The role of genomic islands in virulence of *Pectobacterium carotovorum* subspecies *brasiliensis* on potatoes

A thesis submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

at

Lincoln University

Preetinanda Panda

Lincoln University
2014

Abstract of a thesis submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

The role of horizontally acquired islands in virulence of *Pectobacterium carotovorum* subspecies *brasiliensis* on potatoes

Preetinanda Panda

Pectobacterium carotovorum subsp. *carotovorum* is primarily responsible for soft rotting of potato tubers, although several strains have been shown to cause blackleg of potato stems. In contrast, *P. atrosepticum* is best known as a seedborne pathogen that causes blackleg. It is also responsible for soft rotting of tubers. Given the taxonomic distance between these two species and that only some strains of *P. carotovorum* subsp. *carotovorum* are able to cause blackleg, it was hypothesised that the capacity of *P. carotovorum* subsp. *carotovorum* to invade potato stems and elicit blackleg may have evolved through independent acquisition of genomic islands (GIs). GIs are large chromosomal regions in bacteria that are acquired by horizontal gene transfer and often encode virulence factors.

To address this hypothesis, the genomes of *P. carotovorum* subsp. *carotovorum* strains associated with blackleg disease of potatoes were compared to those of non-blackleg causing strains, to identify GIs in blackleg causing strains and novel virulence factors encoded on these islands. First, the identity of a highly aggressive *P. carotovorum* subsp. *carotovorum* strain ICMP19477 and other New Zealand isolates known to cause blackleg, was re-assessed using molecular and phylogenetic assays. These assays identified the isolates as *P. carotovorum* subsp*. brasiliensis*.

Comparative genomics using the genome sequence obtained for ICMP19477 and three other *Pectobacterium* in this study, as well as the genomes of 10 soft rot erwiniae (SRE) obtained from public databases, identified a total of 69 genomic islands and 10 gene islets in the genome of ICMP19477. Many of these islands and islets harboured genes predicted to be associated with the virulence of this pathogen (e.g. genes encoding putative plant cell wall degrading enzymes, phytotoxins, secretion systems, sugar utilisation, etc.), yet only one was present exclusively in the *P. atrosepticum* and *P. carotovorum* subsp. *brasiliensis* isolates historically known to cause blackleg. This islet (*sim* gene islet) encodes a

sugar:phosphotransferase system, which is important in the pathogenicity of vascular plant pathogens such as *Erwinia amylovora*. The majority of other GIs and islets carrying putative virulence genes were either not present in the genomes of all blackleg causing strains or were present in strains that did not cause blackleg (e.g. GI PbN1_GI24, which encoded a putative non-ribosomal peptide with similarity to syringomycin).

To confirm the role of several putative virulence factors in virulence of *P*. *carotovorum* subsp. *brasiliensis* ICMP19477, knockout mutants were constructed by allelic exchange mutagenesis. Mutations in the non-ribosomal peptide synthetase cluster of *P*. *carotovorum* subsp. *brasiliensis* (on PbN1_GI24) did not result in significant differences in virulence of the wild type and the mutant in pathogenicity assays on potato. In contrast, single crossover mutants in the islets encoding the sugar:phosphotransferase system and phenolic acid decarboxylase significantly reduced virulence of the pathogen. Complementation studies still need to be conducted to prove their role in disease.

In summary, this study has provided new insights into the impact of GIs on the virulence of *P*. *carotovorum* subsp. *braisliensis* and also on the mechanisms by which SRE cause diseases. These data suggest that the accumulation of multiple virulence factors on these elements might play a larger role in virulence of *Pectobacterium* than the acquisition of a specific gene or gene cluster.

Keywords: *Pectobacterium*; blackleg; soft rot; genomic island (GI); gene islet; horizontal gene transfer; virulence factor; comparative genomics; sugar:phosphotransferase system; phenolic acid decarboxylase

Acknowledgements

First of all I would like to express my sincere thanks to my primary supervisor Dr Andrew Pitman, for providing me with the opportunity to undertake my PhD studies. You have always been there pointing (and sometime pushing!) me in the right direction, sharing your knowledge and keeping me on track. Without his continuous support, guidance, patience and enthusiasm, this thesis would have not been successfully completed.

I thank my research co-supervisor Dr Karen Armstrong for her continual support, encouragement and valuable guidance throughout the course of this PhD research. I would also like to thank Dr Mark Fiers and Dr Andrew Pitman for giving me a chance to enter into the world of Bioinformatics where I gained an enormous amount of knowledge throughout this study. I thank Dr Ian Toth and Dr Peter Fineran for their valuable inputs throughout my PhD study.

Besides my supervisors, I would also like to express my gratitude to Ashley Lu for readily assisting me with my bioinformatics queries and analyses (during nights and weekends). Thank you! I thank Dr Ruth Butler for her biometric advice and Donna Gibson for her assistance with photography for presentations and for this thesis. I also wish to acknowledge the Tertiary Education commission for the financial support for this project.

Thanks to members of labs at the Bioprotection Research Centre, Lincoln University and the New Zealand Institute of Plant & Food Research, Lincoln. I also wish to acknowledge my contemporaries, Dr Bhanupratap Vanga and Dr Subha Das for their continuous support and for the many discussions on science and life in general (including UFOs and Aliens).

Special thanks to my dearest friends, Lovepreet Singh and Unni Krishnan Gopinathan, who have shared this journey with me and always had faith that I would successfully complete my study. All this would have not been possible without your continuous support. Lastly, I would like to thank my Dad (Mr Banaj Nayan Panda) and Mum (Mrs Geetanjali Panda) and and my loving sisters (Madhu Chhanda Panda and Swarna Gandha Panda) for their unconditional,

everlasting love and support. Special thanks to my parents for giving me courage and letting me live my dream!!! Without you both, I am nothing and I dedicate this thesis to you.

