

A molecular study of the causal agent of bacterial leaf streak disease

AS Kraemer
22904433

Dissertation submitted in fulfilment of the requirements for the
degree *Magister Scientiae* in *Microbiology* at the Potchefstroom
Campus of the North-West University

Supervisor: Prof CC Bezuidenhout
Co-supervisor: Dr C Mienie

November 2016



ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to my supervisors for their assistance throughout the structuring and development of this project and guiding me through its challenges.

Firstly, to Prof. Carlos Bezuidenhout for providing me with this study opportunity as well as for his insightful guidance. It was a privilege to have been given the opportunity to tread the path of next generation sequencing technologies. Secondly, to Dr. Charlotte Mienie for executing the whole genome sequencing run as well as for her technical assistance.

I would like to thank Dr. Tomasz Sanko for his guidance with the bioinformatics component of this project. I am grateful for the help and provision of extensive knowledge.

Also a word of thanks to Abraham Mahlatsi and Lee-Hendra Julies for their general assistance and taking care of the orders and administration.

I am greatly thankful to my parents and brother for their unceasing motivation and encouragement throughout my studies as well as for their financial support that has enabled me to complete this project.

Lastly, I would like to thank my fellow M.Sc. students: Bren Botha, Rohan Fourie, Carissa van Zyl and Vivienne Visser. All your support, motivation, late-night working sessions, companionship in the laboratory and countless coffee breaks would forever stay with me.

DECLARATION

I declare that this dissertation submitted for the degree of *Master of Science in Microbiology* at the North-West University, Potchefstroom Campus, has not been previously submitted by me for a degree at this or any other university, that it is my own work in design and execution, and that all material contained herein has been duly acknowledged.

A.S. Kraemer

Date

ABSTRACT

Bacterial leaf streak disease (BLSD) of maize is caused by the pathogen currently classified as *Xanthomonas campestris* pv. *zeae*. The disease has only been reported in South Africa and is a potential threat to commercial maize practices. Limited information is available regarding the taxonomy of the causal agent. A specific and reliable diagnostic tool has not yet been developed for effective detection of the pathogen. The aim of this study was to use a molecular and bioinformatics approach in order to gain more insight into the taxonomy and for detection of the pathogen. The molecular study involved whole genome sequencing and analysis of the genomic composition of the causal agent. Genome sequences of the BLSD causal agent were assembled and compared to reference genomes of *Xanthomonas* relatives. Six *Xanthomonas* housekeeping genes (*atpD*, *dnaK*, *fyuA*, *recA*, *rpoB* and *rpfB*) were selected from the whole genome sequence data. These complete gene sequences were compared to gene sequences of other *Xanthomonas* species to determine nucleotide variation between the BLSD causal agent and relatives. Housekeeping gene sequences were used to construct phylogenetic relationships between the causal agent and close relatives. Phylogenetic trees for individual genes revealed different clustering of the BLSD pathogen within the genus. Based on phylogenetic analysis, closest relatives of the causal agent were *X. vasicola*, *X. oryzae*, *X. axonopodis* and *X. hortorum* pv. *hederae*. The BLSD causal agent never clustered with the *X. campestris* species, regardless of the gene that was analysed. Results from this study supported the conclusion that *X. campestris* pv. *zeae* is not part of the *X. campestris* species. As the pathogen was not placed consistently with the same related species in phylogenetic analysis of different genes, the pathogen is likely a novel *Xanthomonas* species. Regions of the *dnaK*, *recA* and *rpoB* genes were mined for DNA markers that were potentially unique to the pathogen. Six primer pairs, targeting variable regions within the genes, were designed to amplify these sequences and specificity was assessed via PCR. Five primer pairs amplified DNA of all BLSD isolates but also one or both of the reference strains and were thus deemed not sufficiently discriminatory. One primer set (XanZ_rpoB) allowed for completely specific amplification of DNA from the BLSD isolates. This primer pair targets a sufficiently variable region within the *rpoB* gene and was a potential molecular marker of the BLSD organism for use in detection of the pathogen. This collection of new molecular data regarding the BLSD causal agent will help to establish the pathogen's taxonomy within the *Xanthomonas* genus as well as enable the development of reliable detection techniques for identification of the pathogen.

Keywords: Bacterial leaf streak disease; housekeeping gene; maize; molecular marker; PCR; phylogenetic tree; whole genome; *Xanthomonas*; *X. campestris* pv. *zeae*

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	ii
DECLARATION	iii
ABSTRACT	iv
LIST OF FIGURES	viii
LIST OF TABLES	x
LIST OF ABBREVIATIONS	ii
LIST OF UNITS AND SYMBOLS	iii
CHAPTER 1: INTRODUCTION	1
1.1. Background and research rationale.....	1
1.2. Research aim and objectives	3
CHAPTER 2: LITERATURE REVIEW	4
2.1. History and importance of maize	4
2.2. The <i>Xanthomonas</i> genus – A general overview	6
2.3. Taxonomy of the <i>Xanthomonas</i> genus	8
2.4. Pathogenicity of <i>Xanthomonas</i>	10
2.5. Diseases of maize	11
2.6. Bacterial leaf streak disease of maize and the causal agent (<i>X. campestris</i> pv. <i>zeae</i>) .	12
2.7. Bacterial leaf streak disease of other host plants.....	14
2.8. Related pathovars of the causal agent of bacterial leaf streak disease.....	15
2.9. Molecular studies of the <i>Xanthomonas</i> genus	16
2.10. Next generation sequencing and whole genome sequencing.....	17
2.11. Gene assembly	18
2.12. Phylogenetics as a tool in molecular studies	20
2.13. Housekeeping genes as molecular markers.....	20
2.13.1. <i>atpD</i>	21
2.13.2. <i>dnaK</i>	22
2.13.3. <i>fyuA</i>	22

2.13.4. <i>recA</i>	22
2.13.5. <i>rpfB</i>	22
2.13.6. <i>rpoB</i>	23
2.14. Primer design.....	23
2.15. Summary of literature.....	24
CHAPTER 3: MATERIALS AND METHODS.....	26
3.1. Bacterial isolation and cultivation of the causal agent.....	26
3.2. Cultivation of reference strains.....	27
3.3. DNA extractions.....	27
3.4. Whole genome sequencing.....	28
3.5. Genome assembly and mapping to reference genomes.....	29
3.6. Extraction of housekeeping gene sequences.....	29
3.7. Comparative gene analysis.....	30
3.8. Phylogenetic analysis.....	31
3.9. Primer design.....	31
3.10. Primer specificity testing with PCR amplification.....	32
CHAPTER 4: RESULTS.....	33
4.1. Cultivation of isolates and Gram-staining.....	33
4.2. Genomic DNA extractions.....	33
4.3. Sequence assembly.....	34
4.4. Genome assembly.....	36
4.5. Comparative gene analysis.....	38
4.5.1. <i>atpD</i>	40
4.5.2. <i>dnaK</i>	40
4.5.3. <i>fyuA</i>	40
4.5.4. <i>recA</i>	41
4.5.5. <i>rpfB</i>	41
4.5.6. <i>rpoB</i>	41
4.6. Phylogenetic analysis.....	42

4.6.1. <i>atpD</i>	43
4.6.2. <i>dnaK</i>	44
4.6.3. <i>fyuA</i>	46
4.6.4. <i>recA</i>	47
4.6.5. <i>rpfB</i>	49
4.6.6. <i>rpoB</i>	49
4.7. Primer design and specificity.....	52
CHAPTER 5: DISCUSSION	56
5.1. Isolate cultivation and Gram-staining.....	56
5.2. Whole genome sequencing.....	56
5.2. Comparative housekeeping gene analysis	57
5.3. Phylogenetic analysis.....	58
5.3.1. <i>atpD</i>	59
5.3.2. <i>dnaK</i>	60
5.3.3. <i>fyuA</i>	61
5.3.4. <i>recA</i>	62
5.3.5. <i>rpfB</i>	62
5.3.6. <i>rpoB</i>	64
5.3.7. Summary of phylogenetic analyses	66
5.4. Clarifications of general trends observed in phylogenetic analyses.....	67
5.5. Possible reclassification of the BLSO causal agent	68
5.6. Primer design and specificity testing	69
CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS.....	71
6.1. Conclusions	71
6.2. Recommendations	72
REFERENCES	74
APPENDIX A: MEDIA COMPOSITION	91
APPENDIX B: SETTINGS FOR BIOINFORMATICS PROGRAMMES.....	92
APPENDIX C: GENBANK SEQUENCES.....	94

LIST OF FIGURES

- Figure 1:** An example of the yellow colonies characteristic to the *Xanthomonas* genus after cultivation of the isolates on GYC agar plates. 33
- Figure 2:** Genomic DNA bands of BLSD samples in 1 % (w/v) agarose gel with a 1 kb ladder (M) to indicate band sizes. Samples loaded in the gel were as follows: lane 1 = X1, lane 2 = X2, lane 3 = X9, lane 4 = X15, lane 5 = X22, lane 6 = X26, lane 7 = X33, lane 8 = X35, lane 9 = X45 and lane 10 = XGP. 34
- Figure 3:** Neighbour-joining tree for the *atpD* gene of various *Xanthomonas* species and the clustering of the BLSD strain (X45) as well as reference strains of *X. axonopodis* pv. *axonopodis* (HKG *Xanthomonas atpD*:X85) and *X. campestris* pv. *campestris* (HKG *Xanthomonas atpD*:X86) within the phylogenetic tree. 43
- Figure 4:** Neighbour-joining tree of the BLSD strain (X45) as well as reference strains of *X. axonopodis* pv. *axonopodis* (HKG *dnaK*:X85) and *X. campestris* pv. *campestris* (HKG *dnaK*:X86) in the *Xanthomonas* genus using phylogenetic analysis of *dnaK* gene sequences. 45
- Figure 5:** Neighbour-joining tree based on *fyuA* gene sequences indicating relationships between the BLSD strain (X45) as well as reference strains of *X. axonopodis* pv. *axonopodis* (HKG *fuyA*:X85), *X. campestris* pv. *campestris* (HKG *fyuA*:X86) and other *Xanthomonas* species. 47
- Figure 6:** Neighbour-joining tree of the BLSD strain (X45) as well as reference strains of *X. axonopodis* pv. *axonopodis* (HKG *recA*:X85) and *X. campestris* pv. *campestris* (HKG *recA*:X86) in the *Xanthomonas* genus using *recA* gene sequences. 48
- Figure 7:** Neighbour-joining analysis of *rpfB* gene sequences and the position of the BLSD strain (X45) as well as reference strains of *X. axonopodis* pv. *axonopodis* (HKG *rpfB*:X85) and *X. campestris* pv. *campestris* (HKG *rpfB*:X86) among reference sequences. 50

Figure 8: Neighbour-joining tree based on *rpoB* gene sequences indicating relationships between the BLSD strain (X45) as well as reference strains of *X. axonopodis* pv. *axonopodis* (HKG *rpoB*:X85) and *X. campestris* pv. *campestris* (HKG *rpoB*:X86) and other *Xanthomonas* species. 51

Figure 9: PCR amplicons of *dnaK* gene fragments as targeted by the XanX_*dnaK* primer set in a 1.5 % agarose gel with a 1 kb ladder (M) to indicate band size. PCR products of samples loaded in the gel in Figure 9 were as follows: lane 1 = X1, lane 2 = X2, lane 3 = X9, lane 4 = X22, lane 5 = X33, lane 6 = X, lane 7 = 35, lane 8 = X45, lane 9 = empty, lane 10 = X85 and lane 11 = X86. 54

Figure 10: Amplified DNA fragments of partial *recA* gene sequences as targeted with PCR by the XanZ_*recA* primer set, in a 1.5 % agarose gel with a 1 kb ladder (M) to indicate amplicon sizes. PCR products of samples loaded in the gel were as follows: lane 1 = X1, lane 2 = X2, lane 3 = X9, lane 4 = X22, lane 5 = X33, lane 6 = X, lane 7 = empty, lane 8 = empty, lane 9 = X85 and lane 10 = X86. 54

Figure 11: Amplified PCR products of genomic DNA targeted by XanZ_*rpoB* primers for amplification of *rpoB* gene fragments in a 1.5 % agarose gel with 1 kb ladder (M) as reference for amplicon sizes. PCR products of samples loaded in the gel were as follows: lane 1 = X1, lane 2 = X2, lane 3 = X9, lane 4 = X22, lane 5 = X33, lane 6 = X35, lane 7 = X45, lane 8 = XGP, lane 9 = X26, lane 10 = empty, lane 11 = empty, lane 12 = X85 and lane 13 = X86. 55

LIST OF TABLES

- Table 1:** Index codes and sequences for i7 and i5 indices added to index 1 and 2 **29**
- Table 2:** Sequence assembly results of the collected assembly of reads for X85 + X86 and X1 – 45 + GP, respectively after sequence trimming in CLC Genomics Workbench **35**
- Table 3:** Whole genome sequences from the two reference strains, X85 and X86, as well as the isolates, X1-45+GP, mapped against two complete genome reference sequences, NC010717 (*X. oryzae* pv. *oryzae* strain PXO99A) and NC003902 (*X. campestris* pv. *campestris* strain ATCC 33913). **37**
- Table 4:** Complete gene sequence comparisons of reference strains with BLSD isolates indicating overall gene similarity of each species with the BLSD causal agent. (X.c = *X. campestris* pv. *campestris* (ATCC 33913); X.o = *X. oryzae* pv. *oryzae* (KACC 10331); X.a = *X. axonopodis* (Xac29-1) and X. alb = *X. albilineans* (GPE PC73). **39**
- Table 5:** Primer pairs designed for amplification of DNA markers potentially unique to the BLSD pathogen with their optimal annealing temperatures..... **52**

LIST OF ABBREVIATIONS

A ₂₆₀	Spectrophotometric light absorbance at 260 nm
A ₂₈₀	Spectrophotometric light absorbance at 280 nm
BLSD	Bacterial leaf streak disease
bp	Base pair
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DSMZ	German collection of cell cultures and microorganisms
GC	Guanine and cytosine
GYC	Glucose yeast extract calcium carbonate
HKG	Housekeeping gene
Hrp	Hypersensitive response and pathogenicity
ITS	Internal transcribed spacer
kb	Kilobase
MEGA	Molecular evolution genetics analysis
NCBI	National centre for biotechnology information
no	Number
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
pv	Pathovar
qPCR	Quantitative real-time polymerase chain reaction
rDNA	Ribosomal deoxyribonucleic acid
RFLP	Restriction fragment length polymorphism
T _m	Melting temperature
X	<i>Xanthomonas</i>

LIST OF UNITS AND SYMBOLS

°C	Degrees Celsius
%	Percentage
>	Greater-than
µg	Microgram
µl	Microliter
cm	Centimeter
ha	Hectare
Mbp	Million base pairs
mg	Milligram
min	Minute
ml	Milliliter
mm	Millimeter
ng	Nanogram
nm	Nanometer
pH	Potential hydrogen
s	Second
V	Volt
w / v	Weight over volume

CHAPTER 1: INTRODUCTION

1.1. Background and research rationale

Members of the *Xanthomonas* genus are primarily pathogenic and cause a diverse range of plant diseases which have led them to become of huge economic importance. Consequently organisms within the *Xanthomonas* genus have been the subject of numerous taxonomic and determinative studies (Berthier *et al.*, 1993). The genus has a world-wide distribution and is found in association with major groups of plants. It is divided into various species and different pathovar strains based on their host specificity and pathogenicity.

Although *Xanthomonas* is of biological, economic and industrial importance, the phylogenetic relationships and classifications among organisms of the genus are still unclear and under constant debate. Due to confusion over relationships between species and pathovars, pathogens in the genus are often reclassified and old pathovars become new species (Rodriguez-R *et al.*, 2012). Reasons for the unsettled taxonomy within the *Xanthomonas* genus include insufficient past descriptions, adjustments in taxonomic concepts, contradictory data and complications in establishing the value of pathogenicity in taxonomy (Palleroni, 1984; *cited by* Janse, 1992; Downing, 1998).

Bacterial leaf streak disease of maize (*Zea mays L.*) has shown a frequent occurrence and an increasing severity on maize crops since the late 1980's (Coutinho & Wallis, 1991). It was first reported in South Africa in 1949 and rose awareness in the 1950's but the disease was not considered economically important (Dyer, 1949). The disease has only been detected in South Africa and is a growing threat to the country's maize industry (Qhobela *et al.*, 1990; Niemann, 2015; Nienaber, 2015). Bacterial leaf streak disease may have influences on maize export activities as well as local commercial practices. Periodic epidemics of bacterial leaf streak disease occur which has the potential to cause considerable yield loss in agricultural maize production (Coutinho & Wallis, 1991; Niemann, 2015; Nienaber, 2015). Efficient control measures for the disease have not yet been developed.

Bacterial leaf streak disease leads to symptoms such as lesions and necrotic regions on maize leaves that decrease the leaf area available for photosynthesis (Qhobela *et al.*, 1990). These symptoms directly impair photosynthetic activity, resulting in premature defoliation and yield losses. Yield losses of up to 40 % have been reported due to the presence of the disease in maize fields (Nowell, unpublished; *cited by* Qhobela *et al.*, 1990).

The pathogen that causes bacterial leaf streak disease of maize is currently classified as *Xanthomonas campestris* pv. *zeari*. However, information regarding the complete taxonomy of this organism is lacking and shortcomings exist in the differentiation of the organism from its close relatives (Niemann, 2015). In the past, detection and identification of the bacterial leaf streak agent has been hindered by the incomplete knowledge of its epidemiology and taxonomy. To date information and research regarding the causal agent of bacterial leaf streak disease are still very limited (Niemann, 2015; Nienaber, 2015). No reports have been published regarding the complete molecular composition of the bacterium or pathovar-specific molecular markers of *X. campestris* pv. *zeari*.

It is still unknown whether or not this pathogen is seed-borne. The movement of contaminated seed over international borders may lead to long-distance dissemination if the causal agent is indeed seed-borne. Before introducing seedlings to a new location it is important to assess the health status of the plants to prevent movement of diseased or infected crops (Cho *et al.*, 2011). Early detection of diseases present in maize plants is necessary to guide selective transplantation of healthy seedlings. A study of this organism on molecular level will help to establish its taxonomy within the *Xanthomonas* genus as well as enable the development of reliable techniques to detect the pathogen within plant material.

The rapid development of genomics in recent years has led to highly improved bacterial pathogenesis research (Wren, 2000). A full genome sequence provides an outline of genetic information for analysis of biological characteristics as well as gene function. Comparative and functional genomics are all-important approaches towards understanding the biology of plant diseases, disease diagnosis and surveillance and developing effective preventative methods (Simpson *et al.*, 2000; Mirmajlessi *et al.*, 2015).

Developments of molecular assays would be beneficial to serve as suitable strategies for quick identification and monitoring of bacterial leaf streak disease of maize. Although the use of 16S rDNA gene sequence analyses is widely employed to identify bacteria, this gene fragment does not allow for entirely sufficient discrimination between *Xanthomonas* species and pathovars (Hauben *et al.*, 1997; Moore *et al.*, 1997; Niemann, 2015). Other housekeeping genes with functions essential to all *Xanthomonas* bacteria are often investigated as an alternative to the 16S rDNA gene (Martens *et al.*, 2008).

Housekeeping genes with greatest potential include the ATP synthase subunit *beta* encoding gene, *atpD*, the molecular chaperone, *dnaK*, TonB-dependent receptor, *fyuA*, *recA*, encoding recombinase A, *rpfB*, coding for long-chain fatty acyl CoA ligase, and the DNA-directed RNA polymerase subunit *beta*, *rpoB* (Simões *et al.*, 2007; Young *et al.*, 2008; Albuquerque *et al.*, 2012; Rodriguez-R *et al.*, 2012). Regions of the housekeeping genes

containing variable sequences between the causal agent and relatives could be used to design primers that are potentially diagnostic to the causal agent. PCR primers could be tested to determine the validity of the use of the specific region as a novel molecular marker (Stepkowski *et al.*, 2003).

1.2. Research aim and objectives

The aim of this research was to conduct a molecular study of the causal agent of bacterial leaf streak disease. By taking a molecular approach, based on the full genome of the pathogen, housekeeping gene analysis and PCR methods, the study aimed to generate valuable data that could assist in elucidating taxonomic disputes and resolving current questions regarding the bacterial leaf streak agent.

The objectives for this study were:

- To perform a whole genome sequence analysis of *Xanthomonas* bacterial leaf streak disease strains and compare it to closely related relatives;
- To determine the closest relatives of the bacterial leaf streak disease agent by sequence analyses of relevant housekeeping genes;
- To “mine” the gene sequence data for DNA markers unique to the pathogen;
- To develop primers for amplification of the molecular markers and test them for their specificity to the causal agent using a PCR assay.

CHAPTER 2: LITERATURE REVIEW

2.1. History and importance of maize

Maize (*Zea mays*) is a plant that belongs to the grass family *Poaceae*. The word “Zea” was taken from an ancient Greek name for grass food (zela). Although the *Zea* genus consists of four species, *Zea mays* L. is the only species of economic importance. It is widely accepted that *Zea mays* has originated from Mexico and Central America, also known as the Meso-American area (Watson & Dallwitz, 1992). Maize is a versatile crop that has the ability to grow in diverse agro climatic zones. It is grown globally and is one of the most important grain crops across the world (Doebley, 1990).

Maize is undoubtedly South Africa’s most important grain crop and the staple food of the majority of the population (National Agricultural Marketing Council, 2003). It is a major source of carbohydrates for human consumption and also serves as major feed grain. It is the largest locally-produced grain crop and the country’s most valuable agricultural commodity (National Agricultural Marketing Council, 2003; Department of Agriculture, Forestry and Fisheries, 2012). Roughly 3.1 million ha of land is used for maize production which yields approximately 8 million tons of maize grain annually (Du Plessis, 2003). Maize is grown in most parts of the country in a variety of different environments. Main regions with highest percentages of maize production in South Africa include the Free State, North West, Mpumalanga and Kwazulu-Natal provinces with contributions of 34 %, 32 %, 24 % and 3 % to the country’s total maize production, respectively.

The commercial maize industry plays a prominent role in the economy, contributing to input industries as well as milling, food processing and animal product industries. In South Africa an average of 3.5 million and 2.6 million tons of maize are milled annually for food and feed respectively. In the past ten years an average of roughly 9 million tons per annum were produced by more than 9 000 commercial maize producers. Around 150 000 farm workers are employed in the commercial industry (Department of Agriculture, Forestry and Fisheries, 2012). Millers employ over 5 300 workers and make use of 80 % of the installed milling capacity for maize.

Maize contributes approximately 46 % of the 6 million tons of total feed produced per annum. Starch and glucose producers use over 500 000 tons of maize annually. The “manufacturing” category that includes the grain milling industry presently contributes the largest portion (18.2 %) to the gross domestic product of South Africa (National Agricultural Marketing Council, 2003). The “food and food products” division within this category makes

the largest contribution (13.6 %) towards this amount of which grain milling is responsible for 2.8 %. The quantity of maize produced is a major factor that influences the gross production value (Department of Agriculture, Forestry and Fisheries, 2012). Increases in total production are mainly responsible for substantial increases in economic income during production years.

Maize is grown in regions with an annual rainfall of more than 350 mm (Du Plessis, 2003; Ministry of Environments and Forests, 2010). Maize plants acquire water from soil moisture reserves and an even distribution of rain during the growing season is necessary for successful production. The planting time for maize is between October and December, depending on the soil, temperatures and rainfall patterns. Days of bright sunlight are necessary to accelerate photosynthetic activity and ensure fast growth of plants. Soil pH between 7.5 and 8.5 is ideal for healthy crop growth. Intermittent sunlight, clouds and rain is the desired conditions for growth (Ministry of Environments and Forests, 2010).

Maize plants are monocotyledonous, determinate, annual plants. Plant height varies between one and four metres in length. Large, narrow and opposing leaves are borne alternately along the length of one solid stem (Ministry of Environments and Forests, 2010). Maize plants possess between eight and twenty leaves that are organised in a spiral manner on the stem with leaves occurring alternately in two opposite rows. Leaves characteristically consist of a blade, sheath, ligules and auricles. Leaf blades are long, narrow and tapered in shape and have a smooth to hairy texture. Stomata are present in rows along the entire surface of the leaf with the most stomata occurring on the underside of the leaf (Du Plessis, 2003). The total leaf area of a mature plant can be more than one square metre (Du Plessis, 2003).

Maize crops have a preference for warm temperatures and prolonged cloudy intervals are detrimental to plant growth (Ministry of Environments and Forests, 2010). The crops are generally grown in areas where mean temperatures during summer months are higher than 23°C and daily temperatures higher than 19°C. Temperatures higher than 32°C may negatively influence the crop yield. Germination will take place fast and consistently in soil temperatures ranging between 16°C and 18°C and maize can be expected to emerge within five to six days.

Maize is sensitive to frost at all growth stages; hence a continuous frost-free length of time is required for optimal unaltered growth (Du Plessis, 2003). Natural reproduction of maize exclusively occurs through seed production and dissemination. As a domesticated plant, maize is dependent on human assistance for its seed dissemination since the species has

lost its ability to distribute the seeds. Disease control is a key factor for the successful and sustainable production of maize (Du Plessis, 2003).

2.2. The *Xanthomonas* genus – A general overview

The genus name *Xanthomonas* consists of two Greek words, *xanthos* (yellow) and *monas* (unit), referring to the characteristic yellow pigment (Saddler & Bradbury, 2005). Bacteria belonging to the *Xanthomonas* genus are members of the *Xanthomonadaceae* family and the *gamma*-Proteobacteria class (Pieretti *et al.*, 2012). The bacteria in this class share three common qualities. They possess (i) the ability to colonize intercellular spaces of host plants, (ii) the ability to destroy living plant cells and (iii) *Hrp* genes as part of their pathogenicity system (Alfano & Collmer, 1997).

Members of the *Xanthomonas* genus are primarily plant associated pathogens. They are found in association with major groups of plants and are responsible for synthesizing exudates at lesion sites on infected plants (Lilly *et al.*, 1958; Goszczynska *et al.*, 2000). Members of the genus are known to cause a range of diseases in at least 124 monocotyledonous and 268 dicotyledonous crop plants of economic and agronomic importance worldwide (Leyns *et al.*, 1984; Mansfield *et al.*, 2012; Pieretti *et al.*, 2012).

Disease symptoms of *Xanthomonas* infections include necrosis and gummosis on stems, fruits and leaves (Leyns *et al.*, 1984). The diseases caused by *Xanthomonas* can lead to serious economic losses of crops such as beans, maize and rice as well as cotton, citrus fruit, crucifers and sugarcane. The genus has a world-wide distribution. The substantial losses in the cultivation of a diversity of crops worldwide have led the *Xanthomonas* genus in becoming of large economic importance (Vauterin *et al.*, 1990; Berthier *et al.*, 1993).

Xanthomonas has the ability to infect intact plants as well as grow and survive on plant leaf surfaces. These bacteria survive on plant surfaces as epiphytes before entering the plant through openings such as stomata, wounds or hydathodes (Büttner & Bonas, 2010). After the hydathodes are colonized, these bacteria consequently gain access to the xylem elements where they rapidly multiply.

Alternatively these bacteria can proliferate locally in intercellular spaces. As the disease intensifies they can spread systematically to the mesophyll tissue within the plant upon being released from the vascular network (Dow & Daniels, 2000). *Xanthomonas* are host-specific hemibiotrophic pathogens which infect living plant tissue during the early stages of infection but later kill the infected plant cells (Boch & Bonas, 2001; Büttner & Bonas, 2010). These

pathogens cause foliar chlorosis and late necrosis on the leaves of host plants (Boch & Bonas, 2001).

Xanthomonas has a temperature range between 10°C and 37°C with optimal growth occurring at a temperature of 28°C (Couthino & Wallis, 1991). Cells are Gram-negative and rod shaped with a single polar flagellum for motility. They are aerobic, catalase positive, oxidase negative or weakly positive and do not reduce nitrate. *Xanthomonas* commonly produce extracellular polysaccharides (Goszczyńska *et al.*, 2000). These extracellular polysaccharides have an important role in the protection of these bacteria from factors such as desiccation and damage by wind and rain.

Xanthomonas bacterial pathogens produce yellow, circular, smooth, convex and mucoid colonies with entire margins (Qhobela & Claflin, 1988; Goszczyńska *et al.*, 2000). The yellow colour is due to the presence of the pigment, xanthomonadin, in the bacterial membrane. It is suggested that this pigment protects the cells from photobiological damage (Rajagopal *et al.*, 1997). Xanthomonadins are an exclusive feature of all yellow xanthomonads and are used as diagnostic and chemotaxonomic markers to identify members of the *Xanthomonas* genus (Starr & Stephens, 1964; Goszczyńska *et al.*, 2000). This yellow pigment is present in the outer membrane of bacterial cell walls and consists of brominated aryl polyene esters that resembles the long polyene moiety found in carotenoids (Stephens & Starr, 1963). The polyene ring provides a protective action which enables the bacteria to survive epiphytically in a broad range of conditions (Chun, 2002). These pigments are essential for the bacteria to successfully cause infection.

Xanthomonadin pigments are insoluble in water, but soluble in petroleum ether, methanol and benzene (Stephens & Starr, 1963). They have maximum absorption in methanol at the wavelength values of 420, 441 and 468 nm (Schaad *et al.*, 2001b). Since various yellow-pigmented bacteria are often isolated from the tissues of plants and soil debris, it may be difficult to visually distinguish colonies of *Xanthomonas* species from colonies of saprophytic bacteria based solely on their colour (Irey & Stall, 1982).

The identification of the unique yellow xanthomonadin pigment can be used for confirmation of *Xanthomonas*. It plays a central role in defining the *Xanthomonas* genus and serves as substantiate data for speciation within the genus. Xanthomonadin darkens with the aging of the cultures, turning from light yellow after three days growth to orange-yellow after fourteen days (Coutinho & Wallis, 1991). Xanthomonadins found in different members of the genus varied in bromination and methylation. Studies on different *Xanthomonas* species

established that the bromination and methylation patterns of xanthomonadin pigments assisted in the identification of xanthomonad members (Stephens & Starr, 1963).

