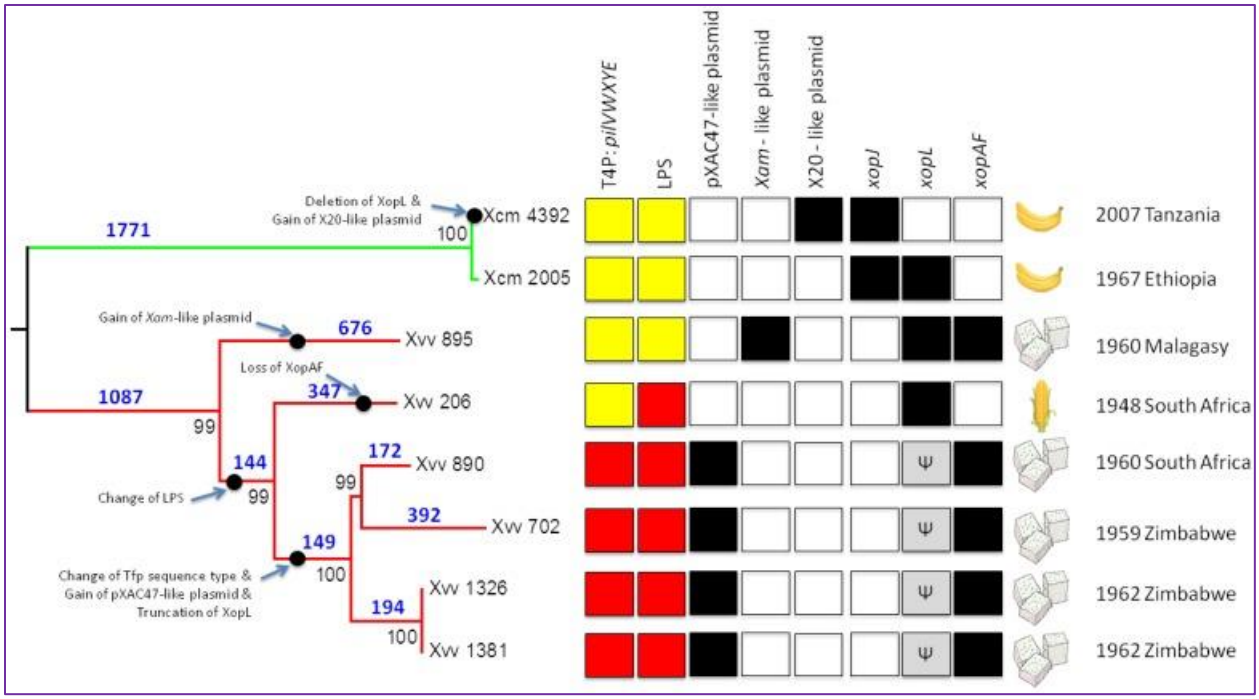

Figure 5.1: Overview of some key genetic changes during the evolution of Xvv and Xcm. The tree is drawn to scale with branch lengths calculated using the average pathway method. For clarity, the numbers of single-nucleotide differences are indicated on the branches in blue boldface. The square boxes indicate presence (black) or absence (white). The grey boxes marked "Ψ" denote that the xopL gene is interrupted by a premature stop codon. The two distinct types of the TFP pilVWXYE gene cluster (fig. 5.4) are indicated, respectively, by yellow and red shading. The two different sequence types of the lipopolysaccharide (LPS) gene cluster (fig. 5.6) are indicated, respectively, by yellow and red shading. The illustrations on the right-hand side (right) indicate the origin of each bacterial isolate: banana/ensete, sugarcane or maize.

5.3.1 Overview of Sequence Data

Table 5.1 gives a brief description of the bacterial isolates from which the sequencing data used in this study originate. Table 5.2 lists summary statistics for the raw sequence data and Table 5.3 summarizes the *de novo* assembly statistics. The sequence data from *Xcm* and for one *Xvv* isolate (*Xvv* 702) were described in previous publications (Studholme et al., 2010; Wasukira et al., 2012). The sequence data for four of the *Xvv* isolates were mentioned in a previous publication (Wasukira et al., 2012), but assembly statistics were not given. Therefore, here provided are details of the assemblies here (Table 5.3). The contiguity of the assemblies for the two newly presented isolates (*Xvv* 890 and *Xvv* 895) are much lower than those of the previously presented assemblies (see N_{50} in Table 5.3). The single exception is fig. 5.2, in which the comparison consists of alignments between *de novo* assemblies; it is possible that some gaps in the alignments in fig. 5.2 could arise

through the incompleteness of the *de novo* assemblies. The fig. 5.4 and 5.5 are based on alignments between *de novo* assemblies, but the findings were also validated by the inspection of alignments of raw (unassembled) sequence reads.

Five out of the six *Xvv* isolates were originally isolated from sugarcane; the exception is 206, which was isolated from maize. Genomes of *Xvv* 890 and 895 were newly sequenced for this study. Genome sequences of *Xvv* 206, 1326 and 1381 were reported in a previous publication (Wasukira et al., 2012), but with only very limited analysis, since the focus of that study was on single-nucleotide polymorphism (SNP) in *Xcm*. The genome sequence of *Xvv* 702 was previously reported and compared against that of *Xcm* 4381(Studholme et al., 2010). Genome sequence data from several isolates of *Xcm* that have previously been published were also included (Studholme et al., 2010; Wasukira et al., 2012), because these are the sequenced genomes most closely related to *Xvv* and probably belong to the same species, *X. vasicola* (V Aritua et al., 2008).

Table 5.1: Bacterial isolates and sequence datasets used in this study

Isolate (NCPPB number)	**Source and date of isolation**	**Source**
Xvv 206 [a]	South Africa 1948	*Zea mays*
Xvv 702 [b]	Zimbabwe 1959	*Saccharum officinarum*
Xvv 890 [c]	South Africa 1960	*S. officinarum*
Xvv 895 [c]	Malagasy Republic 1960	*S. officinarum*
Xvv 1326 [a]	Zimbabwe 1962	*S. officinarum*
Xvv 1381 [a]	Zimbabwe 1962	*S. officinarum*
Xcm 2005 [a]	Ethiopia 1967	*Ensetee ventricosum*
Xcm 2251 [a]	Ethiopia 1969	*Musa* sp.

Xcm 4379 [a]	Uganda 2007	*Musa* sp.
Xcm 4380 [a]	Uganda 2007	*Musa* sp.
Xcm 4381 [b]	Uganda 2007	*Musa* sp.
Xcm 4383 [a]	Uganda 2007	*Musa* sp.
Xcm 4384 [a]	Uganda 2007	*Musa* sp.
Xcm 4387 [a]	D. R. Congo 2007	*Musa* sp.
Xcm 4389 [a]	Rwanda 2007	*Musa* sp.
Xcm 4392 [a]	Tanzania 2007	*Musa* sp.
Xcm 4394 [a]	Tanzania 2007	*Musa* sp.
Xcm 4395 [a]	Tanzania 2007	*Musa* sp.
Xcm "Kenyan" [d]	Kenya (year not known)	*Musa* sp.

[a] These sequence data were previously reported in (Wasukira et al., 2012). [b] These sequence data were previously reported in (Studholme et al., 2010). [c] These sequences were newly generated for this study. [d] This sequence assembly was submitted to GenBank by the International Institute of Tropical Agriculture in 2011, but no accompanying manuscript has been published to the best of our knowledge. NCPPB.

Table 5.2: Summary of sequence data for all sequenced Xanthomonas strains in this study

Isolate (NCPPB number)	Number of read-pairs	Read length	Coverage	SRA accession
Xvv 206 [a]	2,579,404	76	70 x	SRR494500.3
Xvv 702 [b]	2,913,785	36	35 x	SRR020202.3
Xvv 890 [c]	4,843,028	67	58 x	SRR1045340
Xvv 895 [c]	2,867,513	67	50 x	SRR1045341
Xvv 1326 [a]	2,365,912	76	63 x	SRR494491.5
Xvv 1381 [a]	2,450,234	76	66 x	SRR494499.3
Xcm 2005 [a]	2,536,030	76	72 x	SRR489154.7
Xcm 4379 [a]	3,652,875	76	102 x	SRR494484.2
Xcm 4380 [a]	4,069,509	76	113 x	SRR494485.2
Xcm 4381 [b]	5,052,905	36	56 x	SRR020203.3
Xcm 4384 [a]	1,976,797	76	55 x	SRR494488.2
Xcm 4392 [a]	2,554,161	76	72 x	SRR494498.3
Xcm 4394 [a]	3,295,956	76	92 x	SRR494489.1
Xcm 4395 [a]	4,117,662	76	117 x	SRR494490.2

[a] These sequence data were previously reported in (Studholme et al., 2010). [b] These sequence data were previously reported in (Wasukira et al., 2012). [c] These sequences were newly generated for this study.

Table 5.3: Summary statistics for de novo genome sequence assemblies

Isolate (NCPPB number)	Number of Scaffolds	Scaffolds N_{50}	Total length (b.p.)	GenBank accession
Xvv 206 [a]	177	103,376	4,825,935	AKBM00000000.1
Xvv 702 [b]	97	205,000	5,478,002	ACHS00000000.1
Xvv 890 [c]	2,565	3,627	4,951,053	AKBN00000000.1
Xvv 895 [c]	1,229	8,362	4,803,807	AKBO00000000.1
Xvv 1326 [a]	253	74,455	4,951,570	AKBK00000000.1
Xvv 1381 [a]	137	108,088	4,958,101	AKBL00000000.1
Xcm 2005 [a]	156	61,233	4,692,764	AKBE00000000.1
Xcm 4379 [a]	95	147,554	4,758,198	AKBF00000000.1
Xcm 4380 [a]	87	149,435	4,751,644	AKBG00000000.1
Xcm 4381 [b]	115	143,688	4,793,900	ACHT00000000.1
Xcm 4384 [a]	84	157,780	4,741,777	AKBH00000000.1
Xcm 4392 [a]	162	90,974	4,728,564	AKBI01000000.1
Xcm 4394 [a]	85	151,917	4,792,617	AKBJ00000000.1
Xcm "Kenyan" [d]	510	36,401	4,907,936	AGFQ01000000.1

[a] These sequence data were previously reported in (Wasukira *et al.*, 2012). [b] These sequence data were previously reported in (Studholme *et al.*, 2010). [c] These sequences were newly generated for this study. [d] This sequence assembly was submitted to GenBank by the International Institute of Tropical Agriculture in 2011, but no accompanying manuscript has been published to the best of our knowledge.

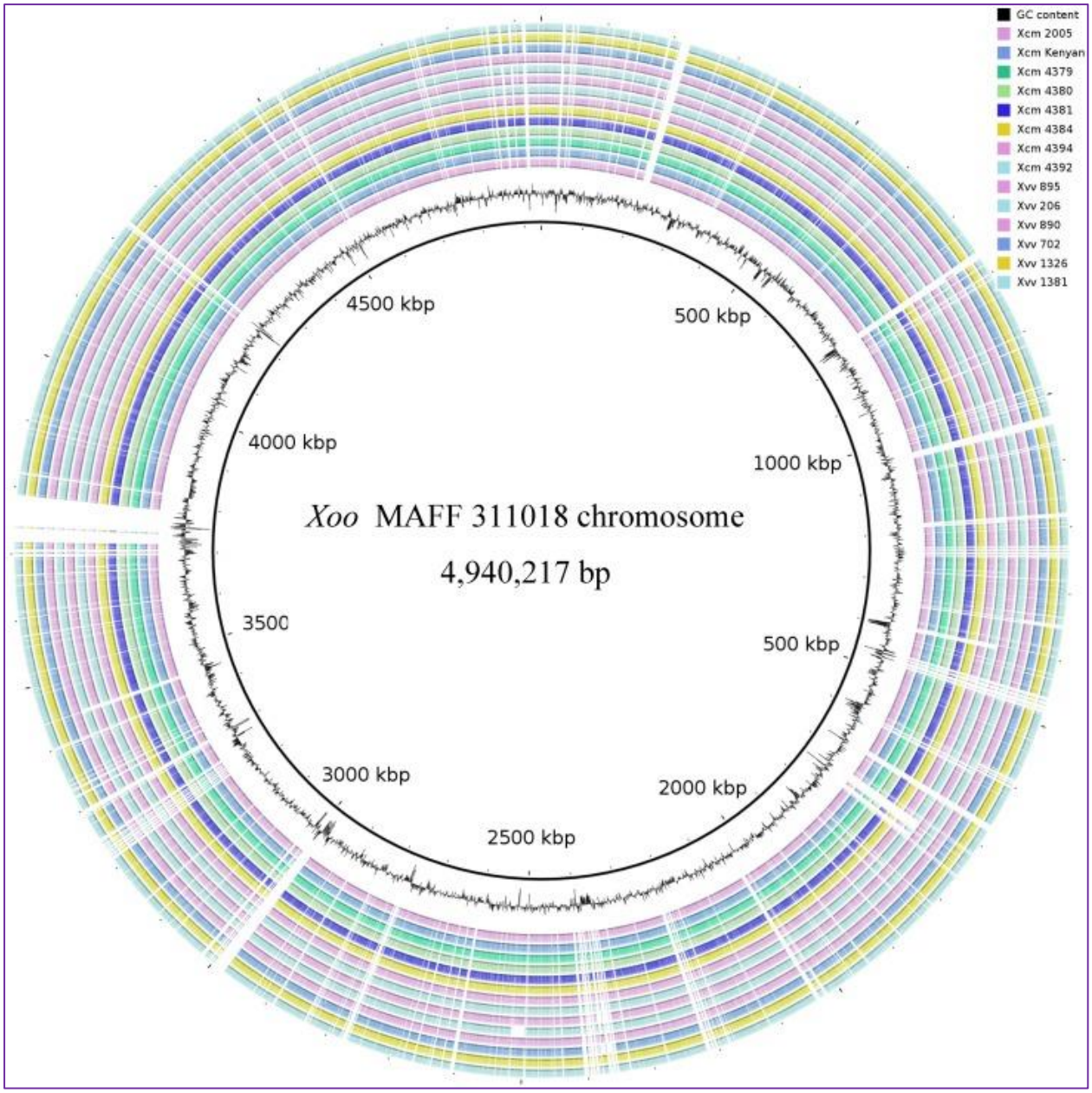

Figure 5.2: Global comparison of Xvv and Xcm genomes with reference to the Xoo chromosome sequence. Alignment of each of the genome sequence assemblies against the Xoo MAFF 311018 chromosome sequence. The innermost ring indicates the genomic position. The next ring is a plot of G + C content. The remaining 14 concentric rings indicate the presence or absence of BLASTN hits at that position, with one ring corresponding to each of 14 Xcm and Xvv genome assemblies. Positions covered by BLASTN alignments are indicated with a solid colour; whitespace gaps represent genomic regions not covered by the BLASTN alignments.

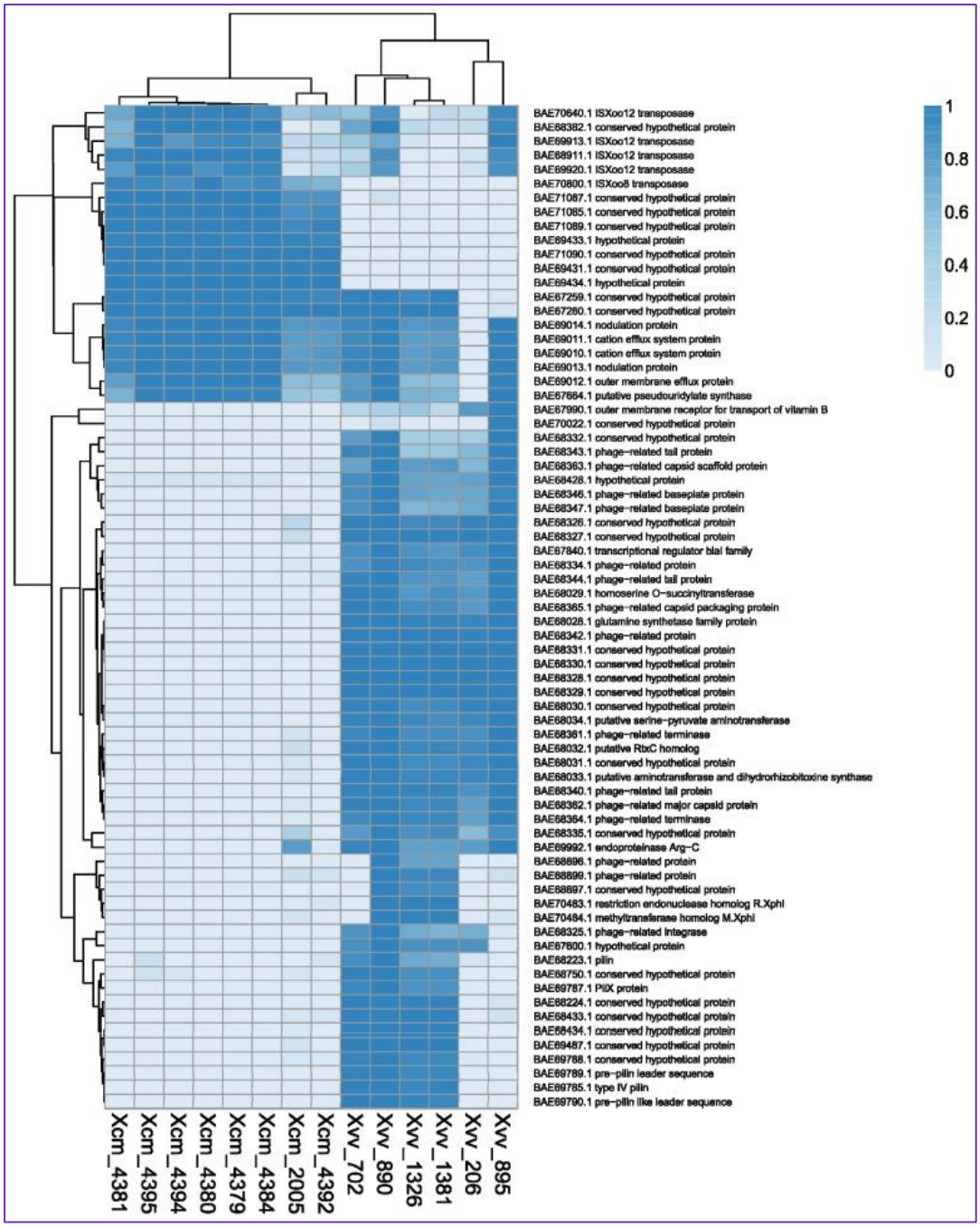

Figure 5.3: Heatmap depicting Xvv and Xcm gene-contents with reference to the Xoo chromosome sequence.

In the chromosome of *Xoo* MAFF 311018, this cluster falls between Positions 3436570 and 3443633 and includes locus tags XOO3030 to XOO3035. It includes the accessions BAE6788, BAE6789, BAE69785, BAE69787 and BAE69790 that appear in the Heatmap in fig. 5.3; accession numbers: AP008229.1 (*Xoo* MAFF 311018) and AEQV01000175.1 (*X. vesicatoria* ATCC 35937). Note, however, that the orientation of the BLASTN alignment (and hence, red

or blue coloring) is not biologically significant, but rather just reflects the direction in which the double-stranded genome sequence has been represented in the database. Positions of annotated genes are indicated on the genome sequence of *X. vesicatoria* ATCC 35937. Note that *Xvv* 890, *Xvv* 702, *Xvv* 1326 and *Xvv* 1381 share sequence similarity along the whole length of this genomic region, and for most of the region, they also share sequence similarity with *Xoo* MAFF 311018. They differ, over the central part of this genomic region, from *Xvv* 206, *Xvv* 895 and *Xcm* 2005, all of which show sequence similarity to *X. vesicatoria* ATCC 35937 over this region.

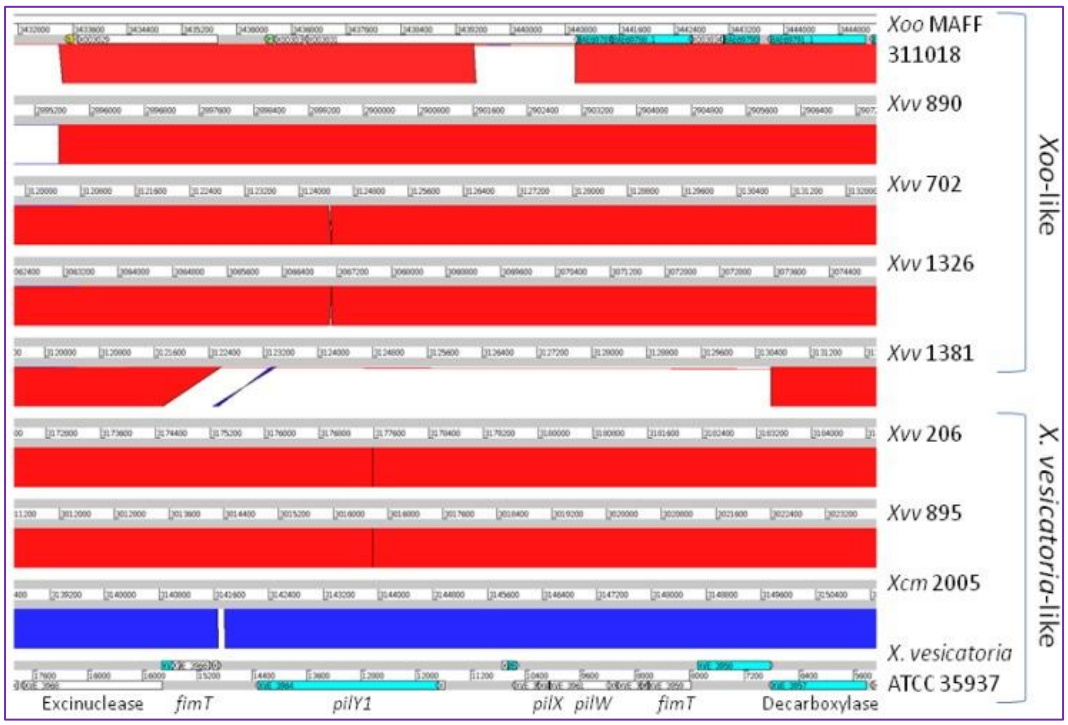

Figure 5.4: Two sequence types of the pilVWXYE gene cluster in Xvv: Xoo-like and X. vesicatoria-like. Each genome is represented by a horizontal white bar, labelled with the genomic position, flanked above and below by grey bars. Between each adjacent pair of genomes, BLASTN alignments are indicated by red blocks (for alignments between the positive strands of both sequences) or blue blocks (for alignments between the opposite strands of both sequences).