Table of Contents

List of Tables

List of Figures

List of abbreviations

°C	degrees Celsius
µL	microlitre
µM	micromolar
AFLP	amplified fragment length polymorphism
AHL	N-acylhomosrine lactones
ANI	average nucleotide identity
AT	autotransporter
ATCC	American Type Culture Collection
att	attachment site
*att*B	bacterial attachment site
*att*P	phage attachment site
BBH	bidirectional best hit
bp	base pair
cDNA	complementary DNA
CDS	coding sequence
CFA	coronafacic acid
CFL	coronafacic ligase
cfu	colony forming units
Chl^R	chloramphenicol resistant
cm	centimetre
CMA	coronamic acid
COR	Coronatine
cv.	cultivar
CVP	crystal violet pectate
DKI	5-keto-4-deoxyuronate
DKII	2, 5-diketo-3-deoxygluconate
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
dpi	days post inoculation
DR	direct repeats
EPS	exopolysaccharide
Evr	*Erwinia* virulence regulator
g	gram
GI	genomic island
GSP	general secretion system
h	hour
HAI	horizontally acquired island
Hcp	hemolysin coagulated-like
HGT	horizontal gene transfer
HMM	hidden markov model
HR	hypersensitive response
hrc	hypersensitive response and conserved
hrp	hypersensitive response and
Hrp	hairpin or helper
ICE	integrative conjugative element
ICMP	International Collection of
IS	insertion sequence
Kb	kilobase pairs
KDG	2-keto-3-deoxygluconate
Kn^R	kanamycin resistant
kV	kilo volt
L	litre
LB	Luria-Bertani
LPS	lipopolysaccharide
LSD	least significant difference
M	molar
Mb	megabase pairs
MeJA	methyl jasmonate

MGE	mobile genetic element
min	minutes
mL	millilitre
MLSA	multi locus sequence analysis
mM	millimolar
mRNA	messenger RNA
NGS	next generation sequencing
Nip	necrosis-inducing protein
NRPS	non-ribosomal peptide synthetase
nt	nucleotide
OD	optical density
OHHL	N-(3-oxohexanoyl)-L-homoserine
oriT	origin of transfer
PAD	phenolic acid decarboxylase
PAI	pathogenicity island
PAMPs	pathogen-associated molecular patterns
PCR	polymerase chain reaction
PCWDE	plant cell wall degrading enzyme
PGAAP	prokaryotic genomes automatic
PTS	phosphoenolpyruvate-dependent
QS	quorum sensing
RAST	rapid annotation server
RBH	reciprocal best hit
RFLP	restriction fragment length polymorphism
R-M	restriction-modification
RNA	ribonucleic acid
rpm	revolutions per minute
rrn	rRNA operon
rRNA	ribosomal RNA
s	seconds
SDW	sterile distilled water
Sec	secretion-dependent
SIM	sucrose isomerase
SPI	Salmonella pathogenicity island
SRE	soft rot erwiniae
Str^R	streptomycin resistant
Svx	avirulence protein
T1SS	Type I secretion system
T2SS	Type II secretion system
T3SS	Type III secretion system
T3aSS	the translocation-associated T3SS
T3bSS	bacterial flagellum T3SS
T4SS	Type IV secretion system
T5SS	Type V secretion system
T6SS	Type VI secretion system
TPS	two-partner secretion system
tra	transfer gene
TRAP	ATP independent periplasmic
tRNA	transfer RNA
UV	ultraviolet
VgrG	valine glycine rich
Vic	virulence cluster
w/v	weight to volume ratio
WGS	whole genome shotgun

Chapter 1

Introduction

1.1 Potato

Potato (*Solanum tuberosum* L.), a cultivated tuber-bearing plant, is considered to be the third most important food crop after rice and wheat (Birch *et al.*, 2012). In 2010, worldwide annual production of potatoes was estimated to be about 324 million tonnes, which was grown on 19 million hectares of land (Food and Agriculture Organization, United Nations http://faostat.fao.org/, verified 26 October 2012). In New Zealand, where potato remains one of the most important export crops for both the fresh food and processed food sectors, approximately 525,000 tonnes of potatoes are cultivated on around 10,590 hectares of arable land annually. Potato production has a combined domestic and export value of 560 million NZ dollars per annum (Potatoes New Zealand, http://www.potatoesnz.co.nz/Overview/Our-Industry/Industry-profile.htm).

There are more than 200 active potato growers in New Zealand, who collectively cultivate up to 50 different varieties of potatoes. The main commercial cultivars are Russett Burbank, Innovator, Rua, Nadine, Agria, Moonlight, Desiree, Ilam Hardy and Red Rascal. Their average yields are around 40 tonnes per hectare (Potatoes, New Zealand, http://www.potatoesnz.co.nz/Overview/Our-Industry/Industry-profile.htm). Potatoes are not only harvested as a food source for fresh market and processed products, but also as seed tubers for planting of crops locally. Average seed potato yields are around 20 tonnes per hectare.

Plant diseases, weeds and insects account for an annual loss in agricultural production worldwide of 25-80% (Oerke, 2006). Of this, an estimated 22% of potatoes alone are lost to viral, bacterial and fungal diseases and to pests each year, which is equivalent to an annual loss of over 65 million tonnes (Ross, 1986; International Potato Centre, Lima, Peru http://www.cipotato.org/; Food and Agriculture Organization, United Nations http://www.fao.org/). Since, potatoes are vegetatively propagated; the salient source of

inoculum for many fungal, viral and bacterial diseases is either latently infected potato tubers or contaminated parent plants (De Boer, 2002; Hélias *et al.*, 2000). As a result, certification systems (that guarantee the quality of seed potatoes) have been established to minimize disease dissemination. In most countries, the seed potato classification schemes are based on set tolerance levels encountered during visual inspections of growing crops and harvested tubers (Elphinstone & Toth, 2007). Unfortunately, certification systems are only partially effective and work only to the extent that disease symptoms can be observed.