2.3. Taxonomy of the *Xanthomonas* genus

The genus *Xanthomonas* forms an extremely tight phylogenetic lineage (Vandamme *et al.*, 1996). The taxonomy of *Xanthomonas* is in some cases still unclear and unsettled. Reasons for this include insufficient past descriptions, adjustments in taxonomic concepts, contradictory data and complications in establishing the value of pathogenicity in taxonomy (Palleroni, 1984; *cited by* Janse, 1992).

The diverse range of plant diseases and consequent economic impacts caused by *Xanthomonas* are the main reasons why the genus has been the subject of numerous taxonomic and determinative studies (Berthier *et al.*, 1993). Different strains exhibit high degrees of host specificity (Kingsley *et al.*, 1993). *Xanthomonas* pathovars possess similar physiological and biochemical characteristics. Certain strains can only be differentiated at a taxonomic level below that of a species. Consequently, the term “pathovar” was suggested in order to refer to bacterial strains with similar characteristics that can be differentiated at an intra-specific level based on their pathogenicity to one or more host plants (Berthier *et al.*, 1993).

The correlation of genomic groups with the specific host they attack is variable. The pathogenic specialization often does not correlate in a satisfactory manner with genetic relationships of genus members. Vauterin *et al.* (1995) discovered that *Xanthomonas* pathovars infecting grasses, crucifers and cereals, which are related hosts, fall within the same or close genomic groups. However, it is not always the case that pathovars from the same genomic group will infect plant hosts that are related. Pathovars infecting dissimilar unrelated hosts may also belong to the same genomic group as the pathovars of grasses, cereals and crucifers. Classification is made difficult as pathogenic specialization can in many cases not successfully be linked to genomic groups (Vauterin *et al.*, 1995).

All current *Xanthomonas* species were previously considered as part of the *X. campestris* group. They are currently recognised as new species with new pathovars. *X. axonopodis* is currently the largest group with the most heterogeneity and includes the most pathovars (Vauterin *et al.*, 1995; Goszczynska *et al.*, 2000). Researchers have used a number of biochemical and physiological tests for identification of Xanthomonads. Genus members were usually characterised based on their host specificity and before 1980, with the methods described by Dye (1962). Nomen-species exhibited physiological and bacteriological

characteristics that were insufficient to quantify these bacteria as different species. Consequently many of the bacterial strains were grouped into a single species, *X. campestris*, and have been assigned unique pathovar names according to their specific host (Norman *et al.*, 1997).

A diverse range of methods have been applied for extensive examination of the *Xanthomonas* genus (Vandamme *et al.*, 1996). The aims of these studies were to address the delineation of the genus, the species and pathovar system as well as the issue regarding the identification of non-virulent strains. Different genotypic and phenotypic methods were used to clarify the complex taxonomy of the pathogen (Vandamme *et al.*, 1996). The species *X. campestris* in particular is considered the most complex species in the *Xanthomonas* genus and has been divided into more than 125 pathovars as described by Dye *et al.* (1980).

Leyns *et al.* (1984) issued a comprehensive report of host range studies of *Xanthomonas*. The authors found that certain strains had strict host specificity while others exhibited a wider host range. It was recommended that a large number of strains should be analysed with a combination of DNA-DNA homology and conventional tests in order to identify *X. campestris* pathovars. The amount of pathovar names have accumulated rapidly due to insufficient information regarding host range and the use of improperly characterised cultures (Norman *et al.*, 1997).

An index of accepted pathovar names was published by Dye *et al.* (1980) together with guidelines to name plant-pathogens in order to establish credibility of the pathovar list. Even with this list it became evident that additional methods should be pursued for further *Xanthomonas* species and pathovar identification (Norman *et al.*, 1997). Vauterin *et al.* (1990) revised the taxonomy of the genus using DNA-DNA homology studies and also suggested a systematic polyphasic approach with the definite inclusion of DNA-DNA homology to improve the understanding of phylogenetic relationships among different *Xanthomonas* pathovars and species.

Serological techniques (Qhobela & Claflin, 1990), restriction fragment-length polymorphism analysis (Qhobela *et al.*, 1991), genomic DNA fingerprinting (Chidamba & Bezuidenhout, 2012), membrane proteins (Qhobela *et al.*, 1991), 16S rDNA sequences (Finkmann *et al.*, 2000; Adhikari *et al.*, 2012), DNA homology (Vauterin *et al.*, 1995) and fatty acid analysis (Vauterin *et al.*, 1996; Massomo *et al.*, 2003) have proved useful to characterise *Xanthomonas* pathovars.

2.4. Pathogenicity of *Xanthomonas*

The *Xanthomonas* genus has a complex phytopathogenicity where 20 to 100 genes are involved in pathogenic functions (Arlat *et al.*, 1991). Pathogenesis involves several phases, starting with bacterial penetration of the host plant through openings. Multiplication of bacterial cells occurs in the intercellular spaces while extracellular polysaccharide is secreted by the bacteria. Results of initial infection include water-soaking and plant cells becoming more permeable which lead to increased nutrient loss from plant cells (Chan & Goodwin, 1999).

In the case of some *Xanthomonas* members, they enter the vascular tissue to multiply and spread throughout the plant. Plant cells in close proximity to bacterial cells degrade and the plant organelles degenerate which cause cell walls to fragment. This enables bacterial cells to enter the cells to multiply further. Upon infection of a susceptible host, symptoms of the plant disease will become visible after several days. Indications of plant infection by *Xanthomonas* pathogens include wilting, chlorosis, necrosis, rotting, cankers and hypertrophy. Xanthomonads are described as hemibiotrophic as the bacteria begin to infect living host tissue but later kill the cells and consume the nutrients present in the dead cells (Chan & Goodwin, 1999).

The outermost molecules of Gram-negative bacteria include extracellular polysaccharide and lipopolysaccharide. Extracellular polysaccharides are secreted to form a slimy outer layer while lipopolysaccharides are components of the cell wall. The extracellular polysaccharide layer determines if bacteria will be able to colonize a particular niche since it is responsible for protecting the pathogen from desiccation, harmful metals as well as defence responses from the host. It also aids in the survival of the bacteria under less favourable conditions (Boch & Bonas, 2001).

The lipopolysaccharide in the outer membrane serves as a permeability barrier to hydrophobic antibiotics and dyes, detergents and other toxic molecules (Kingsley *et al.*, 1993). Extracellular polysaccharides furthermore enhance disease development and have important roles in disease epidemiology (Denny, 1995). Exopolysaccharides may significantly contribute to the virulence of pathogens but are generally not considered a central pathogenicity factor. In many cases phytopathogenic bacteria synthesise various types of exopolysaccharides, depending on single genes or large gene clusters encoding a multifunctional enzyme (Denny, 1995).

Xanthomonads possess numerous pathways to move proteins across the cell envelope. These secretion systems are essential for pathogenic bacteria for the infection of eukaryotic host cells and vary from simple structures such as the type I secretion system which only secretes one substrate protein to complex secretion systems such as the type III and IV which can translocate many effector proteins into host cells. *Xanthomonas* has six secretory systems (Gerlach & Hensel, 2007). These secretion systems are responsible for the secretion of toxins and effector proteins either into the extracellular milieu or directly into the host cells (Remaut & Waksman, 2004).

Many *Xanthomonas* genes have fixed roles either in connection with virulence or in the biosynthesis of the extracellular polysaccharide, xanthan (Dow & Daniels, 2000). Certain pathovars of *X. oryzae* and *X. campestris* are considered among the top ten most important bacterial plant pathogens in the world based on their scientific and economic impacts (Mansfield *et al.*, 2012). The fourth, fifth and sixth positions of this top ten bacterial plant pathogen list, are held by *Xanthomonas* species, all clearly distinctive in their pathology and host targets (Mansfield *et al.*, 2012).

Some members of the *Xanthomonas* genus, in particular *X. campestris* pv. *campestris* produce a unique extracellular polysaccharide known as xanthan. This compound is a polymer of repeating units of pentasaccharide consisting of a cellulose, β -1,4-linked D-glucose, backbone with trisaccharide side chains. Xanthan gives these bacteria their characteristic mucoid appearance (Büttner & Bonas, 2010). Xanthan is commercially utilized as a thickener and viscosifier during the manufacturing processes of nutritional and pharmaceutical products (Becker *et al.*, 1998).

2.5. Diseases of maize

Maize is the target of approximately 110 fungal, bacterial and viral diseases worldwide (Ministry of Environments and Forests, 2010). A large number of bacterial species and pathovars are capable of causing plant diseases. Diseases such as leaf blight, leaf spot, wilt and stalk rot are caused by bacterial infections of maize (CIMMYT Maize Program, 2004). Plant-pathogenic bacteria commonly known to affect maize include *Pseudomonas*, *Erwinia*, *Clavibacter*, *Corynebacterium*, *Spiroplasma* and *Xanthomonas*.

Dissemination of phyto-bacteria mainly occur through human activities, flowing, splashing or wind-blown water, animals such as livestock but also arthropods and nematodes, and by transferring parts of infected plants or soil to other regions. Bacteria become dormant in unfavourable environmental conditions and remain on plant material, soil particles, surfaces

of animal bodies and agricultural instruments until conditions change to promote bacterial growth and proliferation (Bhawsar, 2011). Bacteria are able to sustain themselves epiphytically for long periods of time before entering the plant through wounds. Wet and moist conditions together with warm temperatures generally promote disease development as it increases the susceptibility of the plant to bacterial invasion (Duveiller *et al.*, 1997).

2.6. Bacterial leaf streak disease of maize and the causal agent (*X. campestris* pv. *zeae*)

In bacterial nomenclature the term “pathovar” (pv.) was established in order to classify pathogens based on proven differences in their pathogenicity, either in terms of the specific host they infect or the disease they cause (Dye *et al.*, 1980). Pathovars share nearly identical characteristics and can only be distinguished on an intra-subspecific level on the basis of their pathogenicity to their specific plant host (Dye *et al.*, 1980; Vauterin *et al.*, 1990).

Pathovar names are usually derived from the name of the host plant that the pathogen infects (Vandamme *et al.*, 1996). Dye and Lelliot (1974) demonstrated that it would not be possible to retain species names for all the phytopathogenic bacteria belonging to the *Xanthomonas* genus. Accordingly, the special purpose nomenclature of different pathovars was proposed for plant-pathogenic bacteria which fail to meet the standards for distinct species classification. This designation of pathovars is recognised as a compromise in nomenclature imposed by the current state of knowledge (Young *et al.*, 1991).

Bacterial leaf streak disease of maize caused by *X. campestris* pv. *zeae* has had an increasingly frequent occurrence and continuous incidence over the past few years which raised concern (Coutinho & Wallis, 1991). The disease was first reported in South Africa in 1949 and has since not been reported in any other maize growing country. Although awareness of the disease rose in the 1950's it was not considered economically important (Dyer, 1949; Qhobela *et al.*, 1990). The disease is currently mainly found in the maize producing areas of the country known for their hot and drier climates (Qhobela *et al.*, 1990).

Bacterial leaf streak disease of maize is recognized by symptoms such as lesions extending the length of the plant leaf and large necrotic regions visible on the leaf. Symptoms are intensified when daily temperatures rise above 32°C. Disease incidence seems to increase during irrigation in hot weather which can lead to a 40 % loss of leaf tissue. *X. campestris* pv. *vasculorum* was initially identified as the specific bacterium responsible for causing the

disease but was subsequently proven to only infect maize upon artificial inoculation (Qhobela *et al.*, 1990).

It was demonstrated that *X. campestris* pv. *zeari* represents a new pathovar of the *X. campestris* species based on the distinctive host range of the pathogen (Coutinho, 1988; Qhobela *et al.*, 1990). *X. campestris* pv. *zeari* is evidently distinguishable from other pathovars capable of infecting maize, namely *X. campestris* pv. *vasculorum* and *X. campestris* pv. *holcicola*, based on RFLP analysis (Qhobela *et al.*, 1990). *X. campestris* pv. *zeari* is the only *Xanthomonas* member known to produce severe symptoms on leaves through natural maize infection. *X. campestris* pv. *vasculorum* and *X. campestris* pv. *holcicola* only infect maize through artificial inoculation and exhibit less severe symptoms in infected leaves.

Comparisons of biochemical and physiological data of *X. campestris* pv. *zeari* and *X. campestris* pv. *vasculorum* and *X. campestris* pv. *holcicola* proved to be insufficient for distinguishing the three pathovars with similar host specialization from one another. Investigations of the possibilities of using dot-immunobinding assays and polyacrylamide gel electrophoresis of the membrane proteins for separation of *X. campestris* pv. *zeari* from other pathovars were done but failed to differentiate all three relevant pathovars (Unpublished data; cited by Qhobela *et al.* 1990). Although PAGE analysis could distinguish *X. campestris* pv. *zeari* from *X. campestris* pv. *vasculorum*, it was not successful in differentiating *X. campestris* pv. *zeari* from *X. campestris* pv. *holcicola*. RFLP analysis done by Qhobela *et al.* (1990) succeeded in differentiating all three pathovars from one another by revealing different polymorphic bands. *X. campestris* pv. *zeari* yielded eight bands, while *X. campestris* pv. *holcicola* and *X. campestris* pv. *vasculorum* revealed two and six polymorphic bands respectively after Southern blot analysis.

Maize breeding for the purpose of developing resistance to this pathogen has been successful but these practices have only been done to a point where the disease is no longer considered a major threat to maize crops. Resistance breeding practices have indicated a highly specific pathogen host interaction. Breeders are able to readily select for resistance in host plants as the pathogenicity factor produced by the pathogen appears to be very specific. Unique characteristics of the pathogen include its apparent restriction to South Africa and its ability to infect maize naturally. Although the pathogen has only been reported in South Africa, it is unlikely that this is indeed the case due to frequent exchange of plant material among countries (Qhobela *et al.*, 1990; Downing, 1998).

2.7. Bacterial leaf streak disease of other host plants

Bacterial leaf streak disease occurs on different host plants. *X. translucens* pv. *undulosa* is responsible for causing leaf streak disease of wheat while other members of the *X. translucens* species cause leaf streak disease of other cereal and grasses. The disease is generally regarded as unimportant although it is widespread. In cases where reports of the disease become more frequent however, disease importance may increase (Duveiller, 1997).

Bacteria causing leaf streak disease on plant hosts are mainly disseminated by contaminated plant residues from previous crop seasons, irrigation water systems and splashing or wind-blown rain drops. The plant leaf is typically infected through hydathodes at the tip of the leaf and at leaf wounds and margins. The bacteria multiply in the intercellular spaces and penetrate the xylem vessels. Beads of exudate develop on the leaf surface which is a diagnostic sign of the disease as well as a source of secondary inoculum (Mew *et al.*, 1993).

An early symptom of leaf streak disease is the presence of translucent stripes while later symptoms are characterised by distinct, elongated, light brown lesions of several centimetres long that merge together to cover large solid areas of the leaf (Duveiller *et al.*, 1997). Bacterial leaf streak outbreaks seem to be sporadic and farmers usually observe these occurrences late in the growing season (Forster *et al.*, 1986). Humidity and temperature are considered to have major influences on the epidemiology of the disease (Duveiller & Maraite, 1995).

High humidity conditions in crop fields caused by factors such as rainfall seasons and sprinkler irrigation are expected to increase the incidence of bacterial leaf streak disease. Moisture facilitates the pathogen's invasion and colonization of the leaf tissue. Excessive water running over plant leaves further contributes to the spreading of the pathogen over the leaves which leads to an increase in the number of lesions. The main factors that contribute to the spreading of the inoculum present in the exudates from leaf to leaf and through entire crop fields, are rain and wind. Hail may cause small injuries to the leaves which result in leaves being more susceptible to bacterial penetration (Duveiller *et al.*, 1997).

After penetration, the moisture of the leaf does not constrain further bacterial growth. Bacteria can continue multiplication even under dry conditions (Duveiller *et al.*, 1997). Severity of the disease and expression of symptoms depend on the environmental conditions, cultivar resistance and aggressiveness of the pathogen (Ricaud & Ryan, 1989).

Bacterial leaf streak disease occurs in areas with diverse climatic conditions ranging from temperate to warmer environments with cool nights, subtropical highlands with high rainfall, and regions with frequent fluctuations in temperature or climate. Some pathogens responsible for leaf streak diseases, such as *X. translucens* have been found to be seed-borne (Tsilosani *et al.* 1969, cited by Duveiller *et al.*, 1997).

2.8. Related pathovars of the causal agent of bacterial leaf streak disease

X. oryzae pv. *oryzae* and *X. oryzae* pv. *oryzicola* are the only two pathovars in the *X. oryzae* species (Swings *et al.*, 1990). Both these seed-borne organisms are pathogenic to rice and primarily transmitted through infected seed. They are close relatives of each other and were previously classified as members of the *X. campestris* species. The two pathovars have different methods of infecting their host which can be observed in the clearly distinct symptoms each of the pathovars causes on rice leaves.

X. oryzae pv. *oryzae* is known as the causal agent of bacterial leaf blight of rice. Rice leaves become infected when the bacterium enters the hydathodes and proceed to xylem vessels where bacterial cells multiply (European and Mediterranean Plant Protection Organization, 2007). The disease is recognized by pale leaves with brown to white lesions that are located in line with the leaf veins (Ezuka & Kaku, 2000). Symptoms initially start as stripes that are water-soaked which then enlarge to develop into wavy and elongated lesions. Infected leaves wilt and roll up which may lead to the demise of the entire plant (European and Mediterranean Plant Protection Organization, 2007).

X. oryzae pv. *oryzicola* causes bacterial leaf streak disease of rice. Transparent, water-soaked streaks between leaf veins are the first signs of infection which later darken and enlarge in length. The colour of the streaks can be used to distinguish *X. oryzae* pv. *oryzae* and *X. oryzae* pv. *oryzicola* from one another as *X. oryzae* pv. *oryzae* exhibits opaque lesions in contrast to the transparent lesions of *X. oryzae* pv. *oryzicola*. Lesions may fuse to form larger regions which cause the leaves to wither (Ou, 1985, *cited by* European and Mediterranean Plant Protection Organization, 2007).

The species *X. vasicola* consist of *X. vasicola* pv. *vasculorum*, *X. vasicola* pv. *holcicola* and *X. vasicola* pv. *musacearum* (Karamura *et al.*, 2015). *X. vasicola* pathovars are known to cause diseases such as gumming disease and wilt in sugarcane, banana and maize plants (Carter *et al.*, 2010; Karamura *et al.*, 2015). The pathogens are known to naturally infect maize and sorghum among other monocotyledonous plants (Dookun *et al.*, 2000). *X.*

vasicola pv. *vasculorum* is responsible for deforming maize plants while *X. vasicola* pv. *holcicola* induces stunted growth of plants (Karamura *et al.*, 2015).

X. campestris pv. *campestris* is pathogenic to cruciferous plants of the *Brassicaceae* family including vegetable crops and weeds. The disease caused by the pathogen is a seed-borne, systematic vascular disease known as black rot which has a world-wide distribution (Da Silva *et al.*, 2002). Black rot disease is recognised by yellow V-shaped lesions that stretch from leaf margins and black leaf veins. A warm, humid climate is optimal for disease development and the pathogen is easily spread by rain and irrigation water (Vincente & Holub, 2013). Several genes associated with pathogenicity and virulence have been identified within the genome of *X. campestris* pv. *campestris*, including the *rpf*, *wxc* and *gum* genes. These genes are responsible for regulating the synthesis of xanthan gum, lipopolysaccharides and extracellular degrading enzymes (Vincente & Holub, 2013).

X. axonopodis species cause a wide range of diseases on different hosts. Plant diseases caused by *X. axonopodis* pathovars include bacterial pustule disease of soybean, bacterial blight of cassava and cotton, leaf blight of pomegranate, and bacterial spot of citrus (Arrieta-Ortiz *et al.*, 2013; Lee *et al.*, 2014). *X. citri* is an extremely close relative to *X. axonopodis* and is often pathogenic on the same host plants such as cotton and citrus (Jalan *et al.*, 2013).

2.9. Molecular studies of the *Xanthomonas* genus

Developments in molecular methods have allowed for accurate and quick identification of pathogenic bacteria up to strain level using powerful molecular tools (Duveiller *et al.*, 1997). The precision and sensitivity of molecular tools have made it more suitable than conventional methods for detecting bacteria within contaminated material. It is especially difficult to determine if plant parts that exhibit no visible symptoms are infected with a disease in which case molecular tools can assist in pathogen detection. Various molecular techniques have been applied to detect plant pathogenic bacteria and determine their diversity (Louws *et al.*, 1999).

Analyses of nucleotide sequences are done to gain information regarding the nucleotide composition of a specific DNA fragment. Both conventional and real-time PCR are especially applicable molecular-based methods for the detection of plant pathogens and confirming their presence within infected material. These methods have proved to be rapid, sensitive and accurate in these functions. Newly designed PCR-based assays continue to become

more user-friendly and reliable tools for detection of pathogens and estimating disease outbreaks (Lu *et al.*, 2014).

Real-time quantitative PCR (qPCR) methods are sensitive, accurate and reproducible in quantitative investigations of nucleic acids. qPCR assays have already been developed to detect various plant pathogens of the *Xanthomonas* genus, including *X. fragariae* (Vandroemme *et al.*, 2008) *X. arboricola* pv. *pruni* (Palacio-Bielsa *et al.*, 2011), *X. albilineans* (Garces *et al.*, 2014), *X. oryzae* pv. *oryzae* (Cho *et al.*, 2011; Lee *et al.*, 2014), and *X. oryzae* pv. *oryzicola* (Lee *et al.*, 2014). The use of qPCR assays are likely a superior approach for reliable detection of plant pathogens. The method proved to be highly specific and sensitive for the detection of plant pathogens.

2.10. Next generation sequencing and whole genome sequencing

The experimental determination of the nucleotide sequence of a DNA molecule is known as DNA sequencing (Mount, 2001). Prior to sequencing of the individual bases the genome of interest is first required to be sheared into many fragments. Depending on the sequencing technique applied, the sheared DNA fragments are subjected to different kinds of procedures to identify their nucleotide sequences. Each resulting sequenced fragment is referred to as a read. Sanger sequencing was introduced in 1977 (Sanger *et al.*, 1977). Considering throughput quantities and related costs, sequencing of genomes and large amounts of genetic material have been costly long-term experiments (Bentley *et al.*, 2008).

For the sequence analysis of complete genomes high throughput technologies with low associated costs were required. These demands were met in the form of massive parallel sequencing technologies which were introduced in 2005 (Rogers & Venter, 2005). Massive parallel sequencing technologies operate on a different principle than the Sanger method. The term “next generation sequencing” was designated to the newly developed sequencing technology to differentiate it from the Sanger method. New sequencing platforms have enabled DNA sequencing to increase the throughput as well as reduce costs (Mardis, 2013; Li, 2015). Different platforms are distinct in their sequential natures, read length and throughput. Nevertheless, they share the common function of generating large amounts of different DNA sequence reads in a parallel manner by reading through large quantities of DNA templates simultaneously (Bentley *et al.*, 2008; Rothberg *et al.* 2011).

Whole genome sequencing refers to the process of obtaining a complete sequence of all the DNA of an organism’s full genome. Recent developments in genomics have greatly

improved the possibilities of bacterial pathogenesis research (Wren, 2000). The goal of genome sequencing studies is to establish the exact sequence of nucleotide bases of which each DNA molecule within a genome consists (Reece, 2004). Information obtained from the genetic profile is valuable to explore biological characteristics, gene product activity and unique DNA markers of a specific organism (Reece, 2004; Qian *et al.*, 2005). The nucleotide composition of a pathogen's whole genome enables a better understanding of the pathogenesis mechanisms of the organism as well as the processes that confine the host range of the specific strain (Lee *et al.*, 2005).

Data generated from whole-genome sequence analyses offers larger resolution and closure regarding phylogeny than analysis of individual gene fragments. Where single gene fragments of different close relatives may be identical, sequences of full genomes can be used to reveal all genetic differences (Wasukira *et al.*, 2014). Molecular comparison of whole genome data between organisms with close phylogenetic relationships is important to understand the evolution and adaptation to a specific host.

Due to the large genotypic variety and plasticity among members of the *Xanthomonas* genus, the use of only a few molecular markers for reliable detection is a risk (Feng *et al.*, 2015). Assays encompassing the full genome of an organism as well as comparative genomics have been developed to distinguish dominant molecular markers to identify strains of pathovars (Lang *et al.*, 2010). Genomic comparisons have an important role in taxonomic reclassification (Salzberg *et al.*, 2008). It serves as a valuable tool to open up a new field of accurate and confident diagnoses (Lang *et al.*, 2010). As whole genome sequencing becomes increasingly popular, this approach can later be extended over more diverse plant pathogenic groups to gather molecular data across many species.

2.11. Gene assembly

Genome assembly is the computational component of DNA sequencing (Dale & Von Schantz, 2002). Experimental sequencing technologies have to be combined with computational methods in order to determine whole genome sequences. Computational technology allows for the assembly and subsequent analysis of sequences generated by DNA sequencing platforms. The output data generated by the high throughput sequencing process is generally in FASTA format which is a text-based format representing the nucleotide sequences as single-letter codes (Dale & Von Schantz, 2002).

Complete genomes generally consist of anything from a few million to billions of base pairs. Currently sequencing technologies are limited to reading fragments of only a couple hundred

base pairs resulting in short read lengths for DNA templates (Fullwood *et al.*, 2009). As the genome of interest is sheared into smaller pieces for sequencing it yields large amounts of short reads and not one complete contiguous sequence. The process of breaking up DNA into many small pieces and reading these short base pair fragments through sequencing is referred to as shotgun sequencing (Dale & Von Schantz, 2002). Shotgun sequencing is fundamentally a nonspecific sampling of the genome (Donmez, 2012). This type of sequencing involves over-sampling of the genome sequence to a certain extent in order to ensure that every base present in the genome is present in at least one of the reads (Scheibye-Alsing *et al.*, 2009).

The ratio of the complete amount of base pairs in resulting reads to the length of the actual genome is referred to as “coverage” and describes the level of over-sampling (Bentley *et al.*, 2008). Through genome assembly the whole genome is reconstructed with the single fragmented reads. Shotgun sequencing involves automated sequencing of short random DNA fragments and longer sequences are created by joining these short sequenced fragments together by alignment and overlapping.

Overlapping fragments are combined to form contigs. The contigs will grow longer as sequence assembly continues and overlapping of contigs will occur with assembly progress. Overlapping contigs are consequently joined together to produce one single consensus sequence containing all the individual sequenced fragments and covering the complete originally sequenced DNA (Dale & Von Schantz, 2002). The consensus sequence generated by the various assembling phases is the end product relevant for post-assembly purposes (Chevreux *et al.* 1999; Scheibye-Alsing *et al.*, 2009).

Sequencing assembly can be done either with the re-sequencing approach or with a *de novo* assembly procedure (Albers *et al.*, 2011). For the re-sequencing approach a reference genome of a closely-related strain should be available for mapping purposes. Sequenced reads are mapped to the reference genome and positions of variant sequences such as single nucleotide polymorphisms as well as indels are identified (Bentley *et al.*, 2008; Albers *et al.*, 2011). For *de novo* assembly no reference genome for a related strain is used as guidance when assembling sequence reads. *De novo* assembly focuses on the identification of novel genes or fragments and would enable the construction of reference genomes for unexplored genomic strains (Li *et al.*, 2010). Mapping of short reads to known reference genomes is computationally less complicated than the challenge of *de novo* assembly (Butler *et al.*, 2008; Scheibye-Alsing *et al.*, 2009).

2.12. Phylogenetics as a tool in molecular studies

Molecular phylogenetics consists of a collection of techniques that allow for the understanding of evolutionary relationships between nucleotide sequences through the comparison of these sequences (Brown, 2002). The popularity and applications of molecular phylogenetics has grown together with the increased availability of DNA sequence data generated by PCR analysis as well as genome projects. Phylogenetic studies aim to reconstruct evolutionary relationships among organisms in the form of tree-like patterns. Phylogenetic trees portray patterns of descent and indicate the history of specific organisms or molecules (Brown, 2002; Speed *et al.*, 2006). Genomes constantly evolve through the slow yet continuous process of accumulating mutations. The number of variations in nucleotide sequences between genomes serves as an indication of how recently the genomes have evolved from a common ancestor.

The first step in the reconstruction of phylogenetic trees is to gather comparative nucleotide sequence data to include in the reconstructed tree and aligning all relevant DNA sequences. Alignment of multiple comparative sequences can be a difficult task due to dissimilarities and the presence of indels or point mutations. Computational programs such as Clustal are employed for multiple alignments to produce correct alignments (Jeanmougin *et al.*, 1998). The comparative data in the alignments is then converted to produce a reconstructed tree based on nucleotide differences.

An out-group is generally used to root a reconstructed phylogenetic tree. The out-group is an organism or nucleotide sequence that is known to be distantly related to the other organisms under study (Brown, 2007). The purpose of the out-group is to provide a located root for the tree to enable tracing of the correct evolutionary course. Information from multiple sequence alignments are interpreted mathematically for the reconstruction of a phylogenetic tree. Points of divergence are represented by the tree's internal nodes (Speed *et al.*, 2006). The amount of dissimilar nucleotides between sequences is used to determine the evolutionary distances and is expressed in the lengths of the connected branches.

2.13. Housekeeping genes as molecular markers

The term "housekeeping genes" refer to genes that are essential for cell survival and are synthesised in all cells of an organism (Thellin *et al.*, 1999). These genes are responsible for necessary maintenance of cells and basic cellular functions required by the cell. Based on their key functions, relative constant levels of gene expression can be predicted in all cells

regardless of the conditions (Eisenberg & Levanon, 2013). Fewer fluctuations within the molecular composition of these molecules occur during their synthesis when compared to less vital genes.