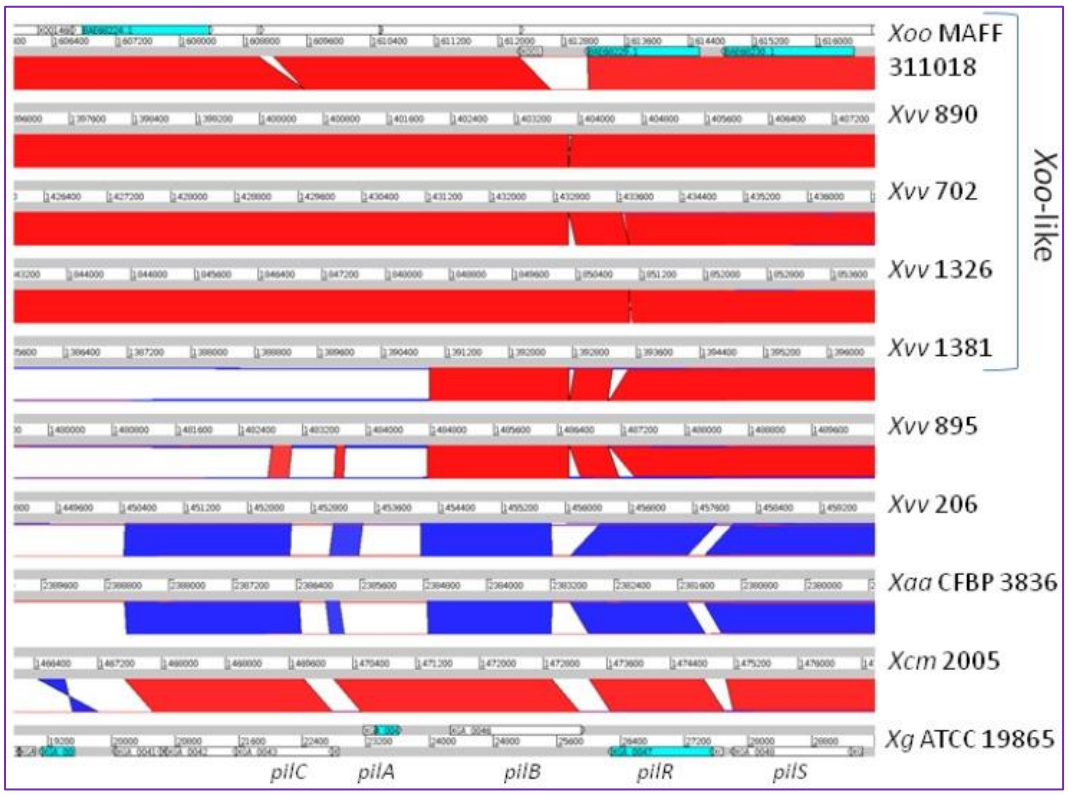

Figure 5.5: Multiple sequence types of the pilCABRS gene cluster in Xvv. Each genome is represented by a horizontal white bar, labelled with the genomic position, flanked above and below by grey bars. Between each adjacent pair of genomes, BLASTN alignments are indicated by red blocks (for alignments between the positive strands of both sequences) or blue blocks (for alignments between the opposite strands of both sequences)

In *Xoo* MAFF 311018, this cluster is located between Positions 1613866 and 1616597 and encompasses Loci XOO1467 to XOO1475; accession numbers: AP008229.1 (*Xoo* MAFF 311018); AUWN01000008.1 (*X. alfalfae* subsp. *alfalfae* (*Xaa*) CFBP 3836); and AEQX01000002.1 (*X. gardneri* (*Xg*) ATCC 19865). The orientation of the BLASTN alignment (and hence, blue or red colouring) is not biologically significant, but rather just reflects the direction in which the double-stranded genome sequence has been represented in the database. Positions of annotated genes are indicated on the genome sequence of *X. vesicatoria* ATCC 35937. Note that *Xvv* 890, *Xvv* 702, *Xvv* 1326 and *Xvv* 1381 share sequence similarity along the whole length of this genomic region, and for most of the region, they also share sequence similarity with *Xoo* MAFF 311018. However, the *pilC* and *pilA* genes are highly divergent among *Xvv* 895, *Xvv* 206 and *Xcm* 2005, with *Xcm* 2005 having *X. gardneri*-like *pilC* and *pilA* genes. *Xvv* 895 and *Xvv* 206 have *Xaa*-like *pilC* and *pilA* genes.

5.3.2 Phylogenetic Relationships among Xvv Isolates

A phylogenetic reconstruction of the sequenced *Xvv* and *Xcm* isolates based on single-nucleotide polymorphisms called against the reference sequence of the *X. oryzae* pathovar *oryzae* (*Xoo*) MAFF 311018 was generated (Salzberg et al., 2008). The maximum parsimony phylogenetic tree is shown in fig. 5.1; the maximum likelihood method produced identical topology, and the topology is consistent with the tree that was previously presented (Wasukira *et al.*, 2012). The *Xvv* isolates form a monophyletic clade closely related to, but distinct from, *Xcm* and the genetic distances among these sequenced *Xvv* isolates are considerably larger than those among *Xcm* isolates.

Within the sequenced *Xvv* isolates, the single isolate from maize falls within the diversity of sugarcane isolates; in other words, there are not separate monophyletic groups for isolates from the two different hosts. Two of the sugarcane-derived isolates (*Xvv* 1326 and 1381) are indistinguishable on the basis of the SNPs used to generate the tree in fig. 5.1. They had both been collected from sugarcane in Zimbabwe in 1962 and may be essentially two isolates of the same bacterial population. However, these two isolates are genetically distinct from the isolate collected from sugarcane in Zimbabwe three years earlier (*i.e.*, *Xvv* 702).

5.3.3 Global Genomic Comparison of Genomes of Xvv and Xcm Isolates

Alignment of our genome assemblies against a closely related reference chromosome sequence (*Xoo* MAFF 311018) suggested numerous differences in gene content both between *Xvv* and *Xcm* and also among *Xvv* isolates (fig. 5.2). Particularly noticeable in fig. 5.2 is a region of the *Xoo* genome (centered at position 2,508 Mb) that is absent from *Xvv* 206, the isolate from maize. This consists of an 18-kb region of the genome, including the eight loci XOO2253–XOO2263 (GenBank accession numbers: BAE69008.1–BAE69018.1) that includes several predicted efflux proteins of unknown function. The absence of this region in *Xvv* 206 was confirmed by the alignment of sequence reads against the *Xoo* reference genome, independently of any *de novo* assembly.

Xoo genes that are differentially present or missing in each of the *Xvv* and *Xcm* isolates were identified (fig. 5.3). Note that this analysis involved the alignment of raw sequence reads against the reference; it was not dependent on any *de novo* assembly of our sequence data. Also confirmed differential presence/absence by PCR amplification from genomic DNA (fig. 5.4). Breadths of coverage for each gene are indicated by the Heatmap in fig. 5.3, which reveals numerous genes, whose presence distinguishes *Xcm* from *Xvv* and also several that distinguish among *Xvv* isolates. However, this approach is limited to the analysis of genes that are in the chromosome of *Xoo*. It excludes plasmids, as well as chromosomal genes in *Xvv* or *Xcm* that are not conserved in *Xoo*.

Therefore, also performed similar analyses based on using genomic assemblies of *Xvv* and *Xcm* genomes instead of the *Xoo* chromosome sequence. Some of these differences in gene content are discussed in more detail below.

5.3.4 Plasmid sequences in Xvv Isolates NCPPB 890, 702, 1326 and 1381 and Isolate NCPPB 895 are variable

Evidence of plasmids in the *Xvv* and *Xcm* genomes were searched by aligning the Illumina reads against all bacterial plasmid sequences in the RefSeq database as of the October 31, 2013. The *Xvv* Isolates 890, 702, 1326 and 1381 all yielded Illumina sequence reads that cover about 70% of the length of plasmid pXAC47 from *Xanthomonas axonopodis* pathovar *citri* (*Xac*) 29-1 (the RefSeq accession number is NC_020798.1). The genomes of *Xvv* Isolates 890, 702, 1326 and 1381 all contain sequences with extensive similarity to pXAC47, while the other isolates of *Xvv* (and *Xcm*) do not. Plasmids are stably transmitted through generations via certain maintenance systems such as partition systems, multimer resolution, and toxin-antitoxin systems (Baxter & Funnell, 2014).

There is extensive sequence similarity between *Xvv* 895 and sequences from several recently published draft genome sequences of *Xam*, a pathogen of cassava (Ferreira-Tonin et al., 2012). These sequences are not found in the other sequenced *Xvv* isolates nor in the sequenced *Xcm* isolates. Specifically, a 34.7-kb *Xvv* 890 contig (GenBank: AKBO01000002.1) shared 99% nucleotide sequence identity with contigs from *Xam* isolates NG1, UA306, ORST17 and ORST X27 originally isolated from Nigeria, Columbia, Congo and Togo (the GenBank accession numbers for these *Xam* sequences are, respectively: AKEG01000263.1, AKEN01000114.1, AKEH01000160.1 and AKEJ01000091.1).

Some other strains of *Xam* also contained some slightly less-similar sequences, sharing up to 89% nucleotide sequence identity, including strains IBSBF725, IBSBF436, IBSBF285, IBSBF2820 and UA324. Most of the 41.4-kb sequence from *Xam* ORST X27 is conserved in *Xvv* 895, but absent from the other sequenced *Xvv* and *Xcm* isolates. This sequence is annotated as a (partial) plasmid (Bart et al., 2012), presumably because it contains several conjugative transfer genes. Overall, these observations indicate that *Xvv* 895 has acquired a plasmid that is closely related to plasmids that are circulating in *Xam* strains, suggesting the inter-species exchange of plasmids over a wide geographic range.

5.3.5 Xcm Isolates NCPPB 4379, 4380, 4383, 4384, 4392 and 4395 Contain Sequences Similar to a Plasmid from *X. citri pv. malvacearum* Strain X20

Previously, also found no evidence for the presence of plasmids in *Xcm* 4381 (Studholme *et al.*, 2010). However, examination of data from *Xcm* Isolates 4379, 4380, 4383, 4384, 4392 and 4395 revealed extensive sequence similarity to a recently reported plasmid sequence from *X. citri* pv. *malvacearum* strain X20, a highly virulent pathogen of cotton isolated in Burkina Faso (Cunnac et al., 2013, 2004; Midha & Patil, 2014). This 39-kb plasmid sequence (GenBank: CM002030.1) was not present in any of the sequenced *Xvv* isolates nor in any of the *Xcm* isolates belonging to Sub-lineage I (as defined by Wasukira *et al.*, 2012); it appears to be restricted to *Xcm* Sub-lineage II among the isolates of which it is present in all, except *Xcm* 4381. The most parsimonious explanation for this pattern of distribution is that this plasmid was acquired by the common ancestor of *Xcm* Sub-lineage II and subsequently lost in *Xcm* 4381. The GenBank accession number for the corresponding plasmid sequence in the *Xcm* 4384 *de novo* assembly is AKBH01000036.1. The average nucleotide sequence identity between the *X. citri* pv. m*alvacearum* plasmid and *Xcm* was 92%, somewhat lower than the 99% identity between *Xvv* 895 and *Xam* plasmid sequences.

5.3.6 Genetic Variation in the Type IV Pilus (TfP) Among Isolates of Xvv

The global analyses of gene content (see fig. 5.3) revealed differences among the *Xvv* isolates with respect to their TFP genes (figs. 5.4 and 5.5). For example, fig. 5.3 shows that *Xvv* Isolates 702, 890, 1326 and 1381 are distinguished from the other sequenced isolates by the presence of several TFP-related genes (including GenBank accessions: BAE68223, BAE6788, BAE6789, BAE69785, BAE69787 and BAE69790).

5.3.6.1 The *pilVWXYE* Gene Cluster

In *Xanthomonas* species, the TFP apparatus is encoded by clusters of genes scattered over several genomic locations, including two large clusters containing *pilVWXYE* and *pilCABRS*, respectively. The first of these two clusters falls between a gene encoding an excinuclease ABC subunit B and a gene encoding a decarboxylase-family protein. Two distinct sequence types at this locus among *Xvv* isolates were discovered. The corresponding gene cluster in *Xvv* 702 encodes homologues of FimT, PilV, PilW, PilX, PilY and PilE and is highly conserved (at least 99% identical nucleotide sequence) in *Xvv* Isolates 890, 1326 and 1381 and *Xoo* (92% identity). However, the corresponding gene cluster is quite different in the other two sequenced *Xvv* isolates, namely 206 and 895. These

two isolates share 99% nucleotide sequence identity with *Xcm* NCPPB 4381 and the other sequenced isolates of *Xcm* and 90% identity with *X. vesicatoria* ATCC 35937.

In summary, there are two distinct sequence types at this *pilVWXYE* locus among *Xvv* isolates:

(i) *Xoo*-like (in *Xvv* 890, 702, 1326 and 1381); and (ii) *X. vesicatoria*-like (*Xvv* 895, 206 and all *Xcm*); see fig. 5.4. The most parsimonious explanation for this pattern is horizontal transfer of the locus in a common ancestor of *Xvv* Isolates 890, 702, 1326 and 1381 after splitting from the *Xvv* 206 lineage (see fig. 5.1). A less parsimonious explanation would be multiple acquisitions of the same sequence.

5.3.6.2 The *pilCABRS* Gene Cluster

In addition to the *pilVWXYE* gene cluster described above, there is also variation at the *pilCABRS* gene cluster, with *pilA* (locus tag XOO1468 and accession number BAE68223 in *Xoo*) being particularly variable (fig. 5.5). The pattern of variation at this locus is similar to that at *pilVWXYE*, insofar as *Xvv* Isolates 890, 702, 1326 and 1381 have a *Xoo*-like sequence at this locus (94% nucleotide sequence identity between *Xvv* and *Xoo*), whereas *Xvv* Isolates 206 and 895 have a different sequence type.

The *pilA* genes are of different sequence types in *Xvv* 206, *Xvv* 895 and *Xcm*. The nucleotide sequence of *Xvv* 206 *pilC* and *pilA* is 87% identical to those of *Xcm*. This degree of sequence identity is significantly lower than for the core genome; most orthologous genes share at least 99% nucleotide sequence identity between *Xvv* and *Xcm*. Apart from *Xcm*, the next most similar sequence to *Xvv* 206 *pilC* and *pilA* comes from *X. alfalfae* subsp. *alfalfae* (*Xaa*) CFBP 3836 (79% identity). In *Xvv* 895, *pilB* is 95% identical to *Xvv* 206 and 94% identical to *Xcm*. However, at the *pilA* locus, there is further variation between *Xvv* 206, *Xvv* 895 and *Xcm*. The most similar sequence in the public databases to the *pilA* gene of *Xvv* 895 is *X. alfalfae* subsp. *alfalfae* CFBP 3836 (90% identity). The *pilA* of *Xcm* shares 94% nucleotide sequence identity with *X. gardneri* ATCC 19865 (Potnis et al., 2011) and shows no detectable nucleotide sequence similarity with any other sequence in the public databases.

The pattern of sequence variation in the *pilCABRS* gene cluster indicates multiple superimposed horizontal transfer events, resulting in *Xvv* having three distinct sequence types of *pilA*: (i) the *Xoo*-type in *Xvv* 702, 890, 1326 and 1381; (ii) the *Xvv* 895 type that is 94% identical to *Xaa*; and (iii) the *Xvv* 206 type that is 79% identical to *Xaa*. The *pilA* of *Xcm* belongs to a fourth, *X. gardneri*-like type.

5.3.6.3 Variation in TFP Genes

The TFP is a key virulence factor for phytopathogenic bacteria (Burdman et al., 2011; Qi et al., 2012; Taguchi & Ichinose, 2011; Wairuri et al.,2012). It performs a range of functions, including twitching motility (Burdman et al., 2011; Doorn et al., 2001; Hardoim et al., 2008; Jin et al., 2011; Shen et al., 2001) and cell-to-cell adhesion, thereby playing a role in the formation of micro-colonies and biofilms (Chagnot et al., 2013; Darsonval et al., 2008; Roine et al., 1997; Romantschuk et al., 2001; Zimaro et al., 2013), and PilA has been implicated in the transmission of the pathogen to seed (Darsonval et al., 2008). It is clear from these results that there have been several horizontal genetic transfers resulting in the replacement of TFP genes with alternative alleles in *X. vasicola*. It is not clear what functional significance, if any, arises from such allele exchanges, and there is no clear-cut correlation between TFP sequence type and host plant species. However, given the key role of the TfP in bacteria—plant interactions and the previously reported observation that some TFP genes are under selection in *Xanthomonas* species (Dunger et al., 2014; Giltner et al., 2012) this may warrant further investigation. It is also possible that a phage might exert a selective pressure on TfP; for example, some phages require the TfP to infect *Pseudomonas aeruginosa* (Giltner et al., 2012).

5.3.7 Distinct Sequence Types of LPS Biosynthesis Clusters in Xvv

The lipopolysaccharide (LPS) molecules that cover the outer membranes of gram-negative bacteria can be an important virulence factor in plant pathogens (Li & Wang, 2011; Yamashita, et al., 2014; Zheng et al., 2012; Zheng et al., 2012b). Horizontal transfer has led to hyper-variability among different strains within individual *Xanthomonas* species (Patil et al., 2007; Patil & Sonti, 2004).

Studholme *et al.*, (2010) showed that the LPS locus in *Xcm* 4381 most closely matches that of *X. axonopodis* pv. *citri* (*Xac*) strain 306 (da Silva *et al.*, 2002), whereas half of the LPS locus in *Xvv* 702 was not detectably similar to *Xcm* 4381, but rather resembled that of *X. albilineans* strain GPE PC73 (Studholme *et al.*, 2010, Pieretti *et al.*, 2009). Studholme *et al.*, 2010 and and subsequently Pieretti *et al.*, 2012 pointed out that this pattern of sequence similarity is incongruent with the close phylogenetic relationship between *Xcm* 4381 and *Xvv* 702 and indicates recent horizontal transfer in one or both strains.

Sequencing of additional isolates indicated that the LPS biosynthesis gene cluster DNA sequence is highly conserved among *Xcm* isolates (Wasukira *et al.*, 2012). However, in the

present study, additional genome sequencing revealed variation in this locus among isolates of *Xvv*. As illustrated in fig. 5.6, *Xvv* 895 shares 99% nucleotide sequence identity with isolates of *Xcm*, which, in turn, share 97% identity with *Xac* 306. However, in *Xvv* 206, 890, 1326 and 1381, the LPS cluster shares at least 99% identity to that of *Xvv* 702. Approximately one half of the LPS cluster in these *Xvv* isolates (adjacent to *etfA*) is 99% identical to that of *Xvv* 895 and *Xcm*. However, the other half (adjacent to *metB*) shares no detectable sequence similarity with *Xvv* 895, but it does share 84% identity with *X. sacchari* (Studholme *et al.*, 2011) and 86% with *X. albilineans* (Studholme *et al.*, 2010, Pieretti *et al.*, 2009).

As Pieretti and colleagues noted (2012), it is interesting that the *Xvv* 702-type LPS cluster is common to three distinct *Xanthomonas* species that all inhabit the xylem of sugarcane, and thus, there might be opportunities for these species to come into contact with each other and exchange genetic material. However, it is unlikely that this type of LPS is uniquely adapted to evading recognition as a pathogen-associated molecular pattern by sugarcane (Ranf, 2016), since *Xvv* 895, isolated from sugarcane, has a *Xac*-type LPS cluster, as does *Xcm*, which can also infect sugarcane (Aritua et al., 2008). Furthermore, *Xvv* 206 has the *Xvv* 702-type LPS cluster and was originally isolated from maize; this further emphasizes that there is not a clear correlation between LPS cluster type and host plant species. Drivers for variation in the LPS might be interactions with phages (Ahern, 2013; Debroy et al., 2004; Moreira et al., 2005; Webb et al., 2003) or with insect vectors (Rapicavoli et al., 2015).

In all Xanthomonads sequenced to date, the LPS cluster lies between the highly conserved etfA and metB genes; these are indicated by grey arrows (fig. 5.6). Pairwise nucleotide sequence identities between blocks of genes are given as percentages. Sequences from each of the Xcm sequences share at least 99% identity with Xcm NCPPB 2005 over the full length of the LPS cluster. The sequences from Xvv NCPPB 206, 890, 1326 and 1381 all share at least 99% identity with that of Xvv NCPPB 702.

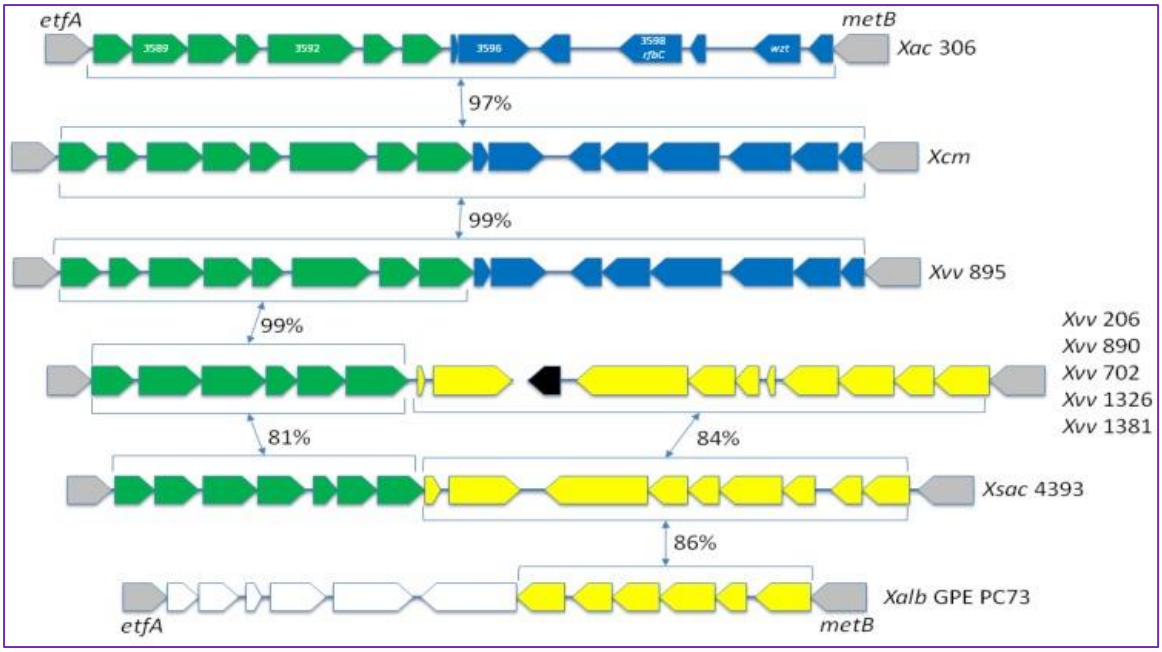

Figure 5.6: LPS biosynthesis gene clusters found in sequenced genomes of Xvv and related LPS clusters from other Xanthomonads. The arrows represent predicted genes and are colored according to blocks of sequence similarity. The black arrow indicates a gene encoding a mobile element protein that is not conserved in Xvv NCPPB 895 nor in X. sac NCPPB 4393.