1.2 Potato diseases associated with pectolytic bacteria

Amongst the most important bacterial diseases of potatoes is seed piece decay or soft rotting of tubers and rotting or wilting of stems on growing potato plants, often referred to as aerial stem rot or blackleg. The primary pathogens responsible for this maceration and decay of potato are a group of bacteria known collectively as the pectolytic erwiniae or soft rot erwiniae (SRE). The SRE were named after plant pathologist, Erwin F. Smith (De Boer, 2003).

1.2.1 Tuber soft rot

Tuber soft rot (Figure 1.1) can occur in the field or in storage depending on environmental conditions. It is characterized by a wet, grainy, white or brown rot (De Boer, 2004), which is generally separated from the healthy tissue by a black margin (Agrios, 1997). The symptoms can range, however, from a slight vascular discolouration at the stolon end, which turns reddish tan, to brown to black in the presence of air. Under inadequately ventilated cool storage conditions, rotting can spread to adjoining tubers leading to massive rotting pockets in stored tuber lots (Czajkowski *et al.*, 2011).

1.2.2 Blackleg

Blackleg disease of potato plants in fields is often preceded by soft rotting of tubers during storage. Blackleg symptoms are caused by bacterial invasion of mother tubers and progress upwards, externally or internally, towards the vascular tissue as the pathogen migrates (Lumb *et al.*, 1986). The disease is generally characterized by a slimy, inky-black discoloration of

previously healthy stems, accompanied by wilting, chlorosis and upward curling of the leaves, especially under wet conditions (Figure 1.1). Blackleg disease develops soon after the plant emerges or later during the potato growing season under conditions of high moisture.

Plants grown from larger tubers are more susceptible to blackleg (del Pilar Marquez-Villavicencio *et al*., 2011). The level of tuber contamination, however, directly influences the probability of blackleg occurring in the field (Smid *et al*., 1995). In particular, the incidence of tuber soft rot and blackleg is correlated with the number of latent bacteria per tuber. The threshold level for disease development is 10^2 - 10^3 colony forming units (cfu) per mL seed tuber peel extract (Bain *et al.*, 1990; Pérombelon, 2002).

1.2.3 Aerial stem rot

Aerial stem rot (Figure 1.1) is much like blackleg, but is generally differentiated from blackleg due to its lighter brown colour (De Boer, 2004). Although the decay moves up the stem, it is a secondary soft rot of stem and petiole and does not originate from the mother tuber (Roberts *et al*., 2007). Instead, infection is initiated on succulent stems through natural openings or on wounded stems as a result of hail, windblown sand, insect feeding, or cultivation practices (Pérombelon, 2002). Initially, symptoms on the stem appear as a soft green decay that later develops quickly into brownish to inky-black spots. These symptoms then enlarge to a soft, watery rot, which progresses until the entire stem wilts and dies (Roberts *et al*., 2007; Rowe *et al*., 1995). This is a significant disease in the Netherlands, where it often cannot be distinguished from blackleg (De Boer, 2004).