These genes are highly conserved within the genome. They are less prone to undergo horizontal gene transfer as they are part of the core genome with only a single copy of the gene present in the complete genomic composition (Marcelletti *et al.*, 2010). Due to the consistent molecular composition of housekeeping genes they often serve as secure internal standards in genomic studies and biotechnological applications (Thellin *et al.*, 1999). The investigation of housekeeping genes is important to infer evolutionary dynamics and phylogenetic relationships (Marcelletti *et al.*, 2010). Housekeeping genes possess the potential to be used as molecular markers for characterising genetic and phylogenetic associations.

Six housekeeping genes that have proven to be useful in previous studies for the inference of phylogenetic relationships within the *Xanthomonas* genus were selected for this study. The housekeeping genes chosen were *atpD* (Simões *et al.*, 2007; Pieretti *et al.*, 2009; Albuquerque *et al.*, 2012), *dnaK* (Pieretti *et al.*, 2009; Albuquerque *et al.*, 2012; Fischer-Le Saux *et al.*, 2015), *fyuA* (Young *et al.*, 2008; Rodriguez-R *et al.*, 2012, Fischer-Le Saux *et al.*, 2015) *recA* (Pieretti *et al.*, 2009), *rpfB* (Simões *et al.*, 2007) and *rpoB* (Ferreira-Tonin *et al.*, 2012; Rodriguez-R *et al.*, 2012). These genes are typically chosen for *Xanthomonas* phylogenetic analysis and are congruent in reflecting species relationships (Young *et al.*, 2008; Pieretti *et al.*, 2009).

2.13.1. *atpD*

The ATP synthase *beta* subunit encoding gene (*atpD*) is a housekeeping gene that has been used in previous studies for the analysis of phylogenetic relationships among *Xanthomonas* bacteria (Simões *et al.*, 2007; Pieretti *et al.*, 2009; Albuquerque *et al.*, 2012; Gardiner *et al.*, 2014). The *atpD* gene is involved in the synthesis of ATP and is often included in multilocus sequence analyses (MLSA) to create concatenated nucleotide sequences for phylogenetic assessment (Pieretti *et al.*, 2009).

2.13.2. *dnaK*

The *dnaK* gene encodes for a molecular chaperone protein that has a functional role in hyperosmotic shock responses. The chaperone is known to prevent aggregation of proteins as well as assist in refolding damaged protein molecules (Stepkowski *et al.*, 2003). Many heat shock proteins including *dnaK* have remained conserved across evolution (Liberek *et al.*, 1992). Due to the vital functions and consequent sequence conservation the *dnaK* gene has been extensively used to gain information of higher resolution in phylogenetic studies.

2.13.3. *fyuA*

The *fyuA* gene encodes for a ton-B-dependent receptor that has been linked to the virulence of several members of the *gamma*-proteobacteria class (Young *et al.*, 2008). The *fyuA* protein is an outer-membrane protein that encodes the ferric yersiniabactin receptor which is responsible for Fe-Ybt siderophore uptake (Schubert *et al.*, 2002). Conditions within a host are often limiting free iron availability and pathogens require highly effective systems to acquire iron. Mechanisms for obtaining iron are considered to be important virulence factors. *Gamma*-proteobacteria genera possess a high-pathogenicity island gene cluster responsible for synthesis of the Fe-Ybt system (Schubert *et al.*, 2002). The *fyuA* protein is an important part of the HPI cluster for Fe-Ybt siderophore uptake.

2.13.4. *recA*

The recombinase A protein encoding gene, *recA*, has vital physiological roles in the prokaryotic SOS response, repairing of DNA and during homologous recombination processes (Rabibhadana *et al.*, 1998). This protein can serve as an ATPase, protease or recombinase. The *recA* gene is ubiquitous and extremely conserved among diverse prokaryotic species (Maréchal *et al.*, 2000).

2.13.5. *rpfB*

The *rpfB* gene is coding for long-chain fatty acyl CoA ligase and is involved in quorum sensing and responsible for regulating pathogenicity factors (Simões *et al.*, 2007; Bi *et al.*, 2014). Proteins involved in controlling synthesis and transduction of a fatty acid signal molecule known as the diffusible signaling factor are produced by the *rpf* gene cluster (Ryan,

2013). The *rpf* gene cluster is made up of nine genes, *rpfA* to *rpfI*, responsible for biosynthesis of the diffusible signal factor signalling molecule. Removal of the *rpfB* gene from the genome of *X. campestris* pv. *campestris* lead to incapability of the bacteria to exploit fatty acids as carbon sources (Bi *et al.*, 2014). The *rpf* gene cluster is specific to the *Xanthomonas* genus as well as *Xylella fastidiosa* and *Stenotrophomonas maltophilia* (Deng *et al.*, 2011). The *rpfB* gene has been employed by Pieretti *et al.* (2012) and Simões *et al.* (2014) as molecular marker to differentiate between *Xanthomonas* species.

2.13.6. *rpoB*

The *rpoB* gene is responsible for encoding the DNA-directed RNA polymerase subunit beta. The gene *rpoB* is considered an alternative molecular marker to distinguish between *Xanthomonas* species and to construct phylogenetic associations (Adékambi *et al.*, 2008; Ferreira-Tonin *et al.*, 2012). Sequence similarity of the *rpoB* gene among bacterial species provides a valuable supplement to DNA-DNA hybridization techniques for delineation of related organisms (Adékambi *et al.*, 2008). The *rpoB* gene contains phylogenetically important information due to its long nucleotide length. A strong correlation exists between the *rpoB* gene and average amino-acid identity which indicates relatedness at complete genome level (Adékambi *et al.*, 2008).

2.14. Primer design

Designing primers for PCR reactions involves the selection of short nucleotide sequences of a DNA region of interest to amplify copies of this specific region. The design of optimal primers for PCR is crucial in order to acquire high-quality sequence results for target regions (Li *et al.*, 2008). The sequence preceding the region of interest as well as the sequence after the region should be known to successfully select primers for amplification. Primers are short single stranded oligonucleotides that selectively amplify genomic DNA fragments through PCR. Amplification of the fragment occurs in both forward and reverse directions as primers anneal to the beginning of the forward strand as well as the beginning of the reverse strand (Li *et al.*, 2008).

Various factors determine whether or not a primer will be effective. Primers for target sequences should meet specific criteria regarding primer length, G + C content, melting temperature, the 3' end nucleotides, palindromes and possibilities of primer-dimer and hairpin formation (Abd-Elsalam, 2003). The sequence should end with at least two G-C

bases at the 3' end known as the GC-clamp, for example GCC or CGC (Dieffenbach *et al.*, 1993). For ideal primers the G + C content should be between 50 % and 80 % (Reece, 2004; Li *et al.*, 2008). Melting temperature should be between 55°C and 65°C. Annealing temperatures of the primer pair must ideally be very similar to each other (Reece, 2004).

The sequence should have no spontaneous ability to produce primer-dimers or hairpin formations. The length of a primer sequence is generally between 17 and 30 nucleotides to ensure selective annealing to a single sequence within the genome (Dieffenbach *et al.*, 1993; Reece, 2004). Sequences containing many of one specific nucleotide would not be ideal as this may lead to the primer binding to repetitive sequences in the DNA. A primer should not contain complementary sequences within its sequence as this may lead to palindromic sequences and hairpin structure formation. Secondary structures are disadvantageous as the primer will be removed from the PCR reaction and consequently the target DNA will not be amplified (Reece, 2004).

Management of plant pathogens rely on detection and diagnosis of diseases, detailed epidemiological studies and strategic control plans. Such management requires reliable quality diagnostic tools that are low-cost and rapid. The potential of comparative genomics with development of diagnostic primers unique to specific plant pathogens have been demonstrated (Lang *et al.*, 2010). The limited availability of pathogen-specific primers is an obstacle for PCR-based pathogen detection and disease diagnosis. Sequencing techniques are becoming more advanced and an increasing amount of bacterial genome sequences are being deposited in the GenBank database which enables broad genomic comparisons (Lang *et al.*, 2010). Based on these comparisons specific primers can be designed based on the polymorphic DNA regions of genes specific to one pathovar (Lu *et al.*, 2014).

2.15. Summary of literature

This chapter provided a general overview of the history and importance of maize as a major food source for the country. Furthermore, the taxonomic structure and pathogenicity of the *Xanthomonas* genus were investigated to explore pathogenicity systems and different methods of plant infection. Other diseases of maize and their causal agents were reviewed briefly to provide a background of plant-pathogenic bacterial dissemination of maize pathogens. Information currently available on the causal agent of bacterial leaf streak disease of maize was provided.

Bacterial leaf streak disease of other host plants, especially of wheat, was described to gain more insight into possible methods of dissemination, climate conditions and disease symptoms. A background of *Xanthomonas* species and pathovars that are closely related to the BLSD causal agent were provided that included a review of their host plants and diagnostic symptoms. The usefulness of molecular studies and next-generation sequencing methods were summarised.

Computational gene assembly was described followed by the use of phylogenetic reconstruction of gene sequences for resolving taxonomic disputes. Information of housekeeping genes relevant to the *Xanthomonas* genus used in the present study were given to demonstrate their molecular functions and their usefulness as molecular markers. Guidelines for designing primers to serve as potentially diagnostic markers were provided as well as applications of diagnostic primers in plant-pathogen detection.

CHAPTER 3: MATERIALS AND METHODS

3.1. Bacterial isolation and cultivation of the causal agent

The medium that was used to cultivate isolates was glucose yeast extract calcium carbonate (GYC) agar (Biolab, South Africa; Appendix A: Medium A) (Schaad *et al.*, 2001a; Atlas, 2010). Isolates of the causal agent used during this study came from two main sources. Firstly, frozen glycerol cultures that were sampled, isolated and long-term stored at - 80°C during previous studies were revived for use in this study (Niemann, 2015; Nienaber, 2015). Bacteria were recovered from the glycerol stock cultures by scraping the frozen surface of the cell culture with a sterile inoculating loop and immersing it in 9 ml of glucose yeast extract calcium carbonate (GYC) broth. After 72 hours of growth at 28°C the growth media was streaked onto GYC agar plates and incubated at 28°C for another 72 hours to obtain single colonies (Tjou-Tam-Sin *et al.*, 2012).

The second manner in which strains of the causal agent were obtained was through direct isolation of the causal agent from infected maize leaves showing symptoms of the disease. Maize leaves were provided by farmers from unknown locations in Grootpan and Ventersdorp, North West Province, South Africa. Direct isolations were done by performing both of the isolation methods recommended by the ARC Grain Crops Institute as described below (Niemann, 2015; Nienaber, 2015).

3.1.1. Direct plating method

Bacterial strains were isolated from infected maize leaves using the following method. Leaves possessing distinct streaks were cut into pieces ($\pm 3 \times 5$ cm). The rectangular pieces were surface sterilized by soaking them firstly in 8.25 % sodium hypochlorite solution for 5 min, followed by 3 min in absolute ethanol. The leaf pieces were then washed in sterilized water to remove residue from the disinfectants after which they were plated onto GYC agar plates. A sterilized spreader was used to flatten the leaf onto the agar for maximum contact between leaf and media.

3.1.2. Serial dilution method

The streaked areas of an infected leaf were removed and surface sterilised as described above. After being rinsed in sterilized water, the leaf pieces were crushed using a sterilized mortar and pestle in 10 ml of sterile water. Serial dilutions up to 10^{-3} were made with the crushed lesion extracts. 0.1 ml of each dilution was pipetted and spread onto a GYC agar plate.

The yellow mucoid colonies were selected and streaked out onto fresh GYC agar plates. Gram-staining as described by Kaplan & Kaplan (1933) was performed on single colonies to determine the purity and morphology of the individual cells. The purity of the isolates was maintained by frequently streaking out single colonies onto freshly made GYC agar plates throughout the study.

3.2. Cultivation of reference strains

Bacterial strains of *X. axonopodis* pv. *axonopodis* (DSM3585) and *X. campestris* pv. *campestris* (DSM3586) were obtained from the German Collection of Cell Cultures and Microorganisms (DSMZ) to serve as reference strains. A freeze-dried pellet of each strain was revived using media and growth conditions specified by the DSMZ Catalogue of Microorganisms (Appendix A: Media B). Freeze-dried pellets of the strains were revived using media and growth conditions specified by the DSMZ Catalogue of Microorganisms. The two reference samples were included in the same whole genome sequencing run as the fresh isolates to obtain standardised comparable results.

3.3. DNA extractions

Prior to DNA extractions, isolates were cultivated in 9 ml nutrient broth under aerobic conditions (Sewariya *et al.*, 2012). Liquid cultures were placed in a temperature-controlled rotary shaker and incubation proceeded for 48 to 72 hours at 28°C. Bacterial pellets were harvested with an Eppendorf 5452 centrifuge (Merck Millipore, Germany). Extraction of total genomic DNA was carried out with the use of a Macherey Nagel NucleoSpin Tissue Kit (REF 740952.250) according to the instructions of the manufacturer. To ensure digestion of all RNA present within the samples, 10 µl of 10 mg/ml RNaseA was added and incubated for 30 minutes at room temperature as an optional step in the manufacturer's protocol.

After extraction, a 1 % (w/v) agarose gel was prepared and gel electrophoresis was performed on the DNA products. One microliter of a 1 kb molecular weight ladder (Fermentas, USA) was mixed with 2 µl 6x Orange loading dye (Thermo Scientific, USA) that contained GelRed (Biotium, US) and was loaded into the first well of the agarose gel to serve as an indicator of DNA band size. Three microliters of DNA from each sample was mixed with 2 µl loading dye and loaded into the wells. Electrophoresis was carried out at 80 V for 45 min with Bio-Rad electrophoresis equipment (USA) to determine the presence and quality of the isolated DNA. Electrophoresis results were viewed using a ChemiDoc MP

imaging system (Bio-rad, USA). Concentrations as well as A_{260} / A_{280} quality of the DNA was determined with a NanoDrop ND-1000 spectrophotometer (Inqaba Biotech, SA).

3.4. Whole genome sequencing

Genomic DNA was prepared for whole genome sequencing as described in Section 3.3. Accurate DNA concentrations were determined with a Qubit 3.0 Fluorometer (Invitrogen, USA) following the protocol provided by the manufacturers. Whole genome sequencing of the DNA of each of the two reference strains, *X. axonopodis* pv. *axonopodis* (DSM3585) and *X. campestris* pv. *campestris* (DSM3586), as well as for eight BLSD isolates was performed in-house with an Illumina Miseq system (US).

The Illumina Nextera XT library preparation kit was used according to the instructions of the manufacturers. Eight picomoles of prepared DNA supplemented with 5 % PhiX (Illumina, US) were loaded for sequencing. The Nextera transposase adapter sequences used for Nextera tagmentation were as follows:

Read 1:

5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG

Read 2:

5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG

PCR primers in the Nextera index kit had the following sequences:

Index 1 Read:

5' CAAGCAGAAGACGGCATACGAGAT [i7] GTCTCGTGGGCTCGG

Index 2 Read:

5' AATGATACGGCGACCACCGAGATCTACAC [i5] TCGTCGGCAGCGTC

Table 1: Index codes and sequences for i7 and i5 indices added to index 1 and 2 sequences.

Sample ID	i7 index	i7 sequence	i5 index	i5 sequence
X85	N703	AGGCAGAA	S503	TATCCTCT
X86	N704	TCCTGAGC	S503	TATCCTCT
XGP	N705	GGA CTCT	S503	TATCCTCT
X1	N706	TAGGCATG	S503	TATCCTCT
X2	N707	CTCTCTAC	S504	AGAGTAGA
X9	N708	CAGAGAGG	S504	AGAGTAGA
X15	N709	GCTACGCT	S504	AGAGTAGA
X22	N710	CGAGGCTG	S504	AGAGTAGA
X23	N711	AAGAGGCA	S504	AGAGTAGA
X45	N712	G TAGAGGA	S504	AGAGTAGA

3.5. Genome assembly and mapping to reference genomes

The genomes were assembled from paired end Illumina Miseq sequence reads using the CLC Genomics Workbench version 9.0 (Qiagen, Aarhus, Denmark). Trimming for quality filtering and final assembly were done in CLC Genomics Workbench selecting the program settings listed in Appendix B.1 – B.3.

The read sets of the reference strains, X85 and X86, as well as for the BLSD isolates, X1-45+GP, were mapped against two complete genome reference sequences of *X. oryzae* pv. *oryzae* PXO99A (NC010717) and *X. campestris* pv. *campestris* ATCC 33913 (NC_003902) obtained from GenBank, NCBI (<http://www.ncbi.nlm.nih.gov/>). Mapping to reference genomes was done to assemble all reads into representative genomes. Contigs for assembled results for X1-45+GP were extracted as a FASTA file.

3.6. Extraction of housekeeping gene sequences

Six housekeeping genes relevant to the *Xanthomonas* genus were selected. The genes used in the study were *atpD* (Simões *et al.*, 2007; Albuquerque *et al.*, 2012), *dnaK* (Albuquerque *et al.*, 2012; Fischer-Le Saux *et al.*, 2015), *fyuA* (Young *et al.*, 2008; Rodriguez-R *et al.*, 2012, Fischer-Le Saux *et al.*, 2015) *recA* (Pierretti *et al.*, 2009), *rpfB* (Simões *et al.*, 2007) and *rpoB* (Ferreira-Tonin *et al.*, 2012; Rodriguez-R *et al.*, 2012). These

housekeeping genes were chosen as they are known to have assisted in providing useful information regarding inter- and intra-species relationships within the *Xanthomonas* genus in previous studies. Sequences of all six housekeeping genes that were available for various currently known *Xanthomonas* species and pathovars were extracted from GenBank as listed in Appendix C: Table 6.

For the *atpD*, *fyuA* and *rpfB* genes only one complete reference gene sequence for each, belonging to *X. campestris* pv. *campestris* (ATCC 33913), were available from GenBank. Four complete reference gene sequences, belonging to *X. oryzae* pv. *oryzae* (KACC 10331), *X. axonopodis* (Xac29-1), *X. albilineans* (GPE PC73) and *X. campestris* pv. *campestris* (ATCC 33913), were used for extracting *dnaK*, *recA* and *rpoB* sequences from the BLSD whole genome sequences. Contigs of the *Xanthomonas* samples were mapped against the reference sequences for the specific genes in CLC Genomics Workbench. Contigs corresponding to gene sequences of interest were extracted. Sequences were loaded into separate databases using the genome assembly program GAP5 (Bonfield & Whitwham, 2010) that is part of the Staden Package (Staden *et al.*, 1998). Each contig sequence set was saved as an individual new database.

Contig sequences were inspected in GAP5 and manual correction of errors and sequence shifts in each contig were done. The edited contigs were used to create consensus sequences that were extracted in FASTA file format. Low quality reads were excluded as well as single or double short sequences present in separate contigs. In the case where more than one complete reference gene sequence was available (*dnaK*, *recA* and *rpoB*), the different sequences of each gene were merged to create one consensus sequence. FASTA files of the consensus sequences were imported into the Molecular Evolution Genetics Analysis program, MEGA6 (Tamura *et al.*, 2013). Shifted positions of nucleotide sequences within all contigs of the same gene were fixed by manually aligning them to reference sequence genes.

3.7. Comparative gene analysis

Sequences for each gene from the various different organisms were imported into MEGA6 and aligned with the corresponding gene sequences extracted for the BLSD isolates. By aligning the gene sequences of the isolates against the gene sequences of other *Xanthomonas* species, the sequence data could be mined for match and mismatch sequence fragments. Identification of unique gene sequences indicating variable regions

among the different *Xanthomonas* species was done for further designing and generating possible diagnostic genomic-based markers.

The numbers of variable and conserved nucleotides within the complete gene sequences were calculated with MEGA6 by comparing selective sequences with one another. Total nucleotide differences between the eight BLSD isolates were determined. Sequences were also compared to reference sequences for each reference species obtained from Genbank. Gene similarity of the BLSD strains and reference species were determined for each of the six housekeeping genes.

3.8. Phylogenetic analysis

Complete and partial housekeeping gene sequences for different species and pathovars of the *Xanthomonas* genus were obtained from GenBank. Additionally, sequences of the relevant housekeeping genes of *Stenotrophomonas* species were obtained to serve as out-group during phylogenetic analysis (Albuquerque *et al.*, 2012; Ferreira-Tonin *et al.*, 2012). A complete list of all *Xanthomonas* strains that were used for the phylogenetic component of this study is provided in Appendix C: Table 7.

Partial sequences of the housekeeping genes were aligned and used to infer the phylogenetic relationships between BLSD and other genus members. The alignment of nucleotide sequences was done with the ClustalW algorithm (Thompson *et al.*, 1994) in MEGA6 using default parameters. Sequences in the alignments were trimmed at the ends to create sequences of consistent lengths. The neighbour-joining method was used to assess topology of the constructed phylogenetic trees with 1 000 bootstrap replications for tree consistency (Ferreira-Tonin *et al.*, 2012). The maximum likelihood model was used to construct separate neighbour-joining trees for each of the six genes. Settings for phylogenetic tree constructions in MEGA6 are provided in Appendix B.4.

3.9. Primer design

The multiple alignments of complete genes containing the sequences of the *Xanthomonas* samples and closely related species (Section 3.5) were analysed in MEGA6. Basic composition analysis was done using the MEGA6 Sequence Data Explorer tool. Conserved and variable sites within and among the sequences were determined through sequence comparisons. Gene sequences of the most closely related species were selected to which

sample sequences were compared to, in order to find the most variable housekeeping gene sequences among the *Xanthomonas* strains examined.

Nucleotide sequences that were conserved among BLSD isolates but different from closely related species were identified. Sequences of interest were documented and assessed with Vector NTI 10 software (Invitrogen, USA) to test their potential to serve as optimal primers for amplification of sequence regions unique to BLSD isolates. Six primer pairs were designed to amplify different variable regions present in the *dnaK*, *recA* and *rpoB* genes of the BLSD causal agent. Sequence regions larger than 500 bp were used to design primers for the variable fragments.

3.10. Primer specificity testing with PCR amplification

All primers were synthesised by Inqaba Biotech (SA). Individual PCR solutions, each with a total volume of 25 µl, were prepared. Each solution contained 12.5 µl DreamTaq Master Mix (DreamTaq DNA polymerase; 2X DreamTaq buffer; 0.4 mM dATP, dCTP, dGTP and dTTP; 4 mM MgCl₂) (Thermo Scientific, USA), 9.5 µl nuclease-free water (Thermo Scientific, USA), 1 µl of each primer (at concentration of 0.4 µM) and 1 µl genomic DNA from each sample.

Primer pairs were tested with PCR reactions in a C1000 thermal cycler (Bio-Rad, US). Cycling conditions were an initial denaturation temperature of 94°C for 300 seconds, 30 cycles of denaturation at 94°C for 30 seconds, primer-specific annealing temperatures for each primer set for 30 seconds, primer extension at 72°C for 60 seconds, followed by final extension at 72°C for 600 seconds (Young *et al.*, 2008). Approximate annealing temperatures as determined by Vector NTI 10 software, applying the settings provided in Appendix B.5, were used to set up a 10°C gradient to determine the annealing temperature that yields optimal PCR results. PCR products were resolved with agarose gel electrophoresis as described in Section 3.3 using 3 µl of each PCR product.

CHAPTER 4: RESULTS

4.1. Cultivation of isolates and Gram-staining

After isolation and cultivation on GYC agar plates, the BLSD isolates yielded single colonies as shown in Figure 1. Colonies appeared convex, smooth, mucoid and glistening with a bright creamy yellow colour. This yellow colour darkened as the colonies aged.



Figure 1: An example of the yellow colonies characteristic to the *Xanthomonas* genus after cultivation of the isolates on GYC agar plates.

When agar plates with cultures were incubated for longer than 72 hours, overgrowth became visible with single streaks merging together due to the mucoid nature of the colonies. Gram-staining of the single colonies showed Gram-negative rod-shaped bacterial cells. No other morphological types were observed in the bacterial smears during microscopy studies. Isolated colonies thus appeared to be successfully purified.

4.2. Genomic DNA extractions

Figure 2 represents an image of the isolated DNA for the BLSD samples. Visualisation of the isolated genomic DNA after gel electrophoresis yielded clear bands with molecular weight above 10 000 bp. No RNA or fragmented DNA was present in any of the lanes.

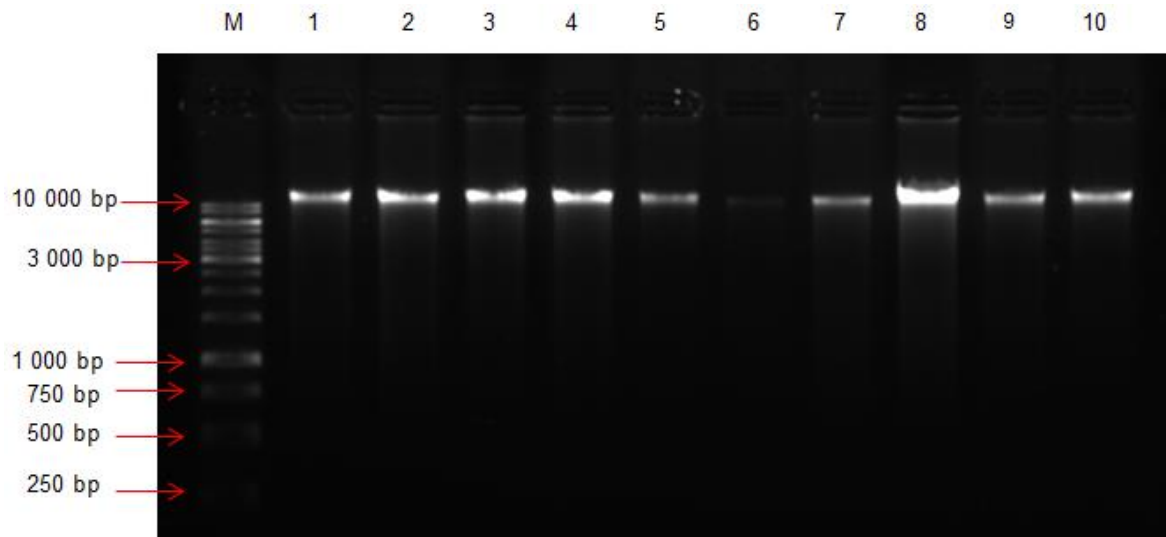


Figure 2: Genomic DNA bands of BLS samples in 1 % (w/v) agarose gel with a 1 kb ladder (M) to indicate band sizes. Samples loaded in the gel were as follows: lane 1 = X1, lane 2 = X2, lane 3 = X9, lane 4 = X15, lane 5 = X22, lane 6 = X26, lane 7 = X33, lane 8 = X35, lane 9 = X45 and lane 10 = XGP.

Based on the appearance and clarity of the bands in the gel as well as the NanoDrop results the samples that exhibited the best DNA quality were selected for subsequent whole genome sequencing. DNA concentrations ranged from 15.74 ng/ μ l to 80.48 ng/ μ l and A_{260}/A_{280} ratio varied between 1.65 and 1.99. Samples were selected based on their A_{260}/A_{280} values. Samples with A_{260}/A_{280} values between 1.6 and 1.8 or which were the closest to this range were selected. RNA-free DNA was the most important criterion for selection of samples. Concentrations of DNA were not of much importance as Illumina whole genome sequencing can be performed with even extremely low DNA concentrations. The isolates representing the BLS causal agent will be collectively referred to as X1-45+GP in all discussions that follow in this study. Sample 85 contained pure DNA material of *X. axonopodis* pv. *axonopodis* (DSM3585) and will be further referred to as X85 while sample 86 represented *X. campestris* pv. *campestris* (DSM3586) and will be shortened to X86.

4.3. Sequence assembly

Sequencing of the genomes of samples with the Illumina Miseq platform yielded paired-end reads that came in two separate files, one for the forward and reverse end respectively. Sequences were in FASTQ format for subsequent input into assembly programs. After trimming of sequences in CLC Genomics Workbench the following trimming results were obtained: 4 902 822 sequenced reads required trimming under the specified settings and

439 742 reads were removed. A total of 438 789 of the removed reads were discarded based on short lengths and 953 read sequences were removed due to poor quality. Sequence assembly results after necessary trimming of reads in CLC Genomics Workbench are provided in Table 2.

Table 2: Sequence assembly results of the collected assembly of reads for X85 + X86 and X1 – 45 + GP, respectively after sequence trimming in CLC Genomics Workbench.

Assembly options from CLC Workbench	Samples of which reads were assembled together	
	X85 + X86	X1 – 45 + GP
Word size	20	22
Bubble size	205	204
Paired distance size	110 – 590 bp (average)	110 – 590 bp (average)
Amount of reads not mapped	8 471	236 692
Reads input	384 725	3 655 045
Average read length	205	204
Reads matched	376 254	3 631 353
Amount of contigs	936	931
Average contig length	4 850	5 697

Word size, bubble size and paired end distances were similar for the isolates (X1-45+GP) and the reference strains (X85+X86). For X1-45+GP 3 631 353 reads were available for input. The total number of reads collectively represented the whole genome sequences of the eight samples. For the two reference strains, X85 and X86, 384 725 reads were available for input to represent the whole genome sequences of the two samples. A total of 3 631 353 input reads of X1-45+GP were matched with other reads from these samples. This yielded a matching percentage of 99.35 %. A total of 376 254 reads matched other read sequences for X85 and X86 which were 97.80 % of the total input reads. The number of contigs generated for X1-45+GP were 931 while 936 contigs were constructed for X85+X86. Average length of the contigs for X1-45+GP were 5 697 nucleotides and for X85+X86 4 850 nucleotides.

4.4. Genome assembly

Collective reads representing the whole genomes of isolates, X1-45+GP, X85 and X86 were mapped against two reference genomes of *X. oryzae* pv. *oryzae* strain PXO99A (NC010717) and *X. campestris* pv. *campestris* strain ATCC 33913 (NC003902) obtained from GenBank. The read mapping results are summarised in Table 3. After reads had been assembled, mapped and contig sequences constructed, a genome sequence of approximately 5 Mbp was obtained for the BLSD pathogen.