The LPS cluster falls into two distinct scaffolds in the genome assembly for *Xvv* NCPPB 702. The scaffolds include: accession numbers: NC_003919.1 (Xac 306); AKBE01000022.1 (*Xcm* NCPPB 2005); AKBO01000066.1 (*Xvv* NCPPB 895); ACHS01000119.1 and ACHS01000380.1 (*Xvv* NCPPB 702); AGDB01000011.1 (*Xsac* NCPPB 4393); FP565176.1 (*Xalb* GPE PC73); AKBM01000022.1 and AKBM01000058.1 (*Xvv* 206); AKBN01000985.1, AKBN01000620.1 and AKBN01000538.1 (*Xvv* 890); AKBK01000238.1 and AKBK01000490.1 (*Xvv* 1381); and AKBL01000107.1 and AKBL01000264.1 (*Xvv* 1826).

5.3.8 Variability in repertoires of TTSS effector genes of Xvv Isolates

The repertoire of TTSS effectors can significantly influence a bacterial phytopathogens host range (Lemaire et al., 2009, Hajri *et al.*, 2012). Therefore, differences in TTSS repertoires between *Xcm* and *Xvv* are of great interest, as they might partly explain the ability of *Xcm* to cause disease in banana, whilst *Xvv* appears to be non-pathogenic in banana. A set of TTSS effectors encoded by *Xcm* 4381 were previously compared against the set encoded by *Xvv* 702 (Studholme *et al.*, 2010). In that previous study, it identified only a few differences.

Specifically, *Xcm* encodes two homologues of XopJ that are absent from the genome of *Xvv* 702, and *Xvv*702 encodes a homologue of XopAF that is absent from *Xcm* 4381 (Studholme *et al.*, 2010). However, in the present study, utilizing additional sequence data and found that XopAF is absent from the genome of *Xvv* 206, though it is present in *Xvv* 895, 890, 1326 and 1381, as well as in *Xvv* 702. It is absent from all sequenced isolates of *Xcm*.

5.3.8.1 Absence of *XopAF* gene from Xvv 206

It was previously reported that *Xvv* 702 encodes a homologue of XopAF (GenBank: ACHS01000051.1, bases 7184–7783; RefSeq: WP_010364039) that is not present in the genome of *Xcm* 4381 (Studholme *et al.*, 2010). This sequence shares 86% amino acid sequence identity with the XopAF (also known as AvrXv3) from *X. euvesicatoria* that was originally identified as an avirulence factor, inducing the hypersensitive response (HR) in resistant tomato and pepper plants (Astua *et al.*, 2000). It is also identical at the amino acid sequence level to proteins encoded by *X. translucens* pv. *translucens* (*Xtt*) DSM 18974 (RefSeq: WP_003475568). In the present study it has been found that this *xopAF* gene is present in *Xvv* 895, 890, 1326 and 1381, as well as in *Xvv* 702. It is not present in *Xcm* nor in *Xvv* 206. Thus, XopAF is encoded by *Xvv* isolates from sugarcane, but not by the *Xvv* isolate from maize and not by *Xcm* isolates from banana and ensetee.

This begs the question of whether *XopAF1* contributes to the limitation of host range in *Xvv*; that is, one might hypothesized that *XopAF* confers avirulence in banana and that the absence of *XopAF* in *Xcm* enables its pathogenicity in banana. In a recent comparative study of the genomes of different pathotypes of *X. citri* pv. *citri* (*Xca*) (Jalan *et al.*, 2013), the authors noted that XopAF was encoded in the genomes of a narrow-host range strain, but absent from a closely related broad-host range strain. Therefore, they hypothesized that XopAF might confer avirulence and contribute to the limitation of the host range. However, their mutational analysis showed that *XopAF* did not affect host range, but it did contribute to the ability of *X. citri* pv. *citri* pathotype A^w (*Xcaw*) to grow in a Mexican lime host plant.

It should be noted that the predicted XopAF protein in *Xcaw* (RefSeq: WP_007652722.1) is much more divergent from the originally described sequence from *X. euvesicatoria* (WP_008577605.1), sharing only 31% amino acid sequence identity, whereas the *Xvv* protein shares 86% identity with *X. euvesicatoria* XopAF. Therefore, the *Xvv* XopAF protein is likely to interact with plants differently. Furthermore, the host plants in question are very different

with *X. euvesicatoria* and *Xcaw* infecting dicots and *Xvv* and *Xcm* infecting monocots. However, it is reasonable to suppose that the *Xvv* protein is likely to be a TTSS effector and a potential avirulence factor, given the 86% identity between it and the experimentally characterized *X. euvesicatoria* XopAF (Astua-monge et al., 2005; Astua-monge et al., 2000). *XopAF* contains a DNA-binding domain at its C terminus and may allow the pathogen to manipulate its host by affecting the expression of plant genes. It would be interesting to test whether heterologous expression of *XopAF* in *Xcm* would cause avirulence in banana and whether deletion of *XopAF* in *Xvv* would have any impact on virulence in sugarcane or maize.

In *Xvv* 702, the *XopAF* gene falls between positions 7181 and 7543, on the reverse strand in GenBank accession ACHS01000051.1. This resides within a genomic region that also encodes several phage-associated proteins, including phage-related lytic enzyme, phage-tail protein, baseplate assembly protein J, phage-tail fibres and phage-tail fibre protein, and may result from the integration of a pro-phage into the *Xvv* genome. In the wheat pathogen, *Xtt* DSM 18974, the *xopAF* gene encoding an identical protein sequence is located in a different genomic context; it resides at positions 40277 to 40933 in GenBank accession CAPJ01000122.1 (locus tag: BN444_00905). This region of the *Xtt* genome does not contain any obvious phage-related genes, but does contain a predicted transposase for insertion sequence element IS629 (locus tag: BN444_00906), suggesting a mechanism for the mobility of this gene.

5.3.8.2 Truncation in gene encoding homologue of XopL in Common Ancestor of *Xvv* 890, 702, 1326 and 1381

The study also found polymorphism with respect to a *xopL*-like gene in *Xvv*. The genomes of *Xvv* isolates 206 and 895 each encode a protein with 70% amino acid sequence identity to XopL from *X. campestris* pv. *vesicatoria* (*Xcv*) strain 85-10 (RefSeq: YP_364951.1) (Noël et al., 2003; Thieme et al., 2005).The GenBank accession numbers for the *Xvv* 895 and 206 sequences are AKBO01000570.1 (Positions 1804–3768) and AKBM01000211.1 (Positions 29591–31555), where they encode a full length protein. However, in *Xvv* Isolates 890, 702, 1326 and 1381, there is a single-nucleotide C→A substitution resulting in a TCA codon being transformed to a TAA stop codon. This substitution occurs at Position 14191 in GenBank accession ACHS01000315.1 and is predicted to result in the protein being truncated to 265 amino acids (compared to the full length 654 amino acids).

This X*opL*-like sequence is completely absent from *Xcm* isolates belonging to Sub-lineage II

(*Xcm* 4379, 4380, 4381, 4383, 4384, 4392 and 4395). However, it is present in *Xcm* isolates belonging to Sub-lineage I (*Xcm* 2005, 2251, 4387 and 4389), where each genome encodes a full length protein. Thus, it appears that this X*opL* homologue may have been lost twice: (i) completely deleted in a common ancestor of *Xcm* Sub-lineage II; and (ii) truncated by a premature stop codon in a common ancestor of *Xvv* 890, 702, 1326 and 1381.

It has recently been demonstrated that *XopL* possesses E3 ubiquitin ligase activity, induces plant cell death and subverts plant immunity and that the ligase activity is associated with the C-terminal region of the protein (Singer et al., 2013). The premature stop codon found in *xopL* of some *Xvv* isolates has split the *XopL*-encoding open reading frame (ORF) into two.

Interestingly, there is a candidate PIP box (Büttner & Bonas, 2003; Jiang et al., 2009; Thieme et al., 2005; Wengelnik & Bonas, 1996) upstream of the second ORF, which corresponds to the C-terminal region of *XopL*, in which the E3 ubiquitin ligase resides. This suggests the hypothesis that this truncated ORF still has the potential to be expressed and be induced *in planta* and that it might still have biochemical activity, though it is unclear whether it would be a substrate of the TTSS. This potential PIP box (sequence TTCCGgcgaacatgcagcaaTTCGC) is located at positions 14107 to 14137 in ACHS01000315.1, which is approximately 160 bp upstream of the C-terminal XopL ORF at 14296 to 15366. There is another PIP box (sequence TTCGCtacgataaagatgacTTCGC) located at 13300 to 13347, which is approximately 50 bp upstream of the ORF homologous to the *XopL* N terminus located at 13395 to 14192.

5.3.8.3 Absence of Genes Encoding Homologues of XopJ Distinguishes *Xcm* from *Xvv*

In addition to the homologues of *XopAF* and *XopL*, a further two potential TTSS show differential presence/absence among our sequenced strains. Previous study reported that *Xcm* 4381 encodes two homologues of *XopJ* that are absent from *Xvv* 702 (Studholme *et al.*, 2010). Our subsequent analyses have confirmed that these are conserved in all the sequenced *Xcm* isolates and are absent from all the sequenced *Xvv*. Therefore, these predicted TTSS effectors are ideal candidates for molecular markers in host range between *Xvv* and *Xcm*.

5.4 Conclusions

Here, we analyzed draft genome sequences for two isolates of *Xvv* to augment the four previously published (Wasukira *et al.*, 2012, Studholme *et al.*, 2010) draft genome sequences of *Xvv*. Comparative analyses of these genome sequences and previously published genome

sequences of the closely related pathovar, *Xcm*, have revealed extensive differences in gene content among *Xvv*. This study described some of these differences in detail, including differences in plasmid content, LPS biosynthesis clusters, TFP and TTSS effectors. The main evolutionary events are summarized graphically in fig. 5.1. As well as providing some insight into evolutionary events within *Xvv*, these sequence analyses also further refine our understanding of the genomic differences between *Xvv* and the very closely related *Xcm*, which is a recently emerging pathogen in banana and ensetee; the availability of the sequence from multiple isolates allows us to distinguish between inherent variation within *Xvv* that might confound attempts to identify important genetic differences between the two pathovars and for functional analysis of important virulence factors.

It is clear that *Xcm* is genetically highly monomorphic (Wasukira *et al.*, 2012); here, it shows that, apart from several phage-related genes and the SNPs in the core genome described previously, the few genetic differences among *Xcm* isolates can be explained by the acquisition of a plasmid in Xcm Sub-lineage II, which is not present in Sub-lineage I, nor in at least one isolate of Sub-lineage II (*Xcm* 4381). Additionally, the two *Xcm* sub-lineages differ in that members of Sub-lineage II have lost a gene encoding a homologue of *XopL*; interestingly, this gene has acquired a premature stop codon in four of the six *Xvv* isolates, suggesting that isolates of both *Xcm* and *Xvv* have independently converged on eliminating *XopL*.

In contrast to the limited genetic diversity within *Xcm*, there is considerable diversity within *Xvv*, both at the level of SNPs in the core genome and at the level of gene content. Some of the differences in gene content are ascribable to the acquisition of two different plasmids (one in *Xvv* 895 and one in *Xvv* 890, 702, 1326 and 1381), but there are also differences in chromosomally located gene clusters, such as those encoding LPS biosynthesis and TFP.

Overall, this work suggests hypotheses for future work towards understanding the molecular basis for the ability of *Xcm* to emerge as an important pathogen of banana and ensete. One consistent difference is that all sequenced *Xcm* isolates encode two homologues of XopJ that are absent from all sequenced isolates of *Xvv*. In *X. campestris* pv. *vesicatoria*, this TTSS effector has been shown to interfere with salicylic acid-dependent defense responses to attenuate the onset of necrosis and to alter host transcription (Ustun *et al.*, 2013); it will be enlightening to test the contribution of the two XopJ homologues in *Xcm* interaction with banana and ensetee. Furthermore, understanding the emergence of Xcm will require the study of genome sequences of a wider range of isolates within the species, *X. vasicola*, to which *Xcm*

probably belongs (Aritua *et al.*, 2008); several isolates are available in strain collections for *X. vasicola* pv. *holcicola* (Aritua et al., 2008; Lang et al., 2017), but there is also a need to survey other as yet unknown members of the species that might inhabit the center of origin, perhaps colonizing other monocot plants.

Chapter Six

Variability of two banana associated Group 1 Clade Xanthomonads and an insect isolated *Xanthomonas sacchari*

Abstract:

This study presents draft genome sequences for three isolates of *Xanthomonas* species, each of which was associated with banana plants (*Musa* species) but is not closely related to the previously sequenced banana pathogen *Xanthomonas campestris* pathovar *musacearum*. Strain NCPPB4393 had been deposited as *Xanthomonas campestris* pathovar *musacearum* but in fact falls within the species *Xanthomonas sacchari*. Isolate NCPPB1132 is more distantly related to *Xanthomonas sacchari* whilst isolate NCPPB 1131 grouped in a distinct species-level clade related to *X. sacchari*, along with strains from ginger, rice, cotton and sugarcane. These three newly sequenced strains share many genomic features with the previously sequenced *Xanthomonas albilineans*, for example possessing an unsual *metE* allele and lacking the Hrp type III secretion system. However, they are distinct from *Xanthomonas albilineans* in many respects, for example showing little evidence of genome reduction. They also lack the SPI-1 type III secretion system found in *Xanthomonas albilineans*. Unlike *X. albilineans*, all three strains possess a *gum* gene cluster. The data reported here provide the first genome-wide survey of non-Hrp *Xanthomonas* species other than *Xanthomonas albilineans*, which is an atypical member of this group. The availability of complete sequence data for this group of organisms is the first step towards understanding their interactions with plants and identifying potential virulence factors.

Keywords: banana; *Xanthomonas*; genome; Clade 1; non-hrp

6.1 Introduction

The genus *Xanthomonas* includes some 27 species of plant-associated gram-negative bacteria. Collectively these species, and their constituent pathovars, cause disease on several hundred plant species, including many economically important crops. Phylogenetic analyses of the genus *Xanthomonas* consistently reveal that it comprises two distinct groups (Simões et al., 2007, Rademaker *et al.*, 2000, Parkinson *et al.*, 2007, Young *et al.*, 2008, Parkinson et al.,

2009). Young et al., (2008) recently proposed that Group 1 and Group 2 might represent two distinct genera.

Complete genome sequences are available for several members of Group 2, including *X. campestris* pv. *campestris*, *X. campestris* pv. *raphani*, *X. campestris* pv. *vesicatoria*, *X. citri*, *X. fuscans* subsp. *aurantifolii*, *X. oryzae* pv. *oryzae* and *X. oryzae* pv. *oryzicola* (Ryan et al., 2011). Draft genome sequences are available for several further members of the Group 2 *Xanthomonas* species (Ryan *et al.*, 2011). The genomes of all of the members of Group 2 investigated so far encode a type III secretion system (TTSS), known as the Hrp TTSS, which is required for delivery of effector proteins into the plant host in order to overcome the host's defenses. The name "Hrp" is derived from "hypersensitive response and pathogenicity".

Of the members of *Xanthomonas* Group 1, the only published genome sequence (Pieretti et al. 2009) is that of *X. albilineans*, a xylem-limited pathogen that causes leaf scorch in sugarcane (*Saccharum* species). Analysis of the *X. albilineans* genome sequence revealed that this species displays several interesting features that are quite distinct from those of the Group 2 species. For example, *X. albilineans* lacks the Hrp TTSS that is universally conserved and central to pathogenicity in Group 2, but it encodes an alternative non-Hrp TTSS that shares sequence similarity to the *Salmonella* SPI-1 TTSS (Rott et al., 2010). This raises the question of whether these features of the *X. albilineans* genome are also shared with the genomes of other Group 1 *Xanthomonas* species. Furthermore, *X. albilineans* appears to have undergone significant genome reduction, perhaps as a consequence of, or adaptation to, its xylem-limited lifestyle (Pieretti *et al.*, 2009). Therefore, it would be interesting to compare its genome with those of other members of Group 2 that do not share this highly specialized lifestyle.

Until recently, the only known *X. albilineans* virulence factor was the toxin albicidin. The complete genome sequence of *X. albilineans* enabled the identification of several new candidate virulence factors via screening of a transposon mutagenesis library (Rott et al., 2010). Many of these were not shared with the Group 2 *Xanthomonas* species and may reflect the distinctiveness of the pathogenic strategies adopted by Groups 1 and 2. Therefore, it raises the question of whether these newly discovered virulence factors are also found in Group 1 species other than *X. albilineans*.

Most recently the genome of an isolate of *X. campestris* pv. *musacearum*, which is the causative agent of banana *Xanthomonas* wilt, a disease currently causing devastation to the banana crop

in East Africa was sequenced (Studholme *et al.*, 2010) and is a member of *Xanthomonas* Group 2. Subsequently, follow-up genome sequencing studies on additional isolates that had been deposited in the NCPPB as *X. campestris* pv. *musacearum* were also performed. It was discovered that NCPPB4393 shared little sequence similarity with the previously sequenced isolate; rather, it showed very close sequence similarity with *X. sacchari*, a member of *Xanthomonas* Group 1. On further investigation it was learned that NCPPB4393 had in fact been isolated from an insect on a diseased banana plant but that there was no evidence that the strain is actually pathogenic on banana. Two additional *Xanthomonas* isolates (NCPPB1131 and NCPPB1132) that had been isolated from banana plants in Eastern and Western Samoa were sequenced to identify virulence factors related to Xcm.

6.2 Materials and Methods

Bacterial isolates were obtained from the NCPPB at Food and Environmental Research Agency (FERA). Sequence alignments were performed using MAFFT (Katoh et al., 2005), BWA (Li and Durbin 2009), BLAST (Altschul *et al.*, 1990) and MUMMER (Delcher *et al.*, 2002). DNA preparation and genome sequencing using the Illumina GA2 were performed as previously described (Studholme *et al.*, 2010). We used CGview (Stohard and Wishart, 2005) to visualize whole-genome alignments and used MEGA5 for phylogenetic analyses. Literature references for previously sequenced bacterial genomes used in this study are listed cited in (Rott et al., 2010). *De novo* assembly of Illumina sequence reads was performed using Velvet 1.1.03 (Zerbino *et al.*, 2008). Any sequence reads that contained one or more 'N' prior to assembly were discarded. The following values for the hash length parameter were used: 25 for NCPPB1131, 41 for NCPPB1132 and 49 for NCPPB4493. The coverage cut-off parameter was set to 4 in all three assemblies. For NCPPB1132 and NCPPB4393 read-pairs, and used Velvet's scaffolding step. Did not perform scaffolding on the NCPPB1131 data as the reads were not paired (*i.e.*, we used single-end sequencing). The RAST server (Aziz *et al.*, 2008) was used for automated annotation of draft assemblies.

Note that in the genome-wide alignments (figs 6.3 and 6.4), the pattern of coverage by aligned raw reads does not exactly coincide with the coverage by aligned contigs/scaffolds. This inconsistency is inevitable since the two alignment methods use different criteria for assigning a match. The BWA alignment tool tolerates mismatches so long as the edit distance does not exceed two between the raw read and the reference genome sequence. On the other hand,

BLAST uses an E-value threshold (1e-10 in this case) as the criterion for whether to accept a match. Furthermore, the process of assembly leads to the correction of errors by consensus of multiple sequence reads.

Initial tree(s) for the heuristic search were obtained automatically as follows. When the number of common sites was <100 or less than one fourth of the total number of sites, the maximum parsimony method was used; otherwise BIONJ method with MCL distance matrix was used. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 219 nucleotide sequences, taken from the studies by Young and colleagues (2008) and Parkinson and colleagues (2009) as well as from the three newly sequenced strains. GenBank accession numbers are indicated for the sequences. However, for clarity, only the sub-tree corresponding to Group 1 is shown. The full length of each GenBank sequence entry was used for all of the Young et al., (2008) and Parkinson et al., (2009) sequences. For the sequences taken from our data, the coordinates of the subsequence are given in the parentheses, following the GenBank accession. All positions containing gaps and missing data were eliminated. There were a total of 517 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Tamura et al., 2011).

6.3 Results and Discussion

6.3.1 Bacterial Isolates

Bacterial isolates (Table 6.1) were obtained from the NCPPB in the United Kingdom. NCPPB1131 and NCPPB1132 had been deposited in 1961 by Hayward A.C. after isolation from banana plants. NCPPB4393 was one of several isolates deposited by one of the authors of the present study (V.A.) in 2007. It was deposited in the isolate collection as *X. campestris* pv. *musacearum*. However, the results of this study suggest that it is actually a member of the species *X. sacchari*. It was originally isolated by Mgenzi Byabachwezi from an insect on a diseased banana plant (Muleba district, Kagera region, North Western Tanzania, on the shores of Lake Victoria). Although the insect was collected from diseased banana, sugarcane is commonly grown in that district of Tanzania and so it is possible that the insect acquired the bacterium from sugarcane. Pathogenicity of isolate NCPPB4393 has not been tested.

Table 6.1: Bacterial isolates sequenced in this study

Isolate	Host plant	Country	Year
NCPPB1131	*Musa paradisiaca*	American (Eastern) Samoa	1961
NCPPB1132	*Musa canksii* var. *samoensis*	Western Samoa	1961
NCPPB4393	*Musa* species Isolated from insect on diseased plant	Tanzania	2007

6.3.2 Genome-Wide Sequence Data

Genome-wide sequence data for three isolates listed in Table 6.1 was generated using the Illumina GA2. After removing bar-code adaptors, the sequence reads were 70 nt long. 1.7 million non-paired reads for NCPPB1131. NCPPB1132 and NCPPB4393 were generated, with 1.9 million and 2.1 million paired reads respectively. These Whole-Genome Shotgun project data have been deposited at DDBJ/EMBL/GenBank under the accession numbers AGHY00000000, AGHZ00000000 and AGDB00000000 respectively.

6.3.3 Phylogenetic Position of the Sequenced Xanthomonas Isolates

To ascertain the phylogenetic position of the three newly sequenced isolates, a series of phylogenetic trees based on the nucleotide sequences of house-keeping genes were generated. The same set of seven genes that were used by Pieretti et al., (2009) were used. For four of these genes *atpD*, *dnaK*, *groEL* and *recA* multiple sequence alignments using the maximum likelihood method were built.

The phylogenetic reconstruction (Tamura *et al.*, 2011) based on *atpD* is shown in fig. 6.1. The trees for *atpD*, *dnaK* and *groEL* are all toplogically consistent with each other, though there is some variation in branch lengths and the precise position of *X. albilineans* within Group 1 is less well resolved in the *atpD* tree than in the others. However, analysis of the *recA* sequences yielded a different branching pattern in which *X. albilineans* falls within the *Xylella fastidiosa* lineage rather than within the *Xanthomonas* Group 1. This phylogenetic analyses clearly and consistently indicated that all three isolates fell within the phylogenetic range of the genus

Xanthomonas. Specifically, all three isolates are more closely related to *X. albilineans* than to the Group 2 *Xanthomonas* species and therefore likely belong to Group 1.