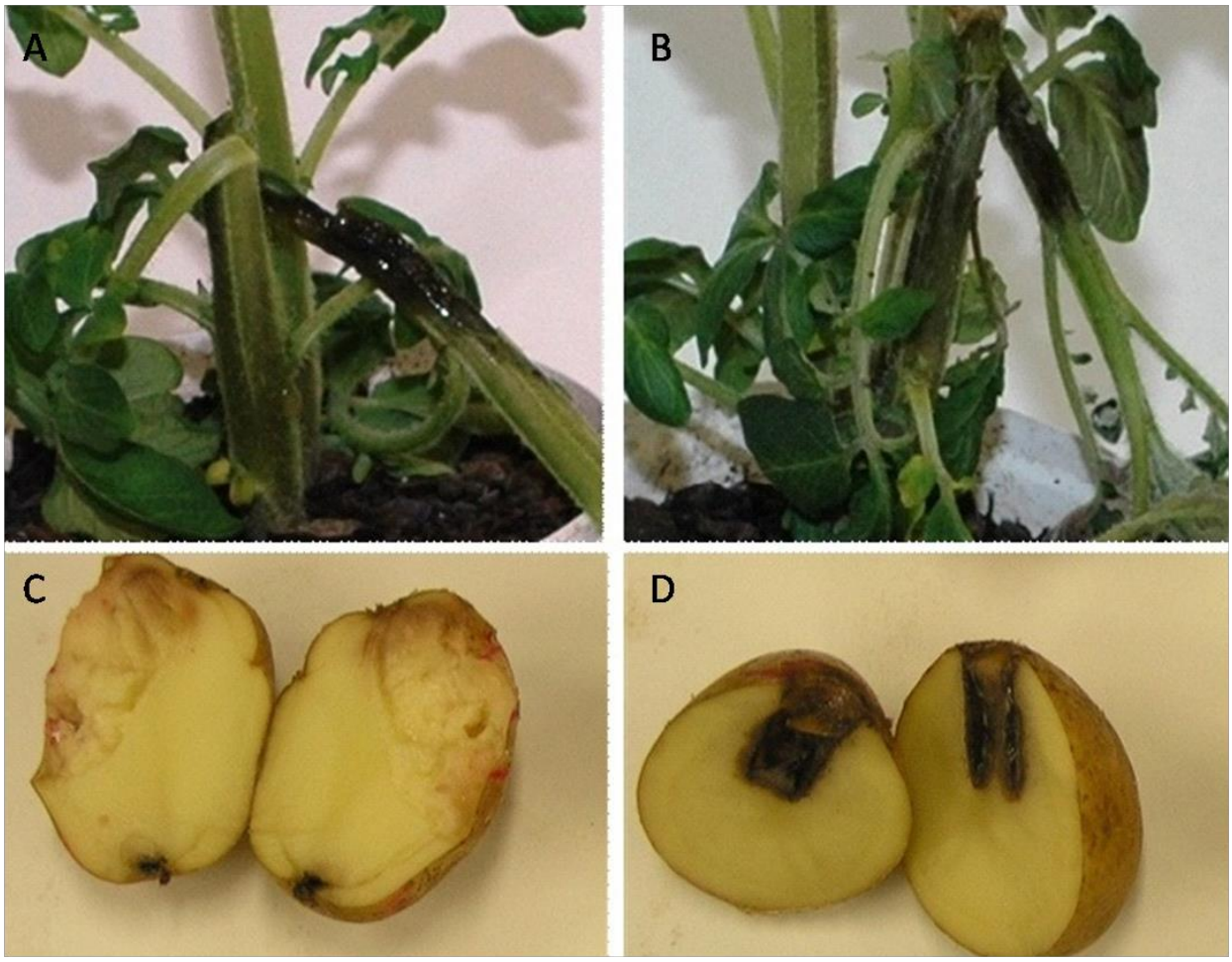


Figure 1.1 Potato diseases caused by soft rot erwiniae. A, Blackleg disease of potato stem characterized by inky black decay of the stem; B, aerial stem rot of potato stem identified by brownish, soft, mushy stem lesions; C and D, soft rot of potato tubers caused due to maceration of tuber tissue.

1.3 The SRE

1.3.1 Taxonomy and host range

The SRE belong to the *Enterobacteriaceae*, a large family of Gram-negative, rod-shaped bacteria that includes many pathogens of plants and animals. The SRE are primarily divided into two genera, *Pectobacterium* and *Dickeya*, but their taxonomic status has been in flux since they were first classified as *Erwinia* in 1901 and 1953, respectively (Burkholder *et al.*, 1953; Jones, 1901) (Figure 1.2).

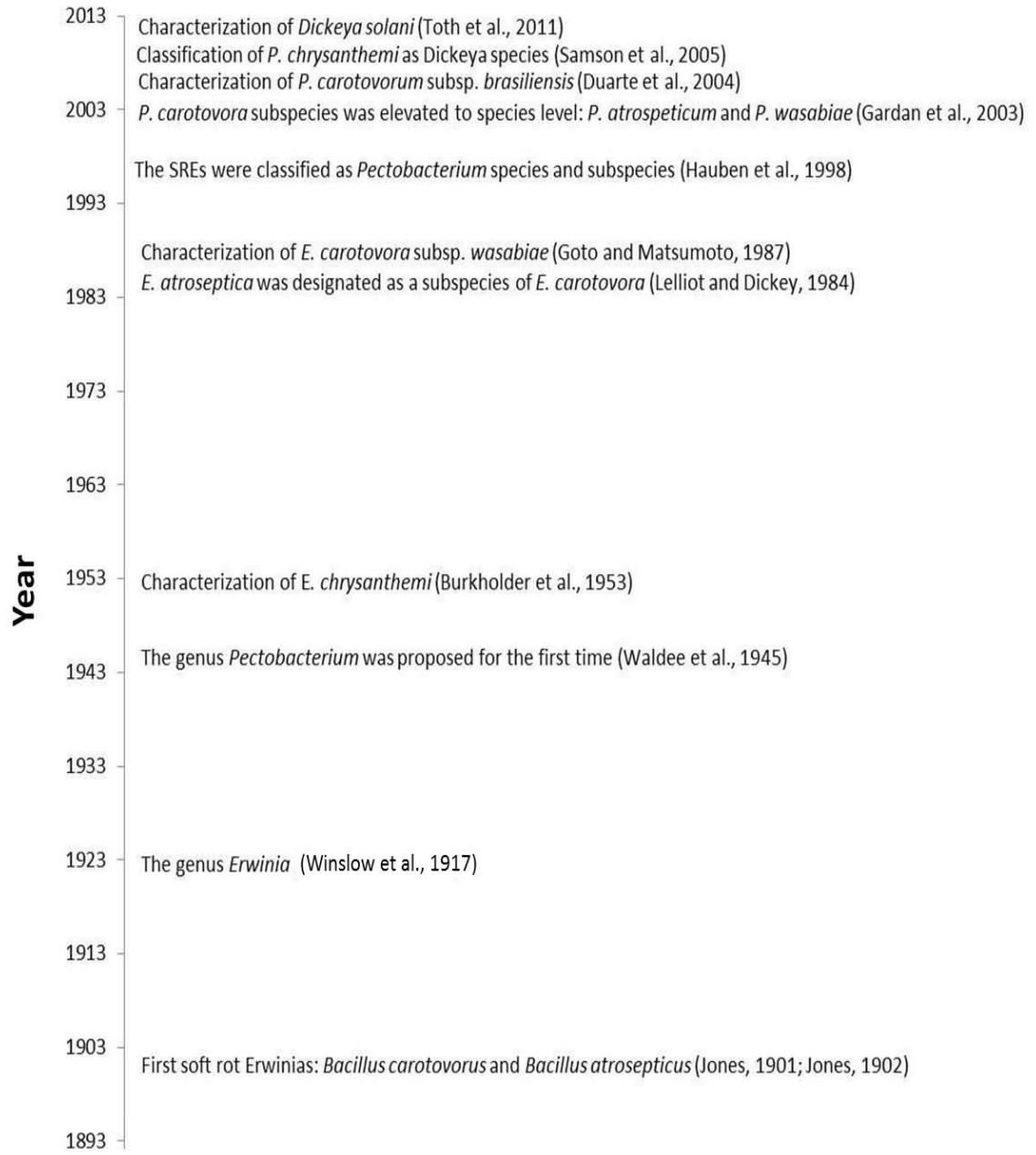


Figure 1.2 The taxonomic evolution of plant pathogenic soft rot erwiniae (SRE). The genus, species and subspecies status of the SRE has been revised several times during the years. All SRE are now divided into two genera, *Pectobacterium* and *Dickeya*.