Several large gaps were observed within the whole genome sequences of all the BLSD samples. It might be expected to detect some gaps between the reference sequences and the BLSD isolate genomes as genome compositions are likely to vary between newly sequenced organisms and reference species. The isolates were thus expected to possess an amount of unique attributes that would not be mapped to existing reference genomes.

However, the few hundred large gaps that were present in the genome sequence were unusual and indicated possible incomplete genome sequencing. This may be due to degeneration of the DNA before sequencing commenced or other problems during the sequencing run. Some of the gaps may have been present due to the specific region not existing within the genome. Other gaps may have represented fragments that actually existed within the genome but were not sequenced due to the above mentioned potential problems. Because of the presence of the gaps in the data, the whole genome sequences could not be assumed to be complete. It would be impractical to use the data for full genome comparison purposes as it would likely lead to incorrect results and conclusions. The data was still useful to indicate possible sequence similarities between the organisms and guide further analyses.

Table 3: Whole genome sequences from the two reference strains, X85 and X86, as well as the isolates, X1-45+GP, mapped against two complete genome reference sequences, NC010717 (*X. oryzae* pv. *oryzae* strain PXO99A) and NC003902 (*X. campestris* pv. *campestris* strain ATCC 33913).

Sample(s) mapped to GenBank reference sequence	Amount of total reads mapped to reference sequence (bp)	Percentage (%) of total reads mapped to reference sequence
X85 to NC010717	280 922 of 384 725	73.02
X85 to NC003902	253 607 of 384 725	65.91
X86 to NC010717	235 967 of 423 310	55.74
X86 to NC003902	422 867 of 423 310	99.90
X1-45+GP to NC010717	2 624 218 of 3 655 045	71.80
X1-45+GP to NC003902	2 337 335 of 3 655 045	63.95

The sequenced reads of sample X85 mapped to the complete genome sequence of *X. oryzae* pv. *oryzae* strain PXO99A (NC_010717) yielded a total of 73.02 % of the reads that matched the whole genome sequence. X85 mapped to *X. campestris* pv. *campestris* strain ATCC 33913 (NC_003902) revealed only 65.91 % of the reads matched the genomic composition of *X. campestris* pv. *campestris*. These results indicate that the complete nucleotide sequence of X85 is more similar to *X. oryzae* pv. *oryzae* strain PXO99A (NC_010717) than to *X. campestris* pv. *campestris*.

Only 55.74 % of the sequenced reads of X86 could successfully be mapped to *X. oryzae* pv. *oryzae* demonstrating that the two share only slightly more than half of their nucleotide composition. X86 mapped to *X. campestris* pv. *campestris* strain ATCC 33913 (NC_003902) revealed that 99.9 % of the reads of X86 were present in the reference genome sequence of *X. campestris* pv. *campestris*. This could be expected as the two differ only on strain level but are the same organism up to pathovar level both being *X. campestris* pv. *campestris*.

All Illumina reads generated for all eight BLS D samples were combined before mapping them to the reference genomes. These collective reads are referred to as X1-45+GP in Table 3. A total of 71.8 % of the sequenced reads generated for the isolates X1-45+GP that collectively represented the genome composition of the BLS D causal agent matched the genome sequence of *X. oryzae* pv. *oryzae* strain PXO99A (NC_010717). Only 63.95 % of the reads could be mapped to the genome of *X. campestris* pv. *campestris* strain ATCC 33913 (NC_003902) which is approximately 12 % less matching sequences than X1-45+GP

to *X. oryzae* pv. *oryzae*. These results might suggest that the full genomic composition of the BLSD causal agent is more similar to *X. oryzae* pv. *oryzae* than to *X. campestris* pv. *campestris*.

4.5. Comparative gene analysis

Complete sequences of relevant housekeeping genes were extracted from the whole genome sequence data. Gene sequences of *atpD*, *dnaK*, *fyuA*, *recA*, *rpfB* and *rpoB* for the BLSD causal agent were obtained. The complete gene sequences of each housekeeping gene were compared to all available corresponding complete sequences for *Xanthomonas* species obtained from GenBank by creating multiple sequence alignments. Table 4 provides a summary of the gene comparison results including the number of similar and different nucleotides between BLSD isolates and included reference species. The analysed reference strains X85 and X86 were also included in the comparisons.

Table 4: Complete gene sequence comparisons of reference strains with BLSD isolates indicating overall gene similarity of each species with the BLSD causal agent. (X.c = *X. campestris* pv. *campestris* (ATCC 33913); X.o = *X. oryzae* pv. *oryzae* (KACC 10331); X.a = *X. axonopodis* (Xac29-1) and X. alb = *X. albilineans* (GPE PC73).

Gene length (bp)	Reference sequence	BLSD isolates	Different nucleotides	Similar nucleotides	% Similarity
atpD 1 407 bp	X85	X1-45+GP	53	1 354	96.2
	X86	X1-45+GP	52	1 355	96.3
	X. c	X1-45+GP	52	1 355	96.3
dnaK 1 926 bp	X85	X1-45+GP	68	1 858	96.6
	X86	X1-45+GP	80	1 846	95.9
	X. c	X1-45+GP	80	1 846	95.9
	X. o	X1-45+GP	80	1 846	95.9
	X. a	X1-45+GP	60	1 866	96.9
	X. alb	X1-45+GP	247	1 679	87.2
fyuA 2 217 bp	X85	X1-45+GP	130	2 087	94.1
	X86	X1-45+GP	130	2 087	94.1
	X. c	X1-45+GP	212	2 005	90.4
recA 1 035 bp	X85	X1-45+GP	93	942	91.0
	X86	X1-45+GP	100	935	90.3
	X. c	X1-45+GP	100	935	90.3
	X. o	X1-45+GP	39	996	96.2
	X. a	X1-45+GP	77	958	92.6
	X. alb	X1-45+GP	144	891	86.1
rpfB 1 683 bp	X85	X1-45+GP	115	1 568	93.2
	X86	X1-45+GP	198	1 485	88.2
	X. c	X1-45+GP	198	1 485	88.2
rpoB 4146 bp	X85	X1-45+GP	143	4 003	96.5
	X86	X1-45+GP	2 743	1 403	33.8
	X. c	X1-45+GP	2 743	1 403	33.8
	X. o	X1-45+GP	141	4 005	96.6
	X. a	X1-45+GP	116	4 030	97.2
	X. alb	X1-45+GP	472	3 674	88.6

4.5.1. *atpD*

The sequence for this gene was 1 407 bp in length. No nucleotide differences could be distinguished between *atpD* sequences of the BLSD isolates (X1-45+GP). Gene similarity among BLSD isolates was thus 100 %. Alignment of the sequences of the BLSD isolates against the reference strain *X. axonopodis* pv. *axonopodis* (X85) revealed 53 nucleotide differences which is the equivalent of 96.2 % gene similarity. Nucleotide similarity of 96.3 % was found between BLSD isolates and *X. campestris* pv. *campestris* (X86 & ATCC 33913) where 52 variable nucleotides were present.

4.5.2. *dnaK*

The *dnaK* gene was 1 926 bp long with 56 nucleotide differences between the BLSD isolates. All of the BLSD isolates exhibited one or more gaps at some positions within the gene sequence. Gaps were present at different positions for different isolates and were thus not consistent at the sites for each affected sample. These inconsistent gaps could be due to incomplete genome sequencing as described in Section 4.4. A representative consensus sequence was created for the isolates using the gene sequence of each isolate. Between the gene sequence of the BLSD isolates and *X. axonopodis* pv. *axonopodis* (X85) 68 nucleotides were variable, giving a gene composition similarity of 96.6 %.

Gene sequences of both *X. campestris* pv. *campestris* (X86) and *X. campestris* pv. *campestris* (ATCC 33913) differed from the consensus sequence of the isolates with 80 variable nucleotides and had 95.9 % gene similarity. The alignment of the BLSD sequence against the *dnaK* gene sequence of *X. oryzae* pv. *oryzae* (KACC 10331) had 80 variable nucleotides and 95.9 % gene similarity. *X. axonopodis* (Xac29-1) differed from the BLSD isolates sequence with 60 mismatched nucleotides, revealing 96.9 % similarity between their *dnaK* genes. The sequences of the isolates compared with the *dnaK* sequence of *X. albilineans* (GPE PC73) revealed a large number of 247 variable bases with gene similarity of 87.2 %.

4.5.3. *fyuA*

The gene sequence was 2 217 bp long. Sequences for the gene were highly conserved between the BLSD isolates with only one variable nucleotide site present. Sequence similarity among the isolates was 99.95 %. Nucleotide sequence similarity between the

BLSD isolates and *X. axonopodis* pv. *axonopodis* (X85) was 94.1 % with 130 variable nucleotide sites present. A similarity of 94.1 % with 130 variable sites were also observed between the BLSD samples and *X. campestris* pv. *campestris* (X86). Gene comparisons between *X. campestris* pv. *campestris* (ATCC 33913) and BLSD isolates indicated 212 variable nucleotides with overall nucleotide similarity of 90.4 %.

4.5.4. *recA*

The *recA* gene sequence was 1 035 bp in length. Between the sequences of the BLSD isolates only one nucleotide difference could be observed, revealing 99.9 % gene similarity. Comparisons of the sequences of BLSD isolates with the reference *recA* gene sequence of *X. axonopodis* pv. *axonopodis* (X85) showed 91 % gene similarity. There were only 93 variable sites. Nucleotide similarity of 90.3 % with 100 nucleotide differences were found between the BLSD sequences and those of both *X. campestris* pv. *campestris* (X86) and *X. campestris* pv. *campestris* (ATCC 33913). Thirty-nine variable nucleotides existed between the *recA* gene sequences of *X. oryzae* pv. *oryzae* (KACC 10331) and the BLSD isolates. Gene similarity between BLSD isolates and *X. oryzae* pv. *oryzae* was 96.2 %. The BLSD isolate *recA* sequences differed from *X. axonopodis* (Xac29-1) with 77 variable sites, sharing 92.6 % nucleotide similarity. BLSD isolate sequences did not match 144 bases of *X. albilineans* (GPE PC73) resulting in 88.6 % gene similarity between the two organisms.

4.5.5. *rpfB*

The *rpfB* gene sequence had a length of 1 683 nucleotides. Three variable nucleotides were present among all BLSD isolate sequences for the *rpfB* gene. BLSD isolates shared 99.82 % nucleotide similarity between different samples. The sequence of reference strain *X. axonopodis* pv. *axonopodis* (X85) exhibited 115 nucleotide differences when compared to sequences of the BLSD samples, revealing gene similarity of 93.2 %. *X. campestris* pv. *campestris* (X86) as well as *X. campestris* pv. *campestris* (ATCC 33913) each shared *rpfB* gene similarities of 88.2 % with the BLSD isolates.

4.5.6. *rpoB*

The complete gene sequence was 4 146 bp long. Gene similarity among isolates was 99.82 % with the occurrence of 7 nucleotide differences among BLSD isolate sequences. *X. axonopodis* pv. *axonopodis* (X85) and the BLSD isolate sequences had 143 variable

nucleotide sites which yielded a genetic similarity of 96.5 %. Gene similarity of only 33.8 % with 2 743 differences was observed between both *X. campestris* pv. *campestris* (X86) and *X. campestris* pv. *campestris* (ATCC 33913) and BLSD sequences. The *rpoB* sequence of *X. oryzae* pv. *oryzae* (KACC 10331) aligned against the gene sequences of BLSD isolates showed 141 mismatch nucleotides and 96.6 % similarity. Nucleotide similarity of 97.2 % with 116 variable sites was present between *X. axonopodis* (Xac29-1) and the gene sequences of the BLSD isolates. The comparison of the *rpoB* sequence of *X. albilineans* (GPE PC73) and the *rpoB* sequences of the BLSD samples revealed 472 differences in the nucleotide bases resulting in 88.6 % gene similarity.

For the genes where more than one complete reference sequence was available (*dnaK*, *recA* and *rpoB*), the gene sequence for *X. albilineans* (GPE PC73) exhibited the smallest nucleotide similarity with the BLSD isolates. Sequence similarity between the BLSD samples and *X. albilineans* (GPE PC73) were in each case below 90 %, while all other included species generally shared sequence similarities of above 90 % with the BLSD samples. *X. albilineans* (GPE PC73) could thus be considered to be the *Xanthomonas* species that is genetically the most different from the BLSD causal agent, from the included species. For each gene the reference species that shared the highest nucleotide similarity with BLSD isolates varied. One specific reference species did not consistently reveal highest gene similarity. One species could thus not be pin-pointed as the closest relative to the BLSD causal agent. It was evident that the BLSD causal agent was not closely related to the *X. campestris* pv. *campestris* species.

4.6. Phylogenetic analysis

Neighbour-joining trees were constructed to illustrate the phylogenetic relationship of BLSD isolates to other *Xanthomonas* members based on the selected housekeeping genes. The maximum composition likelihood model was used for analysis. One thousand bootstrap replicates were selected to assure consistent branch values. Phylogenetic analysis of all the housekeeping genes revealed clear groupings of all the BLSD isolates together, with high bootstrap support values (>90 %). Because of this high similarity among isolates the sample X45 was selected to serve as representative type strain of the BLSD causal agent in all phylogenetic trees.

4.6.1. *atpD*

Phylogenetic analysis of the *atpD* gene was based on partial sequences of 700 nucleotides in length obtained after trimming the ends of the multiple sequence alignment. BLSD type strain (X45) represents the clustering of the BLSD causal agent among other species in Figure 3 based on *atpD* gene sequences. HKG *Xanthomonas atpD*:X85 represents the position of the reference sample of *X. axonopodis* while HKG *Xanthomonas atpD*:X86 indicates the position of the reference sample *X. campestris* pv. *campestris*.

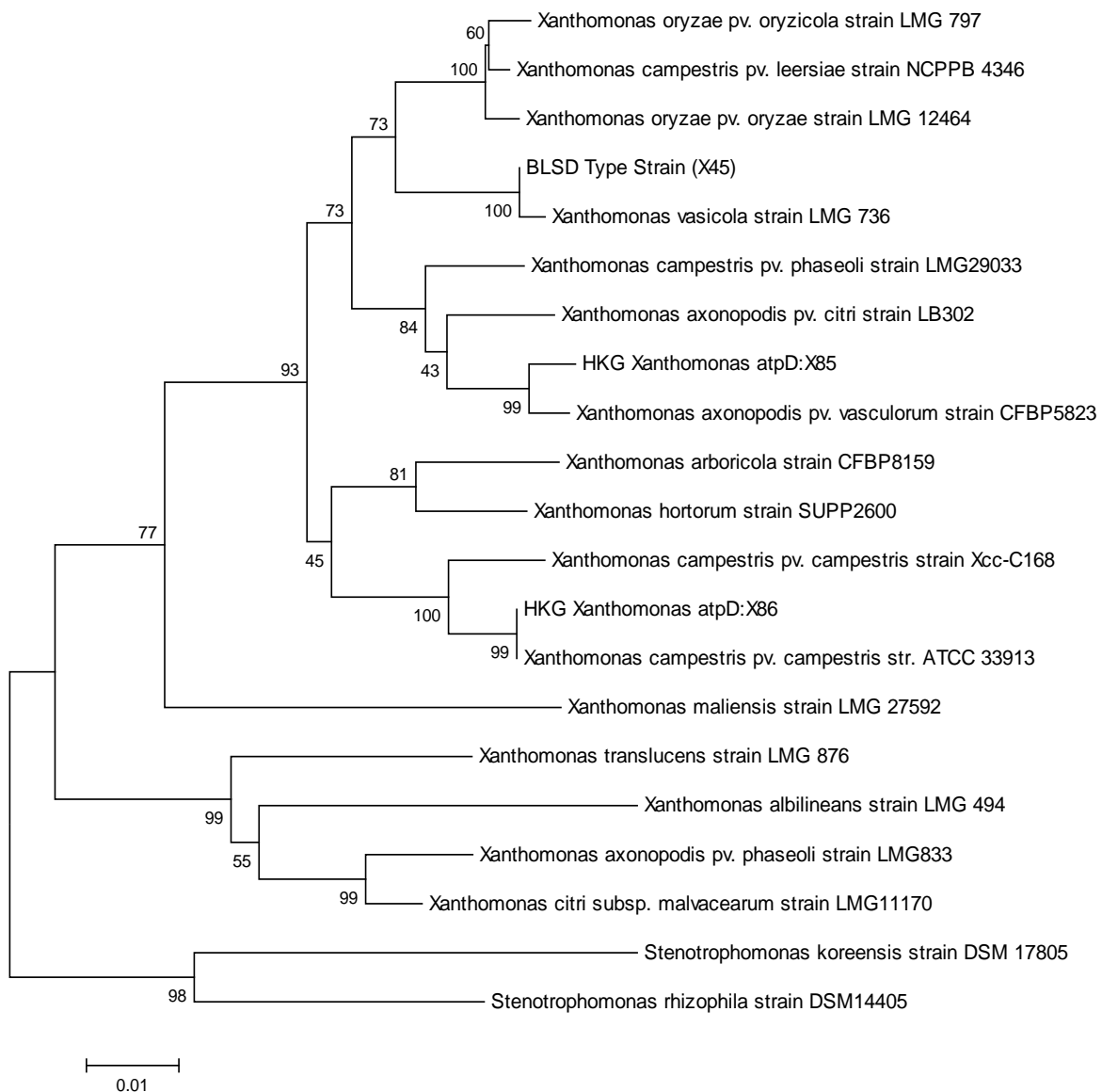


Figure 3: Neighbour-joining tree for the *atpD* gene of various *Xanthomonas* species and the clustering of the BLSD strain (X45) as well as reference strains of *X. axonopodis* pv. *axonopodis* (HKG *Xanthomonas atpD*:X85) and *X. campestris* pv. *campestris* (HKG *Xanthomonas atpD*:X86) within the phylogenetic tree.

Figure 3 shows that the BLSD *atpD* gene sequence clustered with *Xanthomonas vasicola* strain LMG 736 with 100 % bootstrap confidence. *X. vasicola* had a longer branch length than X45 that indicates a slightly higher number of nucleotide substitutions when compared to the BLSD strain. The *atpD* gene of the reference strain of *X. axonopodis* (X85) included in the analysis grouped together with *X. axonopodis* pv. *vasculorum* with 99 % bootstrap support confirming a close relationship. The reference strain of *X. campestris* pv. *campestris* (X86) revealed high similarity to the gene sequence of *X. campestris* pv. *campestris* ATCC 33913 with a bootstrap value of 99 %. There is also a close clustering with *X. campestris* pv. *campestris* strain Xcc-C168 with 100 % bootstrap support. This should be expected as all three organisms only differ at the strain level but all belong to the same pathovar.

All included strains could be differentiated from one another based on their *atpD* gene sequences, with no indiscriminate clustering of different species. *X. axonopodis* strains revealed an unusual scattered phylogeny (Figure 3) with the various pathovars not grouping together with other species. *X. axonopodis* pv. *phaseoli* and *X. citri* subsp. *malvacearum* shared an early-branching cluster with *X. translucens* and *X. albilineans* while the other *X. axonopodis* species formed a later-branching group.

4.6.2. *dnaK*

Sequences of different species included in the *dnaK* sequence alignment were trimmed to 812 bp. The *dnaK* gene sequence of the BLSD isolates grouped together with *X. vasicola* pv. *vasculorum* with 100 % bootstrap accuracy (Figure 4). The reference strain of *X. axonopodis* (X85) clustered with *X. axonopodis* pv. *vasculorum* strains with 100 % bootstrap confidence. This clustering confirms the identity of the reference strain, X85, as a member of the *X. axonopodis* species.

The reference strain of *X. campestris* pv. *campestris*, X86, clustered with the *X. campestris* pv. *campestris* ATCC 33913 sequence and other *X. campestris* pv. *campestris* strains as well as *X. campestris* pv. *incanae*. The branch node where *X. campestris* species grouped together had a bootstrap value of 100 %. This clustering confirms the identity of sample X86 as *X. campestris* pv. *campestris*. *Stenotrophomonas maltophilia* formed a distinct root on the tree while *X. translucens* pathovars grouped together, forming a separate early-branching cluster. All different species could be distinguished from one another, each forming a distinct group. Variations in different pathovars from the same species could be observed such as the discrimination of different *X. oryzae* pathovars (Figure 4). Differentiation of different strains of the same pathovar were also possible by *dnaK* sequence comparisons, such as

the close grouping of different *X. campestris* pv. *campestris* strains closely together but with a notable degree of variation.

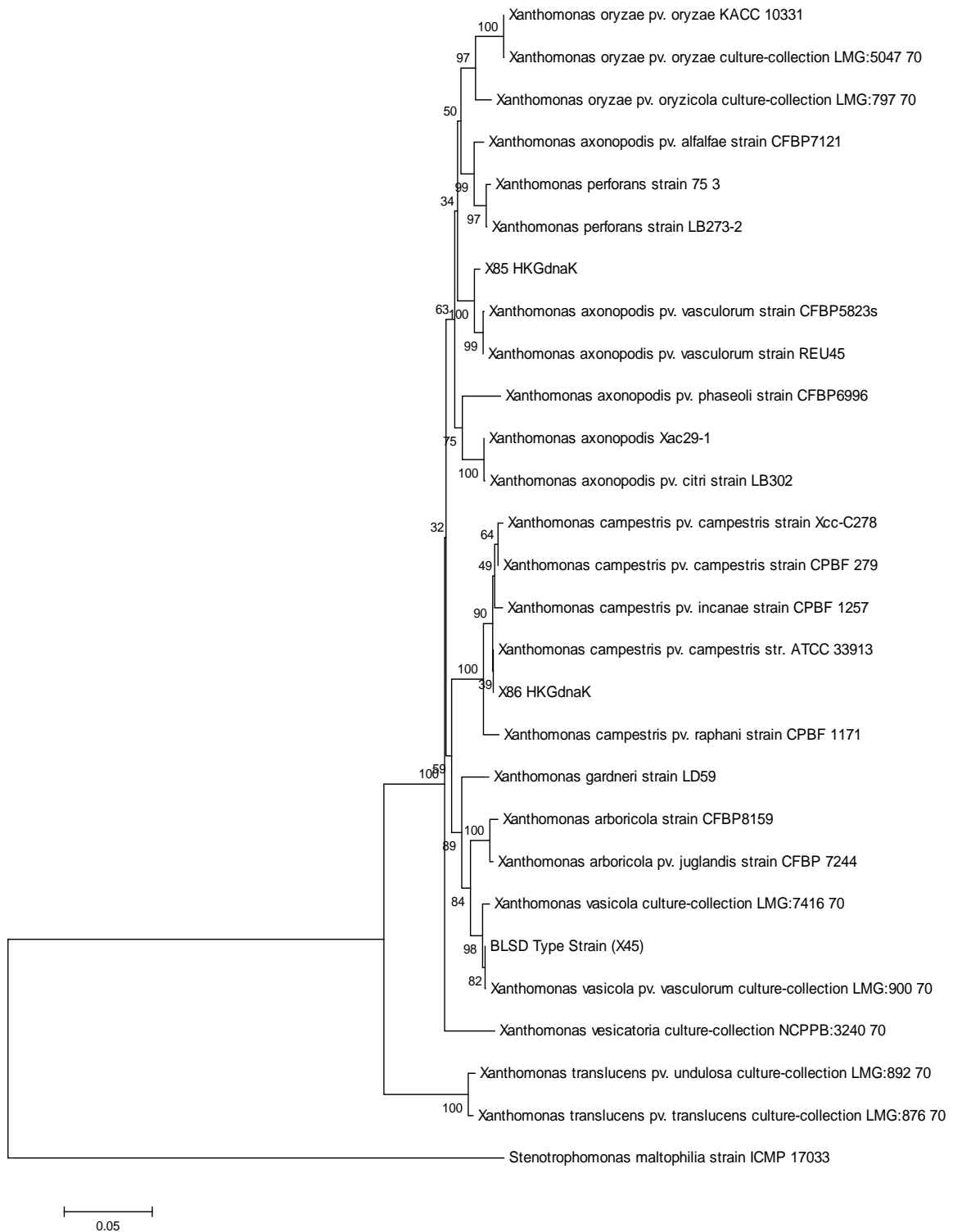


Figure 4: Neighbour-joining tree of the BLSD strain (X45) as well as reference strains of *X. axonopodis* pv. *axonopodis* (HKG dnaK:X85) and *X. campestris* pv. *campestris* (HKG dnaK:X86) in the *Xanthomonas* genus using phylogenetic analysis of *dnaK* gene sequences.

4.6.3. *fyuA*

Nucleotide sequences for phylogenetic analysis based on the *fyuA* gene were trimmed to lengths of 749 nucleotides. The phylogenetic tree obtained using partial *fyuA* gene sequences of various *Xanthomonas* species is shown in Figure 5. Figure 5 shows the clustering of BLSD among other available species based on their *fyuA* gene sequences. The BLSD grouped separately but close to *X. oryzae* pv. *oryzae* and *X. oryzae* pv. *oryzicola* strains with 92 % bootstrap confidence.

X. axonopodis, X85, and reference strain of *X. campestris* pv. *campestris*, X86, grouped together with a bootstrap value of 100 %. Although forming a separate cluster the two representatives (X85 and X86) grouped near *X. axonopodis* pv. *citrumelo* and other known close relatives of *X. axonopodis* (*X. citri* and *X. fuscans*) as well as *X. campestris* pv. *vesicatoria*. All included species and pathovars could successfully be differentiated by *fyuA* gene analysis except for *X. citri* pv. *citri* and *X. axonopodis* which formed a cluster with bootstrap accuracy of 100 %. *X. translucens* pv. *undulosa* and *X. albilineans* formed their own early-branching group with notable divergence from the other species.

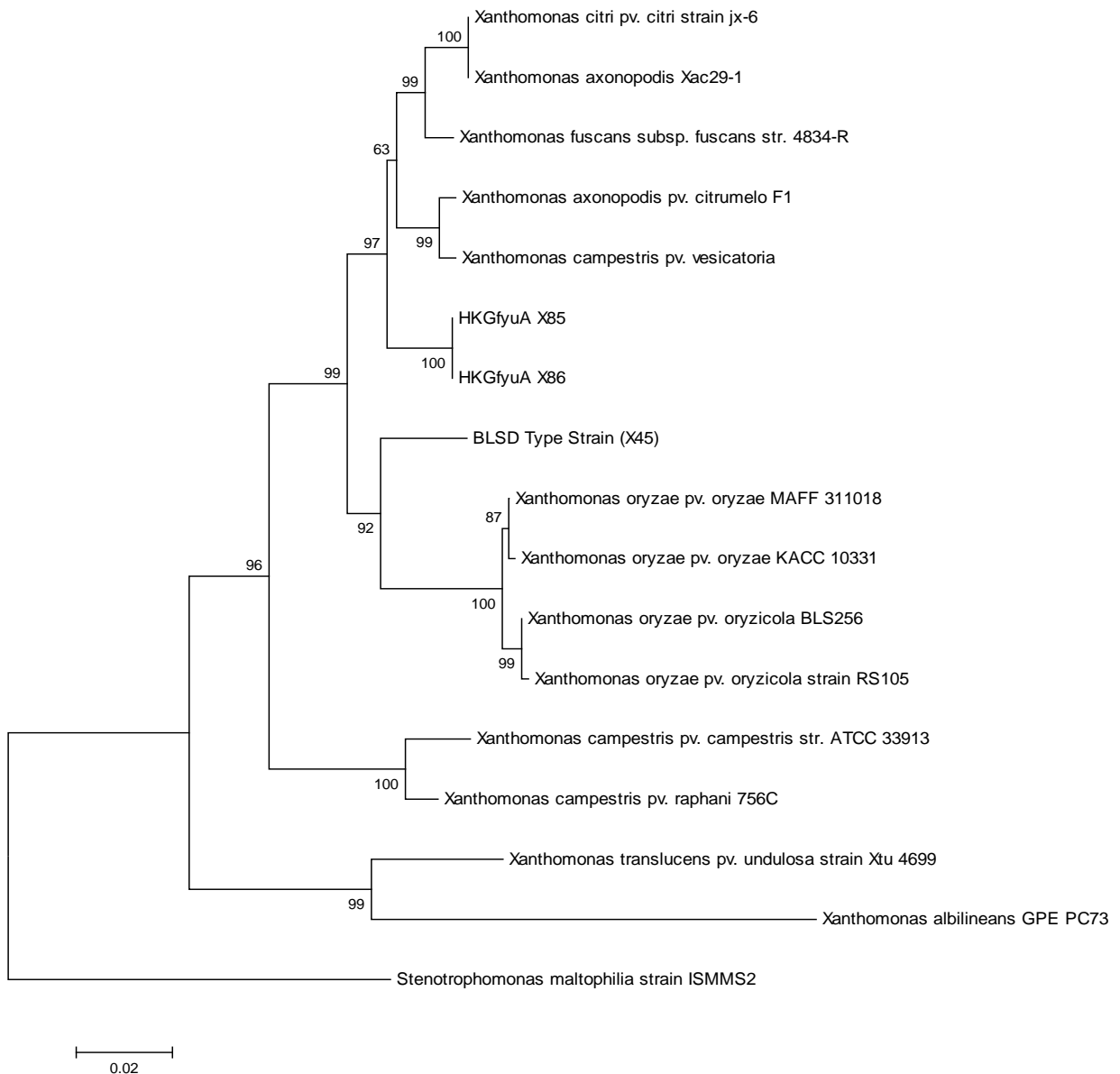


Figure 5: Neighbour-joining tree based on *fyuA* gene sequences indicating relationships between the BLSD strain (X45) as well as reference strains of *X. axonopodis* pv. *axonopodis* (HKG fuyA:X85), *X. campestris* pv. *campestris* (HKG fyuA:X86) and other *Xanthomonas* species.

4.6.4. *recA*

The length of sequences included in the alignment for the *recA* gene was 980 bp. Phylogenetic relationships of different *Xanthomonas* species and the position of the BLSD causal agent among other species based on *recA* gene sequences are illustrated in Figure 6. Gene sequences of reference samples *X. axonopodis* pv. *axonopodis* (HKG *recA*:X85)

and *X. campestris* pv. *campestris* (HKG recA:X86) were included during phylogenetic tree construction as references to monitor accuracy of clustering of samples with specific species.