It should be noted that in the analyses based on three out of four house-keeping genes, the genus *Xanthomonas* comprises a single monophyletic clade, distinct from the related genera *Stenotrophomonas* and *Xylella*. This is consistent with previous studies (Ryan et al., 2011, Lu *et al.*, 2008) but contradicts recent claims that *X. albilineans* and *Xylella fastidiosa* form a monophyletic clade distinct from the Group 2 *Xanthomonas* species (Pieretti *et al.*, 2009, Rott et al., 2010). On the other hand, the analyses based on the *recA* gene were consistent with Pieretti's hypothesis that *X. albilineans* and *Xylella fastidiosa* form a monophyletic clade distinct from the Group 2. The incongruity between *atpD*, *groEL* and *dnaK* on the one hand and *recA* on the other implies that there has been recombination and that not all of these house-keeping genes truly reflect the vertical descent of the core genome. The most possible explanation is that *recA* has undergone horizontal transfer in either the *Xylella* lineage or in the *X. albilineans* lineage. The reasons for discrepancy between our phylogenetic reconstructions for *atpD*, *groEL* and *dnaK* compared with that of Pieretti and colleagues (Pieretti *et al.*, 2009) are two-fold. First, the analysis presented by Pieretti is a composite of genes displaying at least two distinct phylogenetic histories.

The results of this analysis are also inconsistent with those of Sharma & Patil, (2011). In their study they present a phylogenetic tree in which *X. albilineans* is the out-group and *Stenotrophomonas* species fall within a monophyletic group along with the other *Xanthomonas* species. However, this discrepancy is explained by their misplacing the root of their tree. (Sharma & Patil, 2011) do not explain how they chose the position of the root in their tree; they simply assume that *X. albilineans* is the out-group without offering any justification for this. On the other hand, a phylogenetically distinct species (*P. aeruginosa*) as the out-group in order to root our tree was used. If the tree of Sharma and Patil (2011) is re-rooted with *Stenotrophomonas* species as the out-group, then their tree is topologically congruent with this study. Note that Sharma and Patil do not include *Xylella fastidiosa* in their analysis.

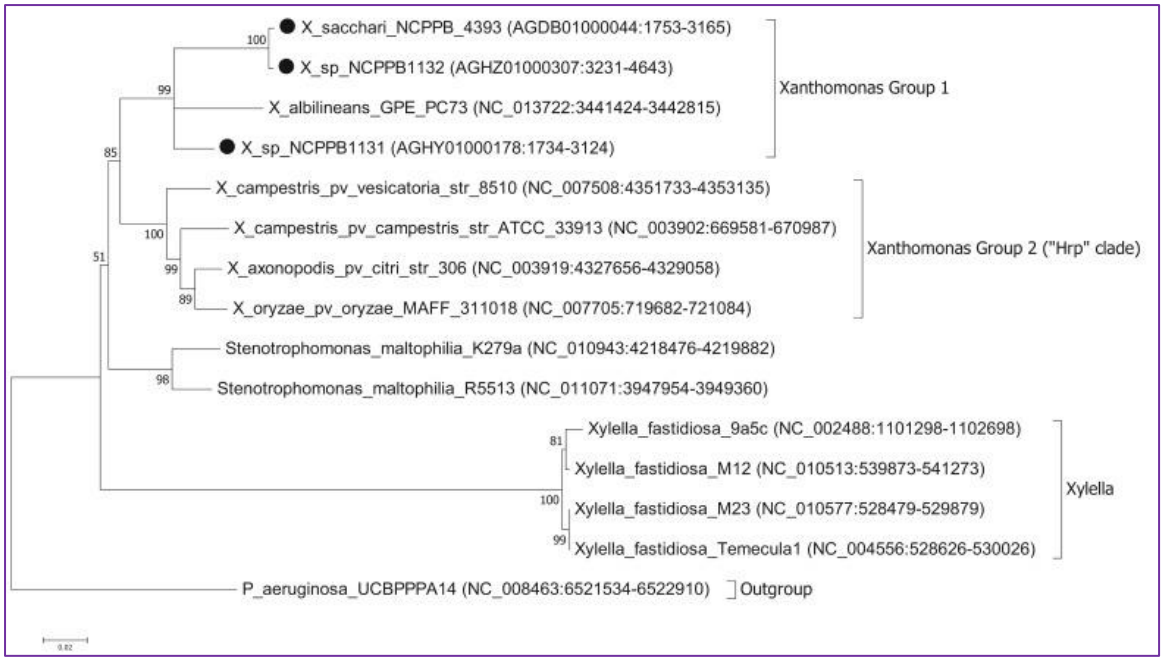

Figure 6.1: Molecular phylogenetic anaylsis of atpD gene of newly sequenced Xanthomonads by Maximum Likelihood method. The newly sequenced isolates are indicated by black circles (●)

The phylogenetic tree was inferred by using the Maximum Likelihood method based on Tamura & Nei, (1993). The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically. When the number of common sites was <100 or less than one fourth of the total number of sites, the maximum parsimony method was used; otherwise BIONJ method with MCL distance matrix was used. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 15 nucleotide sequences. Codon positions included were 1st + 2nd + 3rd + Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 1,373 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Tamura *et al.*, 2011). The newly sequenced bacterial isolates are indicated with black circles. For each nucleotide sequence, RefSeq accession numbers and coordinates are given in parentheses.

To more precisely resolve the newly sequenced isolates' positions within *Xanthomonas* Group 1, phylogenetic analyses based on the gyrase B (*gyrB*) gene (fig. 5.2) was performed; partial sequences of *gyrB* are available from two studies (Young et al., 2008, Parkinson *et al.*, 2009) including many more *Xanthomonas* strains than those for which there are fully sequenced genomes. It was found that the *gyrB* sequence from NCPPB4393 was identical to those from strains of *X. sacchari*. Previously, *X. sacchari* was described as comprising isolates isolated from diseased sugarcane (Vauterin *et al.*, 1995). Therefore, the description may need to be modified to include isolates isolated from insects.

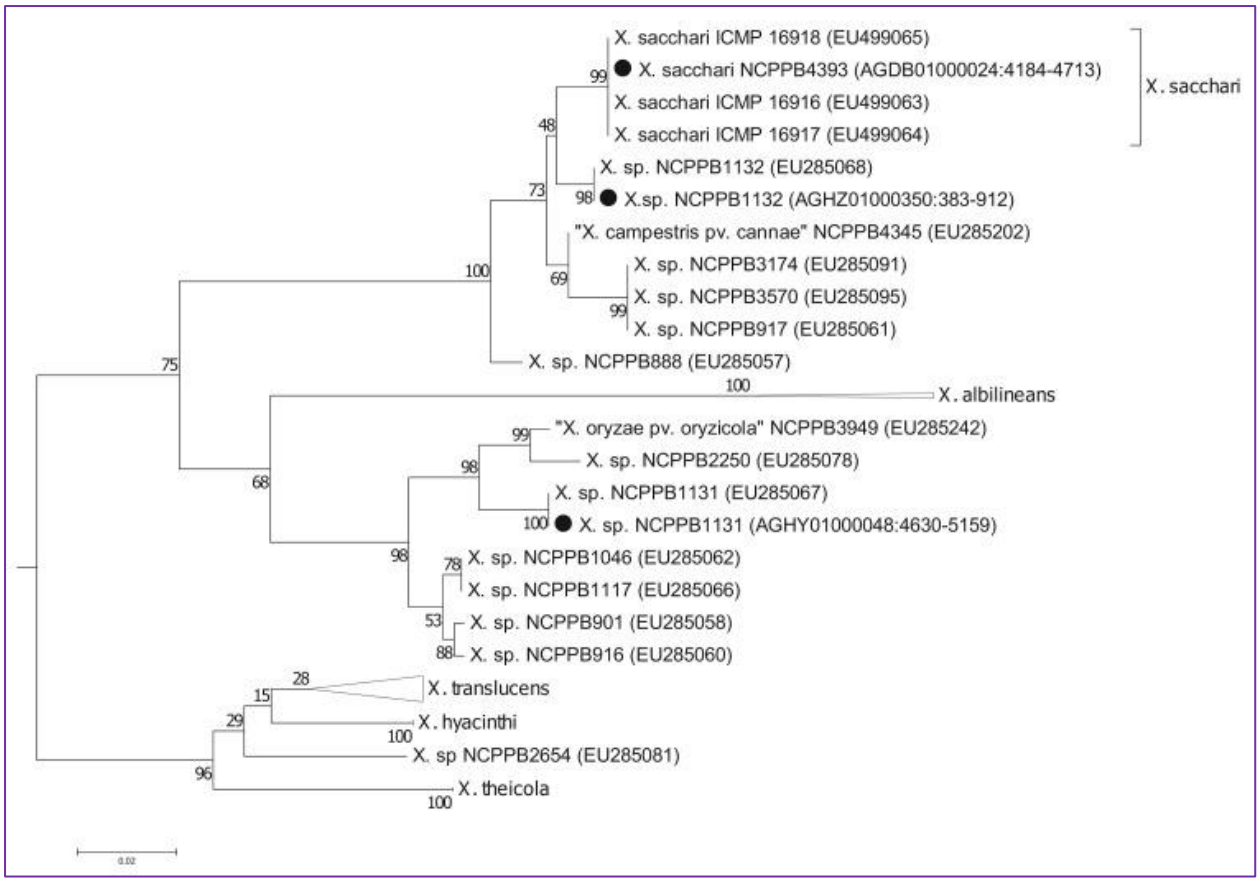

Figure 6.2: Phylogenetic positions of the three newly sequenced isolates within Xanthomonas Group 1. The figure shows the divergence of the gyrB gene as inferred by using the Maximum Likelihood method based on the Tamura-Nei model (1993). The tree with the highest log likelihood (_7358.2201) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The newly sequenced isolates are indicated by black circles (●).

Isolate NCPPB1132 is also closely related to *X. sacchari* but is more divergent than NCPPB4393 (fig. 6.2). It falls within a clade that included *X. sacchari sensu strictu* as well as

X. "*campestris*" pv. *cannae* (from canna lilly, a relative of banana) and several unnamed isolates isolated from sugarcane (NCPPB888 and NCPPB917), foxtail millet (NCPPB3174) and arrow leaf elephant ear (NCPPB3570).

Isolate NCPPB1131 is more closely related to *X. albilineans* than to *X. sacchari*, but shows closest affinity with NCPPB2250 (fig. 6.2), which was originally isolated from ginger, a relative of banana. Isolate NCPPB1131 is also similarly closely related to NCPPB3949, isolated from rice and erroneously deposited as *X. oryzae* pv. *oryzicola*. Parkinson and colleagues (2009) describe a species-level clade (Slc 7) that included NCPPB1131, NCPPB3949 and isolates isolated from ginger, cotton and sugarcane.

6.3.4 Comparison of the Three Genomes Versus X. albilineans

The total sizes of the genome assemblies were 3.8 Mb for NCPPB1131, 4.7 Mb for NCPPB1132 and 4.9 Mb for NCPPB4393. These can be used as estimates of genome size, but may be inaccurate since we have not closed the gaps in the draft assembly. The size for NCPPB1131 should be treated with particular caution, since the use of non-paired sequence reads yielded a very fragmented assembly. Contiguity of the assemblies can be represented by N_{50} scaffold lengths which were 1.1 Kb (NCPPB1131), 4.8 Kb (NCPPB1132) and 51.5 Kb (NCPPB4393). The numbers of scaffolds in each assembly were 4,158 (NCPPB1131), 1,652 (NCPPB1132) and 259 (NCPPB4393). Nevertheless, these estimates are congruent with sizes of previously sequenced *Xanthomonas* genomes, which range from 4.8 to 5.3 Mb for Group 2, whilst *X. albilineans* has a genome of less than 3.8 Mb.

The three genome assemblies against the genome sequence of *X. albilineans* species were aligned including also the reads, without performing *de novo* assembly. The fig. 6.3 illustrates a genome-wide overview of these alignments. It should be noted that the sequencing depths obtained for the three isolates ensures complete coverage over the entire breadth of the genomes. This means that by examining alignments of sequence reads against a reference sequence, independently from any *de novo* assembly, we can confidently determine the presence or absence of genes. The depths of coverage of each genome, as determined by depths of alignments of raw reads against assemblies, were 20× (NCPPB1131), 67× (NCPPB1132) and 72× (NCPPB4393).

Clearly, a significant proportion of the *X. albilineans* genome is not conserved (at the nucleotide sequence level) in the three newly sequenced Group 1 isolates (fig. 6.3). Prominent among the

non-conserved regions is the gene cluster encoding the *X. albilineans* SPI-1 TTSS (positions 1,703,391-1,730,688). No evidence for any non-flagellar TTSS in any of the three isolates was found. All three isolates also lack the albicidin biosynthesis cluster at positions 1,740,869-1,788,517 and so they likely do not produce albicidin.

Pieretti and colleagues (2009) observed that the genomes of two xylem-limited pathogens,

X. albilineans and *Xylella fastidiosa*, both share an unusual allele of the *metE* gene, which encodes 5-methyltetrahydropterolyl-triglutamate-homocysteine methyltransferase, an enzyme required for methionine biosynthesis. They infer that the ancestor of *X. albilineans* and *X. fastidiosa* lost *metE*, while the rest of the Xanthomonads retained it, and then this ancestor gained a new allele of *metE* by horizontal transfer. This is contrary to other observations since the last common ancestor of *X. albilineans* and *Xylella fastidiosa* mays also have been the ancestor of the other Xanthomonads, including *Stenotrophomonas* and all *Xanthomonas* species (fig. 6.1). It was found that NCPPB1131, NCPPB1132 and NCPPB4393 all contain a *metE* gene that most closely resembles (at least 90% amino acid sequence identity) that of *X. albilineans* rather than those of Group 2 *Xanthomonas* species. This suggests that the *X. albilineans metE* occurs widely in the Group 1 *Xanthomonas* species and is not restricted only to xylem-limited species. The incongruence between the phylogeny of *metE* genes and the core house-keeping genes indicates that *metE* has been replaced independently in at least two distinct lineages during the evolution of the Xanthomonads.

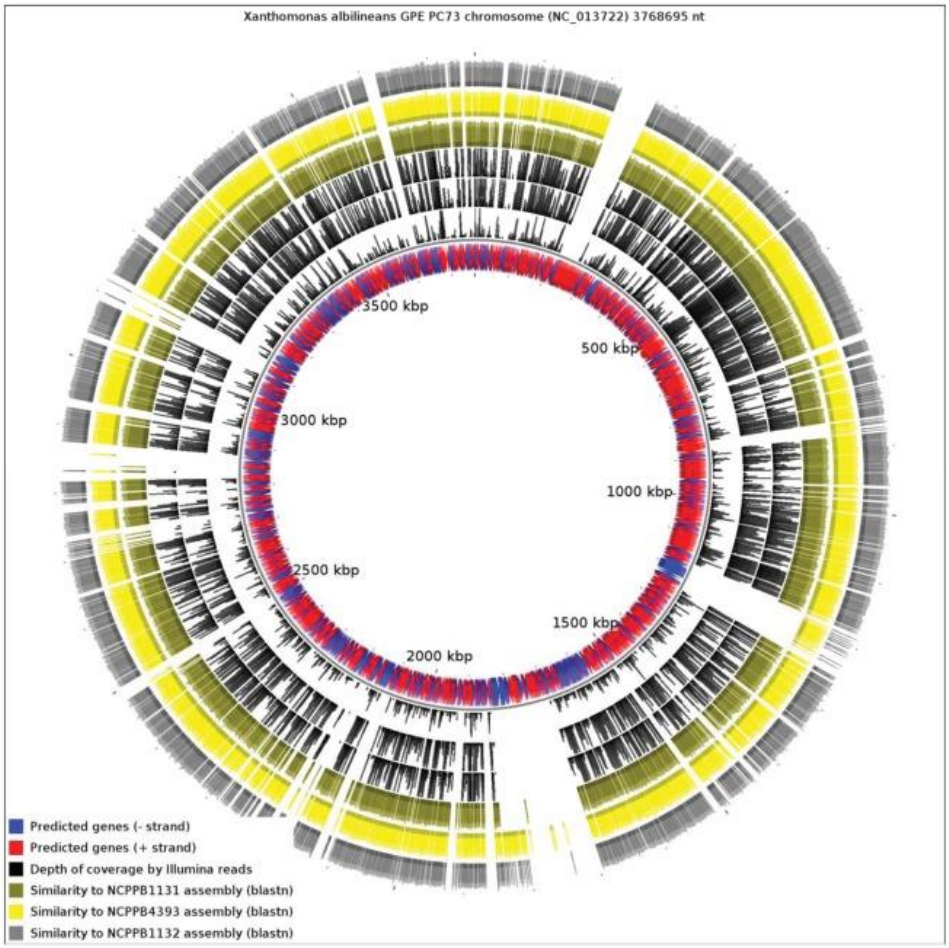

Figure 6.3: Alignment of the three Xanthomonas Group 1 new genome sequences against the chromosome of X. albilineans. The blue and red inner track represents annotated genes. The next three black tracks represent depth of coverage by Illumina sequence reads for NCPPB1131, NCPPB4393 and NCPPB1131 respectively. The four colored outer rings indicate sequence similarity to the genome assemblies of NCPPB1131 (olive), NCPPB4393 (yellow) and NCPPB1132 (grey) respectively.

6.3.5 Genome Reduction

Genome reduction has occurred independently in two separate lineages of xanthomonads that are specialized for a xylem-limited lifestyle. That is, *Xylella fastidiosa* and *X. albilineans* have independently converged on a xylem-limited lifestyle, each having evolved from a different ancestor with a larger genome. The only other xylem-limited bacterial species for which complete genome sequence data are available is *Leifsonia xyli*. Interestingly, *L. xyli* also appears to have a reduced genome, its chromosome being 2.6 megabases long in contrast to its non-xylem-associated relative *Clavibacter michiganensis*, which has a 3.3-megabase chromosome (Chalupowicz et al., 2009). Thus, genome reduction is associated with at least three distinct lineages of xylem-limited bacteria, suggesting that a stripped-down genome may

be adaptive for this relatively stable environment. Complete genome data are not yet available for the other well-known xylem-limited species, *Ralstonia syzygii* (Chang et al., 2012; Garnier, 2002; Kamper et al., 1985), so we cannot yet be sure that this is a universal phenomenon.

Since the only sequenced member of *Xanthomonas* Group 1 is a specialized xylem-limited pathogen with a reduced genome, this raises the question of whether other members of Group 1 show similar evidence of genome reduction. These results reveal that *X. albilineans* has undergone significantly more reduction than NCPPB1131, NCPPB1132 and NCPPB4393. We aligned all four available genomes from Group 1 against the reference sequence of *X. campestris* pv. *vesicatoria* (*Xcv*) 85-10 (RefSeq: NC_007508), a member of Group 2 (fig. 6.4). It was found that only 13.82% of the *Xcv* chromosome was conserved in *X. albilineans*. Significantly larger portions of the *Xcv* were conserved in NCPPB1131, NCPPB1132 and NCPPB4393 (21.69%, 26.85% and 28.44% respectively). This is consistent with *X. albilineans* having lost more of the *Xanthomonas* genome than have the other three isolates.

X. albilineans produces the toxin albicidin, which is a DNA gyrase inhibitor (Hashimi et al., 2007; Zhang, Xu, & Birch, 1999). In addition to its action on host plant chloroplasts, it is also deleterious to most bacteria. Pieretti and colleagues (2009) propose that albicidin played a key role in the erosion of the *X. albilineans* genome. Specifically, they propose that exposure to sub-lethal doses of intracellular albicidin induced recombination and mutagenesis. They note that *X. albilineans* has, in addition to its albicidin transporter (AlbF), an unusual DNA gyrase, containing a 43-amino-acid insertion that confers resistance to albicidin (Hashimi *et al.*, 2007). They go on to propose that "genome erosion induced by albicidin was likely arrested by the evolution of the albicidin-resistant *DNA gyrase*" (Pieretti *et al.*, 2009). Interesting though it is, there is no evidence to support this conjecture. Nor is there any need to invoke such a special mechanism, since genome reduction is a common phenomenon, seen in many parasitic organisms that do not produce albicidin-like antibiotics. Furthermore, sequence insertions similar to that in *X. albilineans* gyrase A are not uncommon among the gamma-Proteobacteria. Therefore, it is quite plausible that albicidin-insensitivity is an ancient and/or common trait that was already present in *X. albilineans* before it acquired the ability to produce albicidin (Rott et al., 2010).

Examination of the genome-wide sequence data revealed that the Group 1 isolates NCPPB1131, NCPPB1132 and NCPPB4393 lack the gene cluster required for albicidin biosynthesis, yet all three isolates encode a gyrase A resembling that of *X. albilineans*,

including the 43 amino acid insertion that is supposed to be responsible for albicidin resistance. These isolates do have other non-ribosomal peptide synthesis genes, so the possibility that the 43 amino acid insertion confers resistance to some other unknown toxin cannot be excluded. Nevertheless, the presence of this insert clearly does not correlate with the incidence of genome reduction. Aside from any involvement in genome reduction, albicidin-like antibiotics probably have had profound effects on the evolution and ecology of xylem-limited parasites. For example, the genome sequence of *L. xyli* encodes a close homologue of AlbF that may allow it to colonise the host simultaneously with *X. albilineans* (da Silva et al., 2002).

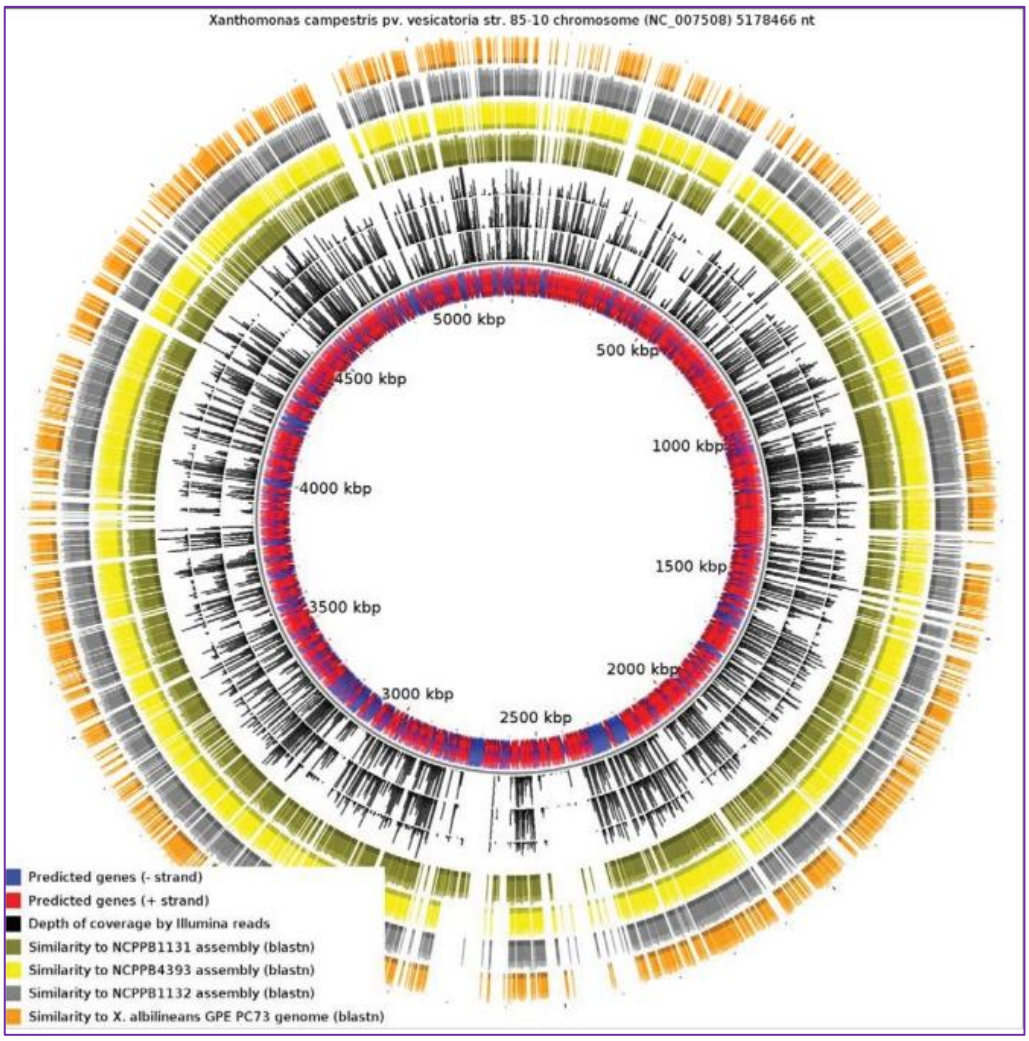

Figure 6.4: Alignment of the four Xanthomonas Group 1 genome sequences against the chromosome of Xanthomonas campestris pv. vesicatoria. The blue and red inner track represents annotated genes. The next three black tracks represent depth of coverage by Illumina sequence reads for NCPPB1131, NCPPB4393 and NCPPB1131 respectively. The four colored outer rings indicate sequence similarity to the genome assemblies of NCPPB1131 (olive), NCPPB4393 (yellow), NCPPB1132 (grey) and X. albilineans (orange) respectively.