The genus *Erwinia* was first described in 1917 to encapsulate all members of the *Enterobacteriaceae* that caused diseases on plants, irrespective of their relatedness to other members of the family (Burkholder, 1957). The genus *Erwinia* was subsequently divided into five major groups based on 16S rDNA sequences: SRE (Young *et al.*, 1996), 'true' *Erwinias*

(Hauben *et al.*, 1998b), *Pantoea* (Gavini *et al.*, 1989), *Enterobacter* (Brenner *et al.*, 1986) and *Brenneria* (Hauben *et al.*, 1998a).

In 1945, the genus *Pectobacterium* was proposed by Waldee and the species included in this group we now know are *Pectobacterium* and *Dickeya* (Skerman *et al.*, 1980). *Pectobacterium atrosepticum*, *P. carotovorum* subsp. *carotovorum* and *Dickeya* spp. are the primary SRE pathogenic to potato. These species were formerly classified as *Erwinia carotovora* subsp. *atroseptica*, *Erwinia carotovora* subsp. *carotovora*, and *Erwinia chrysanthemi* respectively, before several taxonomic re-constructions led to movement of the soft rotting '*carotovora*' species to the genus '*Pectobacterium*' and the re-classification of the *Erwinia chrysanthemi* strains as various *Dickeya* species (Hauben *et al.*, 1998a; Ma *et al.*, 2007a; Samson *et al.*, 2005). *Erwinia carotovorum* subsp. *atrosepticum* was subsequently elevated to species level and was renamed *P. atrosepticum* (Gardan *et al.*, 2003). The SRE are genetically closely related but differ in their host ranges and geographical distributions as well as the symptoms they cause on potato.

Pectobacterium atrosepticum is considered a distinct species, largely restricted to potatoes in cool and temperate regions (Pérombelon, 1992; Pérombelon & Kelman, 1987). It is seed tuber-borne and responsible both for soft rot of potato tubers and for blackleg disease. As a result of its narrow host range, *P. atrosepticum* has limited genetic diversity (De Boer, 2003), with the majority of strains typing into a single clade whether analysis is based on serology (De Boer *et al.*, 1979), biochemical tests (Sledz *et al*., 2000), phage specificity (Toth *et al*., 1999), or molecular assays (Toth *et al*., 1999; Ward *&* De Boer, 1994). Avrova *et al*. (2002) used amplified fragment length polymorphism (AFLP) fingerprinting to examine genetic diversity in *P. carotovorum* strains, confirming that *P. carotovorum* subsp. *carotovorum* is significantly more diverse than *P. atrosepticum*. Similarly, Yap *et al*. (2004) showed considerable genomic diversity in *P. carotovorum* based on comparing the rRNA genomic skeletons of multiple closely related *E. carotovora* subsp. *carotovora* genomes.

Pectobacterium carotovorum and *Dickeya* spp. are ubiquitous in nature and infect a much broader range of plants including potatoes, brussel sprouts, carrots, celery, cucumbers, bananas, tomatoes and capsicums (Avrova *et al.*, 2002; De Boer *et al.*, 1987; Kang *et al.*, 2002;

Kang *et al.*, 2003b; Ma *et al.*, 2007a; Rashid *et al.*, 2012; Samson *et al.*, 2005; Smith & Bartz, 1990). Isolates of *P*. *carotovorum* can be divided on the basis of differences in host preference and their ability to survive in different geographical environments (De Boer, 2003). These differences, in addition to separation of the SRE using DNA-DNA hybridization, numerical taxonomy, phylogenetic analysis and serology, has resulted in the recognition of several subspecies of *P*. *carotovorum*: subspecies *carotovorum*, subsp. *odoriferum*, subsp. *betavasculorum*, and more recently, subsp. *brasiliensis* (De Boer *et al.*, 2012; Gardan *et al.*, 2003; Hauben *et al.*, 1998a; Ma *et al.*, 2007a; Nabhan *et al.*, 2012b). *Pectobacterium carotovorum* subsp. *carotovorum* is generally involved in soft rotting of potato tubers, although some strains have also been shown to cause blackleg. Extensive surveys conducted in 2001 and 2005 in the Netherlands revealed the presence of strains of *P*. *c*. subsp. *carotovorum* on 20% of blackleg-diseased plants. Vaccum infiltration of potato tubers with several of the strains resulted in typical blackleg symptoms in 50% of plants in temperate climates (de Haan *et al.*, 2008). Several other studies also showed that isolates of *P*. *c*. subsp. *carotovorum*, when inoculated into tubers or stems, can cause typical blackleg symptoms in the field (Molina & Harrison, 1977; Molina & Harrison, 1980; Peltzer & Sivasithamparam, 1985).