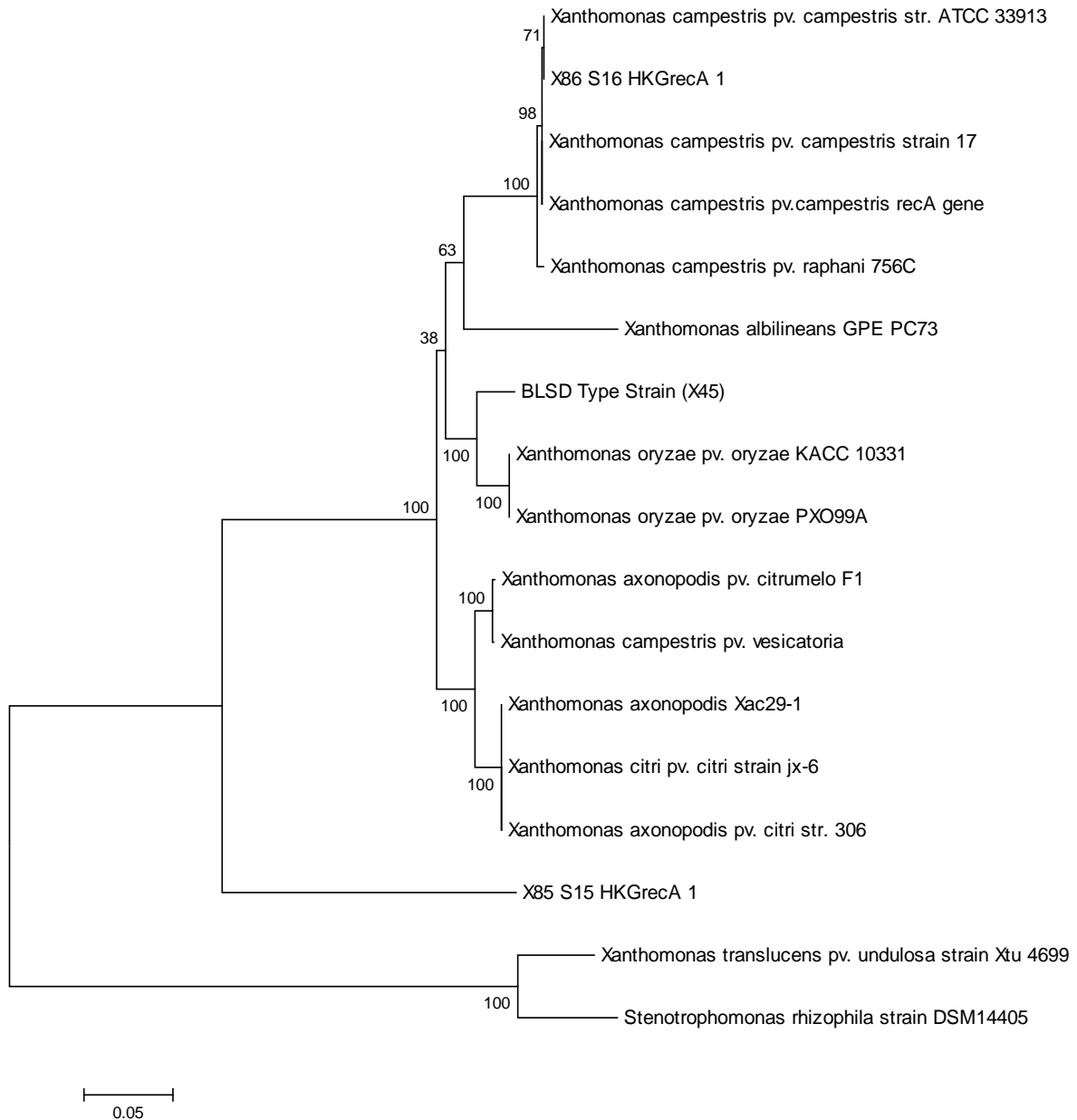


Figure 6: Neighbour-joining tree of the BLSD strain (X45) as well as reference strains of *X. axonopodis* pv. *axonopodis* (HKG recA:X85) and *X. campestris* pv. *campestris* (HKG recA:X86) in the *Xanthomonas* genus using *recA* gene sequences.

The representative gene sequence for the *recA* gene of BLSD isolates once again formed a cluster with *X. oryzae* pv. *oryzae* pathovars with a 94 % bootstrap value. The sequence of *X. axonopodis*, X85, did not group with any of the reference sequences included in Figure 6 and formed its own cluster. A possible explanation for this is that some of the nucleotide data

of the gene may have been absent due to the gaps that occurred within the whole genome sequences. The reference strain of *X. campestris* pv. *campestris*, X86, clustered with *X. campestris* pv. *campestris* ATCC 33913, confirming high sequence similarity as could be expected. All representatives revealed distinct clustering, allowing for discrimination between different species based on the *recA* gene. Different pathovars from the same species could be differentiated as can be seen in the grouping of different *X. campestris* pathovars.

Different strains from the same pathovar could not be successfully differentiated from each other as *X. oryzae* pv. *oryzae* strains grouped together with no nucleotide variation between them and with 100 % bootstrap confidence. *Stenotrophomonas rhizophila* formed the root of the tree but were accompanied by *X. translucens* pv. *undulosa* which appeared to be very distinct from other *Xanthomonas* species based on the *recA* gene.

4.6.5. *rpfB*

All nucleotide sequences included in phylogenetic analysis of the *rpfB* gene (Figure 4.7) were trimmed to 1 686 bp before constructing the tree. Based on sequences included the *rpfB* gene sequence of the BLSD isolates clustered with *X.citri* subsp. *citri* and the reference sequence of *X. axonopodis* (X85). *X. oryzae* pv. *oryzae* also had a similar nucleotide composition to the BLSD isolates. *X. campestris* pathovars, including reference sample (X86) clustered in a single separate group with a 98 % bootstrap value. The *X. campestris* species was distantly placed from the BLSD causal agent, demonstrating no extremely close relation between the BLSD agent and the *X. campestris* species. *X. sacchari* and *X. translucens* were the most divergent species based on *rpfB* gene analysis as can be observed in Figure 7.

4.6.6. *rpoB*

Partial *rpoB* sequences used for construction of a neighbour-joining tree (Figure 8) were trimmed to 615 bp to ensure consistent sequence length for all included species. The *rpoB* gene sequences of the BLSD isolates clustered together with *X. hortorum* pv. *hederae* with 98 % bootstrap confidence. *X. axonopodis* (X85) showed closest similarity to *X. dyei* pv. *dysoxylis* and *X. axonopodis* Xac 29-1. *X. campestris* pv. *campestris* (X86) were placed with *X. campestris* pv. *campestris* ATCC 33913 with 99 % bootstrap support. *X. albilineans*, *X.theicola*, *X. sacchari*, *X. hyacinthi* and *X. translucens* species together formed a distinct early-branching cluster.

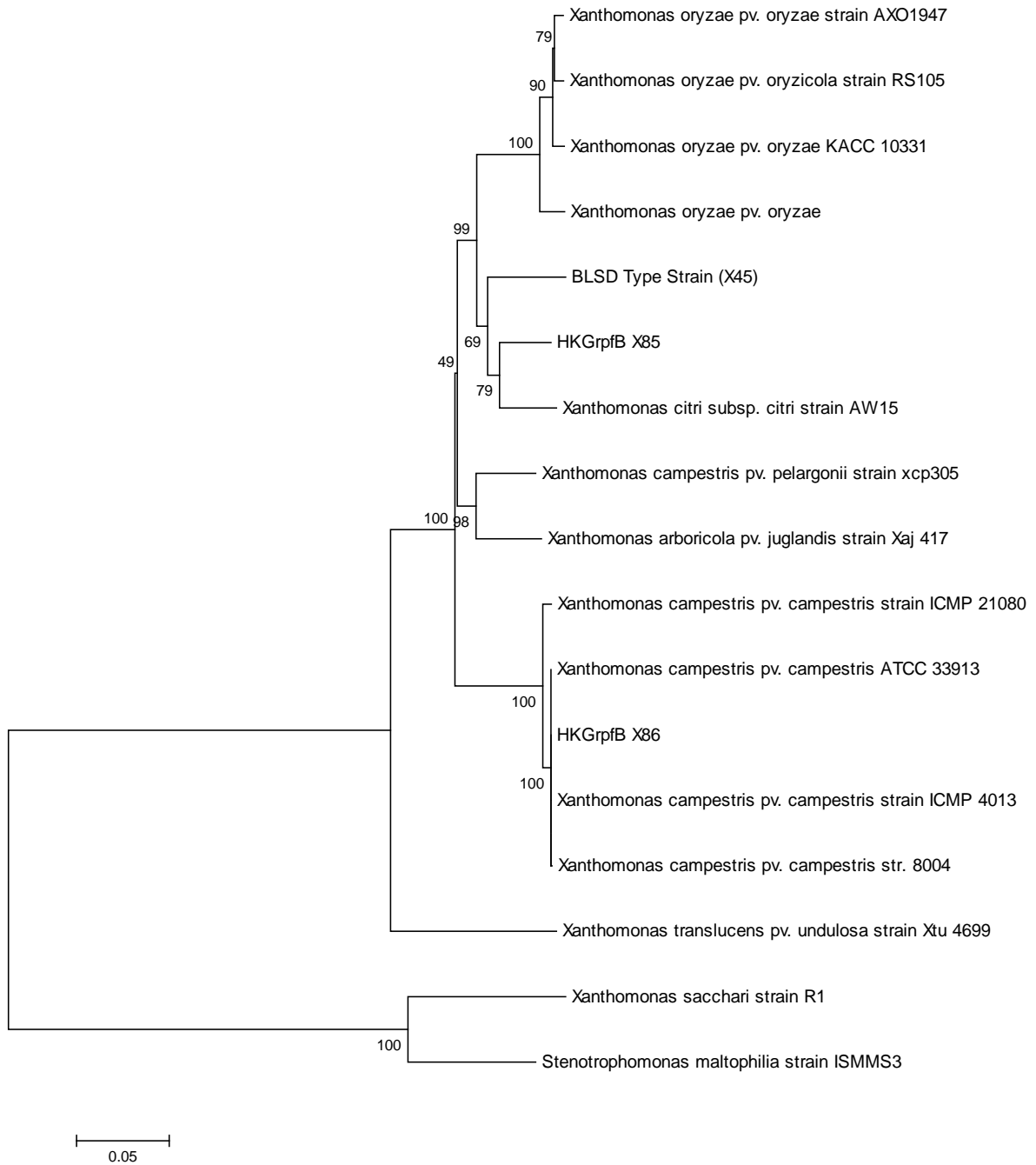


Figure 7: Neighbour-joining analysis of *rpfB* gene sequences and the position of the BLSD strain (X45) as well as reference strains of *X. axonopodis* pv. *axonopodis* (HKG rpfB:X85) and *X. campestris* pv. *campestris* (HKG rpfB:X86) among reference sequences.

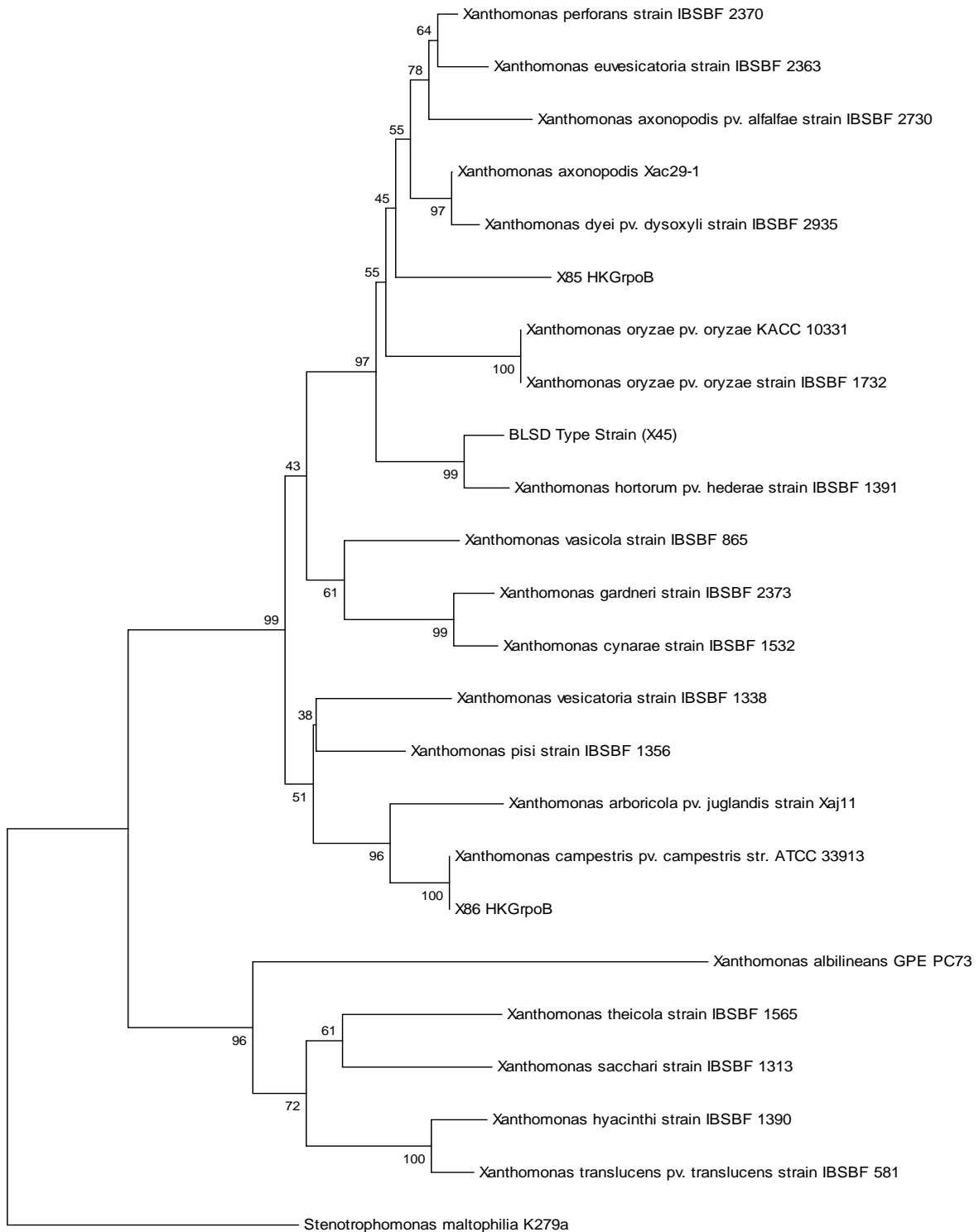


Figure 8: Neighbour-joining tree based on *rpoB* gene sequences indicating relationships between the BLSD strain (X45) as well as reference strains of *X. axonopodis* pv. *axonopodis* (HKG *rpoB*:X85) and *X. campestris* pv. *campestris* (HKG *rpoB*:X86) and other *Xanthomonas* species.

4.7. Primer design and specificity

Sequences of the variable regions of the housekeeping genes, *recA*, *dnaK* and *rpoB*, were used to design PCR primers. Alignment of the gene sequences of the BLSD isolates against the sequences of other *Xanthomonas* species revealed variable regions containing polymorphic sites. The presence of variable regions allowed for designing and generating potentially diagnostic genomic-based markers. Two primer pairs were designed for each gene, targeting different variable regions within the genes to determine if variability was sufficient for diagnostic purposes. Sequences for the selected primer pairs are provided in Table 5.

Table 5: Primer pairs designed for amplification of DNA markers potentially unique to the BLSD pathogen with their optimal annealing temperatures.

Gene	Primer name	Amplicon length (bp)	Primer sequence [5' > 3']	Annealing temperature (°C)
<i>dnaK</i>	Xan_dnaK-F	1 703	GTGACCAACCCGAAGAACACC	58
	Xan_dnaK-R		GGCGTCCACCACGTCATCG	
	XanZ_dnaK-F	895	GTCGAGCAGCCAGCAGACCG	56
	XanZ_dnaK-R		TCCAGATCCGACAATGCC	
<i>recA</i>	Xan_recA-F	544	GAYGAHYTGCTGCTGTTCGC	58
	Xan_recA-R		GTCCTTGCCCTGVCCGATGC	
	XanZ_recA-F	594	CAGTGTCAGAAGCTGGGC	56.5
	XanZ_recA-R		GCCGAACCGCAACCTGTGGG	
<i>rpoB</i>	Xan_rpoB-F	959	CGCCCGCATCATTCCGTACCG	67
	Xan_rpoB-R		CCTTGACCGCRCGCTCGACGC	
	XanZ_rpoB-F	703	TTCGCCAGGTGTGTTCTTCG	56.5
	XanZ_rpoB-R		GGAACAGGTTCTGTGCGGC	

Two regions within the *dnaK* gene sequence were targeted by primer pairs Xan_dnaK and XanZ_dnaK. Primer pair Xan_dnaK targeted a sequence region from nucleotide position 175 to nucleotide position 1 878 within the complete gene sequence. Primer pair XanZ_dnaK targeted a region from nucleotide position 826 to nucleotide position 1 721 in the *dnaK* gene.

For the *recA* gene, Xan_recA was designed to target a variable region from nucleotide position 349 to nucleotide position 893. XanZ_recA aimed to amplify a region from nucleotide position 241 to nucleotide position 835 in the *recA* gene sequence. Xan_rpoB was designed for amplification of a variable region from nucleotide position 546 to 1 505 nucleotide position in the *rpoB* gene. Primer pair XanZ_rpoB targeted a variable region from base pair position 468 to 1 171 in the complete *rpoB* gene sequence.

Amplicon lengths of the gene regions designed in this study varied from 544 bp to 1 703 bp. Forward and reverse primers were designed to amplify each of the chosen regions. Primers were between 18 and 21 nucleotides in length. Optimal annealing temperatures for primer pairs as determined through PCR optimization are listed in Table 5. Annealing temperatures varied between 56°C to 67°C. All six developed primer pairs were tested for their specificity to the BLSD isolates with a PCR assay to determine their potential as diagnostic genetic markers. Genomic DNA of BLSD isolates were used for PCR as well as DNA of two reference strains of close relatives, *X. axonopodis* (X85) and *X. campestris* pv. *campestris* (X86). A no-template control (NT) containing no DNA was included in each case to monitor contamination.

PCR amplification using primer pairs Xan_recA and XanZ_dnaK amplified all targeted gene fragments in all samples including the two reference strains. PCR amplification with the Xan_recA primer pair revealed a single band present at approximately 544 bp amplified for all the samples except in the no-template control lane. The XanZ_dnaK primer pair amplified a single band with approximate size of 895 bp in all samples. No product was amplified in the lane containing the no-template, confirming that no contamination was present. The primer pairs Xan_recA and XanZ_dnaK are thus not suitable as diagnostic genetic markers to detect the BLSD causal agent as it does not differentiate between BLSD isolates and related *Xanthomonas* strains.

Primer pairs Xan_rpoB and Xan_dnaK amplified gene fragments in all BLSD isolates as well as the *X. axonopodis* reference strain but not for the *X. campestris* pv. *campestris* strain. Figure 9 illustrates amplification of PCR products targeted by the primer set Xan_dnaK. The Xan_dnaK primer pair amplified gene fragments that were approximately 1 500 bp as can be observed in Figure 9.

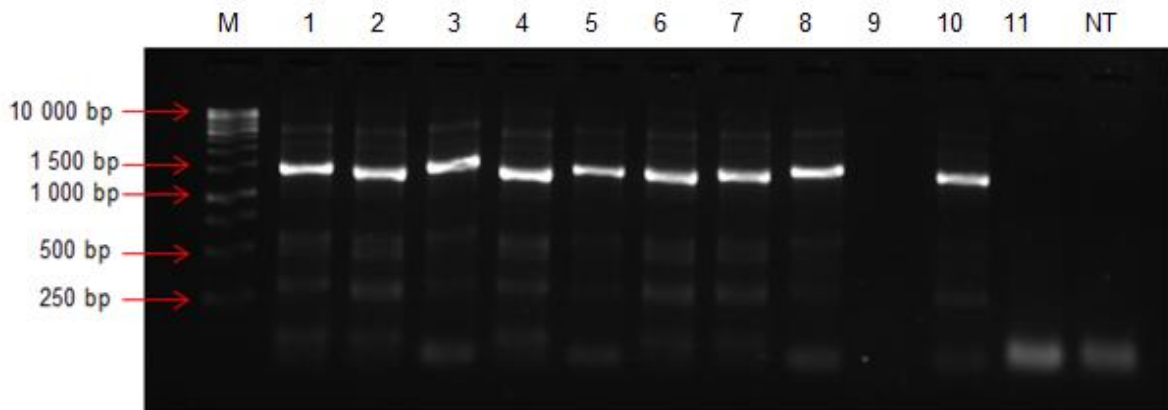


Figure 9: PCR amplicons of *dnaK* gene fragments as targeted by the XanX_*dnaK* primer set in a 1.5 % agarose gel with a 1 kb ladder (M) to indicate band size. PCR products of samples loaded in the gel in Figure 9 were as follows: lane 1 = X1, lane 2 = X2, lane 3 = X9, lane 4 = X22, lane 5 = X33, lane 6 = X, lane 7 = 35, lane 8 = X45, lane 9 = empty, lane 10 = X85 and lane 11 = X86.

Similar results were obtained during testing of the Xan_*rpoB* primer set. Gene fragments amplified for the BLSD samples were approximately 960 bp in size for primer pair Xan_*rpoB*. However, an amplified fragment of the same size was also visible in the *X. axonopodis* reference sample. Although not specific to BLSD isolates alone, primer pairs Xan_*rpoB* and Xan_*dnaK* could still be useful in differentiating BLSD samples from *X. campestris* pv. *campestris*. The XanZ_*recA* primer pair allowed amplification of gene fragments in all BLSD isolates, but also in the *X. campestris* pv. *campestris* (X86) reference sample as presented in Figure 10.

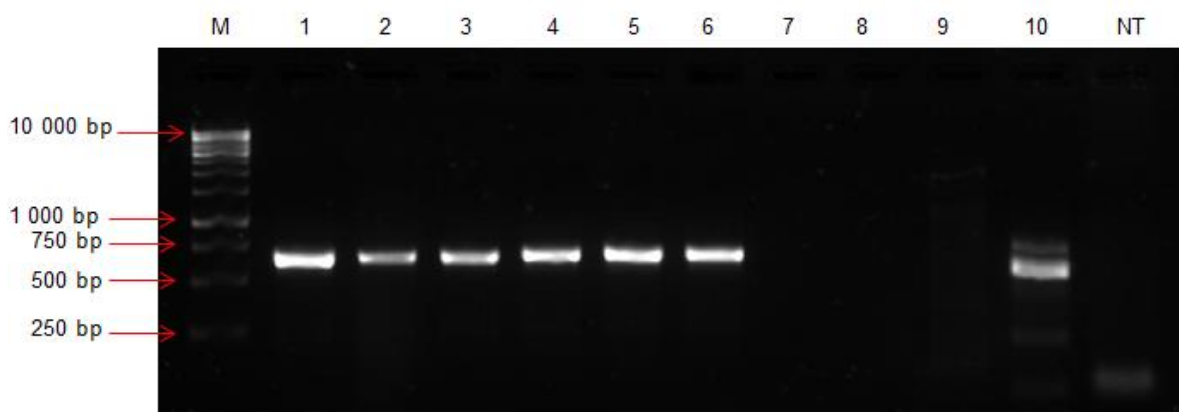


Figure 10: Amplified DNA fragments of partial *recA* gene sequences as targeted with PCR by the XanZ_*recA* primer set, in a 1.5 % agarose gel with a 1 kb ladder (M) to indicate amplicon sizes. PCR products of samples loaded in the gel were as follows: lane 1 = X1,

lane 2 = X2, lane 3 = X9, lane 4 = X22, lane 5 = X33, lane 6 = X, lane 7 = empty, lane 8 = empty, lane 9 = X85 and lane 10 = X86.

A gene fragment with approximate size of 600 bp was amplified in all BLSD samples and in *X. campestris* pv. *campestris* (lane 10; X86) but did not amplify *X. axonopodis* DNA (lane 9; X85). The primer pairs thus failed to amplify regions in BLSD isolates alone with undesired amplification of a similar fragment size in sample X86. The primer set XanZ_recA was not specific to BLSD strains, however, the primer was able to distinguish the BLSD isolates from *X. axonopodis*. PCR testing of the primer set XanZ_rpoB revealed selective amplification of gene fragments of the expected size in BLSD samples exclusively as shown in Figure 11.

Primer pair XanZ_rpoB succeeded in amplifying the desired gene fragment in all BLSD samples exclusively without any amplification of gene fragments for both reference strains. Bands were detected at the desired basepair size of approximately 700 bp. The gene product was not amplified in both of the reference samples *X. axonopodis* (lane 12; X85) and *X. campestris* pv. *campestris* (lane 13; X86). Primer pair XanZ_rpoB thus appears to be specific to BLSD isolates and may therefore be considered a potential molecular marker for the BLSD causal agent.

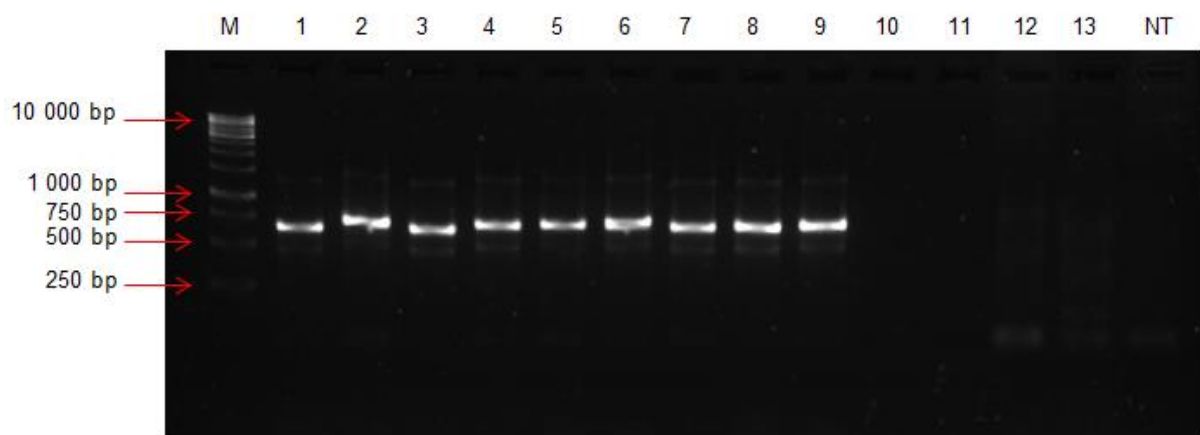


Figure 11: Amplified PCR products of genomic DNA targeted by XanZ_rpoB primers for amplification of rpoB gene fragments in a 1.5 % agarose gel with 1 kb ladder (M) as reference for amplicon sizes. PCR products of samples loaded in the gel were as follows: lane 1 = X1, lane 2 = X2, lane 3 = X9, lane 4 = X22, lane 5 = X33, lane 6 = X35, lane 7 = X45, lane 8 = XGP, lane 9 = X26, lane 10 = empty, lane 11 = empty, lane 12 = X85 and lane 13 = X86.

CHAPTER 5: DISCUSSION

5.1. Isolate cultivation and Gram-staining

Colonies from the isolates cultivated and streaked onto GYCA plates all yielded identical *Xanthomonas*-like colonies as described by Qhobela & Claflin (1988). Gram-staining results indicated that cells were Gram-negative and rod-shaped corresponding with the findings of previous phenotypical and morphological studies done on *Xanthomonas* and the BLSD causal agent (Coutinho & Wallis, 1991; Goszczynska *et al.*, 2000; Nienaber, 2015). All the isolates thus appeared to have the same colony characteristics as well as morphology and could be positively confirmed as pure *Xanthomonas* cultures based on these findings.

5.2. Whole genome sequencing

The nucleotide composition of the genome of the *Xanthomonas* BLSD causal agent was determined with whole genome sequencing and subsequent sequence assembly. Sequenced reads for the newly sequenced genome were mapped to the reference genomes of *X. oryzae* pv. *oryzae* PXO99A and *X. campestris* pv. *campestris* ATCC 33913. *X. oryzae* pv. *oryzae* PXO99A possesses a single circular chromosome with a length of 5 240 075 bp (Salzberg *et al.*, 2008). The complete genome of *X. campestris* pv. *campestris* ATCC 33913 constitutes 5 076 188 bp within a single circular chromosome (Da Silva *et al.*, 2002). The size of the assembled genome sequence of the BLSD causal agent was approximately 5 Mbp. This size was similar to other *Xanthomonas* species as described by Pieretti *et al.* (2009). However, the large gaps within the complete assembled genome sequence would reduce the total number of present nucleotides, which possibly indicates a reduced genome of the BLSD pathogen.

Although the gaps observed within the BLSD genome might be due to incomplete genome sequencing and sequencing complications, reduced genomes have been reported for specific *Xanthomonas* species. Data confirming that *X. albilineans* presents a reduced genome size have been published by Pieretti *et al.* (2009) and Rodriguez-R *et al.* (2012). The species has a genomic size of 3.77 Mbp as opposed to approximate sizes of 5 Mbp of other species (Pieretti *et al.*, 2009). Pieretti *et al.* (2009) have identified 131 ancestral genes that have potentially been lost in *X. albilineans* due to short deletions or pseudogenization. Rodriguez-R *et al.* (2012) have further discovered that *X. vasicola* also revealed notable

genomic gains and losses. In *X. vasicola*, genomic gains compensated for genomic losses that were twice as much as genomic losses observed in *X. albilineans*.

It is suggested that *X. vasicola* genomes have the ability to maintain a genome size similar to other *Xanthomonas* species but are highly dynamic with significant levels of genomic gains and losses. The genome size of *X. vasicola* pv. *vasculorum* NCPPB 702 is 5.4 Mbp which is an expected size for a *Xanthomonas* genome (Pieretti *et al.*, 2009; Studholme *et al.*, 2010). Genome reduction in *Xanthomonas* is not restricted to the divergent species, *X. albilineans*, but may also affect other genus members such as *X. vasicola*. It is thus possible that some degree of reductive genome evolution may exist in the genome of the BLSD pathogen.