6.3.6 Novel Genes in Xanthomonas Group 1

Until the present study, *X. albilineans* was the only member of Group 1 for which genome-wide sequence data were available. However, *X. albilineans* is an atypical example of the group, since it has undergone significant genome reduction associated with adaptation to life within the xylem. Therefore, this study examined the genome sequence data from NCPPB1131, NCPPB1132 and NCPPB4393 for genomic regions that are absent from *X. albilineans* GPE PC73. Comprehensive lists of these regions include heavy metal resistance proteins, haemolysins, haemagglutinins, citrate synthase, persistence factor HipA, sigma54-dependent activators, type IV pili, drug efflux pumps, vancomycin B-type resistance protein VanW, xylanase, chitinase, beta-lactamase, restriction modification systems and many others.

There are, of course, some differences among the genomes of the Group I *Xanthomonas* isolates. The genomes of NCPPB1132 and NCPPB4393 are alignable with each other over 88.64% and 84.33% of their respective lengths. That is, 12–16% of their genomes are variable and presumably subject to relatively recent horizontal transfer. They share 96.85% nucleotide sequence identity over the alignable conserved core portion of their genomes. Similarly they share 94.47% and 94.45% identity respectively with NCPPB1131.

The genomes of NCPPB1131, NCPPB1132 and NCPPB4393 each encode a protein with 64% amino acid sequence identity to AvrXca (also known as AvrXccA1), a protein that confers avirulence on *Arabidopsis* in *X. campestris* pv. *raphani* (Buell, 2002; Buell et al., 2003). Based on this avirulence activity, it has been speculated that AvrXca might be an effector. However, all three isolates lack a TTSS, so it seems unlikely that it is secreted via the TTSS. Interestingly, at least two homologues of AvrXca are effectors secreted by the type II secretion system (T2SS) (Corbett et al., 2005; Jha et al 2005), suggesting that AvrXca might also be a T2SS-dependent effector.

Each of the three newly sequenced genomes encode two cellulose-degrading enzymes that are absent from *X. albilineans.* Specifically, these include an endoglucanase that shares 61% amino acid sequence identity with XCC0028 from *X. campestris* pv. *campestris* ATCC 33913 and a cellulase (glycosyl hydrolase family 5) that shares 68% identity with XCV0358 from *X. campestris* pv. *vesicatoria* 85-10. Both enzymes are widely distributed among the Group 2 *Xanthomonas* species as well as in the three newly sequenced Group 1 isolates, but absent from *X. albilineans*. It is possible that they play a role in degrading plant cell walls.

The *gum* genes, found in *Xylella fastidiosa* and all studied Group 2 *Xanthomonas* species, play a role in producing extracellular polysaccharides and forming biofilms and are implicated in pathogenicity. However, no *gum* genes have been found in *X. albilineans* (Pieretti *et al.*, 2009). The study found homologues of these (*gumBGCEKDHIMJL*) conserved in all three newly sequenced isolates. Therefore, it appears that the *gum* cluster was probably present in the common ancestor of *Xylella*, *Stenotrophomonas* and *Xanthomonas* and was subsequently lost by *X. albilineans* and *Stenotrophomonas* but retained in *X. sacchari* and the other Group 1 *Xanthomonas* species.

6.4 Conclusions

The ability to survive on banana plants has evolved more than once within the genus *Xanthomonas*, with strains isolated from banana falling within both major phylogenetic lineages: Group 1 (NCPPB1131 and NCPPB1132) and Group 2 (*X. campestris* pv. *musacearum*). Clearly their strategies are different. *Xanthomonas campestris* pv. *musacearum* encodes an apparently intact Hrp TTSS and a suite of effectors that it presumably uses to overcome the host's defenses. On the other hand, the Group 1 strains, related to *X. sacchari* and *X. albilineans*, lack the TTSS and must use some other strategy to avoid triggering host defenses. In the case of *X. albilineans*, the strategy appears to be one of stealth, where the pathogen restricts its colonization to the dead xylem. As a result, it has undergone substantial genome reduction, shedding genes unnecessary for this restricted niche. Very little information is available about the endophytic lifestyles of *X. sacchari* and other related members of *Xanthomonas* Group 1 and there is no evidence that they are limited to the xylem. Certainly the lack of genome reduction would be consistent with having to survive in more diverse conditions. The example strain that was sequenced here, NCPPB4393, apparently spent at least part of its life cycle associated with insects. It is expected that the availability of complete sequence data for this group of organisms is the first step towards understanding their interactions with plants and identifying potential virulence factors.

Chapter Seven

General Discussion, Conclusion and Recommendations

7.1 General Discussion

Studholme et al (2010) sequenced and predicted some differences in virulence factors of Xvv702 and Xcm4381. This thesis used PCR to determine the presence or absence of some of the predicted virulence factors in more strains of both Xvv and Xcm. The study has shown that Xcm strains tested contain some lipopolysaccharides that are not detected in the Xvv strains. This confirmed Studholme et al., (2010) presumption that lipopolysaccharide locus in Xvv702 and Xcm4381 were not significantly similar. Of much importance is the finding that O-antigen is only present in the Xcm strains but neither in Xvv nor in Xsp. This can be used for characterizing other Xcm strains and also for studying the interaction of outer membrane proteins of banana infecting Xanthomonas. The O-antigen has a polysaccharide chain that varies in length with up to 40 repeated units of dideoxyhexoses. At least 20 different sugar molecules may compose the O-antigen, including molecules that are rarely found in nature, such as abequose, colitose, paratose and tyvelose. These components are strain-specific. The O-antigen displays a large degree of inter- species and intra-species variation, which is related to the nature, order and union of the different sugars. The O antigen is the immunodominant part of LPS and therefore is the easiest target for the humoral response of the host. For this reason, the O antigen is the basis for the serological classification of gram-negative bacteria. The O-antigen is recognized by the innate immune response and participates in complement activation and in the inhibition of the formulation of the complex that attacks the membrane (Reyes et al., 2012).

According to Giltner et al., (2012) the minor pilins such as PilE, PilV, PilW, PilX, FimT in pilus assembly play roles in priming of pilus extension or prevention of pilus retraction; in control of pilus length; or in pilus-specific functions including adherence, transformation competence or motility. PilE has been confirmed in Xcm strains and sugar cane infecting Xvv but not Xvv206 from maize. Type 4 fimbrial biogenesis protein FimT and the putative type IV pilus protein PilY1 are also potential virulence factors to understanding the attachment of Xcm and Xvv interaction on hosts and non-hosts.

Among the TTSS factors was YopJ-like1, YopJ-like 2, HopW1 and XopAF (AvrXv3) which

were only in Xcm and not Xvv thus candidates for application in disease management. This is because of their role in inhibition of the mitogen-activated protein kinase (MAPK), nuclear factor kB (NF-kB) signaling for induction of localized cell death in plants. It was previously shown that PstDC3000 strain expressing HopW1-1 had reduced growth and caused weak disease symptoms in the Ws Arabidopsis plants. This study suggests that in-depth evaluation of more virulence factors in the banana infecting strains and other hosts will build up the understanding of the pathogen-host relationship.

PCR-restriction fragment length polymorphism (RFLP)-based analysis, also known as cleaved amplified polymorphic sequence (CAPS), is a popular technique for genetic analysis. It has been applied for the detection of intraspecies as well as interspecies variation. The study on validation of presence/absence of virulence factors using PCR showed similarity of some predicted virulence factors within strains of Xcm and Xvv. However it also indicated some inherent differences within the strains. A genetic evaluation of more strains of Xcm and Xvv were further compared after sequencing.

The ability of pathogenic bacteria to cause disease in a susceptible host is determined by multiple virulence factors acting individually or together at different stages of infection. Virulence factors are often involved in direct interactions with the host tissues or in concealing the bacterial surface from the host's defense mechanisms. Previously only a single genome of Xcm was sequenced. This study has contributed fourteen new genome assemblies namely AKBE00000000, AKBF00000000, AKBG00000000, AKBH00000000, AKBI00000000, AKBJ00000000, AKBK00000000, AKBL00000000, AKBM00000000, AGDB00000000, AKBN00000000, AKBO00000000, AGHY00000000 and AGHZ00000000. These are derived from the strains for NCPPB2005, NCPPB4379, NCPPB4380, NCPPB4384, NCPPB4392, NCPPB4394, NCPPB1326, NCPPB4393, NCPPB1381, NCPPB206, NCPPB1131 and NCPPB1132.

A genetic comparison of Xcm and Xvv resulted into two unique clades and 31 virulence factors that differentiate Xcm from Xvv. It was insinuated that Xcm could have spread from two different sources within the Great lakes of Eastern Africa. In addition non silent SNPs that distinguish the monophyletic Xcm into two clades by origin were identified. Considering that the strains were collected from different geographical locations at different times, it is important to understand the pathogen dispersions and associated genetic changes.

An in-depth exploration of Xcm isolates was necessary to determine the origin and potential genetic factors for understanding its interaction with the host. Udatha et al., (2013) suggested that the non-synonymous SNPs, the so-called non-silent SNPs, which are single-nucleotide variations in the coding regions that give "birth" to amino acid mutations, are often involved in the modulation of protein function. Understanding the effect of individual amino acid mutations on a protein/enzyme function or stability is useful for altering its properties for a wide variety of engineering studies. SNPs developed in this study were used to generate molecular markers that can be used in epidemiological and biogeographical studies. This infers therefore that the SNPs identified in this study can be used for further understanding the interaction of different Xcm clades with their banana host which could result into more sustainable management options. Single-nucleotide polymorphisms (SNPs) have been explored as a high-resolution marker set for accelerating the mapping of disease genes. The homologues of HrpF and HopW1 were identified to be different for banana associated Xcm but ensete associated Xcm. Butner et al. (2002) conducted protein functional studies that indicated that HrpF presumably plays a role at the bacterial-plant interface as part of a bacterial translocon which mediates effector protein delivery across the host cell membrane. Functional annotation of genes identified would provide further information on Xcm.

In an attempt to expound the understanding of banana associated Xvv, further analysis of genome sequences was conducted. The comparative analyses of these genome sequences and previously published genome sequences of the closely related pathovar Xcm, have revealed extensive differences in gene content among Xvv. The study described some of these differences in plasmid content, LPS biosynthesis clusters, T4P and TTSS effectors. The availability of the sequence data from multiple isolates allows to distinguish between inherent variations within Xvv. These might confound attempts to identify important genetic differences between the two pathovars and for functional analysis of important virulence factors. The two Xcm sub-lineages previously described in chapter 5 of this thesis differ in that members of Sub-lineage II have lost a gene encoding a homologue of XopL; this gene however has acquired a premature stop codon in four of the six Xvv isolates, suggesting that isolates of both Xcm and Xvv have independently converged on eliminating XopL. The pilC and pilA genes are highly divergent among Xvv 895, Xvv 206 and Xcm 2005, with Xcm 2005 having X. gardneri-like pilC and pilA genes. Xvv 895 and Xvv 206 have Xaa-like pilC and pilA genes. Type-4 fimbriae mediate attachment to host epithelial tissues and a form of surface translocation called

twitching motility. Analysis of mutants defective in fimbrial function in *Pseudomonas aeruginosa* has allowed the identification of many of the genes involved in the biogenesis of these organelles (Alm et al., 2005). This implies that mutation of these genes can be useful in understanding the strains interaction with banana host through functional alteration. Pieretti et al., (2012) indicated that biogenesis of a type IV pilus involves a large number of proteins that are highly conserved within the Xanthomonas genus, with the exception of the PilA pilin and the PilV-W-X-Y-E operon. The PilA protein is the main structural protein of the type IV pilus. The sequence variability of the pilA gene observed among the sequenced Xanthomonas could be correlated with host specificity. Pieretti et al., (2015) determined that pilA could be specific to sugar cane. There is considerable diversity within Xvv, at the level of SNPs in the core genome and at the level of gene content which shows differences in chromosomally located gene clusters, such as those encoding LPS biosynthesis and T4P. The sequenced Xcm isolates encode two homologues of XopJ that are absent from all sequenced isolates of Xvv. In *X. campestris pv. vesicatoria*, XopJ has been shown to interfere with salicylic acid-dependent defense responses to attenuate the onset of necrosis and to alter host transcription (Üstün et al., 2013, 2015). Effector XopJ suppresses defense- related callose deposition and host protein extracellular secretion (Bartetzko et al., 2009). Another XopJ family member, avrRxv, interacts with host 14-3-3 protein (Whalen et al., 2008) and is also hypothesized to bind to MAPKs and to interfere with the signaling cascade. AvrBsT, a member of the XopJ/YopJ family, triggers an HR in a single landrace of Arabidopsis with a recessive mutation in a gene for carboxylesterase. A model has been proposed that, in the absence of the esterase activity, AvrBsT can suppress recognition by a host R-gene product (Cunnac et al. 2004). Mutations of individual effector genes do not always result in a change in virulence, which might implicate redundancy of effectors (Castañeda et al., 2005). It has been found that XopAF is absent from the genome of all Xcm and Xvv 206, though it is present in Xvv 895, 890, 1326 and 1381, as well as in Xvv 702. XopAF has effector homologue AvrXv3 whose biochemical grouping is transcriptional activator domain. Mutation analysis of xopAF in lime has indicated that the gene contributes to virulence. Some effectors from xanthomonads have been characterized for their contribution towards virulence on their respective hosts and this can be explored further for banana xanthomonas wilt management.

Banana associated xanthomas strains are considered to have genetic elements which can provide information on host specificity. Bacterial strains were isolated from Malagasy on Musa

species which could provide more knowledge on virulence factors in these species. Studholme et al. (2016) also sequenced Xanthomonas strains including *X. arboricola pv. celebensis* isolated from bananas in eastern and western Samoa. It was hoped that further analysis of these strains will contribute to understanding banana associated xanthomonas. Analysis of the recA sequences on NCPPB1131, NCPPB1132 and NCPPB4393 yielded a different branching pattern in which *X. albilineans* falls within the Xylella fastidiosa lineage rather than within the Xanthomonas Group 1. All three strains are more closely related to X. albilineans than to the Group 2 Xanthomonas species and therefore belong to Group 1. Strain NCPPB1132 is also closely related to *X. sacchari* but is more divergent than NCPPB4393. It falls within a clade that included X. sacchari sensu strictu as well as X. "campestris" pv. *cannae* (from canna lilly, a relative of banana) and several unnamed strains isolated from sugarcane, foxtail millet and arrow leaf elephant ear. Strain NCPPB1131 is more closely related to *X. albilineans* than to X. sacchari, but shows closest affinity with NCPPB2250 which was originally isolated from ginger, a relative of banana. Strain NCPPB1131 is also similarly closely related to NCPPB3949, isolated from rice and erroneously deposited as *X. oryzae pv. oryzicola*. Parkinson and colleagues (2009) describe a species-level clade (Slc 7) that included NCPPB1131, NCPPB3949 and strains isolated from ginger, cotton and sugarcane.

7.2 Conclusions

The availability of multiple genomes from closely related organisms has led to the field of comparative genomics, a powerful approach to evaluate genomic diversity. Investigation of genomic diversity can provide insight into the evolutionary history of bacterial species. Luckily complete genome sequencing is now practical and affordable for large strain collections and was used in this project. PCR-based methods were also used to further investigate the genomic diversity of sequenced strains. The work presented in this thesis has contributed to the field of genomics by providing sequence data for fifteen isolates. It also identified SNPs that are good markers for difference between different Xcm and Xvv isolates. It was able to propose virulence factors that can be further investigated for functionality and use in BXW management. In particular this thesis has been able to:-

a) Generated fourteen (14) genome assemblies de novo for NCPPB2005, NCPPB4379, NCPPB4380, NCPPB4384, NCPPB4392, NCPPB4393, NCPPB4394, NCPPB1326, NCPPB1381, NCPPB206, NCPPB890, NCPPB895, NCPPB 1131 and NCPPB1132.

b) Determined that the spread of Xcm in the great lakes could have originated from two different sources thus resulting into two sub-lineages of Xcm. Sub- lineageI is comprised of strains from Burundi, Kenya, Tanzania and Uganda while sub lineageII includes strains from Democratic republic of Congo, Ethiopia and Rwanda.

c) Identified thirty two (32) candidate genes for development of an Xcm-specific PCR based assay to differentiate any Xcm from Xvv.

d) Provided a resource for generating molecular markers that can be used for epidemiological and biogeographical studies on a much larger panel of isolates without the need for whole-genome sequencing using single nucleotide polymorphisms (SNPs). The restriction enzymes and target sites NZ_ACHT01000081: AluI, NZ_ACHT01000124: AluI, NZ_ACHT01000112: FokI and NZ_ACHT01000304: NdeI are necessary markers for further Xcm characterization in the region.

e) Revealed extensive differences in gene content among Xvv with some of these differences in plasmid content, LPS biosynthesis clusters, T4P and TTSS effectors. XopAF is encoded by Xvv isolates from sugarcane, but not by the Xvv isolate from maize and not by Xcm isolates from banana and ensete, also found polymorphism with respect to a xopL-like gene in Xvv. xopL-like sequence is completely absent from Xcm isolates belonging to Sub-lineage II but, it is present in Xcm isolates belonging to Sub-lineage I,.

“Comparative genomics” usually conveys parallel study of genome structure, dynamics, and function among different individuals or groups of individuals (taxa) at the different ranks in the taxonomic hierarchy, from species to kingdom. Genome comparisons at different phylogenetic distances answer different sorts of questions (Hardison, 2003). Phylogenetic or genealogical interpretation of DNA sequence data from multiple genomic regions has become the gold standard for species delimitation and population genetics. Precise species concepts can inform quarantine decisions but are likely to reflect evolutionary events too far in the past to impact disease management. On the other hand, multilocus approaches at the population level can identify patterns of endemism or migration directly associated with episodes of disease, including host shifts and associated changes in determinants of pathogenicity and avirulence. Generating and exploiting a genetic loss of susceptibility is the modern strategy for gaining durable disease resistance. Better understanding of host susceptibility genes (S) and new breeding technologies now enable the targeted mutation of pathogen-derived effectors as

molecular probes to identify S genes for genetic plant protection. Comparative genomics often targets similarities among genomes to identify the genes that are relatively invariant, or conserved, because of their significant function. Mutational variants of genes that do not improve the fitness of the individual tend to be lost through natural selection. Mutations that change the function or protein products of genes that have important functions tend to be weeded out through a kind of natural selection. This thesis therefore has built a foundation for further characterization of genes and their functions in the interaction of xcm/xvv and their hosts.

7.3 Recommendations

This study has been able to identify several predicted virulence factors for both Xcm and Xvv that could be important for understanding their interaction with individual hosts but also be applied in management options during breeding for resistance. The following are suggested to be conducted to further the development of effective banana bacterial wilt management options:-

a) Need to survey genotypes of a much larger collection of isolates from outbreaks in all the ensete and banana-growing areas to uncover the routes of geographical spread at a much higher degree of resolution. This ideally should be conducted on isolates for which precise details are available on the date and the geographic location at which they are collected.

b) Analysis and comparisons of gene expression profiles in different strains is necessary to provide new insight into the pathogen's virulence strategies. Determine the molecular mechanisms underlying Xcm/Xvv effector-triggered susceptibility of banana, maize, sugarcane (Genome wide functional analysis of virulence factors). Therefore there is need to conduct an in-depth transcription expression profile of all the identified virulence factors xcm and xvv.

c) To conduct genetic analysis of the effects of combinations of effector mutations on virulence. Identifying a gene and understanding its function are altogether different matters. At least one-fourth of genes that are identified in bacterial genomes, whether large or small, whether from pathogen or non-pathogen, are "hypothetical" or of unknown function. Grouping genes based on functional similarity can systematically enhance biological interpretation of large lists of genes derived from high throughput studies. Therefore propose to conduct a gene functional analysis of the expressed virulence factors. This work provides resources for

functional studies aiming at understanding host specificity of pathogenic bacteria, functional redundancy and the driving forces shaping virulence effector repertoires.

d) Carryout gene editing and transformation of popular banana varieties using some of the predicted virulence factors

REFERENCES

Abbasi, S. (2010). *Review of Next Generation Sequencing, BBSRC*.

Adams, J. U. (2008). DNA Sequencing Technologies. *Nature*, *1*(2008), 1–6.

Adriko, J., Aritua, V., Mortensen, C. N., Tushemereirwe, W. K., Kubiriba, J., & Lund, O. S. (2012). Multiplex PCR for specific and robust detection of *Xanthomonas campestris* pv. *musacearum* in pure culture and infected plant material. *Plant Pathology*, *61*(3), 489–497. https://doi.org/10.1111/j.1365-3059.2011.02534.x

Adriko, J., Mbega, E., Kubiriba, J., Lund, O. S., Mabagala, R. B., Mondjana, a M., … Mortensen, C. N. (2011). Sampling procedures for the diagnosis of banana Xanthomonas wilt. *Technical Bulletin. Danish Seed Health Centre for Developing Countries, University of Copenhagen*, (2011).

Agritrade. (2013). *Executive brief Update: Banana sector*. *The Technical Centre for Agricultural and Rural Cooperation (CTA):http://agritrade.cta.int/*.

Ahern, S. J. (2013). *Novel Virulent Phages for Xylella fastidiosa and other members of the Xanthomonadaceae*.