Pectobacterium carotovorum subsp. *odoriferum* and *betavasculorum* were originally isolated from chicory and beet, respectively (Gardan *et al*., 2003; Gallois *et al*., 1992). They were subsequently shown to cause soft rotting of potato tubers (Fessehaie *et al.*, 2002; Kang *et al.*, 2002; Ma *et al.*, 2007a; Pitman *et al.*, 2010; Thomson *et al.*, 1981; Toth *et al.*, 1999), but have never been detected on naturally infected potato tubers or from infected plants in the field. *Pectobacterium carotovorum* subsp. *betavsaculorum* was recently re-classified as *P*. *betavasculorum* (Gardan *et al.*, 2003). In contrast, *P*. *carotovorum* subsp. *brasiliensis* was originally isolated from potatoes in Brazil, where it was identified as the major cause of blackleg and soft rot (Duarte *et al.*, 2004). It is also known to cause soft rot in capsicums, perennial plants and wild carrots (Ma *et al.*, 2007a; Nabhan *et al.*, 2012b). It has subsequently been detected in potato cropping systems in South Africa (van der Merwe *et al.*, 2010), Canada (De Boer, 2012), Israel (del Pilar Marquez-Villavicencio *et al.*, 2011), Zimbabwe (Ngadze *et al.*, 2012) and in the Netherlands (Leite *et al.*, 2014). About 20% of the *P*. *carotovorum* strains collected in Syria were also identified as *P*. *carotovorum* subsp.

brasiliensis (Nabhan *et al.*, 2012a). Thus, *P. carotovorum* subsp. *brasiliensis* is predicted to have a global distribution.

Dickeya is divided into eight species: *dadantii*, *zea*, *dianthicola*, *paradiasica*, *dieffenbachiae*, *chrysanthemi*, *aquatica* and *solani* (Parkinson *et al*., 2014; Samson *et al*., 2005; Toth *et al*., 2011). Members of the genera are commonly found in tropical and subtropical climates where they are reported to be the causal agents of aerial stem rot and wilt disease of potatoes. *Dickeya dianthicola* strains are better adapted to temperate conditions, however, and have recently become an economic problem on potato in Europe (Samson *et al.*, 2005).

Pectobacterium wasabiae was first identified on horseradish (Goto & Matsumoto, 1987), although it was subsequently detected in many potato growing countries in the world. These countries included New Zealand, the United States, Iran and South Africa (Baghaee-Ravari *et al*., 2011; Kim *et al*., 2009; Moleleki *et al*., 2013; Pitman *et al*., 2010). The recent detection of this pathogen on potato was in part due to the re-assessment of historic specimens from infected potato, which indicated that this bacterium had been part of the SRE complex on potatoes for some time. *Pectobacterium wasabiae* had probably eluded detection due to an inability to differentiate it from *P*. *carotovorum* subsp. *carotovorum* (Nabhan *et al.*, 2012b; Pitman *et al.*, 2010). Strains previously characterized as *P*. *carotovorum* but re-classified as *P*. *wasabiae* included the well-studied Finnish model strain SCC3193 (Nykyri *et al.*, 2012). *Pectobacterium wasabiae* causes blackleg under experimental conditions, but its capacity to cause blackleg in the field is not understood.

1.3.2 Biochemical and physiological characteristics

The SRE are primarily differentiated from other plant pathogenic enterobacteria by their ability to produce large numbers of pectolytic enzymes and their colony characteristics on crystal violet pectate (CVP) medium (Helias *et al*., 2011; Kelman & Dickey, 1988). Furthermore, all species of *Pectobacterium* and *Dickeya* are catalase positive and oxidase negative (i.e. they have the ability to produce catalase, but not the oxidase enzyme on nutrient agar). *Pectobacterium atrosepticum*, *P*. *carotovorum* subsp. *carotovorum* and *Dickeya* spp. are easily distinguished by the temperature required for their growth on artificial media (CVP) and by

characteristic biochemical properties. Growth of *P. atrosepticum* is restricted at temperatures above 35°C, whereas *P. carotovorum* subsp. *carotovorum* can grow at temperatures up to 39°C and *Dickeya* can survive at even higher temperatures (>39°C). In contrast, *P. atrosepticum* can produce acid from α-methyl glucoside and reducing substances from sucrose, whereas neither *P. carotovorum* subsp. *carotovorum* nor *Dickeya* sp. can produce acid from α-methyl glucoside or reducing substances from sucrose (De Boer, 2003; Perombelon & Kelman, 1980). *Dickeya* species can be differentiated from *P. carotovorum* subsp. *carotovorum* on the basis of indole production, phosphate activity and sensitivity to erythromycin (De Boer, 2003; Perombelon & Kelman, 1980).

Pectobacterium atrosepticum can also be distinguished from many strains of *P. carotovorum* subsp. *carotovorum* and *Dickeya* because of the absence of pili or fimbriae and non-production of bacteriocins (Itoh *et al.*, 1978; Perombelon & Kelman, 1980). Furthermore, the majority of *P. carotovorum* subsp. *carotovorum* and *Dickeya* strains are bacteriocinogenic (strains that can produce bacteriocins, toxins that kill bacteria) (Itoh *et al.*, 1978).

1.3.3 Epidemiology

Generally, SRE enter potato plants through natural openings (lenticels) or via wound sites on the tuber surface. Infection then spreads from contaminated tubers to the stem causing blackleg symptoms or to the progeny tubers causing soft rot disease (Toth *et al*., 2006). The SRE penetrate the vascular tissues and intercellular spaces of suberized or thin-walled parenchymatous tissues, where they remain latent until environmental conditions including free water, oxygen availability and temperature (Perombelon & Kelman, 1980; Pérombelon & Salmond, 1995) become conducive to cell multiplication, movement and disease development (Toth *et al*., 2006). Symptoms of wilting and desiccation are most common under dry conditions and soft rot under wet conditions (Pérombelon, 1992).