5.2. Comparative housekeeping gene analysis

Sequences of the six housekeeping genes were successfully extracted from the whole genome sequencing data for the BLSD isolates. The genes were all present within the genomic composition of the BLSD causal agent and had high nucleotide similarity to complete gene sequences of other *Xanthomonas* species (Young *et al.*, 2008). These findings confirmed relatedness of the BLSD pathogen to the genus. All genes included (*atpD*, *dnaK*, *fyuA*, *recA*, *rpfB* and *rpoB*) are found in all *Xanthomonas* members and serve important cellular functions (Simões *et al.*, 2007; Young *et al.*, 2008; Pieretti *et al.*, 2009; Vandroemme *et al.*, 2013).

Complete gene sequences of *atpD*, *dnaK*, *fyuA*, *recA*, *rpfB* and *rpoB* genes of the BLSD samples revealed highly consistent sequences among isolates for all genes as described in Chapter 4.5. Little to no variation between nucleotide sequence alignments of different BLSD isolate samples was present. These findings were similar to the results found in a study by Niemann (2015) where no nucleotide sequence variations and 100 % alignment similarity were found among BLSD isolates with analyses using the 16S, 23S rDNA and *gryaseB* genes, as well as the *XgumD* genetic marker.

X. albilineans (GPE PC73) was the least similar to the BLSD isolates in terms of nucleotide sequences for the complete housekeeping genes (Table 4). *X. albilineans* is known to be one of the most distinct species in the *Xanthomonas* genus (Moore *et al.*, 1997; Simões *et al.*, 2007). Genome reduction in *X. albilineans* has previously been identified in comparison to other *Xanthomonas* species (Rodriguez-R *et al.*, 2012). From previous studies based on MLSA (Pieretti *et al.*, 2007) and draft genome sequence analysis (Rodriguez-R *et al.*, 2012)

it has been established that the relationship of *X. albilineans* with other *Xanthomonas* species is unclear and that the species may rather be part of the closely related genus *Xylella fastidiosa*. This may serve as explanation for the significantly lower genetic similarity of the complete gene sequences of *X. albilineans* with other species including the BLSD pathogen.

The *Xanthomonas* genus consists of the *Hrp* species group as well as species not possessing the *Hrp* gene cluster (Pieretti *et al.*, 2009). *X. albilineans* is known to have a reduced genome with the absence of the *Hrp* genes. Main representatives of the *Hrp Xanthomonas* group are *X. axonopodis*, *X. oryzae* and *X. campestris*. All analysed housekeeping genes of the BLSD causal agent revealed significantly higher gene similarity to members of the *Hrp Xanthomonas* group.

5.3. Phylogenetic analysis

Genetic characterisation of species belonging to the *Xanthomonas* genus remains a challenge (Simões *et al.*, 2007). In a study by Hauben *et al.* (1997) the 16S rDNA gene of *Xanthomonas* species was used and allowed for discrimination of only three different phylogenetic groups for the genus. The 16S rDNA gene is particularly conserved and is known to exhibit a high degree of similarity among species which is problematic in species distinction. Gonçalves & Rosato (2002) employed the 16S-23S rDNA gene spacer region for phylogenetic analysis of *Xanthomonas* which resulted in distinguishing six phylogenetic groups. However, each cluster still consisted of several species that could not be separated. Analysis of additional or alternative genetic markers continues to be necessary to resolve genetic relatedness within the *Xanthomonas* genus.

Niemann (2015) have determined the phylogenetic position of the BLSD pathogen based on 16S rDNA, 23S rDNA and *gyraseB* sequences. Based on 16S rDNA data the BLSD causal agent clustered with *X. vasicola* pv. *vasculorum*, *X. axonopodis* pv. *vasculorum*, *X. hortorum* and *X. arboricola* pv. *juglandis*. Phylogenetic analysis of the 23S rDNA gene revealed clustering of the BLSD pathogen with *X. axonopodis* pv. *vasculorum* and *X. oryzae* pv. *oryzicola*. The BLSD causal agent clustered with *X. axonopodis* pv. *vasculorum* and *X. vasicola* based on phylogenetic data of *gyraseB* gene sequences. From the three analysed genes, *gyraseB* exhibited greatest ability to discriminate between the BLSD pathogen and other *Xanthomonas* members. BLSD isolates thus repeatedly clustered with *X. axonopodis* pv. *vasculorum* and *X. vasicola*. Phylogenetic analyses from the study by Niemann (2015) have indicated that the BLSD causal agent is not closely related to *X. campestris*.

The aim of phylogenetic analyses done in this work was to gain more insight and higher resolution into the taxonomic position of BLSD isolates within the genus. Six housekeeping genes commonly found in all *Xanthomonas* species were analysed and the efficiency of each gene to differentiate BLSD strains from close relatives were verified.

5.3.1. *atpD*

The BLSD causal agent was phylogenetically mostly closely related to *X. vasicola* based on *atpD* gene sequences. The BLSD strain was grouped together with *X. vasicola* with bootstrap value of 100% indicating high reliability of the close relationship and accurate clustering. Bootstrap scores higher than 90 % at branch points confirm high accuracy of the tree (Patwardhan *et al.*, 2014). *X. vasicola* pathovars are known to be pathogenic to sugarcane, banana and maize plants (Carter *et al.*, 2010; Karamura *et al.*, 2015). According to Qhobela *et al.* (1990) *Xanthomonas* species, other than the BLSD causal agent, with the abilities to infect maize are *X. campestris* pv. *vasculorum* and *X. campestris* pv. *holcicola*. *X. campestris* pv. *vasculorum* has since been reclassified as *X. vasicola* pv. *vasculorum* due to a lack of genetic similarity to the *X. campestris* species (Vauterin *et al.*, 1995; Karamura *et al.*, 2015).

As *X. vasicola* pathovars are known to be pathogenic to monocot hosts such as maize (Qhobela *et al.*, 1990; Karamura *et al.*, 2015) this might explain the extremely close grouping of the BLSD isolates with *X. vasicola*. The high *atpD* gene similarity between the two organisms may be due to their shared capability to infect the same plant host. Gardiner *et al.* (2014) have suggested that *Xanthomonas* species infecting similar host plants can be expected to possess similar pathogenicity mechanisms and genetic compositions in order to infect these plants.

Simões *et al.* (2007) have determined that the *atpD* gene was an applicable marker to discriminate between certain *Xanthomonas* species by dividing the species into five clusters. However, diversity among some species that were within the same clusters could not be distinguished. *X. sacchari* and *X. albilineans* formed a cluster most distinct from the others. *X. campestris*, *X. hortorum* and *X. vasicola* could not be successfully distinguished from one another (Simões *et al.*, 2007). In the present study all *X. campestris* species could be clearly differentiated from *X. vasicola* based on their *atpD* gene sequences (Figure 3).

Analysis of the *atpD* gene sequences of *Xanthomonas* species done by Gardiner *et al.* (2014) revealed two main phylogenetic groups. The first group comprised of early-branching species including *X. albilineans*, *X. sacchari* and *X. translucens*. The second group consisted

of later-branching, less divergent species including *X. campestris*, *X. axonopodis* and *X. oryzae*. Similar groups could be distinguished in Figure 3 with *X. translucens* and *X. albilineans* typically more divergent with longer branch lengths. The BLSD strain was placed among species of the second group illustrating its relatedness to the second, less divergent, group of *Xanthomonas* species.

Phylogenetic analysis of the *atpD* gene in this work has demonstrated close relatedness of the BLSD causal agent to *X. vasicola*. A possible explanation for this occurrence might be connected to the two pathogens shared ability to infect maize plant hosts. All species included in the analysis could be well differentiated and *atpD* tree topology agreed with previous findings of Simões *et al.* (2007) and Gardiner *et al.* (2014).

5.3.2. *dnaK*

Molecular phylogenetics based on the *dnaK* gene has significantly improved resolution below species level in previous studies (Stepkowski *et al.*, 2003; Young *et al.*, 2008). Stepkowski *et al.* (2003) have found that phylogeny of the *dnaK* gene revealed higher resolution of internal phylogenetic branches as well as higher bootstrap confidence between relationships than the use of the 16S rDNA gene. Results found by Young *et al.* (2008) indicated that *dnaK* analysis corresponded well with 16S rDNA data and clearly differentiates species previously established by DNA-DNA hybridization. Furthermore, the study showed that *X. euvesicatoria* and *X. perforans* are likely the same organism as *dnaK* analysis could not differentiate between the two species.

In the present study the BLSD isolates clustered with *X. vasicola* pv. *vasculorum* and a representative *X. vasicola* species strain based on phylogenetic analysis of the *dnaK* gene. This demonstrated very high *dnaK* gene similarity between *X. vasicola* and the BLSD causal agent. The gene similarity and close clustering of the BLSD pathogen and *X. vasicola* was also observed when phylogenetics was based on *atpD* gene sequences. Similarity of the BLSD causal agent and *X. vasicola* could possibly be due to the ability of both pathogens to infect maize as described by Qhobela *et al.* (1990).

General clustering of species in Figure 4 based on *dnaK* gene sequences were similar to results from a neighbour-joining tree generated by Young *et al.* (2008). The *X. axonopodis* clade consisted of closely related species including *X. perforans* and *X. citri*. *X. oryzae* clustered with a group close to the *X. axonopodis* clade. In phylogenetic analyses *X. axonopodis* generally comprises *X. axonopodis* pathovars as well as *X. citri* and *X. fuscans*

(Young *et al.*, 2008; Rodriguez-R *et al.*, 2012). Other major species such as *X. campestris*, *X. oryzae*, *X. vasicola*, *X. translucens*, *X. sacchari* and *X. albilineans* all formed their own clades. Clustering of *Xanthomonas* species in the present study was consequently found to be similar than in previous studies (Young *et al.*, 2008; Rodriguez-R *et al.*, 2012), supporting phylogenetic analysis findings based on the *dnaK* gene as determined in the present study.

Phylogenetic analysis of the *dnaK* gene done in the present study has shown high similarity between the BLSD causal agent and *X. vasicola*. Similar results were observed during analysis of the *atpD* gene. General tree topology of *dnaK* phylogeny in Figure 4 was similar to *Xanthomonas dnaK* phylogeny previously constructed (Young *et al.*, 2008; Rodriguez-R *et al.*, 2012). All included strains could be differentiated with different species clustering in distinct clades.

5.3.3. *fyuA*

The BLSD pathogen grouped alone based on *fyuA* gene sequence analysis but was phylogenetically closest to *X. oryzae* species. *X. oryzae* pathovars infect rice plants exclusively. Both the BLSD pathogen and the *X. oryzae* species are pathogenic to monocot host plants as maize and rice are both monocotyledonous (Du Plessis, 2003; Lee *et al.*, 2005). Genetic similarities among *Xanthomonas* species that infect monocotyledonous hosts could be linked to their pathogenicity systems (Parkinson *et al.*, 2007; Gardiner *et al.*, 2014). Distinct gene fragments among these species can be associated with their unique proteins and gene clusters encoding specific secretion systems for virulence to their plant hosts (Lee *et al.*, 2005; Gardiner *et al.*, 2014).

Parkinson *et al.* (2007) have found that *X. vasicola* (pathogenic to maize) and *X. oryzae* grouped together forming a separate phylogenetic group based on *gyraseB* sequence analysis. In a study on the BLSD pathogen by Niemann (2015) using the *gyraseB* gene for phylogenetic inference, the BLSD causal agent clustered with *X. vasicola*. Consequently, from previous studies a close relationship between *X. oryzae* and *X. vasicola* has been established. Furthermore, data from Niemann (2015) and the present study suggest that the BLSD pathogen could be closely associated with both these species. The majority of other *Xanthomonas* species infect dicotyledonous plant hosts (Gardiner *et al.*, 2014).

The overall topology of the neighbour-joining tree in Figure 5 was similar to phylogenetic relationships among *Xanthomonas* species based on the *fyuA* gene constructed by Vandroemme *et al.* (2013). *X. axonopodis*, *X. fuscans* and *X. citri* grouped close together

and share high genetic similarity. In the present study the use of the *fyuA* gene for phylogenetic purposes was not entirely discriminative. The clustering of reference samples *X. axonopodis* (X85) and *X. campestris* pv. *campestris* (X86) with a 100 % bootstrap value was incongruent with expected results as these two organisms were from different species.

Phylogenetic analysis of the *fyuA* gene indicated that *X. oryzae* is a close relative of the BLSD pathogen. A close relationship between the BLSD causal agent and *X. oryzae* has also been found by Niemann (2015) through analysis of the 23S rDNA gene. The fact that both the BLSD pathogen and *X. oryzae* are capable of infecting monocotyledonous host plants could be a possible explanation for their high gene similarity. The structure of the *fyuA* tree constructed in this work is similar to previous findings but was not able to discriminate between all included strains.

5.3.4. *recA*

Based on *recA* gene sequences the BLSD causal agent exhibited highest similarity to *X. oryzae* strains. This further supports the hypothesis that the high similarity between BLSD isolates and *X. oryzae* may be due to their shared pathogenicity towards monocotyledonous plant hosts. Pieretti *et al.* (2009) included *recA* gene sequences of *Xanthomonas* species in a study to determine the phylogenetic relationship of *X. albilineans* strains within the genus. Similar taxonomic clustering was observed in Figure 6 was found revealing separate groupings of *X. campestris*, *X. oryzae* and *X. axonopodis*. *X. albilineans* possessed a notably longer branch length indicating a larger amount of nucleotide substitutions per site. This demonstrates significant genetic divergence of *X. albilineans* from the other species corresponding to the findings of Pieretti *et al.* (2009).

In the present study, phylogenetic analysis of the *recA* gene revealed high similarity between the BLSD pathogen and the *X. oryzae* species. This close relation was also found with analysis of the *fyuA* gene. All included species formed distinct clusters and could be differentiated from each other. Tree topology for the *recA* gene reflected a similar structure to *recA* phylogeny previously constructed by Pieretti *et al.* (2009).

5.3.5. *rpfB*

Simões *et al.* (2007) have established that the *rpfB* gene is conserved among *Xanthomonas* species with enough variation to serve as a potential marker gene during phylogenetic analyses. The *rpfB* gene is part of the *rpf* gene cluster responsible for regulating

pathogenicity and is unique to *Xanthomonas*, *Stenotrophomonas maltophilia* and *Xylella fastidiosa* (Deng *et al.*, 2011). Data from the present study showed separate grouping of the BLSD causal agent at an individual branch but it was placed closest to the *X. axonopodis* reference sample (X85) and *X. citri* subspecies *citri* which is closely associated with the *X. axonopodis* group. *X. axonopodis* pathovars have a broad host range and infect soybean, cassava, cotton and citrus which are all dicotyledonous plants (Arrieta-Ortiz *et al.*, 2013; Mielewczik *et al.*, 2013; Lee *et al.*, 2014). *X. citri* is known to be an extremely close relative to *X. axonopodis* (Jalan *et al.*, 2013). As the BLSD causal agent is pathogenic to a monocot host plant its close relationship to *X. axonopodis* and *X. citri* could not be explained.

Gardiner *et al.* (2014) have described that *Xanthomonas* strains that are closely related to specific strains may also exhibit genetic similarity to more distantly related strains due to the presence of specific secretion systems shared by the distantly related strains. In the study by Gardiner *et al.* (2014) *X. translucens* strain DAR61454 exhibited closest relation to *X. translucens* strains based on individual gene analysis. However, the DAR61454 strain also encoded a secretion system that is associated with *X. oryzae* and not found in any other species. These findings demonstrated that although a strain may be closest related to a specific species, the strain may also share genetic characteristics with more distantly related members based on specific genes.

A combination of phylogenetic data including *rpfB* gene analysis done by Simões *et al.* (2007) allowed clustering of the following species: *X. campestris*, *X. vasicola*, *X. vesicatoria*, *X. arboricola* and *X. hortorum* were grouped within the same cluster. *X. translucens* did not cluster with any other species while *X. cassavae* and *X. oryzae* shared a cluster. *X. axonopodis* grouped alone and *X. albilineans* exhibited individual clustering, the latter comprising the most distinct cluster. Overall, the genetic variation among most of the species could be explored and diversity well distinguished. According to Simões *et al.* (2007) it is plausible that the *rpfB* gene tolerates a higher mutation rate than other housekeeping genes with more essential functions. Similar topology results as Simões *et al.* (2007) were found in the present study for phylogenetic groupings of *Xanthomonas* species based on *rpfB* gene sequences (Figure 7) considering a reduced amount of sequences used for this study.

X. oryzae pathovars all grouped together in the same way in the present study as in the study by Simões *et al.* (2007). All sequences of the *X. campestris* species grouped together with high bootstrap confidence and similar positions within neighbour-joining trees. *X. citri* subspecies *citri* occupied a similar phylogenetic position as *X. axonopodis* did in the study by Simões *et al.* (2007). These two species are known to be extremely closely related and are known to share a clade (Young *et al.*, 2008; Rodriguez-R *et al.*, 2012). *X. arboricola* is

placed within the same cluster as *X. campestris* species (Figure 7) which corresponds to similar findings of Simões *et al.* (2007). *X. translusens* clustered separately (Figure 7) as was also the case in the study by Simões *et al.* (2007).

Phylogenetic analysis of the *rpfB* gene in the present study indicated closest relatedness of the BLSD agent to the *X. axonopodis* clade. The high gene similarity could not be directly associated with similar plant hosts as the two pathogens infect completely different types of plants. *Xanthomonas* species infecting different types of host plants could still share high similarity within some genes as not all genes are associated with plant host infections and interactions. Analysis of the *rpfB* gene allowed for distinct clustering of all included species that were similar to *rpfB* phylogeny constructed by Simões *et al.* (2007).

5.3.6. *rpoB*

Ferreira-Tonin *et al.* (2012) evaluated the suitability of the *rpoB* gene to serve as molecular marker for successfully distinguishing between *Xanthomonas* species. The *rpoB* gene data indicated that this gene would be useful to serve as an alternative molecular marker to determine genetic relationships in the *Xanthomonas* genus. Data provided by *rpoB* gene analysis with PCR-RFLP suggested that the gene can be used to identify and diagnose closely related species and strains. The *rpoB* gene is suitable to use as a molecular marker to distinguish between *Xanthomonas* species and to use as a molecular tool for assessing phylogenetic relatedness between members.

In a study by Case *et al.* (2007) it was determined that the *rpoB* gene provided higher resolution to allow discrimination between closely related organisms such as species and subspecies in comparison to the 16S rDNA gene. The 16S rDNA gene is only sufficient to distinguish organisms on higher taxonomic levels. As this study focused on determining taxonomic placements on potentially subspecies level, the protein-coding *rpoB* gene was employed in an attempt to refine close relationships.

The *rpoB* gene sequences of the BLSD isolates as well as sequences of other *Xanthomonas* species obtained from GenBank exhibited sufficient variation among each other to cluster each species in a separate group. Phylogenetic analysis based on *rpoB* gene sequences placed the BLSD agent close to *X. hortorum* pv. *hederae*. *X. hortorum* pv. *hederae* is the causal agent of bacterial leaf spot of ivy and is thus not pathogenic to maize or other monocots (Pirc *et al.*, 2016). Phylogenetic analysis of the 16S rDNA gene done by Niemann (2015) had also grouped the BLSD causal agent together with *X. hortorum* along with *X. axonopodis* pv. *vasculorum* and *X. vasicola* pv. *vasculorum*. Although not pathogenic to any

similar hosts, close relations of *X. hortorum* pv. *hederae* to *X. vasicola* have been documented in previous studies inferring a notable degree of genetic similarity between the two organisms (Simões *et al.*, 2007; Niemann, 2015).

Phylogenetic analysis of the *rpoB* gene done by Ferreira-Tonin *et al.* (2012) placed *Xanthomonas* species into two primary groups with high bootstrap confidence of 99 %. The first group included generally more distinct early-branching *Xanthomonas* species: *X. albilineans*, *X. sacchari*, *X. theicola*, *X. translucens* and *X. hyacinthi*. The second group consisted of *X. euvesicatoria*, *X. perforans*, *X. alfalfae*, *X. citri*, *X. dyei*, *X. axonopodis*, *X. oryzae*, *X. hortorum*, *X. bromi*, *X. vasicola*, *X. cynarae*, *X. gardneri*, *X. campestris*, *X. fragariae*, *X. arboricola*, *X. cassavae*, *X. cucurbitae*, *X. pisi*, *X. vesicatoria*, *X. codiae* and *X. melonis* (Ferreira-Tonin *et al.*, 2012). Similar results were obtained by Parkinson *et al.* (2007) for phylogenetic analyses based on *gyraseB* sequences.

Phylogenetic relationships constructed for *rpoB* gene sequences in the present study also revealed two clearly distinct clusters dividing early- and later-branching *Xanthomonas* species representing the first and second *Xanthomonas* groups as described by Ferreira-Tonin *et al.* (2012). The BLSD causal agent clustered within the second *Xanthomonas* group. General topology of *rpoB* taxonomic relationships in Figure 8 correlated well with the taxonomic groupings established by Ferreira-Tonin *et al.* (2012).

In this work, phylogenetic analysis of the *rpoB* gene have suggested a high gene similarity between the BLSD pathogen and *X. hortorum* pv. *hederae*. *X. hortorum* pv. *hederae* also shared a cluster with the BLSD agent in a study by Niemann (2015) based on 16S rDNA data. As with the *rpfB* gene, the high gene similarity between the BLSD pathogen and its most similar organism based on the specific gene, could not be explained by pathogenicity on any similar host plants. It has previously been documented, however, that *X. hortorum* pv. *hederae* has been closely associated with the *X. vasicola* species (Simões *et al.*, 2007; Niemann, 2015).

In the present study the BLSD causal agent showed high gene similarity to *X. vasicola* based on *atpD* and *dnaK* data. This might demonstrate an indirect connection between the BLSD agent and *X. vasicola* based on *rpoB* data as *X. hortorum* pv. *hederae* could be associated with *X. vasicola*. Analysis of the *rpoB* gene allowed for discriminative clustering of all included strains with high resolution. General *rpoB* tree topology was similar to previous studies, supporting the findings of Ferreira-Tonin *et al.* (2012) that the *rpoB* gene is a useful marker to explore *Xanthomonas* genetic relationships.

5.3.7. Summary of phylogenetic analyses

Martens *et al.* (2008) have explored the possibilities of ten different housekeeping genes as an alternative to DNA-DNA-hybridization and 16S rDNA gene analysis for bacterial classification. The ten housekeeping genes included *atpD*, *dnaK*, *recA* and *rpoB* genes that were also employed in the present study. Genetic analyses for all these housekeeping genes had ascertained clear boundaries between species and higher discrimination ability in comparison to 16S and 23S rDNA genes (Martens *et al.*, 2008). It was also determined that data provided by housekeeping gene sequence analysis correlated well with DNA-DNA hybridization data to indicate intraspecies relationships.

In the present study all analysed housekeeping genes demonstrated sufficient potential to discriminate between *Xanthomonas* members on intra-species level. The BLSD causal agent could be differentiated from other *Xanthomonas* species and pathovars and its closest relatives could be pinpointed. All species could be discriminated from each other and were placed in separate groups with no large indiscriminative clusters of various species. When compared to data obtained by Niemann (2015) of both 16S and 23S rDNA analysis to determine genetic relatedness of the BLSD causal agent to known *Xanthomonas* species, all included housekeeping genes have provided better phylogenetic resolution. The six housekeeping genes had similar potential for resolving taxonomic relationships of the BLSD causal agent as *gyraseB* sequences used by Niemann (2015).

Phylogenetic data reported by Niemann (2015) on 16S rDNA, 23S rDNA and *gyraseB* sequences as well as data from all six housekeeping genes analysed in the present study, have clustered the BLSD pathogen as a close relative of *X. vasicola*, *X. axonopodis*, *Xanthomonas hortorum* and *X. oryzae* species. Data from the study by Niemann (2015) as well as from the present study suggested that the BLSD causal agent is not a pathovar of the *X. campestris* species. No phylogenetic evidence supporting close relatedness between the BLSD pathogen and *X. campestris* strains could be found.

5.4. Clarifications of general trends observed in phylogenetic analyses

The *Xanthomonas* genus can be divided into two main groups. The first group consist of early-branching species including *X. albilineans*, *X. translucens*, *X. sacchari*, *X. theicola* and *X. hyacinthi*. The second group contains less divergent species that are phylogenetically notably less distinct from one another (Parkinson *et al.*, 2007; Ferreira-Tonin *et al.*, 2012; Gardiner *et al.*, 2014).

The continuous occurrence of the species *X. albilineans*, *X. translucens*, *X. sacchari* *X. theicola* and *X. hyacinthi* forming early-branching clusters with relative longer branch lengths and notable divergence in the present study were thus congruent with previous findings and matched the general topology of the *Xanthomonas* genus. The BLSD pathogen was closely related to species that clustered within the second primary group based on phylogenetic analysis of all included gene sequences. Consequently, the BLSD causal agent can be assigned to the second *Xanthomonas* group consisting of later-branching species genetically less distinct from one another.

Overall tree topologies for the housekeeping genes did not reveal consistent clustering of different *Xanthomonas* species. Fischer-Le Saux *et al.*, (2015) have found that individual phylogenetic trees for seven housekeeping genes of *Xanthomonas* strains did not reveal congruent tree topologies. Each tree revealed slightly different clustering of different strains. No gene exhibits exactly the same evolutionary history as single genes only reflect that specific gene's evolution. Evolutionary histories of genes may also vary due to horizontal gene transfer (Patwardhan *et al.*, 2014).

The different clustering of strains based on each gene may also be explained by a high degree of interspecies recombination that rearranges the phylogenetic signal (Fischer-Le Saux *et al.*, 2015). Different genes within a single organism may have different rates of evolution (Fischer-Le Saux *et al.*, 2015). Reports of inconsistencies in phylogeny among organisms based on single marker genes have been reported (Patwardhan *et al.*, 2014). Although useful in determining relative phylogenetic placements, each gene individually does not possess enough phylogenetic information to resolve full genetic history. Analysis of multiple genes provides more accurate information regarding the evolutionary background of an organism (Patwardhan *et al.*, 2014).

5.5. Possible reclassification of the BLSD causal agent

Other *Xanthomonas* species that are pathogenic to maize were previously classified as *X. campestris* pv. *vasculorum* and *X. campestris* pv. *holcicola*. These pathogens have since been reclassified and were moved from the *X. campestris* species to the *X. vasicola* species (Vauterin *et al.*, 1995; Karamura *et al.*, 2015). *X. campestris* pv. *zeae* was likely given its classification due to sharing a target host with *X. campestris* pv. *vasculorum* and *X. campestris* pv. *holcicola*, both also capable of infecting maize (Qhobela *et al.*, 1990). As it has since been established that both these pathogens did not belong to the *X. campestris* species, it is unlikely that the BLSD causal agent is a member of the *X. campestris* species.

Different *Xanthomonas* members that infect the same plant host are typically genetically and taxonomically very similar. *X. oryzae* pv. *oryzae* and *X. oryzae* pv. *oryzicola* are both pathogenic to rice and are placed in the same phylogenetic group in previous studies (Gardiner *et al.*, 2014; Fischer-Le Saux *et al.*, 2015). It can thus be expected that the BLSD isolates are very closely related to *X. vasicola* species based on host specificity.

The BLSD causal agent is currently classified as *X. campestris* pv. *zeae* (Qhobela *et al.*, 1990; Coutinho & Wallis, 1991). However, in phylogenetic analyses of all six housekeeping genes included in this study, the BLSD type strain never grouped with *X. campestris* or clustered in a closely related group with the *X. campestris* species. The BLSD causal agent was, in comparison to some other genus members, relatively distant related to the *X. campestris* species and appeared to be more closely related to species such as *X. vasicola* and *X. oryzae*. Whole genome sequencing data has also revealed higher sequencing similarity between the BLSD isolates and *X. oryzae* although this information might not be entirely conclusive.

In a study by Niemann (2015) on the phylogenetic position of the BLSD causal agent among other *Xanthomonas* species based on the 16S rDNA, 23S rRNA and *gyraseB* gene sequences, similar BLSD clustering were observed as in the present study. Analysis of 16S rDNA sequences revealed that the BLSD causal agent shared the highest similarity to *X. axonopodis* pv. *vasculorum* and *X. vasicola* pv. *vasculorum*. The BLSD isolate grouped together with the two above mentioned species as well as *X. hortorum* and *X. arboricola*. *X. campestris* pv. *campestris* and other *X. campestris* pathovars were placed in a different group, distant from the BLSD group.

Comparison of *gyraseB* sequences grouped the BLSD causal agent with *X. vasicola* as the closest relative. *X. campestris* was again placed in a separate group from the BLSD causal

agent. Analysis of the 23S rRNA gene has shown that *X. axonopodis* pv. *vasculorum* and *X. oryzae* pv. *oryzicola* shared a group with the BLSD causal agent revealing high gene similarity among these organisms (Niemann, 2015).

The BLSD causal agent was not consistently grouped with one specific reference species during analysis of the six housekeeping genes. Although closest relatedness of the causal agent could be assigned to *X. vasicola* and *X. oryzae* the strain was not consistently placed with these species in all phylogenetic trees. The BLSD causal agent is possibly a novel *Xanthomonas* species as it could not be consistently associated with one specific *Xanthomonas* species.

5.6. Primer design and specificity testing

Early detection of bacterial plant diseases is known to be the best preventative measure of pathogen dissemination. Molecular-based methods are considered rapid and accurate techniques for plant pathogen detection with increasing popularity (Lopez *et al.*, 2003; Lopez *et al.*, 2010). However, the development of discriminatory genetic markers for enhanced diagnosis of *Xanthomonas* pathogens remains a challenge (Obradovic *et al.*, 2004). PCR based primer pairs are often not specific enough for target strains.