Alfano, J. R., & Collmer, A. (2004). Type III secretion system effector proteins: Double agents in bacterial disease. *Annu. Rev. Phytopathol*, *42*, 385–414. https://doi.org/10.1146/annurev.phyto.42.040103.110731

Ali, A., Soares, S. C., Barbosa, E., Santos, A. R., Barh, D., Bakhtiar, S. M., … David, W. (2013). Bacteriology & Parasitology Microbial Comparative Genomics : An Overview of Tools and Insights Into The Genus Corynebacterium. *Journal of Bacteriology and Parasitology*, *4*(2), 167. https://doi.org/10.4172/2155-9597.1000167

Alikhan, N., Petty, N. K., Zakour, N. L. Ben, & Beatson, S. A. (2011). BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics*, *12*(1), 402. https://doi.org/10.1186/1471-2164-12-402

Alm, E. J., Huang, K. H., Price, M. N., Koche, R. P., Keller, K., Dubchak, I. L., & Arkin, A. P. (2005). The Microbes Online Web site for comparative genomics. *Genome Research*, *15*, 1015–1022. https://doi.org/10.1101/gr.3844805.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic Local

Alignment Search Tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Ansorge, W. J. (2009). Next-generation DNA sequencing techniques. *New Biotechnology*, *25*(4), 195–203. https://doi.org/10.1016/j.nbt.2008.12.009

Aritua, V, Parkinson, N., Thwaites, R., Heeney, J. V, Jones, D. R., Tushemereirwe, W., … 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. https://doi.org/10.1111/j.1365-3059.2007.01687.x

Aritua, Valente, Harrison, J., Sapp, M., Buruchara, R., Smith, J., & Studholme, D. J. (2015). Genome sequencing reveals a new lineage associated with lablab bean and genetic exchange between *Xanthomonas axonopodis* pv. *phaseoli* and *Xanthomonas fuscans* subsp. *fuscans*. *Frontiers in Microbiology*, *6*, 1080. https://doi.org/10.3389/fmicb.2015.01080

Arnold, R., Brandmaier, S., Kleine, F., Tischler, P., & Heinz, E. (2009). Sequence-Based Prediction of Type III Secreted Proteins. *PLoS Pathogens*, *5*(4). https://doi.org/10.1371/journal.ppat.1000376

Astua-monge, G., Freitas-astua, J., Bacocina, G., & Roncoletta, J. (2005). Expression Profiling of Virulence and Pathogenicity Genes of *Xanthomonas axonopodis* pv. *citri*. *Journal of Bacteriology*, *187*(3), 1201–1205. https://doi.org/10.1128/JB.187.3.1201

Astua-monge, G., Minsavage, G. V, Stall, R. E., Davis, M. J., Bonas, U., & Jones, J. B. (2000). Resistance of Tomato and Pepper to T3 Strains of *Xanthomonas campestris* pv. *vesicatoria* Is Specified by a Plant-Inducible Avirulence Gene. *Molecular Plant-Microbe Interactions*, *13*(9), 911–921.

Azhar M, H.-H. J. (2008). Genomes, diversity and resistance gene analogues in Musa species. *Cytogenetic Genome Research*, *121*(1), 59–66.

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., … Zagnitko, O. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*, *9*(1), 75. https://doi.org/10.1186/1471-2164-9-75

Bagamba, F., Burger, K., & Kuyvenhoven, A. (2007). Determinants of Smallholder Farmer Labour Allocation Decisions in Uganda. *The 106th Seminar of the EAAE, 25-27 October*

2007, (October), 1–22.

Bart, R., Cohn, M., Mccallum, E. J., Shybut, M., Petriello, A., Krasileva, K., … Chen, J. (2012). High-throughput genomic sequencing of cassava bacterial blight strains identifies conserved effectors to target for durable resistance. *Proceedings of the National Academy of Sciences*, *109*(32), 13130–13130. https://doi.org/10.1073/pnas.1211014109

Bartetzko, V., Sonnewald, S., Vogel, F., Hartner, K., Stadler, R., Hammes, U. Z., & Börnke, F. (2009). The *Xanthomonas campestris* pv. *vesicatoria* Type III Effector Protein XopJ Inhibits Protein Secretion: Evidence for Interference with Cell Wall–Associated Defense Responses. *Molecular Plant-Microbe Interactions*, *22*(6), 655–664. https://doi.org/10.1094/mpmi-22-6-0655

Baxter, J. C., & Funnell, B. E. (2014). Plasmid Partition Mechanisms. *Microbiology Spectrum*, *2*(6), 1–20.

Bent, A. F., & Mackey, D. (2007). Elicitors, Effectors, and R Genes: The New Paradigm and a Lifetime Supply of Questions. *Annu. Rev. Phytopathol.*, *45*, 399–436. https://doi.org/10.1146/annurev.phyto.45.062806.094427

Biek, R., Pybus, O. G., Lloyd-Smith, J. O., and Didelot, X. (2015). Measurably evolving pathogens in the genomic era. *Trends Ecol Evol*, *30*(6), 306–313. https://doi.org/10.1016/j.tree.2015.03.009.Measurably

Biruma, M., Pillay, M., Tripathi, L., Blomme, G., Abele, S., Bandyopadhyay, R., … Eden-green, S. (2007). Banana Xanthomonas wilt : a review of the disease , management strategies and future research directions. *African Journal of Biotechnology*, *6*(April), 953–962.

Boucher, C., & Roby, D. (2009). Pathogenicity determinants of Ralstonia solanacearum and their plant targets. In *Agrobiosciences, interactions and biodiversity* (p. 148).

Buell, C. R. (2002). *Interactions between Xanthomonas Species and Arabidopsis thaliana*. *The Arabidopsis Book*. https://doi.org/10.1199/tab.0031

Buell, C. R., Joardar, V., Lindeberg, M., Selengut, J., Paulsen, I. T., Gwinn, M. L., … Collmer, A. (2003). The complete genome sequence of the Arabidopsis and tomato pathogen Pseudomonas syringae pv . tomato DC3000. *PNAS*, *100*(18), 10181–10186.

Burdman, S., Bahar, O., Parker, J. K., & Fuente, L. D. La. (2011). Involvement of Type IV Pili

in Pathogenicity of Plant Pathogenic Bacteria. *Genes*, *2*, 706–735. https://doi.org/10.3390/genes2040706

Büttner, D., & Bonas, U. (2003). Common infection strategies of plant and animal pathogenic bacteria. *Current Opinion in Plant Biology*, *6*(4), 312–319. https://doi.org/10.1016/S1369-5266(03)00064-5

Büttner, D., & Bonas, U. (2006). Who comes first? How plant pathogenic bacteria orchestrate type III secretion. *Current Opinion in Microbiology*, *9*(2), 193–200. https://doi.org/10.1016/j.mib.2006.02.006

Büttner, D., Noël, L., Thieme, F., & Bonas, U. (2003). Genomic approaches in *Xanthomonas campestris* pv. *vesicatoria* allow fishing for virulence genes. *Journal of Biotechnology*, *106*(2–3), 203–214. https://doi.org/10.1016/j.jbiotec.2003.07.012

Buttner, D, & Bonas, U. (2002). Getting across--bacterial type III effector proteins on their way to the plant cell. *Embo Journal*, *21*(20), 5313–5322. https://doi.org/10.1093/emboj/cdf536

Buttner, Daniela, & Bonas, U. (2009). Regulation and secretion of Xanthomonas virulence factors. *FEMS Microbiology Reviews*, *34*, 107–133. https://doi.org/10.1111/j.1574-6976.2009.00192.x

Buttner, Daniela, & He, S. Y. (2009). Type III Protein Secretion in Plant Pathogenic Bacteria. *Plant Physiology*, *150*(August), 1656–1664. https://doi.org/10.1104/pp.109.139089

Cai, R., Lewis, J., Yan, S., Liu, H., Clarke, C. R., Campanile, F., … Vinatzer, B. A. (2011). The Plant Pathogen *Pseudomonas syringae* pv. *tomato* Is Genetically Monomorphic and under Strong Selection to Evade Tomato Immunity. *PLoS Pathogens*, *7*(8), e1002130. https://doi.org/10.1371/journal.ppat.1002130

Carter, B. A., Reeder, R., Mgenzi, S. R., Kinyua, Z. M., Mbaka, J. N., Doyle, K., … Smith, J. J. (2010). Identification of *Xanthomonas vasicola* (formerly *X. campestris* pv. *musacearum*) , causative organism of banana xanthomonas wilt , in Tanzania , Kenya and Burundi First report of Xanthomonas hortorum pv . pelargonii causing bacterial blight of geranium , 3059. https://doi.org/10.1111/j.1365-3059.2009.02124.x

Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M. A., Barrell, B. G., & Parkhill, J. (2005). ACT: The Artemis comparison tool. *Bioinformatics*, *21*(16), 3422–3423. https://doi.org/10.1093/bioinformatics/bti553

Casabuono, A., Petrocelli, S., Ottado, J., Orellano, E. G., & Couto, A. S. (2011). Structural Analysis and Involvement in Plant Innate Immunity of *Xanthomonas axonopodis* pv. *citri* Lipopolysaccharide. *The Journal of Biological Chemistry*, *286*, 25628–25643. https://doi.org/10.1074/jbc.M110.186049

Castañeda, A., Reddy, J. D., El-yacoubi, B., & Gabriel, D. W. (2005). Mutagenesis of all Eight avr Genes in *Xanthomonas campestris* pv. *campestris* Had No Detected Effect on Pathogenicity, But One avr Gene Affected Race Specificity. *Molecular Plant-Microbe Interactions*, *18*(12), 1306–1317.

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

Chagnot, C., Zorgani, M. a, Astruc, T., & Desvaux, M. (2013). Proteinaceous determinants of surface colonization in bacteria: bacterial adhesion and biofilm formation from a protein secretion perspective. *Frontiers in Microbiology*, *4*, 303. https://doi.org/10.3389/fmicb.2013.00303

Chalupowicz, L., Dror, O., Eichenlaub, R., Gartemann, K., Sessa, G., & Barash, I. (2009). Sequential Expression of Bacterial Virulence and Plant Defense Genes During Infection of Tomato with Clavibacter michiganensis subsp . michiganensis. *Phytopathology*, *100*(3), 252–261.

Champoiseau, P., Daugrois, J., Pieretti, I., Cociancich, S., Royer, M., & Rott, P. (2006). High Variation in Pathogenicity of Genetically Closely Related Strains of *Xanthomonas albilineans*, the Sugarcane Leaf Scald Pathogen , in Guadeloupe. *Phytopathology*, *96*(10), 1081–1091.

Chan, V. L., Sherman, P. M., & Bourke, B. (2006). *Bacterial Genomes and Infectious Diseases*. Humana Press.

Chang, C., Shih, H., & Su, C. (2012). Diseases of important crops, a review of the causal fastidious prokaryotes and their insect vectors. *Plant Pathology Bulletin*, *21*, 1–10. Retrieved from http://ir.tari.gov.tw:8080/handle/345210000/6024

Chen, L., Xiong, Z., Sun, L., Yang, J., & Jin, Q. (2012). VFDB 2012 update: Toward the genetic diversity and molecular evolution of bacterial virulence factors. *Nucleic Acids Research*, *40*(D1), 641–645. https://doi.org/10.1093/nar/gkr989

Cianciotto, N. P. (2005). Type II secretion: A protein secretion system for all seasons. *Trends in Microbiology*, *13*(12), 581–588. https://doi.org/10.1016/j.tim.2005.09.005

Collmer, A., Schneider, D. J., & Lindeberg, M. (2009). Lifestyles of the Effector Rich : Genome-Enabled Characterization of Bacterial Plant Pathogens 1. *Society*, *150*(August), 1623–1630. https://doi.org/10.1104/pp.109.140327

Corbett, M., Virtue, S., Bell, K., Birch, P., Burr, T., Hyman, L., … Salmond, G. (2005). Identification of a new quorum-sensing-controlled virulence factor in *Erwinia carotovora* subsp. *atroseptica* secreted via the type II targeting pathway. *Molecular Plant-Microbe Interactions*, *18*(4), 334–342. https://doi.org/10.1094/MPMI-18-0334

Cornelis, G. R., & Gijsegem, F. Van. (2000). Assembly and Function of Type III Secretory Systems. *Annu. Rev. Microbiol*, *54*, 735–774.

Costa, T. R. D., Felisberto-rodrigues, C., Meir, A., Prevost, M. S., Redzej, A., Trokter, M., & Waksman, G. (2015). Secretion systems in Gram-negative insights. *Nature Publishing Group*, *13*(6), 343–359. https://doi.org/10.1038/nrmicro3456

Cunnac, S., Bolot, S., Forero Serna, N., Ortiz, E., Szurek, B., Noël, L. D., … Koebnik, R. (2013). High-Quality Draft Genome Sequences of Two *Xanthomonas citri* pv. *malvacearum* Strains. *Genome Announcements*, *1*(4), 4–13. https://doi.org/10.1128/genomeA.00674-13

Cunnac, S., Occhialini, A., Barberis, P., Boucher, C., & Genin, S. (2004). Inventory and functional analysis of the large Hrp regulon in *Ralstonia solanacearum*: Identification of novel effector proteins translocated to plant host cells through the type III secretion system. *Molecular Microbiology*, *53*(1), 115–128. https://doi.org/10.1111/j.1365-2958.2004.04118.x

da Silva, a C. R., Ferro, J. a, Reinach, F. C., Farah, C. S., Furlan, L. R., Quaggio, R. B., … Kitajima, J. P. (2002). Comparison of the genomes of two Xanthomonas pathogens with differing host specificities. *Nature*, *417*(6887), 459–463. https://doi.org/10.1038/417459a

Darsonval, A., Darrasse, A., Meyer, D., Demarty, M., Durand, K., Bureau, C., … Jacques, M.

(2008). The Type III Secretion System of *Xanthomonas fuscans* subsp. *fuscans* Is Involved in the Phyllosphere Colonization Process and in Transmission to Seeds of Susceptible Beans. *Applied and Environmental Microbiology*, *74*(9), 2669–2678. https://doi.org/10.1128/AEM.02906-07

De Langhe, E., Vrydaghs, L., De Maret, P., Perrier, X., & Denham, T. (2009). Why bananas matter: An introduction to the history of banana domestication. *Ethnobotany Research and Applications*, *7*, 165–178.

Debroy, S., Thilmony, R., Kwack, Y., Nomura, K., & He, S. Y. (2004). A family of conserved bacterial effectors inhibits salicylic acid-mediated basal immunity and promotes disease necrosis in plants. *PNAS*, *101*(26), 9927–9932.

Denancé, N., Lahaye, T., & Noël, L. D. (2016). Editorial: Genomics and Effectomics of the crop killer Xanthomonas. *Frontiers in Plant Science*, *7*(February), 2–3. https://doi.org/10.3389/fpls.2016.00071

Doorn, J. Van, Hollinger, T. C., & Oudega, B. (2001). Analysis of the Type IV Fimbrial-Subunit Gene fimA of *Xanthomonas hyacinthi*: Application in PCR-Mediated Detection of Yellow Disease in Hyacinths. *Applied and Environmental Microbiology*, *67*(2), 598–607. https://doi.org/10.1128/AEM.67.2.598-607.2001

Douzi, B., Filloux, A., & Voulhoux, R. (2012). On the path to uncover the bacterial type II secretion system. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1592), 1059–1072. https://doi.org/10.1098/rstb.2011.0204

Dunger, G., Guzzo, C. R., Andrade, M. O., Jones, J. B., & Farah, C. S. (2014). Xanthomonas citri subsp. citri Type IV Pilus Is Required for Twitching Motility, Biofilm Development, and Adherence. *Molecular Plant-Microbe Interactions : MPMI*, *27*(10), 1132–1147. https://doi.org/10.1094/MPMI-06-14-0184-R

Dutilh, B. E., Backus, L., Edwards, R. A., Wels, M., Bayjanov, J. R., & Van Hijum, S. A. F. T. (2013). Explaining microbial phenotypes on a genomic scale: Gwas for microbes. *Briefings in Functional Genomics*, *12*(4), 366–0380. https://doi.org/10.1093/bfgp/elt008

Eddy, S. (2003). *HMMER User's Guide: Biological sequence analysis using profile hidden Markov models*. (S. Eddy, Ed.) (2.3.2). HHMI/Washington University School of Medicine.

Egan, A. N., Schlueter, J., & Spooner, D. (2012). Applications of Next-Generation Sequencing in plant biology. *American Journal of Botany*, *99*(2), 175–185. https://doi.org/10.3732/ajb.1200020

Erbs, G., & Newman, M. A. (2003). The role of lipopolysaccharides in induction of plant defence responses. *Molecular Plant Pathology*, *4*(5), 421–425. https://doi.org/10.1046/j.1364-3703.2003.00179.x

Escalant, J., Markham, R., Roux, N., & Staver, C. (2004). bacterial wilt Info Musa Can model plants help banana improvement through biotechnology ? Diseases and pests : A review of their importance and management. *The International Journal on Banana and Plantain*, *13*(2), 48.

FAO. (2010). Intergovernmental Group on Bananas and Tropical Fruits. *Commitee on Commodity Problems*, (February), 1–6.

Ferreira-Tonin, M., Rodrigues-Neto, J., Harakava, R., & Lanza Destéfano, S. A. (2012). Phylogenetic analysis of Xanthomonas based on partial rpoB gene sequences and species differentiation by PCR-RFLP. *International Journal of Systematic and Evolutionary Microbiology*, *62*(6), 1419–1424. https://doi.org/10.1099/ijs.0.028977-0

Galán, J. E., & Collmer, A. (2007). Type III Secretion Machines : Bacterial Devices for Protein Delivery into Host Cells. *Science*, *1322*(1999). https://doi.org/10.1126/science.284.5418.1322

Garnier, M. (2002). Phloem-and xylem-restricted plant pathogenic bacteria. *Plant Science*, *163*, 1083–1098.

Gilbert, G. S., & Webb, C. O. (2007). Phylogenetic signal in plant pathogen-host range. *Proceedings of the National Academy of Sciences*, *104*(12), 4979–4983. https://doi.org/10.1073/pnas.0607968104

Giltner, C. L., Nguyen, Y., & Burrows, L. L. (2012). Type IV pilin proteins: Versatile molecular modules. *Microbiology and Molecular Biology Reviews : MMBR*, *76*(4), 740–772. https://doi.org/10.1128/MMBR.00035-12

Gomes, L. H., Maria, K., Duarte, R., Andrino, F. G., & Cesar, F. (1988). A simple method for DNA isolation from Xanthomonas spp. *Scientia Agricola*, *57*(3), 553–555.

Grant, J. R., & Stothard, P. (2008). The CGView Server: a comparative genomics tool for

circular genomes. *Nucleic Acids Research*, *36*, 181–184. https://doi.org/10.1093/nar/gkn179

Grant, S. R., Fisher, E. J., Chang, J. H., Mole, B. M., & Dangl, J. L. (2006). Subterfuge and Manipulation : Type III Effector Proteins of Phytopathogenic Bacteria. *Annu. Rev. Microbiol*, *60*, 425–449. https://doi.org/10.1146/annurev.micro.60.080805.142251

Green, S., Studholme, D. J., Laue, B. E., Dorati, F., Lovell, H., Arnold, D., … Kamoun, S. (2010). Comparative Genome Analysis Provides Insights into the Evolution and Adaptation of *Pseudomonas syringae* pv. *aesculi* on Aesculus hippocastanum These authors contributed equally to this work . *PLoS ONE*, *5*(4), 1–34.

Greenberg, J. T., & Vinatzer, B. A. (2003). Identifying type III effectors of plant pathogens and analyzing their interaction with plant cells. *Current Opinion in Microbiology*, *6*, 20–28. https://doi.org/10.1016/S1369-5274(02)00004-8

Gu¨rlebeck, D., Thieme, F., & Bonas, U. (2006). Type III effector proteins from the plant pathogen Xanthomonas and their role in the interaction with the host plant. *Journal of Plant Physiology*, *163*, 233–255. https://doi.org/10.1016/j.jplph.2005.11.011

Hajri, A., Joël, F. P., Marion, F.-L.-S., Bonneau, S., Stéphane, P., Boureau, T., … Charles, M. (2012). Type Three Effector Gene Distribution and Sequence Analysis Provide New Insights into the Pathogenicity of Plant-Pathogenic *Xanthomonas arboricola*. *Applied and Environmental Microbiology*, 371–384. https://doi.org/10.1128/AEM.06119-11

Hakizima, S., & Gallagher, S. (2008). *Fighting Banana Xanthomonas Wilt Disease in Rwanda: A Combination of FFW, Seed Fairs, Eradication, Training & Education Banana*.

Hall, B. G. (2013). Building phylogenetic trees from molecular data with MEGA. *Molecular Biology and Evolution*, *30*(5), 1229–1235. https://doi.org/10.1093/molbev/mst012

Hardison, R. C. (2003). Comparative genomics. *PLoS Biology*, *1*(2), 156–160. https://doi.org/10.1371/journal.pbio.0000058

Hardoim, P. R., van Overbeek, L. S., & Elsas, J. D. Van. (2008). Properties of bacterial endophytes and their proposed role in plant growth. *Trends in Microbiology*, *16*(10), 463–471. https://doi.org/10.1016/j.tim.2008.07.008

Harrison, J., & Studholme, D. J. (2014). Draft genome sequence of *Xanthomonas axonopodis* pv. *vasculorum* NCPPB 900. *FEMS Microbiology Letters*, *360*(2), 113–116.

https://doi.org/10.1111/1574-6968.12607

Hashimi, S. M., Wall, M. K., Smith, A. B., Maxwell, A., & Birch, R. G. (2007). The phytotoxin albicidin is a novel inhibitor of DNA gyrase. *Antimicrobial Agents and Chemotherapy*, *51*(1), 181–187. https://doi.org/10.1128/AAC.00918-06

He, S. Y. (1998). Type III protein secretion systems in plant and animal bacteria. *Annu. Rev. Phytopathol*, *36*, 363–392.

Henao, J., & Baanante, C. (2006). Agricultural Production and Soil Nutrient Mining in Africa Prepared by Summary of the Paper Agricultural Production and Soil Nutrient Mining in Africa Policy Development. *IFDC Technical Bulletin*, (March).

Henderson, I. R., & Navarro-garcia, F. (2004). Type V Protein Secretion Pathway : the Autotransporter Story. *Microbiology and Molecular Biology Reviews : MMBR*, *68*(4), 692–744. https://doi.org/10.1128/MMBR.68.4.692

Ho Sui, S. J., Fedynak, A., Hsiao, W. W. L., Langille, M. G. I., & Brinkman, F. S. L. (2009). The association of virulence factors with genomic islands. *PloS One*, *4*(12), e8094. https://doi.org/10.1371/journal.pone.0008094

Hultgren, S. J., Abraham, S. N., Caparon, M. G., Falk, P., St. Geme III, J. W., & Normark, S. (1993). Pilus and non-pilus bacterial adhesins:assembly and function in cell recognition. *Cell*, *73*(93), 887–901.