The SRE also undergo various alternative life-cycles including an epiphytic life-cycle on plants, an endophytic life-cycle in plants and a saprophytic life-cycle in ground water or soil (Perombelon & Kelman, 1980; Pérombelon & Salmond, 1995). Blackleg causing SRE survive poorly in the soil, however, and the main sources of inocula are latently infected seed tubers,

host plant parts, cull piles, and the rhizosphere of various other plants (Pérombelon, 2002; Perombelon & Kelman, 1980). The soft rot bacteria do not overwinter well in soil either, with survival restricted to one week to six months depending on environmental conditions such as soil temperature, moisture and pH. In any event, the bacteria cannot survive in the soil in a crop rotation system of 3 - 8 years (Pérombelon *et al.*, 1988). Instead, much like blackleg causing strains, non-blackleg causing SRE survive on plant residues (Elphinstone, 1987). The rhizosphere of other hosts and weeds may serve as further reservoirs for SRE during their non-pathogenic lifestyles. Additionally, these bacteria can be disseminated in the potato field by irrigation water, and via insects, nematodes, earthworms, rain, bacterial aerosols or contaminated equipment (Agrios, 1997; Elphinstone, 1987; Perombelon, 1992; Perombelon & van der Wolf, 2002). Consistent with these diverse lifestyles, the SRE genome suggests that this group of pathogens can respond to a wide range of nutrient sources and environmental cues to support their existence in various habitats (Bell *et al.*, 2004). Little is known about the alternative life-cycles of SRE.

Moisture and temperature are important determinants of disease development and determine which pectolytic species has the greatest chance of causing infection. Some species predominate in temperate, cool, moist climates. With increasing temperatures later in the growing season, other species may cause blackleg and soft rotting of potatoes.

1.3.4 Economic importance of SRE

As a variety of SRE have the ability to cause soft rot of tubers during storage and can also result in the occurrence of various field symptoms, identification of the casual organism by visual observation in the field is unreliable. Regulatory field inspections of seed potatoes in many countries, therefore, are based on recognition of a 'blackleg complex' of symptoms without discriminating between the various blackleg-causing organisms (de Haan *et al.*, 2008). The blackleg complex is regarded amongst the most problematic issues resulting from bacterial pathogens on potatoes, with soft rot and blackleg accounting for 20 - 60% of losses in potato production each year worldwide (Abo-Elyousr *et al.*, 2010; Baghaee-Ravari *et al.*, 2011; Czajkowski *et al.*, 2011; Ngadze *et al.*, 2012; Toth *et al.*, 2011). Losses in seed potato production due to the blackleg complex are second only to those associated with bacterial

wilt caused by *Ralstonia solanacearum*, and ahead of ring rot and common scab caused by *Clavibacter michiganensis* subsp. *sepedonicus* and *Streptomyces scabies*, respectively (Van der Wolf & De Boer, 2007).

1.4 Pathogenicity determinants of *Pectobacterium*

Species of *Pectobacterium* have long been considered necrotrophic pathogens, producing a vast array of plant cell wall degrading enzymes (PCWDEs) to actively kill host tissues to promote the quick release of nutrients. Yet more recently, a variety of other virulence determinants have been identified from species of *Pectobacterium*, suggesting that the pathogenicity of these bacteria requires the co-ordination of virulence factors associated with several stages in the lifecycle of the pathogen. These stages include the initial colonisation of the pathogen, the maintenance of latent populations prior to the onset of disease progression, and the necrosis of plant tissues during the disease phase. The co-ordinated production, secretion and expression of virulence factors has led to the adoption of a range of strategies for infecting potato to suppress plant defences and to derive nutrients from the host, extending from brute force to stealth.

Brute force is employed largely by necrotrophic pathogens. Such pathogens produce a plethora of PCWDEs, toxins and necrosis-inducing proteins that kill host tissues. In contrast, pathogens that rely on living plant cells for nutrient acquisition during the latter stages of infection are known as biotrophs. Pathogens such as *Pseudomonas syringae* and *Xanthomonas oryzae*, which have been considered for some time to be biotrophic, use a more stealth-like strategy that involves the suppression of plant defence responses. Defence suppression is driven primarily by the secretion of effector proteins, which enable the bacterium to obtain nutrients and multiply within living plant hosts (Collmer *et al.*, 2009; Göhre & Robatzek, 2008; Kay & Bonas, 2009).

Traditionally, pathogens have been considered either necrotrophs or biotrophs. As the interactions associated with these pathogens are increasingly dissected, however, it seems many pathogenic bacteria have both a biotrophic and a necrotrophic phase in their life cycle on the host. Thus, increasingly, pathogens can be considered to be hemibiotrophs.

Hemibiotrophs are pathogens that obtain nutrients from living tissues that die during the later stage of pathogenesis (Glazebrook, 2005).

Other than the PCWDEs, the virulence determinants identified in *Pectobacterium* include: flagella-based motility, cell membrane structures such as lipopolysaccharide (LPS), various secretion systems including the type III secretion systems (T3SS), type IV secretion systems (T4SS) and type VI secretion systems (T6SS), as well as a necrosis-inducing protein (Nip), a protein similar to an avirulence protein in *Xanthomonas* (Svx), an effector protein (DspE), coronafacic acid (CFA), a plant ferredoxin-like protein (FerE) and citrate uptake, gluconate metabolism and 3-hydroxy-2-butanone pathways (Bud) and genes encoding QseC (involved in biofilm production) and tolC (involved in phytochemical resistance) (Bell *et al.*, 2004; Corbett *et al.*, 2005; del Pilar Marquez-Villavicencio *et al.*, 2011; Evans *et al.*, 2010; Kim *et al.*, 2009; Lee *et al.*, 2013; Liu *et al.*, 2008; Mattinen *et al.*, 2004; Mole *et al.*, 2010; Sjöblom *et al.*, 2008; Toth *et al.*, 2003; Urbany & Neuhaus, 2008). In addition to these determinants, *Pectobacterium* species also produce small molecules that contribute to virulence. These include siderophores involved in iron acquisition under iron limiting conditions, pigments and small phytotoxins that affect virulence, and various other small molecules that regulate key virulence genes (Charkowski *et al.*, 2012; Franza *et al.*, 2002; Franza *et al.*, 1999). The major pathogenicity determinants and virulence factors of *Pectobacterium* are described briefly in the following sections.