Molecular markers are unique DNA fragments present in the genome of an organism which can be used to detect the organism (Liu *et al.*, 2012). Polymorphisms deemed from nucleotide mutations, additions and deletions lead to genomic variations within genomes. These changes result in differences in the base pair composition at a specific locus for different species of organisms. Genetic markers have important applications in bacterial species identification and classification (Mishra *et al.*, 2014). Molecular markers developed in this study were based on sequences associated with *dnaK*, *recA* and *rpoB* genes. Liu *et al.* (2012) have listed these three genes as significantly useful and popular genetic markers together with 16S rRNA, 23S rRNA, *gyrB* and 16S-23S ISR genes.

All six primer pairs succeeded in positively amplifying the respective gene fragments from the BLSD isolates of the expected amplicon sizes. However, most of them were not specific to the BLSD pathogen alone. Regions were not variable enough to completely exclude amplification of other *Xanthomonas* strains. Albuquerque *et al.* (2012) have designed primer pairs to amplify specific regions in *X. euvesicatoria* strains. Three of the ten tested pairs were specific to one strain while the rest positively amplified regions in all tested strains. In many cases intra-specific variability among closely related *Xanthomonas* strains is low and nucleotide differences are few, which complicates the design of highly specific primers.

Five of the six molecular marker regions tested in this study were not completely specific to the BLSD pathogen. The primer pair XanZ_rpoB is potentially a specific genetic marker for the BLSD pathogen. The sequence region that the primer set targeted was variable enough between BLSD isolates and other species for exclusive amplification of the region in BLSD samples. However, it cannot be assumed that this primer pair is completely specific as only two other strains of different species were included as negative controls (*X. axonopodis* and *X. campestris* pv. *campestris*).

As *X. vasicola* species are also capable of infecting maize (Karamura *et al.*, 2015) it would be ideal to test *X. vasicola* strains with the XanZ_rpoB marker. Data from this study have also shown close relation of the BLSD isolates with *X. vasicola*, further supporting the inclusion of *X. vasicola* in analysis with the molecular XanZ_rpoB marker. *X. oryzae* has also repeatedly revealed high similarity and close relation to the BLSD pathogen in this study, hence it would be recommended to include *X. oryzae* as well.

CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS

6.1. Conclusions

The present study aimed to analyse the causal agent of bacterial leaf streak disease at the molecular level. The collection of molecular data compiled in this study would supply valuable information regarding the taxonomic position of the BLSD causal agent in relation to different *Xanthomonas* species, as well as for genetic marker genes for the BLSD pathogen.

Whole genome sequencing was performed to obtain the complete nucleotide composition of the BLSD pathogen. Hundreds of gaps were present within the assembled genome which requires further investigation. However, the data has still suggested that the genomic composition of the BLSD causal agent shared higher similarity to the complete genome sequence of *X. oryzae* than to *X. campestris*. Sequence gaps may have been caused by incomplete sequencing but may also indicate a possible reduced genome of the BLSD pathogen. The whole genome composition was further mined for gene sequences of housekeeping genes associated with *Xanthomonas* species for gene comparison purposes.

Comparative housekeeping gene analyses of six complete gene sequences of *atpD*, *dnaK*, *fyuA*, *recA*, *rpfB* and *rpoB* genes were done. The BLSD organism and other *Xanthomonas* species shared gene similarities of between 86.1 % and 97.2 %. Of the included species, *X. alibilineans* was genetically the least similar to the BLSD pathogen with all gene similarities less than 90 %. Depending on the individual gene analysed, *X. axonopodis* and *X. oryzae* exhibited highest nucleotide similarities (92.6 % – 97.2 %) to the BLSD isolates.

Partial sequences of the six housekeeping genes were used to construct phylogenetic trees to pinpoint the *Xanthomonas* pathogens that are the most similar to the BLSD causal agent. Based on *atpD* and *dnaK* gene sequences, the BLSD pathogen was most closely related to *X. vasicola*. Analyses of the *fyuA* and *recA* genes placed the BLSD causal agent in its own group but with highest similarity to *X. oryzae* pathovars. Based on *rpfB* gene sequences the BLSD type strain was most similar to *X. axonopodis* and *X. citri* pv. *citri*. Analysis of *rpoB* gene sequences exhibited high genetic similarity between the BLSD causal agent and *X. hortorum* pv. *hederae*.

Although the use of individual housekeeping genes does not predict overall taxonomic relatedness and does not discriminate all related taxa, it was still valuable in providing information about taxonomic clustering and determining relatedness. The use of different individual housekeeping genes succeeded in pinpointing the closest relatives of the BLSD

causal agent. These were *X. vasicola*, *X. oryzae*, *X. hortorum* pv. *hederae* and *X. axonopodis* strains. The BLSD causal agent was not closely grouped with the *X. campestris* species, regardless of the gene that was analysed. New information regarding the taxonomic position of the BLSD causal agent within the *Xanthomonas* genus was valuable to support suggestions for possible revision of its current classification.

Data gathered in the present study regarding other *Xanthomonas* species that were the most similar to the BLSD pathogen could be used to determine possible connections between the closely related pathogens as well as their host plants. Revision of the current classification of the BLSD pathogen, *X. campestris* pv. *zeari*, is necessary. Data from the present study has supported the conclusion that the BLSD causal agent is not a member of the *X. campestris* group. It is possibly a novel species as phylogenetic data did not place the BLSD causal agent consistently with any specific known species.

A primer pair was developed for selective amplification of BLSD isolates based on a variable region of the *rpoB* gene. The primer pair XanZ_rpoB with a PCR based method could be used as a potential molecular tool for the detection of the bacterial leaf streak disease pathogen. The primer pair succeeded in amplifying the desired gene fragment of 703 bp present in BLSD isolates with no amplification of the two reference strains being observed. These results suggested that the primer set was specific to the target BLSD pathogen. This could enable possibilities of managing bacterial leaf streak disease through early detection in the future.

6.2. Recommendations

- Whole genome sequencing should be repeated as sequencing results of the genome of BLSD causal agent appeared to be incomplete in this study due to the presence of numerous large gaps in the sequences. The genome obtained in this work was smaller than reference genomes of *X. campestris* pv. *campestris* and *X. oryzae* pv. *oryzae*. Genomic DNA samples were already prepared with sequencing reagents a while before sequencing commenced in this study. This may have caused degeneration of the DNA leading to incomplete sequences. It is thus recommended to repeat whole genome sequencing and adding all sequencing reagents immediately before sequencing to ensure good quality DNA. A second whole genome sequencing run would further be valuable in determining if the BLSD pathogen possibly has a reduced genome in comparison to other *Xanthomonas* species.

- The phylogenetic relationship of the BLSD causal agent needs to be refined by obtaining more sequences of *Xanthomonas* species for all genes. Sequences of suitable lengths available in GenBank for specific genes are still limited but the numbers are increasing as new sequences are deposited. The alternative would be to obtain strains of other *Xanthomonas* species from a culture bank and use primers to extract specific gene sequences. Concatenated data sets of different housekeeping genes should also be created to determine overall genomic relatedness with the combined data. This approach would require individual data sets of all genes to contain exactly the same reference strains and sequence lengths in order to combine them.
- The PCR primer set, XanZ_rpoB, that was found to be a potentially diagnostic molecular marker, needs to be tested with *X. vasicola* and *X. oryzae* strains to establish its specificity. XanZ_rpoB failed to amplify any bands in two relatives of the BLSD pathogen, *X. axonopodis* and *X. campestris pv. campestris*. However, it has not been determined if the primer set is exclusively specific to the BLSD causal agent since the most similar *Xanthomonas* species to the BLSD pathogen (as determined in the present study), *X. vasicola* and *X. oryzae*, were not included in PCR specificity tests. As XanZ_rpoB was the only one of the six tested primer sets that selectively targeted BLSD samples, it is recommended that the primer set is renamed to XanBLSD for future purposes as potential BLSD molecular marker.
- A real-time PCR assay should be developed using the PCR primer set, XanZ_rpoB (XanBLSD), as a diagnostic method for detection of the pathogen within infected material. With application of a real-time PCR method, the pathogen could be detected even where disease symptoms are not visible. This approach would be valuable in determining if BLSD is seed-borne and what other material may potentially carry and spread the disease. Once main sources of inoculum for the disease have been determined and unique DNA markers identified, effective measures can be developed to prevent spreading of the pathogen and control disease outbreaks. By enabling reliable detection of the presence of the BLSD causal agent in plant products through molecular techniques, transmission can be prevented in the future which is of great importance, particularly in commercial and international trade.

REFERENCES

- Abd-El Salam, K.A. 2003. Bioinformatics tools and guidelines for PCR primer design. *African Journal of Biotechnology*, 2(5):91-95.
- Adékambi, T., Shinnick, T.M., Raoult, D. & Drancourt, M. 2008. Complete *rpoB* gene sequencing as a suitable supplement to DNA–DNA hybridization for bacterial species and genus delineation. *International Journal of Systematic and Evolutionary Microbiology*, 58:1807–1814.
- Adhikari, T.B., Gurung, S., Hansen, J.M. & Bonman, J.M. 2012. Pathogenic and genetic diversity of *Xanthomonas translucens* pv. *undulosa* in North Dakota. *Phytopathology*, 102(4):390-403.
- Albers, C.A., Lunter, G., MacArthur, D.G., McVean, G., Ouwehand, W. H. & Durbin, R. 2011. Dindel: Accurate indel calls from short-read data. *Genome Research*, 21:961-973.
- Albuquerque, P., Caridade, C.M.R., Rodrigues, A.S., Marcal, A.R.S., Cruz, J., Cruz, L., Santos, C.L., Mendes, M.V. & Tavares, F. 2012. Evolutionary and experimental assessment of novel markers for detection of *Xanthomonas euvesicatoria* in plant samples. *PLoS One*, 7(5):e37836.
- Alfano, J.R. & Collmer, A. 1997. The Type III (*Hrp*) secretion pathway of plant pathogenic bacteria: trafficking harpins, Avr proteins, and death. *Journal of Bacteriology*, 179(18):5655-5662.
- Atlas, R.M. 2010. Microbiological media. 4th ed. Washington, DC: CRC Press.
- Arlat, M., Gough, C.L., Barber, C.E., Boucher, C. & Daniels, M.J. 1991. *Xanthomonas campestris* contains a cluster of *Hrp* genes related to the larger *Hrp* cluster of *Pseudomonas solanacearum*. *Molecular Plant-Microbe Interactions*, 4(6):593-601.
- Arrieta-Ortiz, M.L., Rodriguez-R, L.M., Perez-Quintero, I.L., Poulin, L., Diaz, A.C., Rojas, N.A., Trujillo, C., Benavides, M.R. Bart, R., Boch, J., Boureau, T., Darrasse, A., David, P., de Bernonville, T., Fontanilla, P., Gagnevin, L., Guerin, F., Jacques, M., Lauber, E., Lefeuvre, P., Medina, C., Medina, E., Montenegro, N., Bodnar, A.M., Noel, L.D., Quinones, J.F., Osorio, D., Pardo, C., Patil, P.B., Poussier, S., Pruvost, O., Robene-Soustrade, I., Ryan, R.P., Tabima, J., Morales, O.G., Verniere, C., Carrere, S., Verdier, V., Szurek, B., Restrepo, S., Lopez, C., Koebnik, R. & Bernal, A. 2013. Genomic survey of pathogenicity

determinants and VNTR markers in the cassava bacterial pathogen *Xanthomonas axonopodis* pv. *manihotis* strain CIO151. *PLoS One*, 8(11): e79704.

Becker, A., Katzen, F., Pühler, A. & Ielpi, L. 1998. Xanthan gum biosynthesis and application: a biochemical/genetic perspective. *Applied Microbiology and Biotechnology*, 50:145–152.

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J. & Brown C.G. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456:53–59.

Berthier, Y., Verdier, V., Guesdon, J., Chevrier, D., Denis, J., Decoux, G. & Lemattre, M. 1993. Characterisation of *Xanthomonas campestris* pathovars by rRNA gene restriction patterns. *Applied Environmental Microbiology*, 59(3):851-861.

Bhawsar, S. 2011. Phytopathogenic bacteria: Bacterial diseases in plants. <http://www.biotecharticles.com/Biology-Article/Phytopathogenic-Bacteria-Bacterial-Diseases-in-Plants-718.html>. Date of access: 15 December 2015.

Bi, H., Yu, Y., Dong, H., Wang, H. & Cronan, J.E. 2014. *Xanthomonas campestris rpfB* is a fatty acyl-CoA ligase required to counteract the thioesterase activity of the *rpfF* diffusible signal factor (DSF) synthase. *Molecular Microbiology*, 93(2):262-275.

Boch, J. & Bonas, U. 2001. Gram-negative plant pathogenic bacteria. (*In* Mühldorfer, I. & Schäfer, K.P., eds. Emerging bacterial pathogens. Basel: Karger. p. 186-196).

Bonfield, J.K. & Whitwham, A. 2010. Gap5—editing the billion fragment sequence assembly. *Bioinformatics*, 26(14):1699-1703.

Brown, T.A. 2002. Molecular phylogenetics. (*In* Brown. T.A. ed. Genomes 2. New York: John Wiley & Sons Inc. p. 391).

Brown, T.A. 2007. Genomes 3. New York: Garland Science Publishing.

Butler, J., MacCallum, I., Kleber, M., Shlyakhter, I. A., Belmonte, M. K., Lander, E. S., Nusbaum, C. & Jaffe, D. B. 2008. ALLPATHS: de novo assembly of whole-genome shotgun microreads. *Genome Research*, 18(5):810–820.

- Büttner, D. & Bonas, U. 2010. Regulation and secretion of *Xanthomonas* virulence factors. *FEMS Microbiology Reviews*, 34:107-133.
- Carter, B.A., Reeder, R., Mgenzi, S.R., Kinyua, Z.M., Mbaka, J.N., Doyle, K., Nakato, V., Mwangi, M., Beed, F., Aritua, V., Lewis Ivey, M.L., Miller, S.A. & Smith, J.J. 2010. Identification of *Xanthomonas vasicola* (formerly *X. campestris* pv. *musacearum*), causative organism of banana *xanthomonas* wilt, in Tanzania, Kenya and Burundi. *Plant Pathology*, 59(2):403.
- Case, R.J., Boucher, Y., Dahllorf, I., Holmstrom, C., Doolittle, W.F. & Kjelleberg, S. 2007. Use of 16S rRNA and *rpoB* genes as molecular markers for microbial ecology studies. *Applied and Environmental Microbiology*, 73:278-288.
- Chan, J.W.Y.F. & Goodwin, P.H. 1999. The molecular genetics of virulence of *Xanthomonas campestris*. *Biotechnology Advances*, 17:489-508.
- Chevreur, B., Wetter, T. & Suhai, S. 1999. Genome sequence assembly using trace signals and additional sequence information. *Computer Science and Technology: Proceedings of the German Conference on Bioinformatics (GCB)*, 99:45-56.
- Chidamba, L. & Bezuidenhout, C.C. 2012. Characterisation of *Xanthomonas campestris* pv. *campestris* isolates from South Africa using genomic DNA fingerprinting and pathogenicity tests. *European Journal of Plant Pathology*, 133:811-818.
- Cho, M.S., Kang, M.J., Kim, C.K., Seol, Y., Hahn, J.H., Park, S.C., Hwang, D.J., Ahn, T-Y., Park, D.H. & Lim, C.K. 2011. Sensitive and specific detection of *Xanthomonas oryzae* pv. *oryzae* by real-time Bio-PCR using pathovar-specific primers based on an *rhs* family gene. *Plant Disease*, 95(5):589-595.
- Chun, W.W.C. 2002. Xanthomonadins, unique yellow pigments of the genus *Xanthomonas*. *The Plant Health Instructor*, DOI: 10.1094/PHI-A-2000-0824-01.
- CIMMYT Maize Program. 2004. Maize diseases: A guide for field identification. 4th ed. Mexico: CIMMYT.
- Coutinho, T.A. 1988. Bacterial streak of maize: A study on host specificity and cultivar susceptibility. M.Sc. thesis, University of Natal.
- Coutinho, T.A. & Wallis, F.M. 1991. Bacterial leaf streak disease of maize (*Zea mays* L.) in South Africa. *Journal of Phytopathology*, 133:112-120.

Da Silva, A.C., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R., Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida, N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C., Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardozo, J., Chambergo, F., Ciapina, L.P., Cicarelli, R.M.B., Coutinho, L.L., Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S., Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C., Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite, R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A., Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, Meidanis, J., Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M., Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira, H.A., Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F., Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trinade dos Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. & Kitajima, J.P. 2002. Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature*, 417:459-463.

Dale, J.W. & Von Schantz, M. 2002. From genes to genomes: Concepts and applications of DNA technology. West Sussex, England: John Willey & Sons Ltd.

Deng, Y., Wu, J., Tao, F. & Zhang, L.H. 2011. Listening to a new language: DSF-based quorum sensing in Gram-negative bacteria. *Chemical Reviews*, 111:160-173.

Denny, T.P. 1995. Involvement of bacterial polysaccharides in plant pathogenesis. *Annual Review of Phytopathology*, 33:173-179.

Department of Agriculture, Forestry and Fisheries. 2012. Maize market value chain profile. <http://www.nda.agric.za/docs/AMCP/Maize2012.pdf>. Date of access: 9 April 2015.

Dieffenbach, C.W., Lowe, T.M.J. & Dvlske, G.S. 1993. General concepts for PCR primer design. *Genome Research*, 3:S30-S37.

Doebley, J.F. 1990. Molecular evidence and the evolution of maize. *Economic Botany*, 44(Supplement 3):6-27.

Donmez, N. 2012. Polymorphism and genome assembly. Ph.D. thesis, University of Toronto.

Dookun, A., Stead, D.E. & Autrey, J. 2000. Variation among strains of *Xanthomonas campestris* pv. *vasculorum* from Mauritius and other countries based on fatty acid analysis. *Systematic and Applied Microbiology*, 23:148-155.

- Dow, J.M. & Daniels, M.J. 2000. *Xylella* genomics and bacterial pathogenicity to plants. *Yeast*, 17:263-271.
- Downing, T.G. 1998. Preliminary investigation of the molecular pathogenicity determinants of *Xanthomonas campestris* pv. *zeae*. M.Sc. thesis, Rhodes University.
- Du Plessis, J. 2003. Maize production. <http://www.arc.agric.za/arc-gci/Fact%20Sheets%20Library/Maize-infopak.pdf>. Date of access: 12 April 2015.
- Duveiller, E. & Maraite, H. 1995. Effect of temperature and air humidity on multiplication of *Xanthomonas campestris* pv. *undulosa* and symptom expression in susceptible and field-tolerant wheat genotypes. *Journal of Phytopathology*, 143:227-232.
- Duveiller, E., Bragard, C. & Maraite, H. 1997. Bacterial leaf streak disease and black chaff caused by *Xanthomonas translucens*. (In Duveiller, E., Fucikovsky, L. & Rudolph, K. eds. The bacterial diseases of wheat. Mexico, D.F: CIMMYT p. 25-47).
- Dye, D.W., Bradbury, J.F., Goto, M., Hayward, A.C., Lelliot, R.A. & Schroth, M.N. 1980. International standards for naming pathovars of phytopathogenic bacteria and a list of pathovar names and pathotype strains. *Review of Plant Pathology*, 59:153-186.
- Dye, D.W. 1962. The inadequacy of the usual determinative tests for identification of *Xanthomonas* spp. *New Zealand Journal of Science*, 4:393-416.
- Dye, D.W. & Lelliot, R.A. 1974. *Xanthomonas* Dowson 1939. (In Buchanan, R.E. & Gibbons, N.E. eds. Bergey's manual of determinative bacteriology. 8th ed. Baltimore: Williams and Wilkins Co. p. 243-249).
- Dyer, R.A. 1949. Botanical surveys and control of plant diseases: Farming in South Africa. *Annual Report of the Department of Agriculture*, 275(24):119-121.
- Eisenberg, E. & Lavanon, E.Y. 2013. Human housekeeping genes, revisited. *Trends in Genetics*, 29(10):569-575.
- European and Mediterranean Plant Protection Organization. 2007. *Xanthomonas oryzae*. *EPPO Bulletin*, 37:543-553.
- Ezuka, A. & Kaku, H. 2000. A historical review of bacterial blight of rice. *Bulletin of The National Institute of Agrobiological Resources*, 15:53-54.

- Feng, W., Wang, Y., Huang, L., Feng, C., Chu, Z., Ding, X. & Yang, L. 2015. Genomic-associated markers and comparative genome maps of *Xanthomonas oryzae* pv. *oryzae* and *X. oryzae* pv. *oryzicola*. *World Journal of Microbiology and Biotechnology*, 31:1353-1359.
- Ferreira-Tonin, M., Rodrigues-Neto, J., Harakava, R. and 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*, 62:1419-1424.
- Finkmann, W., Alterndorf, K., Stackebrandt, E. & Lipski, A. 2000. Characterisation of N₂O-producing *Xanthomonas*-like isolates from biofilters as *Stenotrophomonas nitritireducens* sp. nov., *Luteimonas mephitis* gen. nov., sp. nov. and *Pseudoxanthomonas broegbernensis* gen. nov., sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, 50:273-282.
- Fischer-Le Saux, M., Bonneae, S., Essakhi, S., Manceau, C. & Jacques, M. 2015. Aggressive emerging pathovars of *Xanthomonas arboricola* represent widespread epidemic clones distinct from poorly pathogenic strains, as revealed by multilocus sequence typing. *Applied and Environmental Microbiology*, 81(14):4651-4668.
- Forster, R.L., Mihuta-Grimm, L., Schaad, N.W. 1986. Black chaff of wheat and barley. University of Idaho, College of Agriculture. *Current Information Series No. 784*.
- Fullwood, M.J., Wei, C., Liu, E. T. & Ruan, Y. 2009. Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Research*, 19:521-532.
- Garces, F.F., Gutierrez, A. & Hoy, J.W. 2014. Detection and quantification of *Xanthomonas albilineans* by qPCR and potential characterization of sugarcane resistance to leaf scald. *Plant Disease*, 98:121-127.
- Gardiner, D.M., Narayana, M.U., Stiller, J., Ellis, J.G., Dodds, P.N., Kazan, K. & Manners, M. Genomic analysis of *Xanthomonas translucens* pathogenic on wheat and barley reveals cross-kingdom gene transfer events and diverse protein delivery systems. *PloS One*, 9: e84995.
- Gerlach, R.G. & Hensel, M. 2007. Protein secretion systems and adhesins: The molecular armory of Gram-negative pathogens. *International Journal of Medical Microbiology*, 279:401-415.

Gonçalves, E.R. & Rosato, Y.B. 2002. Phylogenetic analysis of *Xanthomonas* species based on 16S-23S rDNA intergenic spacer sequences. *International Journal of Systematic and Evolutionary Microbiology*, 52:355-361.

Goszczyńska, T., Serfontein, J.J. & Serfontein, S. 2000. Introduction to practical phytobacteriology. 1st ed. Pretoria, South Africa: SAFRINET.

Hauben, L., Vauterin, L., Swings, J. & Moore, E.R.B. 1997. Comparison of 16S ribosomal DNA sequences of all *Xanthomonas* species. *International Journal of Systematic Bacteriology*, 47:328-335.

Irey, M.S. & Stall, R.E. 1982. Value of xanthomonadins for identification of pigmented *Xanthomonas campestris* pathovars. (In Lozano, C. ed. Proceedings of the fifth internal conference on plant pathological bacteria. Cali, Columbia: CIAT. p. 85-95).

Jalan, N., Kumar, D., Yu, F., Jones, J.B., Graham, J.H. & Wang, N. 2013. Complete genome sequence of *Xanthomonas citri* subsp. *citri* strain Aw12879, a restricted-host-range citrus canker-causing bacterium. *Genome Announcements*, 1(3):1-2.

Janse, J.D. 1992. Whole cell fatty acid analysis as a tool for classification of phytopathogenic *Pseudomonas* bacteria. Ph.D. thesis, Landbouwniversiteit, Wageningen.

Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G. & Gibson, T.J. 1998. Multiple sequence alignment with Clustal X. *Trends in Biochemical Sciences*, 10:403-405.

Kaplan, M.L. & Kaplan, L. 1933. The Gram stain and differential staining. *Journal of Bacteriology*, 25(3):309-321.

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(9):385-400.

Kingsley, M.T., Gabriel, D.W., Marlow, G.C. & Roberts, P.D. 1993. The *opsX* locus of *Xanthomonas campestris* affects host range and biosynthesis of lipopolysaccharide and extracellular polysaccharide. *Journal of Bacteriology*, 175(18):5839-5850.

Lang, J.M., Hamilton, J.P., Diaz, M.G.Q., Van Sluys, M.A., Burgos, M.R.G. & Vera Cruz, C.M. 2010. Genomics-based diagnostic marker development for *Xanthomonas oryzae* pv. *oryzae* and *X. oryzae* pv. *oryzicola*. *Plant Disease*, 94:311–319.

Lee, B.M., Park, Y.J., Park, D.S., Kang, H.W., Kim, J.G., Song, E.S., Park, I.C., Yoon, K.O., Kim, J.H., Jung, C., Koh, N.H., Seo, J.S. & Go, S.J. 2005. The genome sequence of *Xanthomonas oryzae* pathovar *oryzae* KACC10331, the bacterial blight pathogen of rice. *Nucleic Acids Research*, 33(2):577-586.

Lee, J., Shin, H., Park, H., Ryu, S. & Han, S. 2014. Draft genome sequence of *Xanthomonas axonopodis* pv. *glycines* 8ra possessing transcription activator-like effectors used for genetic engineering. *Journal of Biotechnology*, 179:15-16.

Leyns, F., De Cleene, M., Swings, J. & De Ley, J. 1984. The host range of the genus *Xanthomonas*. *The Botanical Review*, 50(3):308-356.

Li, K., Brownley, A., Stockwell, T.B., Beeson, K., McIntosh, T.C., Busam, D., Ferriera, S., Murphy, S. & Levy, S. 2008. Novel computational methods for increasing PCR primer design effectiveness in directed sequencing. <http://www.biomedcentral.com/1471-2105/9/191>. Date of access: 5 June 2016.

Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., Li, Y., Li, S., Shan, G., Kristiansen, K., Li, S., Yang, H., Wang, J. & Wang, J. 2010. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Research*, 20:265-272.

Li, J. 2015. Application of next generation sequencing on genetic testing. Ph.D. thesis, University of Copenhagen.

Liberek, K., Galitski, T.P., Zylicz, M. & Georgopoulos, C. 1992. The *dnaK* chaperone modulates the heat shock response of *Escherichia coli* by binding to the σ^{32} transcription factor. *Proceedings of the National Academy of Sciences*, 89:3516-3520.

Lilly, V.G., Wilson, H.A. & Leach, J.G. 1958. Bacterial polysaccharides: II. Laboratory-scale production of polysaccharides by species of *Xanthomonas*. *Applied and Environmental Microbiology*, 6(2):105.

Liu, W., Li, L., Khan, M.A. & Zhu, F. 2012. Popular molecular markers in bacteria. *Molecular Genetics, Microbiology and Virology*, 27(3):103-107.

Louws, F.J., Rademaker, J.L.W. & De Bruijn, F.J. 1999. The three D's of PCR-based genomic analysis of phytobacteria: Diversity, detection and disease diagnosis. *Annual Review of Phytopathology*, 37:80-125.

- Lopez, M.M., Bertolini, E., Olmos, A., Caruso, P. 2003. Innovative tools for detection of plant pathogenic viruses and bacteria. *International Microbiology*, 6(4):233-243.
- Lopez, M.M., Rosello, M. & Palacio-Bielsa, A. 2010. Diagnosis and detection of the main bacterial pathogens of stone fruit and almond. *Journal of Plant Pathology*, 92(Supplement 1):57-66.
- Lu, W., Pan, L., Zhao, H., Jia, Y., Wang, Y., Yu, X. & Wang, X. 2014. Molecular detection of *Xanthomonas oryzae* pv. *oryzae*, *Xanthomonas oryzae* pv. *oryzicola*, and *Burkholderia glumae* in infected rice seeds and leaves. *The Crop Journal*, 2:398-406.
- Mansfield, J., Genin, S., Magori, S., Citovsky, V., Sriariyanum, M., Ronald, P., Dow, M., Verdier, V., Beer, S.V., Machado, M.A., Toth, I., Salmond, G. & Foster, G.D. 2012. Top 10 plant pathogenic bacteria in molecular plant pathology. *Molecular Plant Pathology*, 13(6):613-629.
- Marcelletti, S., Ferrante, P. & Scortchini, M. 2010. Multilocus sequence typing reveals relevant genetic variation and different evolutionary dynamics among strains of *Xanthomonas arboricola* pv. *juglandis*. *Diversity*, 2:1205-1222.
- Mardis, E.R. 2013. Next-generation sequencing platforms. *Annual Review of Analytical Chemistry*, 6:287-303.
- Maréchal, J., Clement, B., Nalin, R., Gandon, C., Orso, S., Cvejic, J.H., Bruneteau, M., Berry, A. & Normand, P. 2000. A *recA* gene phylogenetic analysis confirms the close proximity of *Frankia* to *Acidothermus*. *International Journal of Systematic and Evolutionary Microbiology*, 50(2):781-785.
- Martens, M., Dawyndt, P., Coopman, R., Gilles, M., De Vos, P. & Willems, A. 2008. Advantages of multilocus sequence analysis for taxonomic studies: a case study using 10 housekeeping genes in the genus *Ensifer* (including former *Sinorhizobium*). *International Journal of Systematic and Evolutionary Microbiology*, 58: 200-214.
- Massomo, S.M.S., Nielsen, H., Mabagala, R.B., Mansfeld-Giese, K., Hockenhull, J. & Mortensen, C.M. 2003. Identification and characterisation of *Xanthomonas campestris* pv. *campestris* strains from Tanzania by pathogenicity tests, Biolog, rep-PCR and fatty acid methyl ester analysis. *European Journal of Plant Pathology*, 105:775-789.
- Mew, T., Alvarez, A., Leach, J. & Swings, J. 1993. Focus on bacterial blight of rice. *Plant Disease*, 77:5-12.