Ib, H., Schmitt, L., & Young, J. (2015). Type 1 protein secretion in bacteria , the ABC - transporter dependent pathway (review). PubMed Commons. *Mol Membr Biol*, *22*, 29–39.

Illumina. (2012). Genome Analyzer IIx User Guide (SCS v2.9), (July), 159. Retrieved from https://icom.illumina.com/download/summary/NhgBUf-w1USCWAimIVWgTA

Illumina. (2013). *An Introduction to Next-Generation Sequencing Technology*.

Ivey, M. L. L., Pathology, P., & Ohio, T. (2010). A Polymerase Chain Reaction Assay for the Detection of *Xanthomonas campestris* pv. *musacearum* in Banana. *Plant Disease*, *94*(January), 109–114.

Jalan, N., Kumar, D., Andrade, M. O., Yu, F., Jones, J. B., Graham, J. H., … Wang, N. (2013). Comparative genomic and transcriptome analyses of pathotypes of *Xanthomonas citri*

subsp. *citri* provide insights into mechanisms of bacterial virulence and host range. *BMC Genomics*, *14*(1), 1. https://doi.org/10.1186/1471-2164-14-551

Jha, G., Rajeshwari, R., & Sonti, R. V. (2005). Bacterial Type Two Secretion System Secreted Proteins : Double-Edged Swords for Plant Pathogens. *Molecular Plant-Microbe Interactions*, *18*(9), 891–898.

Jiang, W., Jiang, B.-L., Xu, R.-Q., Huang, J.-D., Wei, H.-Y., Jiang, G.-F., … Tang, J.-L. (2009). Identification of six type III effector genes with the PIP box in *Xanthomonas campestris* pv. *campestris* and five of them contribute individually to full pathogenicity. *Molecular Plant-Microbe Interactions : MPMI*, *22*(11), 1401–1411. https://doi.org/10.1094/MPMI-22-11-1401

Jin, F., Conrad, J. C., Gibiansky, M. L., & Wong, G. C. L. (2011). Bacteria use type-IV pili to slingshot on surfaces. *PNAS*, 2–7. https://doi.org/10.1073/pnas.1105073108

Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. *Nature*, *444*(November), 323–329. https://doi.org/10.1038/nature05286

Kamper, S. M., French, W. J., & De Kloet, S. R. (1985). Genetic relationships of some fastidious xylem-limited bacteria. *International Journal of Systematic Bacteriology*, *35*(2), 185–188. https://doi.org/10.1099/00207713-35-2-185

Kang, Y., Jelenska, J., Cecchini, N. M., Li, Y., Lee, M. W., Kovar, D. R., & Greenberg, J. T. (2014). HopW1 from *Pseudomonas syringae* Disrupts the Actin Cytoskeleton to Promote Virulence in Arabidopsis. *PLoS Pathogens*, *10*(6), 1–10. https://doi.org/10.1371/journal.ppat.1004232

Karamura, D. A. (1998). *Numerical taxonomy studies of the East African Highland bananas (Musa AAA- East Africa) in Uganda*. *Inibap, University of Reading*.

Karamura, E., Osiru, M., Blomme, G., Lusty, C., & Picq, C. (2006). *Developing a regional strategy to address the outbreak of banana Xanthomonas wilt in East and Central Africa*. *Proceedings of the Banana Xanthomonas wilt regional preparedness and strategy development workshop held in Kampala, Uganda - 14-18 February 2005. International Network for the Improvement of Banana and Plantain, Montpellier, France. Graphic*. Retrieved from http://agris.fao.org/agris-search/search/display.do?f=2008/QJ/QJ0802.xml;QJ2007000037

Karamura, E., & Tinzaara, W. (2007). Management of Banana Xanthomonas Wilt in East and Central Africa. In *Proceedings of the workshop on review of the strategy for the management of banana xanthomonas wilt* (p. 95).

Karamura, G., Smith, J., Studholme, D., Kubiriba, J., & Karamura, E. (2015). Comparative pathogenicity studies of the *Xanthomonas vasicola* species on maize , sugarcane and banana. *African Journal of Plant Science*, *9*(September), 385–400. https://doi.org/10.5897/AJPS2015.1327

Katoh, K., Kuma, K. I., Toh, H., & Miyata, T. (2005). MAFFT version 5: Improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research*, *33*(2), 511–518. https://doi.org/10.1093/nar/gki198

Kay, S., & Bonas, U. (2009). How Xanthomonas type III effectors manipulate the host plant. *Current Opinion in Microbiology*, 37–43. https://doi.org/10.1016/j.mib.2008.12.006

Kilimo Trust. (2012). *Banana Value Chain(s) in East Africa : Consumption, Productivity and Challenges*.

Kline, K. A., Folker, S., Dahlberg, S., Normark, S., & Henriques-Normark, B. (2009). Bacterial Adhesins in Host-Microbe Interactions. *Cell Host and Microbe*, *5*(6), 580–592. https://doi.org/10.1016/j.chom.2009.05.011

Kodama, Y., Shumway, M., & Leinonen, R. (2012). The sequence read archive : explosive growth of sequencing data, *40*(October 2011), 2011–2013. https://doi.org/10.1093/nar/gkr854

Kohn, L. M. (2004). Applying comparative genomics to plant disease epidemiology. *Phytoprotection*, *85*(1), 45–48.

Kolde, R. (2015). *Package ‘ pheatmap .’ Implementation of heatmaps that offers more control over dimensions and appearance*.

Kubiriba, J., Karamura, E. B., Tushemereirwe, W. K., & Tinzaara, W. (2012). Community mobilization : A key to effective control of banana xanthomonas wilt. *Journal of Development and Agricultural Economics*, *4*(5), 125–131. https://doi.org/10.5897/JDAE11.098

Kubiriba, J., & Tushemereirwe, W. K. (2014). Approaches for the control of banana Xanthomonas wilt in East and Central Africa. *African Journal of Plant Science Review*,

8(August), 398–404. https://doi.org/10.5897/AJPS2013.1106

Kumar, S., Banks, T. W., & Cloutier, S. (2012). SNP Discovery through Next-Generation Sequencing and Its Applications. *International Journal of Plant Genomics*, 15. https://doi.org/10.1155/2012/831460

Lahaye, T., & Bonas, U. (2001). Molecular secrets of bacterial type III effector proteins. *Trends in Plant S Cience*, *6*(10), 479–485.

Land, M., Hauser, L., Jun, S. R., Nookaew, I., Leuze, M. R., Ahn, T. H., … Ussery, D. W. (2015). Insights from 20 years of bacterial genome sequencing. *Functional and Integrative Genomics*, *15*(2), 141–161. https://doi.org/10.1007/s10142-015-0433-4

Lang, J. M., DuCharme, E., Ibarra Caballero, J., Luna, E., Hartman, T., Ortiz-Castro, M., … Leach, J. E. (2017). Detection and Characterization of *Xanthomonas vasicola* pv. *vasculorum* (Cobb 1894) comb. nov. Causing Bacterial Leaf Streak of Corn in the United States . *Phytopathology*, *107*(11), 1312–1321. https://doi.org/10.1094/phyto-05-17-0168-r

Leach, J. E., & White, F. F. (1996). Bacterial avirulence genes. *Annual Review of Phytopatholgy*, *34*, 153–179.

Lefeuvre, P., Cellier, G., Remenant, B., Chiroleu, F., & Prior, P. (2013). Constraints on Genome Dynamics Revealed from Gene Distribution among the *Ralstonia solanacearum* Species. *PloS One*, *8*(5). https://doi.org/10.1371/journal.pone.0063155

Leinonen, R., Sugawara, H., & Shumway, M. (2011). The Sequence Read Archive, *39*(November 2010), 2010–2012. https://doi.org/10.1093/nar/gkq1019

Lemaire, C., Hajri, A., Brin, C., & Hunault, G. (2009). A « Repertoire for Repertoire » Hypothesis : Repertoires of Type Three Effectors are Candidate Determinants of Host Specificity in Xanthomonas. *PLoS ONE*, *4*(8), e6632. https://doi.org/10.1371/journal.pone.0006632

Leyns, F., De Cleene, M., Swings, J. G., & De Ley, J. (1984). The host range of the genus Xanthomonas. *The Botanical Review*, *50*(3), 308–356. https://doi.org/10.1007/BF02862635

Li, C., Brown, I., Mans, J., Stevens, C., Boureau, T., Romantschuk, M., & Taira, S. (2002). The Hrp pilus of *Pseudomonas syringae* elongates from its tip and acts as a conduit for

translocation of the effector protein HrpZ. *European Molecular Biology Orgnisation*, *21*(8), 1909–1915.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, J., & Wang, N. (2011). Genome-Wide Mutagenesis of *Xanthomonas axonopodis* pv. *citri* Reveals Novel Genetic Determinants and Regulation Mechanisms of Biofilm Formation. *PloS One*, *6*(7). https://doi.org/10.1371/journal.pone.0021804

Lindgren, P. B. (1997). The role of hrp genes during plant bacterial interactions. *Annual Review of Phytopathology*, *35*, 129–152. https://doi.org/10.1146/annurev.phyto.35.1.129

Llaca, V. (2012). Sequencing Technologies and their use in plant biotechnology and breeding. *DNA Sequencing and Applications*, 174. https://doi.org/10.5772/37918

Lorenz, C., Schulz, S., Wolsch, T., Rossier, O., Bonas, U., & Bu, D. (2008). HpaC Controls Substrate Specificity of the Xanthomonas Type III Secretion System. *PLoS Pathogens*, *4*(6). https://doi.org/10.1371/journal.ppat.1000094

Lu, H., Patil, P., Sluys, M. Van, White, F. F., Ryan, R. P., Dow, J. M., … J, A. (2008). Acquisition and Evolution of Plant Pathogenesis – Associated Gene Clusters and Candidate Determinants of Tissue-Specificity in Xanthomonas. *Gene*, *3*(11). https://doi.org/10.1371/journal.pone.0003828

Lu, M., An, H., & Li, L. (2016). Genome survey sequencing for the characterization of the genetic background of rosa roxburghii tratt and leaf ascorbate metabolism genes. *PLoS ONE*, *11*(2), 1–17. https://doi.org/10.1371/journal.pone.0147530

Mackey, D., & McFall, A. J. (2006). MAMPs and MIMPs: Proposed classifications for inducers of innate immunity. *Molecular Microbiology*, *61*(6), 1365–1371. https://doi.org/10.1111/j.1365-2958.2006.05311.x

Manivannan, A., Kim, J. H., Yang, E. Y., Ahn, Y. K., Lee, E. S., Choi, S., & Kim, D. S. (2018). Next-Generation Sequencing Approaches in Genome-Wide Discovery of Single Nucleotide Polymorphism Markers Associated with Pungency and Disease Resistance in Pepper. *BioMed Research International*, *2018*. https://doi.org/10.1155/2018/5646213

Martin, G. B., Bogdanove, A. J., & Sessa, G. (2003). Understanding the functions of plant

disease resistance proteins. *Annu. Rev. Plant Biol*, *54*, 23–61. https://doi.org/10.1146/annurev.arplant.54.031902.135035

Merrell, D. S., & Camilli, A. (2000). Detection and analysis of gene expression during infection by in vivo expression technology. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *355*(1397), 587–599. https://doi.org/10.1098/rstb.2000.0600

Metzker, M. L. (2008). Advances in Next Generation DNA Sequencing Technologies. In J. R. Brown (Ed.), *Comparative genomics: Basic and Applied research* (p. 409). Taylor & Francis Group.

Meyer, D. F., & Bogdanove, A. J. (1995). Genomics-driven Advances in Xanthomonas biology. In *xanthomonas Genomics* (pp. 147–161).

Mhedbi-Hajri, N., Darrasse, A., Pigné, S., Durand, K., Fouteau, S., Barbe, V., … Jacques, M. A. (2011). Sensing and adhesion are adaptive functions in the plant pathogenic xanthomonads. *BMC Evolutionary Biology*, *11*(1), 67. https://doi.org/10.1186/1471-2148-11-67

Mhedbi-Hajri, N., Hajri, A., Boureau, T., Darrasse, A., Durand, K., Brin, C., … Jacques, M.-A. (2013). Evolutionary History of the Plant Pathogenic Bacterium *Xanthomonas axonopodis*. *PloS One*, *8*(3), e58474. https://doi.org/10.1371/journal.pone.0058474

Microbewiki. (2015). *Type Six Secretion System (T6SS)*.

Midha, S., & Patil, P. B. (2014). Genomic Insights into the Evolutionary Origin of *Xanthomonas axonopodis* pv. *citri* and Its Ecological Relatives. *Applied and Environmental Microbiology*, *80*(20), 6266–6279. https://doi.org/10.1128/AEM.01654-14

Moreira, L. M., Laia, M. L. De, Souza, R. F. De, Zaini, P. A., Cr, A., Silva, A. M., & Ferro, J. A. (2010). Development and validation of a Xanthomonas axonopodis pv . citri DNA microarray platform (XACarray) generated from the shotgun libraries previously used in the sequencing of this bacterial genome. *BMC Research Notes*, *3*(150).

Moreira, L. M., Souza, R. F. De, Jr, N. F. A., Setubal, C., Oliveira, J. C. F., Furlan, L. R., … Silva, A. C. R. (2004). Comparative genomics analyses of citrus -associated bacteria. *Annual Review of Phytopatholgy*, *42*, 163–184.

https://doi.org/10.1146/annurev.phyto.42.040803.140310

Moreira, L. M., Souza, R. F. D. E., & Digiampietri, L. A. (2005). Comparative Analyses of Xanthomonas and Xylella Complete Genomes. *A Journal of Integrative Biology*, *9*(1), 43–76.

Morozova, O., Hirst, M., & Marra, M. A. (2009). Applications of New Sequencing Technologies for Transcriptome Analysis. *Annual Review of Genomics and Human Genetics*, *10*, 135–151. https://doi.org/10.1146/annurev-genom-082908-145957

Mudgett, M. B. (2005). New Insights to the Function of Phytopathogenic Bacterial Type III Effectors in Plants. *Annual Review of Plant Biology*, *56*, 509–531. https://doi.org/10.1146/annurev.arplant.56.032604.144218

Mwangi, M. (2006). Banana Xanthomonas wilt : Symptoms , Spread & Management. ELEWA Publications Ltd.

Mwangi, M., Tinzaara, W., Ndungo, V., Namu, F., Ragama, P., & Bandyopadhyay, R. (2006). Comparative study of banana Xanthomonas wilt spread in mid and high altitudes of the Great Lakes region of Africa. *Conference on International Research for Development*.

Namukwaya, B., Tripathi, L., Tripathi, J. N., Arinaitwe, G., Mukasa, S. B., & Tushemereirwe, W. K. (2012). Transgenic banana expressing Pflp gene confers enhanced resistance to Xanthomonas wilt disease. *Transgenic Research*, *21*(4), 855–865. https://doi.org/10.1007/s11248-011-9574-y

Narayanasamy, P. (2008). Molecular Biology of Plant Disease Development. In *Molecular Biology in Plant Pathogenesis and Disease Management* (pp. 99–118).

Ndabamenye, T., Vanlauwe, B., van Asten, P. J. A., Blomme, G., Swennen, R., Uzayisenga, B., … Barnard, R. O. (2013). Influence of plant density on variability of soil fertility and nutrient budgets in low input East African highland banana (Musa spp. AAA-EA) cropping systems. *Nutrient Cycling in Agroecosystems*, *95*(2), 187–202. https://doi.org/10.1007/s10705-013-9557-x

Ndungo, V., Fiaboe, K.K.M. and Mwangi, M. (2008). Banana Xanthomonas wilt in the DR Congo : Impact , spread and management. *Journal of Applied Biosciences*, *1*(1), 1–7.

Ndungo, V., Eden-Green, S., Blomme, G., Crozier, J., & Smith, J. J. (2006). Presence of banana xanthomonas wilt (*Xanthomonas campestris* pv. *musacearum*) in the Democratic

Republic of Congo (DRC). *Plant Pathology*, *55*(2), 294. https://doi.org/10.1111/j.1365-3059.2005.01258.x

Noël, L., Thieme, F., Gäbler, J., Büttner, D., Bonas, U., Ga, J., & Bu, D. (2003). XopC and XopJ , Two Novel Type III Effector Proteins from Xanthomonas campestris XopC and XopJ , Two Novel Type III Effector Proteins from Xanthomonas campestris pv . vesicatoria. *Journal of Bacteriology (2003)*, *185*(24), 7092–7102. https://doi.org/10.1128/JB.185.24.7092

Odipio, J., Tusiime, G., Tripathi, L., & Aritua, V. (2009). Genetic homogeneity among Ugandan isolates of Xanthomonas campestris pv. musacearum revealed by randomly amplified polymorphic DNA analysis. *African Journal of Biotechnology*, *8*(21), 5652–5660. Retrieved from http://www.academicjournals.org/ajb/PDF/pdf2009/2Nov/Odipio et al.pdf

Orth, K. (2007). TTSS from Different Bacteria Function. *Microbe*, *2*(4), 183–186.

Orth, K., Xu, Z., Mudgett, M. B., Bao, Z. Q., Palmer, L. E., Bliska, J. B., … Dixon, J. E. (2000). Disruption of Signaling by Yersinia Effector YopJ , a Ubiquitin-Like Protein Protease. *Science*, *290*(November), 1594–1597.

Pareek, C. S., & Smoczynski, R. (2011). Sequencing technologies and genome sequencing. *Journal of Applied Genetics*, *52*, 413–435. https://doi.org/10.1007/s13353-011-0057-x

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*(12), 2881–2887. https://doi.org/10.1099/ijs.0.65220-0

Parkinson, N., Cowie, C., Heeney, J., & Stead, D. (2009). Phylogenetic structure of Xanthomonas determined by comparison of *gyrB* sequences. *International Journal of Systematic and Evolutionary Microbiology*, *59*(2), 264–274. https://doi.org/10.1099/ijs.0.65825-0

Patil, P. B., Bogdanove, A. J., & Sonti, R. V. (2007). The role of horizontal transfer in the evolution of a highly variable lipopolysaccharide biosynthesis locus in xanthomonads that infect rice, citrus and crucifers. *BMC Evolutionary Biology*, *7*(243). Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18053269

Patil, P. B., & Sonti, R. V. (2004). Variation suggestive of horizontal gene transfer at a lipopolysaccharide (lps) biosynthetic locus in Xanthomonas oryzae pv. oryzae, the bacterial leaf blight pathogen of rice. *BMC Microbiology*, *4*(1), 40. https://doi.org/10.1186/1471-2180-4-40

Pfeilmeier, S., Caly, D. L., & Malone, J. G. (2016). Bacterial pathogenesis of plants: future challenges from a microbial perspective: Challenges in Bacterial Molecular Plant Pathology. *Molecular Plant Pathology*, *17*(8), 1298–1313. https://doi.org/10.1111/mpp.12427

Pieretti, I., Cociancich, S., Bolot, S., Carrèr, S., Morisset, A., Rott, P., & Royer, M. (2015). Full genome sequence analysis of two isolates reveals a novel xanthomonas species close to the sugarcane pathogen Xanthomonas albilineans. *Genes*, *6*(3), 714–733. https://doi.org/10.3390/genes6030714

Pieretti, I., Royer, M., Barbe, V., Carrere, S., Koebnik, R., Cociancich, S., … Rott, P. (2009). The complete genome sequence of *Xanthomonas albilineans* provides new insights into the reductive genome evolution of the xylem-limited Xanthomonadaceae. *BMC Genomics*, *15*, 1–15. https://doi.org/10.1186/1471-2164-10-616

Pieretti, I., Royer, M., Barbe, V., Carrere, S., Koebnik, R., Couloux, A., … Cociancich, S. (2012). Genomic insights into strategies used by *Xanthomonas albilineans* with its reduced artillery to spread within sugarcane xylem vessels. BMC genomics, 13(1). https://doi.org/10.1186/1471-2164-13-658

Pizarro-Cerd, J., & Cossart, P. (2006). Bacterial adhesion and entry into host cells. *Cell*, *124*(4), 715–727. https://doi.org/10.1016/j.cell.2006.02.012

Ponciano, G., Ishihara, H., Tsuyumu, S., & Leach, J. E. (2003). Bacterial Effectors in plant disease and defense: keys to durable resistance? *Plant Disease*, *87*(11).

Potnis, N., Krasileva, K., Chow, V., Almeida, N. F., Patil, P. B., Ryan, R. P., … Jones, J. B. (2011). Comparative genomics reveals diversity among xanthomonads infecting tomato and pepper. *BMC Genomics*, *12*(1), 146. https://doi.org/10.1186/1471-2164-12-146

Prasannath, K. (2013). Review Article Pathogenicity and Virulence Factors of Phytobacteria. *Journal of Biosciences*, *1*(1), 24–33.

Preston, G. M., Studholme, D. J., & Caldelari, I. (2005). Profiling the secretomes of plant

pathogenic Proteobacteria. *FEMS Microbiology Reviews*, *29*(2), 331–360. https://doi.org/10.1016/j.femsre.2004.12.004

Puritz, J. B., Addison, J. A., & Toonen, R. J. (2012). Next-generation phylogeography: A targeted approach for multilocus sequencing of non-model organisms. *PLoS ONE*, *7*(3). https://doi.org/10.1371/journal.pone.0034241

Qhobela, M., & Claflin, L. E. (1988). Characterization of *Xanthomonas campestris* pv. *pennamericanum* pv. *nov*., Causal Agent of Bacterial Leaf Streak of Pearl Millet. *International Journal of Systemic Bacteriology*, *48*(4), 362–366.

Qhobela, M., & Leach, J. E. (1991). Characterisation of strains of *Xanthomonas campestris* pv. *holcicola* by PAGE of membrane proteins and by REA and RFLP analysis of genomic DNA. *Plant Disease*, *75*(1), 32–36.

Qi, Y., Xu, L., Dong, X., Yau, H., Ho, L., Koh, L., & Shochat, G. (2012). Functional Divergence of FimX in PilZ Binding and Type IV Pilus Regulation. *Journal of Microbiology*, *194*(21), 5922–5931. https://doi.org/10.1128/JB.00767-12

Qian, W., Jia, Y., Ren, S. X., He, Y. Q., Feng, J. X., Lu, L. F., … He, C. (2005). Comparative and functional genomic analyses of the pathogenicity of phytopathogen Xanthomonas campestris pv. campestris. *Genome Research*, *15*(6), 757–767. https://doi.org/10.1101/gr.3378705

Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., … Gu, Y. (2012). A tale of three next generation sequencing platforms : comparison of Ion Torrent , Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, *13*, 1471–2164.

Quinlan, A. R., & Hall, I. M. (2010). The BEDTools manual. *Genome*, *16*(6), 1–77. https://doi.org/10.1093/bioinformatics/btq033

Rademaker, J. L. W., Hoste, B., Louws, F. J., Kersters, K., Swings, J., Vauterin, L., … De Bruijn, F. J. (2000). Comparison of AFLP and rep-PCR genomic fingerprinting with DNA-DNA homology studies: Xanthomonas as a model system. *International Journal of Systematic and Evolutionary Microbiology*, *50*(2), 665–677.