Mielewczik, M., Friedli, M., Kirchgessner, N. & Walter, A. 2013. Diel leaf growth of soybean: a novel method to analyze two-dimensional leaf expansion in high temporal resolution based on a marker tracking approach (Martrack Leaf). *Plant Methods*, 9:30.

Ministry of Environments and Forests. 2010. Biology of maize. [http://www.envfor.nic.in/divisions/csurv/geac/Biology_of_Maize\[1\].pdf](http://www.envfor.nic.in/divisions/csurv/geac/Biology_of_Maize[1].pdf). Date of access: 9 April 2015.

Mirmajlessi, S.M., Destefanis, M., Gottsberger, A., Mänd, M. & Loit, E. 2015. PCR-based specific techniques used for detecting the most important pathogens on strawberry: a systematic review. *Systematic Reviews*, 4:9.

Mishra, K.K., Fougat, R.S., Ballani, A., Thakur, V., Jha, Y. & Bora, M. 2014. Potential and application of molecular markers techniques for plant genome analysis. *International Journal of Pure and Applied Bioscience*, 2(1):169-188.

Moore, E.R.B., Kruger, A.S., Hauben, L., Seal, S.E., De Baere, R., De Wachter, R., Timmis, K.N. & Swings, J. 1997. 16S rRNA gene sequence analyses and inter- and intrageneric relationships between *Xanthomonas* species and *Stenotrophomonas maltophilia*. *FEMS Microbiology Letters*, 151:145-153.

Mount, D.W. 2001. Bioinformatics: Sequence and genome analysis. Cold Spring Harbor Laboratory Press: New York.

National Agricultural Marketing Council. 2003. Report on the investigation of the effects of deregulation on the maize industry. http://www.namc.co.za/upload/section_7_reports/ImpactReport_Deregulation_MaizeIndustry_Mar2001.pdf. Date of access: 9 April 2015.

Niemann, N.J.J. 2015. Molecular characterisation of the causal agent of bacterial leaf streak of maize. M.Sc. thesis, North-West University.

Nienaber, J.J. 2015. Phenotypic and biochemical characterisation of the causal agent of bacterial leaf streak of maize. M.Sc. thesis, North-West University.

Norman, D.J., Chase, A.R., Hodge, N.C. & Stall, R.E. 1997. Differentiation of three species of *Xanthomonas* and *Stenotrophomonas maltophilia* using cellular fatty acid analysis. *European Journal of Plant Pathology*, 103:687-693.

Obradovic, A., Mavridis, A., Rudolph, K., Janse, J.D., Arsenijevic, M., Jones, J.B., Minsavage, G.V. & Wang, J-F. 2004. Characterization and PCR-based typing of

Xanthomonas campestris pv. *vesicatoria* from peppers and tomatoes in Serbia. *European Journal of Plant Pathology*, 110: 285.

Palacio-Bielsa, A., Cubero, J., Cambra, M.A., Collados, R., Berruete, I.M. & Lopez, M.M. 2011. Development of an efficient real-time quantitative PCR protocol for detection of *Xanthomonas arboricola* pv. *pruni* in *Prunus* species. *Applied and Environmental Microbiology*, 77:89-97.

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

Patwardhan, A., Ray, S. & Roy, A. 2014. Molecular markers in phylogenetic studies: A review. *Phylogenetics and Evolutionary Biology*, 2(2):131.

Pieretti, I., Royer, M., Barbe, V., Carrere, S., Koebnik, R., Cociancich, S., Couloux, A., Darrasse, A., Gouzy, J., Jacques, M., Lauber, E., Manceau, C., Mangenot, S., Poussier, S., Segurens, B., Szurek, B., Verdier, V., Arlat, M. & 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*, 10:616.

Pieretti, I., Royer, M., Barbe, V., Carrere, S., Koebnik, R., Cououx, A., Darrasse, A., Gouzy, J., Jacques, M., Lauber, E., Manceau, C., Mangenot, S., Poussier, S., Segurens, B., Verdier, V., Arlat, M., Gabriel, D.W., Rott, P. & Cociancich, S. 2012. Genomic insights into strategies used by *Xanthomonas albilineans* with its reduced artillery to spread within sugarcane xylem vessels. *BMC Genomics*, 13:658-681.

Pirc, M., Dreo, T., Sustarsic, M., Erjavec, J. & Ravnkar, M. 2016. First report of *Xanthomonas hortorum* pv. *hederae* causing bacterial leaf spot of *Hedera hibernica* in Slovenia. *Plant Disease*, 96:141.

Qhobela, M. & Claflin, L.E. 1988. Characterization of *Xanthomonas campestris* pv. *pennamericanum* pv. nov., causal agent of bacterial leaf streak disease of Pearl millet. *International Journal of Systematic Bacteriology*, 38(4):362-366.

Qhobela, M., Claflin, L.E. & Nowell, D.C. 1990. Evidence that *Xanthomonas campestris* pv. *zeae* can be distinguished from other pathovars capable of infecting maize by restriction fragment length polymorphism of genomic DNA. *Canadian Journal of Plant Pathology*, 12:183-186.

Qhobela, M., Leach, J.E., Claflin, L.E. & Pearson, D.L. 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*, 75:32-37.

Qian, W., Jia, Y., Ren, S., He, Y., Feng, J., Lu, L., Sun, Q., Ying, G., Tang, D-J., Tang, H., Wu, W., Hao, P., Wang, L., Jiang, B-L., Zeng, S., Gu, W-Y., Lu, G., Rong, L., Tian, Y., Yao, Z., Fu, G., Chen, B., Fang, R., Qiang, B., Chen, Z., Zhao, G-P., Rabibhadana, S., Chamnongpol, S., Sukchawalit, R., Ambulos Jr., N.P., Trempey, J.E. & Mongkolsuk, S. 1998. Characterization and expression analysis of a *Xanthomonas oryzae* pv. *oryzae* recA. *FEMS Microbiology Letters*, 158:195-200.

Qian, W., Jia, Y., Ren, S., He, Y., Feng, J., Lu, L., Sun, Q., Ying, G., Tang, D., Tang, H., Wu, W., Hao, P., Wang, L., Jilang, B., Zeng, S., Gu, W., Lu, G., Rong, L., Tian, Y., Yao, Z., Fu, G., Chen, B., Fang, R., Qiang, B., Chen, Z., Zhao, G., Tang, J. & He, C. 2005. Comparative and functional genomic analyses of the pathogenicity of phytopathogen *Xanthomonas campestris* pv. *campestris*. *Genome Research*, 15:757-767.

Rajagopal, L., Sundari, C.S., Balasubramanian, D. & Sonti, R.V. 1997. The bacterial pigment xanthomonadin offers protection against photodamage. *FEBS Letters*, 415(2):125-128.

Reece, R.J. 2004. Analysis of genes and genomes. England: John Wiley & Sons Ltd.

Remaut, H. & Waksman, G. 2004. Structural biology of bacterial pathogenesis. *Current Opinion in Structural Biology*, 14:161-170.

Ricaud, C. and Ryan C. C. 1989. Leaf scald. (In C. Ricaud, B.T. Egan, A.G. Gillaspie Jr. and C.G. Hughes. eds. Diseases of sugarcane. Major diseases. Amsterdam, The Netherlands: Elsevier Science Publishers B.V. p. 39-58).

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*, 12:43.

Rogers, Y.H. & Venter, J.C. 2005. Genomics: Massively parallel sequencing. *Nature*, 437:326-327.

Rothberg J.M., Hinz, W., Rearick, T.M., Schultz, J., Mileski, W., Davey, M., Leamon, J.H., Johnson, K., Milgrew, M.J., Edwards, M., Hoon, J., Simons, J.F., Marran, D., Myers, J.W., Davidson, J.F., Branting, A., Nobile, J.R., Puc, B.P., Light, D., Clark, T.A., Huber, M., Branciforte, J.T., Stoner, I.B., Cawley, S.E., Lyons, M., Fu, Y., Homer, N., Sedova, M., Miao,

X., Reed, B., Sabina, J., Feierstein, E., Schorn, M., Alanjary, M., Dimalanta, E., Dressman, D., Kasinskas, R., Sokolsky, T., Fidanza, J.A., Namsaraev, E., McKernan, K.J., Williams, A., Roth, G.T. & Bustillo, J. 2011. An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, 475:348–52.

Ryan, R.P. 2013. Cyclic di-GMP signalling and the regulation of bacterial virulence. *Microbiology*, 159:1286–1297.

Saddler, G.S. & Bradbury, J.F. 2005. The Proteobacteria. (In Brenner, D.J., Krieg, N.R., Staley, J.T. and Garrity, G.M., eds. *Bergey's Manual of Systematic Bacteriology*. New York: Springer. p. 63).

Salzberg, S.L., Sommer, D.D., Schatz, M.C., Phillippy, A.M., Rabinowicz, P.D., Tsuge, S. & Furutani, A. 2008. Genome sequence and rapid evolution of the rice pathogen *Xanthomonas oryzae* pv. *oryzae* PXO99^A. *BMC Genomics*, 9:204-220.

Sanger F., Nicklen S. & Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74: 5463–5467.

Schaad, N.W., Jones, J.B. & Chun, W. 2001 a. Laboratory guide for identification of plant pathogenic bacteria. 3rd ed. Minnesota, USA: APS Press.

Schaad, N.W., Jones, J.B. & Lacy, G.H. 2001 b. Gram-negative bacteria - *Xanthomonas*. (In Schaad, N.W., Jones, J.B. & Chun, W. 2001. eds. *Laboratory guide for identification of plant pathogenic bacteria*. 3rd ed. Minnesota, USA: APS Press. p. 175-200).

Scheibye-Alsing, K., Hoffmann, S., Frankel, A.M., Jensen, P. & Stadler, P.F. 2009. Sequencing assembly. *Computational Biology and Chemistry*, 33(2):121-136.

Schubert, S., Picard, B., Gouriou, S., Heesemann, J. & Denamur, E. 2002. *Yersinia* high-pathogenicity island contributes to virulence in *Escherichia coli* causing extraintestinal infections. *Infection and Immunity*, 70:5335–5337.

Sewariya, V.M., Shrivastava, R., Prasad, G. & Arora, K. 2012. In-vitro evaluation of novel synthetic compounds against *Xanthomonas campestris* pv. *campestris*. *International Journal of Pharma and Bio Sciences*, 3:441-453.

Simões, T.H.N., Goncalves, E.R. Rosato, Y.B. & Mehta, A. 2007. Differentiation of *Xanthomonas* species by PCR-RFLP of *rpfB* and *atpD* genes. *FEMS Microbiology Letters*, 271:33-37.

Simpson, A. J. G., Reinach, F.C., Arruda, P., Abreu, F. A., Acencio, M., Alvarenga, R., Alves, L. M. C., Araya, J. E., Baia, G. S., Baptista, C. S., Barros, M. H., Bonaccorsi, E. D., Bordin, S., Bove, J. M., Briones, M. R. S., Bueno, M. R. P., Camargo, A. A., Camargo, L. E. A., Carraro, D. M., Carrer, H., Colauto, N. B., Colombo, C., Costa, F. F., Costa, M. C. R., Costa-Neto, C. M., Coutinho, L. L., Cristofani, M., Dias-Neto, E., Docena, C., El-Dorry, H., Facincani, A. P., Ferreira, A. J. S., Ferreira, V. C. A., Ferro, J. A., Fraga, J. S., Franca, S. C., Franco, M. C., Frohme, M., Furlan, L. R., Garnier, M., Goldman, G. H., Goldman, M. H. S., Gomes, S. L., Gruber, A., Ho, P. L., Hoheisel, J. D., Junqueira, M. L., Kemper, E. L., Kitajima, J.P., Krieger, J. E. Kuramae, E. E., Laigret, F., Lambais, M. R., Leite, L. C. C., Lemos, E. G. M., Lemos, M. V. F., Lopes, S. A., Lopes, C. R., Machado, J. A., Machado, M. A., Madeira, A. M. B. N., Madeira, H. M. F., Marino, C. L., Marques, M. V., Martins, E. A. L., Martins, E. M. F., Matsukuma, A. Y., Menck, C. F. M., Miracca, E. C., Miyaki, C. Y., Monteiro-Vitorello, C. B., Moon, D. H., Nagai, M. A., Nascimento, A. L. T. O., Netto, L. E. S., Nhani, A., Nobrega, F. G., Nunes, L. R., Oliveira, M. A., de Oliveira, M. C., de Oliveira, R. C., Palmieri, D. A., Paris, A., Peixoto, B. R., Pereira, G. A. G., Pereira, H. A., Pesquero, J. B., Quaggio, R. B., Roberto, P. G., Rodrigues, V., de M. Rosa, A. J., de Rosa, V. E., de Sa, R. G., Santelli, R. V., Sawasaki, H. E., da Silva, A. C. R., da Silva, A. M., da Silva, F. R., Silva, W. A., da Silveira, J. F., Silvestri, M. L. Z., Siqueira, W. J., de Souza, A. A., de Souza, A. P., Terenzi, M. F., Truffi, D., Tsai, S. M., Tsuhako, M. H., Vallada, H., Van Sluys, M. A., Verjovski-Almeida, S., Vettore, A. L., Zago, M. A., Zatz, M., Meidanis, J. & Setubal, J. C. 2000. The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature*, 406:151-159.

Speed, T.P., Kechris, K.J. & McWeeney, S.K. 2006. Molecular evolution, substitution models, and phylogenies. http://www.cs.toronto.edu/~brudno/csc2417_15/evophylo.pdf. Date of access: 14 July 2016.

Staden, R., Beal, K.F. & Bonfield, J.K. 1998. The Staden package. (In Misener, S. & Krawetz, S.A., eds. *Methods in molecular biology: Bioinformatics methods and protocols*, Vol. 132. Totowa: The Humana Press Inc. p. 115-130).

Starr, M.P. & Stephens, W.L. 1964. Pigmentation and taxonomy of the genus *Xanthomonas*. *Journal of Bacteriology*, 87:293-302.

Stephens, W.L. & Starr, M.P. 1963. Localization of carotenoid pigment in the cytoplasmic membrane of *Xanthomonas juglandis*. *Journal of Bacteriology*, 86:1070-1074.

Stepkowski, T., Czaplinska, M., Miedzinska, K. & Moulin, L. 2003. The variable part of the *dnaK* gene as an alternative marker for phylogenetic studies of Rhizobia and related alpha *Proteobacteria*. *Systematic and Applied Microbiology*, 26:483-494.

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

Swings, J., Van den Mooter, M., Vauterin, L., Hoste, B., Gillis, M., Mew, T. W., and Kersters, K. 1990. Reclassification of the causal agents of bacterial blight (*Xanthomonas campestris* pv. *oryzae*) and bacterial leaf streak (*Xanthomonas campestris* pv. *oryzicola*) of rice as pathovars of *Xanthomonas oryzae* (ex Ishiyama, 1922) sp. nov., nom. rev. *International Journal of Systematic Bacteriology*, 40:309-311.

Tamura, K., 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.

Thellin, O., Zorzi, W., Lakaye, B., De Borman, B., Courmans, B., Hennen, G., Grisar, T., Igout, A. & Heinen, E. 1999. Housekeeping genes as internal standards: use and limits. *Journal of Biotechnology*, 75:291-295.

Thompson, J.D., Higgins, D.G. & Gibson, T.J. 1994. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22:4673-4680.

Tjou-Tam-Sin, N.N.A., van de Bilt, J.L.J., Bergsma-Vlami, M., Koenraadt, H., Westerhof, J., van Doorn, J., Pham, K.T.K. & Martin, W.S. 2012. First report of *Xanthomonas arboricola* pv. *pruni* in ornamental *Prunus laurocerasus* in the Netherlands. *Plant Disease*, 96(5):759.

Vandamme, P., Pot, B., Gillis, M., De Vos, P., Kersters, K. & Swings, J. 1996. Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiological reviews*, 60(2): 407-438.

Vandroemme, J., Baeyen, S., Van Vaerenbergh, J., De Vos, P. & Maes, M. 2008. Sensitive real-time PCR detection of *Xanthomonas fragariae* in strawberry plants. *Plant Pathology*, 57:438-444.

- Vandroemme, J., Cottyn, B., Bayen, 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.
- Vauterin, L., Swings, J., Kersters, M., Gillis, M. & Mew, T.W. 1990. Towards an improved taxonomy of *Xanthomonas*. *International Journal of Systematic Bacteriology*, 40(3):312-316.
- Vauterin, L., Yang, P., Hoste, B., Vancanneyt, M., Civerolo, E.L., Swings, J. & Kersters, K. 1991. Differentiation of *Xanthomonas campestris* pv. *citri* strains by sodium dodecyl sulphate-polyacrylamide gel electrophoresis of proteins, fatty acid analysis, and DNA-DNA hybridization. *International Journal of Systematic Bacteriology*, 41(4):535-542.
- Vauterin, L., Hoste, B., Kersters, K. & Swings, J. 1995. Reclassification of *Xanthomonas*. *International Journal of Systematic Bacteriology*, 45(3):472-489.
- Vauterin, L., Yang, P. & Swings, J. 1996. Utilization of fatty acid methyl esters for the differentiation of new *Xanthomonas* species. *International Journal of Systematic Bacteriology*, 46:296-304.
- Vauterin, L., Rademaker, J. & Swings, J. 2000. Synopsis of the taxonomy of the genus *Xanthomonas*. *Phytopathology*, 90(7):677-682.
- Vincente, J.G. & Holub, E.B. 2013. *Xanthomonas campestris* pv. *campestris* (cause of black rot of crucifers) in the genomic era is still a worldwide threat to brassica crops. *Molecular Plant Pathology*, 14(1):2-18.
- Wasukira, A., Coulter, M., Al-Sowayeh, N., Thwaites, R., Paszkiewicz, K., Kubiriba, J., Smith, J., Grant, M. & Studholme, D.J. 2014. Genome sequencing of *Xanthomonas vasicola* pathovar *vasculorum* reveals variation in plasmids and genes encoding lipopolysaccharide synthesis, type-IV pilus and type-III secretion effectors. *Pathogens*, 3:211-237.
- Watson, L. & Dallwitz, M.J. 1992. The grass genera of the world. Wallingford: CAB International.
- Wren, B.W. 2000. Microbial genome analysis: Insights into virulence, host adaptation and evolution. *Nature Review Genetics*, 1:30-39.

Young, J.M., Bradbury, J.F., Gardan, L., Gvozdyak, R.I., Stead, D.E., Takikawa, Y. & Vidaver, A.K. 1991. Comment on the reinstatement of *Xanthomonas citri* (ex Hasse 1915) Gabriel et al. 1989 and *X.phaseoli* (ex Smith 1897) Gabriel et al. 1989: Indication of the need for minimal standards for the genus *Xanthomonas*. *International Journal of Systematic Bacteriology*, 41:172-177.

Young, J.M., Park, D.C., Shearman, H.M. & Fargier, E. 2008. A multilocus sequence analysis of the genus *Xanthomonas*. *Systematic and Applied Microbiology*, 31:366-377.

Young, J.M., Wilkie, J.P., Park, D.C. & Watson, D.R.W. 2010. New Zealand strains of plant pathogenic bacteria classified by multi-locus sequence analysis; proposal of *Xanthomonas dyei* sp. nov. *Plant Pathology*, 59:270-281.

APPENDIX A: MEDIA COMPOSITION

Media A: Glucose yeast extract medium with calcium carbonate (Atlas, 2010).
(For the cultivation and maintenance of *Xanthomonas* species)

Composition per litre:

Agar.....	15.0 g
CaCO ₃	7.5 g
Peptone.....	5.0 g
Yeast extract.....	5.0 g
Glucose.....	3.0 g

pH: 7.0 ± 0.2 at 25°C

Preparation of medium: Distilled water should be added to bring overall volume to 1.0 litre. Components should be mixed thoroughly and gently heated until it starts boiling. After heating, pH should be adjusted to 6.3. The mixture should then be autoclaved for 15 min at 121°C and pour into sterile Petri dishes.

Media B: Glucose yeast extract medium (DSMZ – Media 54)

Agar.....	17.0 g
CaCO ₃ (light precipitate).....	20.0 g
Yeast extract.....	10.0 g
Glucose.....	20.0 g

Add 1000.0 ml distilled water to the dry media.

APPENDIX B: SETTINGS FOR BIOINFORMATICS PROGRAMMES

B.1. CLC Genomics Workbench – Sequence trimming settings for genome assembly:

- Trim maximum amount of ambiguities (2)
- Quality trim (low quality threshold) (0.01)
- Minimum amount of nucleotide in read (75 bases)
- Discard short reads

B.2. CLC Genomics Workbench – Mapping of reads to complete genome reference sequences of *Xanthomonas oryzae* pv. *oryzae* PXO99A (NC010717) and *Xanthomonas campestris* pv. *campestris* ATCC 33913 (NC_003902):

- match score (1)
- mismatch cost (3)
- insert cost (2)
- deletion cost (2)
- insert open cost (6)
- insert extend cost (1)
- deletion open cost (6)
- deletion extend cost (1)
- no global alignment

B.3. Assembly of reads for X85 and X86 as well as X1 – 45 + GP:

- map reads back to contigs
- update contigs
- auto bubble size
- min. contig length (100)
- miss match cost (3)
- insert cost (3)
- deletion cost (3)
- auto word size
- scaffolding

B.4. Analysis preferences for construction of Neighbour-joining trees in MEGA6:

- Variance estimation method >> Bootstrap method (1 000)
- Substitutions type >> Nucleotide
- Genetic code table >> Standard

- Model / Method >> Maximum likelihood
- Substitutions to include >> Transitions and transversions
- Rates among sites >> Gamma distribution
- Patterns among lineages >> Different (Heterogeneous)

B.5. Settings for determining approximate annealing temperatures in Vector NTI 10:

Analyses >> Oligo analyses >> Thermodynamic properties

APPENDIX C: GENBANK SEQUENCES

Table 6: GenBank details of complete housekeeping gene sequences of the six housekeeping genes for *Xanthomonas* species used to extract matching gene sequences from whole genome sequence data of BLSD isolates.

Gene	GenBank gene ID	Sequence location	<i>Xanthomonas</i> strain
atpD	999311	NC_003902.1 (669581-670987)	<i>X. campestris</i> pv. <i>campestris</i> (ATCC 33913)
dnaK	3262129	NC_006834.1 (2130125-2132050)	<i>X. oryzae</i> pv. <i>oryzae</i> (KACC 10331)
	14852277	NC_020800.1 (1759434-1761359)	<i>X. axonopodis</i> (Xac29-1)
	8702525	NC_013722.1 (1981091-1983013, complement)	<i>X. albilineans</i> (GPE PC73)
	1001892	NC_003902.1 (1720789-1722717)	<i>X. campestris</i> pv. <i>campestris</i> (ATCC 33913)
fyuA	1001538	NC_003902.1 (1186078-1188294)	<i>X. campestris</i> pv. <i>campestris</i> (ATCC 33913)
recA	3265073	NC_006834.1 (3150244-3151278, complement)	<i>X. oryzae</i> pv. <i>oryzae</i> (KACC 10331)
	14852488	NC_020800.1 (2001234-2002268)	<i>X. axonopodis</i> (Xac29-1)
	8702778	NC_013722.1 (1369531-1370568)	<i>X. albilineans</i> (GPE PC73)
	998392	NC_003902.1 (2000077-2001108)	<i>X. campestris</i> pv. <i>campestris</i> (ATCC 33913)
rpoB	1002270	NC_003902.1 (2163347_2165029, complement)	<i>Xanthomonas campestris</i> pv. <i>campestris</i> (ATCC 33913)
rpoB	3265412	NC_006834.1 (3853543-3857694, complement)	<i>X. oryzae</i> pv. <i>oryzae</i> (KACC 10331)
	14851749	NC_020800.1 (1125483-1129634)	<i>X. axonopodis</i> (Xac29-1)
	8703135	NC_013722.1 (555376-559527)	<i>X. albilineans</i> (GPE PC73)
	1001453	NC_003902.1 (1042071-1046234)	<i>X. campestris</i> pv. <i>campestris</i> (ATCC 33913)

Table 7: GenBank reference sequences of all *Xanthomonas* strains included in the phylogenetic component of this study.

Strain	Strain code	GenBank accession no.
<i>X. campestris</i> pv. <i>phaseoli</i>	LMG29033	gi 992067926 gb KT585785.1
<i>X. campestris</i> pv. <i>campestris</i>	Xcc-C168	gi 869221543 gb KT203401.1
	Xcc-C278	gi 888795797 gb KT071748.1
	CPBF279	gi 746829698 gb KM272649.1
	ATCC33913	gi 21166373 gb AE008922.1
	17	gi 928198021 gb CP011946.1
	ICMP4013	gi 902825171 gb CP012146.1
	ICMP21080	gi 902821146 gb CP012145.1
<i>X. campestris</i> pv. <i>leersiae</i>	NCPPB4346	gi 519767425 gb JX682663.1
<i>X. campestris</i> pv. <i>incanae</i>	CPBF1257	gi 746829736 gb KM272668.1
<i>X. campestris</i> pv. <i>raphani</i>	CPBF1171	gi 746829728 gb KM272664.1
	756C	gi 341934791 gb CP002789.1
<i>X. campestris</i> pv. <i>vesicatoria</i>	-	gi 78033986 emb AM039952.1
<i>X. pelargonii</i>	xcp305	gi 346306032 gb JN165103.1
<i>X. oryzae</i> pv. <i>oryzae</i>	LMG5047	gi 339271395 gb HM569028.1
	IBSBF1732	gi 308156020 gb HM469994.1
	LMG12464	gi 519767397 gb JX682649.1
	KACC10331	gi 58424217 gb AE013598.1
	MAFF311018	gi 84365597 dbj AP008229.1
	PXO99A	gi 854884118 gb CP000967.2
<i>X. oryzae</i> pv. <i>oryzicola</i>	RS105	gi 856757157 gb CP011961.1
	BLS256	gi 854958603 gb CP003057.2
	LMG797	gi 519767429 gb JX682665.1
<i>X. axonopodis</i> pv. <i>citri</i>	306	gi 21240769 gb AE008923.1
<i>X. axonopodis</i> pv. <i>citrumelo</i>	F1	gi 346647687 gb CP002914.1
<i>X. axonopodis</i> pv. <i>vasculorum</i>	CFBP5823	gi 355347103 gb HQ590608.1
	REU45	gi 158904994 gb EU015262.1
<i>X. axonopodis</i> pv. <i>phaseoli</i>	LMG833	gi 748943080 gb KJ559683.1
	CFBP6996	gi 355347451 gb HQ590782.1
<i>X. axonopodis</i> pv. <i>alfalfae</i>	IBSBF2730	gi 326909318 gb JF409701.1
<i>X. citri</i> subsp. <i>malvacearum</i>	LMG11170	gi 748943026 gb KJ559656.1
<i>X. citri</i> pv. <i>citri</i>	jx-6	gi 1009330446 gb CP011827.2
<i>X. citri</i> pv. <i>citri</i>	AW13	gi 780475693 gb CP009031.1
<i>X. translucens</i>	LMG876	gi 550601436 gb KF306111.1
<i>X. translucens</i> pv. <i>translucens</i>	IBSBF581	gi 308156028 gb HM469998.1
<i>X. translucens</i> pv. <i>undulosa</i>	LMG892	339271415 gb HM569038.1
	LMG876	gi 339271413 gb HM569037.1
	Xtu4699	gi 828451590 gb CP008714.1
<i>X. albilineans</i>	LMG494	gi 550601420 gb KF306103.1
	GPEPC73	gi 283472039 emb FP565176.1

<i>X. maliensis</i>	LMG27592	gi 550601444 gb KF306115.1
<i>X. euvesicatoria</i>	IBSBF2363	gi 326909322 gb JF409703.1
<i>X. vesicatoria</i>	IBSBF1338	gi 308156032 gb HM470000.1
<i>X. arboricola</i>	CFBP8159	gi 1004618541 gb KT874411.1
<i>X. arboricola</i> pv. <i>juglandis</i>	CFBP7244	gi 825177591 gb KP669352.1
	Xaj417	gi 910827970 gb CP012251.1
<i>X. vasicola</i>	LMG736	gi 748941212 gb KJ491154.1
	LMG7416	gi 339271419 gb HM569040.1
	IBSBF865	gi 323652574 gb HM469999.2
<i>X. vasicola</i> pv. <i>vasculorum</i>	LMG900	gi 339271429 gb HM569045.1
<i>X. hortorum</i>	SUPP2600	gi 469832121 dbj AB809156.1
<i>X. hortorum</i> pv. <i>hederae</i>	IBSBF1391	gi 323652566 gb HM469991.2
<i>X. cynarae</i>	IBSBF1532	gi 326909320 gb JF409702.1
<i>X. perforans</i>	75_3	gi 355347485 gb HQ590799.1
	LB273-2	gi 315440051 gb GU322710.1
	IBSBF2370	gi 326909324 gb JF409704.1
<i>X. dyei</i> pv. <i>dysoxyl</i>	IBSBF2935	gi 326909328 gb JF409706.1
<i>X. sacchari</i>	R1	gi 743685209 gb CP010409.1
	IBSBF1313	gi 323652570 gb HM469996.2
<i>X. fuscans</i> subsp. <i>fuscans</i>	4834-R	gi 549143970 emb FO681494.1
<i>X. gardneri</i>	IBSBF2373	gi 326909326 gb JF409705.1
<i>X. theicola</i>	IBSBF1565	gi 323652572 gb HM469997.2
<i>X. hyacinthi</i>	IBSBF1390	gi 323652568 gb HM469992.2
<i>X. pisi</i>	IBSBF1356	gi 308156022 gb HM469995.1
<i>S. koreensis</i>	DSM17805	gi 449332216 gb KC209214.1
<i>S. maltophilia</i>	ICMP17033	gi 188090025 gb EU498848.1
	ISMMS2	
	ISMMS3	
	K279a	gi 190572091:942248-946402
<i>S. rhizophila</i>	DSM14405	gi 627787876 gb CP007597.1