Ranf, S. (2016). Immune Sensing of Lipopolysaccharide in Plants and Animals: Same but Different. *PLoS Pathogens*, *12*(6), 1–7. https://doi.org/10.1371/journal.ppat.1005596

Rapicavoli, J. N., Kinsinger, N., Perring, T. M., Backus, E. A., Shugart, H. J., Walker, S., &

Roper, M. C. (2015). O Antigen Modulates Insect Vector Acquisition of the Bacterial Plant Pathogen Xylella fastidiosa. *Applied and Environmental Microbiology*, *81*(23), 8145–8154. https://doi.org/10.1128/AEM.02383-15.Editor

Raskin, D. M., Seshadri, R., Pukatzki, S. U., & Mekalanos, J. J. (2006). Bacterial Genomics and Pathogen Evolution. *Cell*, *124*(4), 703–714. https://doi.org/10.1016/j.cell.2006.02.002

Rediers, H., Rainey, P. B., & Vanderleyden, J. (2005). Unraveling the Secret Lives of Bacteria : Use of In Vivo Expression Technology and Differential Fluorescence Induction Promoter Traps as Tools for Exploring Niche-Specific Gene Expression. *Microbiology and Molecular Biology Reviews : MMBR*, *69*(2), 217–261. https://doi.org/10.1128/MMBR.69.2.217

Ren, C.-P. (2009). *Comparative genomics of selected species of gram negative bacteria*. *University of Birmingham*.

Reyes, R. E., González, C. R., Jiménez, R. C., Herrera, M. O., & Andrade, A. A. (2012). Mechanisms of O-Antigen Structural Variation of Bacterial Lipopolysaccharide (LPS). In *The Complex World of Polysaccharides* (pp. 71–98). https://doi.org/10.5772/48147

Rietveld, A. M., Mpiira, S., Jogo, W., Staver, C., & Karamura, E. B. (2013). The beer banana value chain in central Uganda. *Banana Systems in the Humid Highlands of Sub-Saharan Africa: Enhancing Resilience and Productivity*, (September), 191–201. https://doi.org/10.1079/9781780642314.0191

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*(1), 24–26. https://doi.org/10.1038/nbt0111-24

Rodriguez-r, L. M., Grajales, A., Arrieta-ortiz, M. L., Salazar, C., Restrepo, S., & Bernal, A. (2012). Genomes-based phylogeny of the genus Xanthomonas. *BMC Microbiology*, *12*(1), 43. https://doi.org/10.1186/1471-2180-12-43

Roine, E. L. R., Ei, W. E. W., Uan, J. I. N. G. Y., Iisa, E. E. V. A., Assila, N. U., & Alkkinen, N. I. K. (1997). Hrp pilus: An hrp-dependent bacterial surface appendage produced by *Pseudomonas syringae* pv. *tomato* DC3000. *Proceedings of the National Academy of Sciences*, *94*(April), 3459–3464.

Romantschuk, M., Roine, E., & Taira, S. (2001). Hrp pilus–reaching through the plant cell wall. *European Journal of Plant Pathology*, *107*(2), 153–160. https://doi.org/10.1023/A:1011235101437

Rott, P., Marguerettaz, M., Fleites, L., Cociancich, S., Girard, J.-C., Pieretti, I., … Royer, M. (2010). Unravelling pathogenicity of *Xanthomonas albilineans*, the causal agent of sugarcane leaf scald. In *Proceedings of International Society on Sugar Cane Technololgy* (Vol. 27, pp. 1–11).

Rutschmann, F. (2006). Molecular dating of phylogenetic trees: A brief review of current methods that estimate divergence times. *Diversity and Distributions*, *12*(1), 35–48. https://doi.org/10.1111/j.1366-9516.2006.00210.x

Ryan, R. P., Vorhölter, F., Potnis, N., & Jones, J. B. (2011). Pathogenomics of Xanthomonas : understanding bacterium – plant interactions. *Nature Publishing Group*, *9*(5), 344–355. https://doi.org/10.1038/nrmicro2558

Saier, M. H. (2006). Protein Secretion Systems in Gram-negative Bacteria. *Microbe*, *1*(9), 414–419.

Salzberg, S. L., Sommer, D. D., Schatz, M. C., Phillippy, A. M., Rabinowicz, P. D., Tsuge, S., … Bogdanove, A. J. (2008). Genome sequence and rapid evolution of the rice pathogen. *BMC Genomics*, *16*(May), 1–16. https://doi.org/10.1186/1471-2164-9-204

Samudrala, R., Heffron, F., & Mcdermott, J. E. (2009). Accurate Prediction of Secreted Substrates and Identification of a Conserved Putative Secretion Signal for Type III Secretion Systems. *PLoS Pathogens*, *5*(4). https://doi.org/10.1371/journal.ppat.1000375

Schechter, L. M., Vencato, M., Jordan, K. L., Schneider, S. E., Schneider, D. J., & Collmer, A. (2006). Multiple Approaches to a Complete Inventory of *Pseudomonas syringae* pv. *tomato* DC3000 Type III Secretion System Effector Proteins, *19*(11), 1180–1192.

Schiavo, G., & van der Goot, F. G. (2001). The bacterial toxin toolkit. *Nature Reviews. Molecular Cell Biology*, *2*(7), 530–537. https://doi.org/10.1038/35080089

Schneider, K. L., Marrero, G., Alvarez, A. M., & Presting, G. G. (2011). Classification of Plant Associated Bacteria Using RIF, a Computationally Derived DNA Marker. *PLoS ONE*, *6*(4), 16. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3080875&tool=pmcentrez&r

endertype=abstract

Sharma, V., & Patil, P. B. (2011). Resolving the phylogenetic and taxonomic relationship of xanthomonas and stenotrophomonas strains using complete rpoB gene sequence. *PLoS Currents*, 1–9. https://doi.org/10.1371/currents.RRN1239

Shen, Y., Chern, M., Silva, F. G., & Ronald, P. (2001). Isolation of a *Xanthomonas oryzae* pv. *oryzae* Flagellar Operon Region and Molecular Characterization of flhF. *Molecular Plant-Microbe Interactions*, *14*(2), 204–213.

Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. *Nature Biotechnology*, *26*(10), 1135–1145.

Shrivastava, S., & Mande, S. S. (2008). Identification and Functional Characterization of Gene Components of Type VI Secretion System in Bacterial Genomes. *PLoS ONE*, *3*(8). https://doi.org/10.1371/journal.pone.0002955

Silverman, J. M., Brunet, Y. R., Cascales, E., & Mougous, J. D. (2012). Structure and regulation of the type VI secretion system. *Annual Review of Microbiology*, *66*(June), 453–472. https://doi.org/10.1146/annurev-micro-121809-151619

Simões, T. H. N., Gonçalves, E. R., Rosato, Y. B., & Mehta, A. (2007). Differentiation of Xanthomonas species by PCR-RFLP of rpfB and atpD genes. *FEMS Microbiology Letters*, *271*(1), 33–39.

Simonson, A. B., Servin, J. A., Skophammer, R. G., Herbold, C. W., Rivera, M. C., & Lake, J. A. (2005). Decoding the genomic tree of life. *PNAS*, *102*.

Singer, A. U., Schulze, S., Skarina, T., Xu, X., Cui, H., Eschen-lippold, L., … Bonas, U. (2013). A Pathogen Type III Effector with a Novel E3 Ubiquitin Ligase Architecture, *9*(1). https://doi.org/10.1371/journal.ppat.1003121

Sivashankari, S., & Shanmughavel, P. (2007). Comparative genomics - A perspective. *Bioinformation*, *1*(9), 376–378.

Sluys, M. A. Van, Camargo, L. E. A., Menck, C. F. M., Silva, A. C. R., Ferro, J. A., Oliveira, M. C., … Simpson, A. J. (2002). Comparative Genomic Analysis of Plant-Associated Bacteria. *Annu. Rev. Phytopathol*, *40*, 169–189. https://doi.org/10.1146/annurev.phyto.40.030402.090559

Smith, J. J., Karamura, E., Jones, D. R., Turyagyenda, F. L., & Blomme, G. (2008). *An analysis of the risk from Xanthomonas campestris pv . musacearum to banana cultivation in Eastern , Central and Southern Africa*. *Bioversity International*.

Smith, J. R. (1987). *From Plane to Spheroid: Determining the Figure of the Earth from 3000 Bc to the 18th Century Lapland and Peruvian Survey Expeditions*. Landmark Enterprises.

Soderlund, C. (2009). Computational techniques for elucidating plant-pathogen interactions from large-scale experiments on fungi and oomycetes. *Briefings in Bioinformatics*, *10*(6), 654–663. https://doi.org/10.1093/bib/bbp053

Soto, G. E., & Hultgren, S. J. (1999). Bacterial Adhesins : Common Themes and Variations in Architecture and Assembly. *Journal of Bacteriology*, *181*(4), 1059–1071.

Srinivasan, S., & Batra, J. (2014). Next Generation : Sequencing & Applications Four Generations of Sequencing- Is it Ready for the Clinic Yet ? *Next Generation: Sequencing & Applications*, *1*(1), 1–7. https://doi.org/10.4172/jngsa.1000107

Ssekiwoko, F, Tushemereirwe, W. K., Batte, M., Ragama, P. ., & Komakech, A. (2006). Reaction of Banana Germplasm to innoculation with *Xanthomonas campestris* pv. *musacearum*. *African Crop Science Journal*, *14*(2), 151–155.

Ssekiwoko, Fred, Turyagyenda, L. F., Mukasa, H., Eden-green, S., & Blomme, G. (2006). *Systemicity of Xanthomonas campestris pv . musacearum in inflorescence-infected banana plants in Uganda*.

Stebbins, C. E. (2005). Structural microbiology at the pathogen-host interface. *Cellular Microbiology*, *7*(9), 1227–1236. https://doi.org/10.1111/j.1462-5822.2005.00564.x

Stothard, P., & Wishart, D. S. (2005). Circular genome visualization and exploration using CGView. *Bioinformatics*, *21*(4), 537–539. https://doi.org/10.1093/bioinformatics/bti054

Studholme, D. J., Kemen, E., Maclean, D., Schornack, S., Aritua, V., Thwaites, R., … Jones, J. D. G. (2010). Genome-wide sequencing data reveals virulence factors implicated in banana Xanthomonas wilt. *FEMS Microbiol Letters*, *310*, 182–192. https://doi.org/10.1111/j.1574-6968.2010.02065.x

Taguchi, F., & Ichinose, Y. (2011). Role of Type IV Pili in Virulence of *Pseudomonas syringae* pv. *tabaci* 6605 : Correlation of Motility, Multidrug Resistance, and HR-Inducing Activity on a Nonhost Plant. *Molecular Plant-Microbe Interactions*, *24*(9), 1001–1011.

Tampakaki, A. P., Fadouloglou, V. E., Gazi, A. D., Panopoulos, N. J., & Kokkinidis, M. (2004). Conserved features of type III secretion. *Cellular Microbiology*, *6*(9), 805–816. https://doi.org/10.1111/j.1462-5822.2004.00432.x

Tamura, K, & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, *10*(3), 512–526. https://doi.org/10.1093/oxfordjournals.molbev.a040023

Tamura, Koichiro, Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, *28*(10), 2731–2739. https://doi.org/10.1093/molbev/msr121

Tamura, Koichiro, Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6 : Molecular Evolutionary Genetics Analysis Version 6 . 0. *Molecular Biology and Evolution*, *30*(12), 2725–2729. https://doi.org/10.1093/molbev/mst197

Tang, X., Xiao, Y., & Zhou, J. (2006). Regulation of the Type III Secretion System in Phytopathogenic Bacteria. *Molecular Plant-Microbe Interactions*, *19*(11), 1159–1166.

Thieme, F., Koebnik, R., Bekel, T., Berger, C., Boch, J., Büttner, D., … Schuster, S. C. (2005). Insights into Genome Plasticity and Pathogenicity of the Plant Pathogenic Bacterium *Xanthomonas campestris* pv. *vesicatoria* Revealed by the Complete Genome Sequence Insights into Genome Plasticity and Pathogenicity of the Plant Pathogenic Bacterium Xantho. *Journal of Bacteriology*, *187*(21). https://doi.org/10.1128/JB.187.21.7254

Tinzaara, W., Karamura, E. B., Blomme, G., Jogo, W., Ocimati, W., & Rietveld, A. (2013). Why Sustainable Management of Xanthomonas Wilt of Banana in East and Central Africa Has Been Elusive. In *VII International Symposium on Banana: ISHS-ProMusa Symposium on Bananas and Plantains: Towards Sustainable Global Production and Improved Use* (pp. 1–10).

Todar, K. (2014). Colonization and Invasion by Bacterial Pathogens. In *Text book of bacteriology* (pp. 3–5). Retrieved from http://textbookofbacteriology.net/colonization.html

Tripathi, L., Mwaka, H., Tripathi, J. N., & Tushemereirwe, W. K. (2010). Expression of sweet pepper Hrap gene in banana enhances resistance to *Xanthomonas campestris* pv.

musacearum. *Molecular Plant Pathology*, *11*(6), 721–731. https://doi.org/10.1111/j.1364-3703.2010.00639.x

Tripathi, L., Mwangi, M., Aritua, V., Tushemereirwe, W. K., Abele, S., & Bandyopadhyay, R. (2009). Xanthomonas wilt: A threat to banana production in East and central Africa. *Plant Disease*, *93*(5).

Tripathi, L., Tripathi, J. N., & Tushemereirwe, W. K. (2004). Strategies for resistance to bacterial wilt disease of bananas through genetic engineering, *3*(December), 688–692.

Tseng, T.-T., Tyler, B. M., & Setubal, J. C. (2009). Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiology*, *9 Suppl 1*, S2. https://doi.org/10.1186/1471-2180-9-S1-S2

Tushemereirwe, W. K., Kangire, A., Kubiriba, J., Maureen, N., & Gold, S. C. (2004). Diseases threateneing banana biodiversity in Uganda. *African Crop Science Journal*, *12*(1), 19–26.

Üstün, S., Bartetzko, V., & Börnke, F. (2013). The Xanthomonas campestris Type III Effector XopJ Targets the Host Cell Proteasome to Suppress Salicylic-Acid Mediated Plant Defence. *PLoS Pathogens*, *9*(6). https://doi.org/10.1371/journal.ppat.1003427

Üstün, S., Bartetzko, V., & Börnke, F. (2015). The Xanthomonas effector XopJ triggers a conditional hypersensitive response upon treatment of *N. benthamiana* leaves with salicylic acid. *Frontiers in Plant Science*, *6*(AUG), 1–11. https://doi.org/10.3389/fpls.2015.00599

Van Baarlen, P., Van Belkum, A., Summerbell, R. C., Crous, P. W., & Thomma, B. P. H. J. (2007). Molecular mechanisms of pathogenicity: How do pathogenic microorganisms develop cross-kingdom host jumps? *FEMS Microbiology Reviews*, *31*(3), 239–277. https://doi.org/10.1111/j.1574-6976.2007.00065.x

Vandroemme, J., Cottyn, B., Baeyen, S., De Vos, P., & Maes, M. (2013). Draft genome sequence of *Xanthomonas fragariae* reveals reductive evolution and distinct virulence-related gene content. *BMC Genomics*, *14*, 829. https://doi.org/10.1186/1471-2164-14-829

Varshney, R. K., Nayak, S. N., May, G. D., & Jackson, S. A. (2009). Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends in Biotechnology Vol.27*. https://doi.org/10.1016/j.tibtech.2009.05.006

Vauterin, L., Hoste, B., Kersters, K., & Swings, J. (1995). Reclassification of Xanthomonas.

International Journal of Systematic Bacteriology, *45*(3), 472–489. https://doi.org/10.1099/00207713-45-3-472

Vencato, M., Tian, F., Alfano, J. R., Buell, C. R., Cartinhour, S., Declerck, G. A., … Schneider, D. J. (2006). Bioinformatics-Enabled Identification of the HrpL Regulon and Type III Secretion System Effector Proteins of *Pseudomonas syringae* pv. *phaseolicola* 1448A. *Molecular Plant-Microbe Interactions*, *19*(11), 1193–1206.

Wairegi, L. W. I., Van Asten, P. J. A., Tenywa, M. M., & Bekunda, M. A. (2010). Abiotic constraints override biotic constraints in East African highland banana systems. *Field Crops Research*, *117*, 146–153. https://doi.org/10.1016/j.fcr.2010.02.010

Wairuri, C., van der Waals, J., van Schalkwyk, A., & Theron, J. (2012). *Ralstonia solanacearum* needs Flp pili for virulence on potato Charles. *Molecular Plant-Microbe Interactions*, *1*, 546–556.

Wassenaar, T. M., Bohlin, J., Binnewies, T. T., & Ussery, D. W. (2009). Genome comparison of bacterial pathogens. *Genome Dynamics*, *6*(February), 1–20. https://doi.org/10.1159/000235759

Wasukira, A., Tayebwa, J., Thwaites, R., Paszkiewicz, K., Aritua, V., Kubiriba, J., … Studholme, D. J. (2012). Genome-wide sequencing reveals two major sub-lineages in the genetically monomorphic pathogen *Xanthomonas campestris* pv. *musacearum*. *Genes*, *3*(3), 361–377.

Webb, J. S., Thompson, L. S., James, S., Charlton, T., Tolker-Nielsen, T., Koch, B., … Kjelleberg, S. (2003). Cell death in *Pseudomonas aeruginosa* biofilm development. *Journal of Bacteriology*, *185*(15), 4585–4592. https://doi.org/10.1128/JB.185.15.4585

Weber, E., & Koebnik, R. (2006). Positive Selection of the Hrp Pilin HrpE of the Plant Pathogen Xanthomonas, *188*(4), 1405–1410. https://doi.org/10.1128/JB.188.4.1405

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

Wei, L., Liu, Y., Dubchak, I., Shon, J., & Park, J. (2002). Comparative genomics approaches to study organism similarities and differences. *Journal of Biomedical Informatics*, *35*, 142–

150.

Wei, W., Plovanich-Jones, A., Deng, W. L., Jin, Q. L., Collmer, A., Huang, H. C., & He, S. Y. (2000). The gene coding for the Hrp pilus structural protein is required for type III secretion of Hrp and Avr proteins in *Pseudomonas syringae* pv. *tomato*. *PNAS*, *97*(5), 2247–2252. https://doi.org/10.1073/pnas.040570097

Wengelnik, K. A. I., & Bonas, U. (1996). HrpXv , an AraC-Type Regulator , Activates Expression of Five of the Six Loci in the hrp Cluster of *Xanthomonas campestris* pv. *vesicatoria*. *Journal of Bacteriology*, *178*(12), 3462–3469.

Whalen, M. C., Richter, T., Zakhareyvich, K., Yoshikawa, M., Al-azzeh, D., Adefioye, A., … Ronald, P. C. (2008). Physiological and Molecular Plant Pathology Identification of a host 14-3-3 protein that interacts with Xanthomonas effector AvrRxv. *Physiological and Molecular Plant Pathology*, 1–10. https://doi.org/10.1016/j.pmpp.2008.05.006

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

Wilmes, P., Simmons, S. L., Denef, V. J., & Banfield, J. F. (2009). The dynamic genetic repertoire of microbial communities. *FEMS Microbiology Reviews*, *33*(1), 109–132. https://doi.org/10.1111/j.1574-6976.2008.00144.x

Wray, G. A., Hahn, M. W., Abouheif, E., Balhoff, J. P., Pizer, M., Rockman, M. V, & Romano, L. A. (2002). Review Article The Evolution of Transcriptional Regulation in Eukaryotes. *Molecular Biology*, 1377–1419. https://doi.org/10.1093/molbev/msg140

Wroblewski, T., Caldwell, K. S., Piskurewicz, U., Cavanaugh, K. a, Xu, H., Kozik, A., … Michelmore, R. W. (2009, August). Comparative large-scale analysis of interactions between several crop species and the effector repertoires from multiple pathovars of *Pseudomonas and Ralstonia*. *Plant Physiology*. https://doi.org/10.1104/pp.109.140251

Xia, Y. (2004). Microreview Proteases in pathogenesis and plant defence. *Cellular Microbiology*, *6*, 905–913. https://doi.org/10.1111/j.1462-5822.2004.00438.x

Yamashita, A., Sekizuka, T., & Kuroda, M. (2014). Characterization of antimicrobial resistance dissemination across plasmid communities classified by network analysis. *Pathogens*, *3*(2), 356–376. https://doi.org/10.3390/pathogens3020356

Yang, H., Tao, Y., Zheng, Z., Li, C., Sweetingham, M. W., & Howieson, J. G. (2012). Application of next-generation sequencing for rapid marker development in molecular plant breeding : a case study on anthracnose disease resistance in *Lupinus angustifolius* L. Application of next-generation sequencing for rapid marker development in mole. *BMC Genomics*, *13*(1), 1. https://doi.org/10.1186/1471-2164-13-318

Yang, Y., Zhao, J., Morgan, R. L., Ma, W., & Jiang, T. (2010). Computational prediction of type III secreted proteins from gram-negative bacteria. *BMC Bioinformatics*, *11*(Suppl 1), S47. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20122221

Young, J. ., Park, D. ., & Shearman, H. M. (2008). A multilocus sequence analysis of the genus Xanthomonas. *Systematic and Applied Microbiology*, *31*(5), 366–377.

Zerbino, D. R., & Birney, E. (2008). Velvet : Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, *18*, 821–829. https://doi.org/10.1101/gr.074492.107

Zhang, F., Du, Z., Huang, L., Cruz, C. V., Zhou, Y., & Li, Z. (2013). Comparative Transcriptome Profiling Reveals Different Expression Patterns in *Xanthomonas oryzae* pv. *oryzae* Strains with Putative Virulence-Relevant Genes. *PLoS ONE*, *8*(5). https://doi.org/10.1371/journal.pone.0064267

Zhang, L., Xu, J., & Birch, R. G. (1999). Engineered detoxification confers resistance against a pathogenic bacterium. *Nature Biotechnology*, *17*, 1021–1024. https://doi.org/10.1038/13721

Zheng, L.-L., Li, Y.-X., Ding, J., Guo, X.-K., Feng, K.-Y., Wang, Y.-J., … Chou, K.-C. (2012). A Comparison of Computational Methods for Identifying Virulence Factors. *PLoS ONE*, *7*(8), e42517. https://doi.org/10.1371/journal.pone.0042517

Zheng, Y., Lilo, S., Mena, P., & Bliska, J. B. (2012). YopJ-Induced Caspase-1 Activation in Yersinia-Infected Macrophages: Independent of Apoptosis, Linked to Necrosis, Dispensable for Innate Host Defense. *PloS One*, *7*(4). https://doi.org/10.1371/journal.pone.0036019

Zimaro, T., Thomas, L., Marondedze, C., Garavaglia, B. S., Gehring, C., Ottado, J., & Gottig, N. (2013). Insights into *Xanthomonas axonopodis* pv. *citri* biofilm through proteomics. *BMC Microbiology*, *13*, 186. https://doi.org/10.1186/1471-2180-13-186