1.4.1 PCWDEs

The primary virulence factors in *Pectobacterium* are the PCWDEs. These are activated during the latter stages of infection (i.e during nectrotrophy), in large quantities inside the cell, resulting in extensive maceration of plant tissues (Py *et al.*, 1998). PCWDEs include a wide range of extracellular pectinases including several pectate lyase isozymes (pectate lyase (Pel), pectin lyase (Pnl), pectin methylesterase (Pme) and polygalacturonase (Peh)) and other degradative enzymes including cellulases and xylanases (Barras *et al.*, 1994). These enzymes breakdown plant cell wall components such as pectate and cellulose present in primary and secondary cell walls and in the middle lamella between plant cells, leading to cell leakage. The resulting degradation products are used as a source of energy and nutrients (Barras *et al.*,

1994; Collmer & Keen, 1986). There is significant evidence for the role of PCWDEs in virulence of *Pectobacterium*. For example, *Escherichia coli* transformants containing *pel* genes are capable of causing soft rot while species of *Pectobacterium* with mutations in functional *peh*, *pel* and *pnl* genes exhibit reduced virulence (Flego *et al.*, 1997; Palva *et al.*, 1993).

Genome sequencing of *P*. *atrosepticum* SCRI1043 led to the identification of at least 20 putative PCWDEs in this strain (Bell *et al*., 2004). Recently, comparative genomics of *P*. *atrosepticum*, *P*. *carotovorum* subsp. *carotovorum* and *P*. *carotovorum* subsp. *brasiliensis* have revealed the presence of orthologous PCWDEs in all these species, including 10 pectate lyases, one pectin lyase, four galactouronases, two cellulases and one rhamogalacturonase (Glasner *et al.*, 2008). Glasner *et al*. (2008) also identified a set of putative protease in *Pectobacterium* species by genome analyses. Both *P*. *carotovorum* subsp. *brasiliensis* and *P*. *carotovorum* subsp. *carotovorum* also encode putative permeases, deacetylases, racemases, hydrolases and xylosidases that are absent in *P*. *atrosepticum* (Glasner *et al*., 2008). More recently, genome sequencing of *P*. *wasabiae* SCC3193 identified the presence of 39 known or putative pectinases, cellulases and proteinases in this strain, most of which were shared by all sequenced *Pectobacterium* strains (Nykyri *et al*., 2012). The presence or absence of different PCWDEs in various *Pectobacterium* strains is highlighted in Nykyri *et al*. (2012). The pectate lyases are the main pectinases involved in pathogenesis, although their number varies between species, subspecies and strain (Kotoujansky, 1987; Pérombelon, 2002; Toth *et al.*, 2003). One such example is PelE, which is important for pathogenicity of *Dickeya* spp. on potato tubers. PelB and PelC are the most important PCWDEs required for pathogenicity of *P*. *carotovorum* subsp. *carotovorum* on potato (Barras *et al.*, 1994; Payne *et al.*, 1987).

In addition to PCWDEs, *Pectobacterium* also secrete various proteases that target host proteins (Heilbronn *et al.*, 1995; Marits *et al.*, 1999). While the role of these proteases in pathogenicity is not clear, studies by Heilborne *et al*. (1995) and Martis *et al*. (1999) showed that these proteases may either provide amino acids for bacterial metabolism or digest defence related proteins from the host. Mutations in *prtW* expressed during early exponential phase *in vitro* and within 4 - 6 h *in planta*, significantly altered virulence of *P*. *carotovorum* subsp. *carotovorum* (Marits *et al.*, 2002).

1.4.2 Secretion systems

Gram-negative bacteria, including *Pectobacterium*, possess six distinct secretion systems (Type I, II, III, IV, V, and VI) that play an important role in injecting virulence factors into host cells (Toth *et al.*, 2006). For example, PCWDEs and proteases are secreted to the extracellular environment through the type I (proteases) and type II (pectinases and cellulases) secretion systems (Andro *et al.*, 1984; Condemine *et al.*, 1992; Dahler *et al.*, 1990; Delepelaire & Wandersman, 1989; Pirhonen *et al.*, 1991; Py *et al.*, 1998). These secretion systems mediate export of effector proteins, enzymes, pilus proteins, heavy metals and other virulence proteins through both the inner and outer membranes of *Pectobacterium* species (Charkowski *et al.*, 2012; Glasner *et al.*, 2008; Holeva *et al.*, 2004; Kim *et al.*, 2009; Liu *et al.*, 2008; Marits *et al.*, 1999; Mattinen *et al.*, 2004; Toth *et al.*, 2003; Tseng *et al.*, 2009).

1.4.2.1 Type I secretion system (T1SS)

The T1SS is composed of three subunits, the ABC protein, membrane fusion protein, and outer membrane protein, which transport proteins from the cytoplasm to the extracellular medium in one step (Binet *et al.*, 1997; Delepelaire, 2004). The T1SS proteins secreted via the T1SS apparatus have many functions and include toxins, adhesins, various hydrolases, and nodulation factors (Delepelaire, 2004). Metalloproteases, which contribute to *Pectobacterium* virulence by either attacking plant cell wall production or by degrading enzymes secreted by the pathogen, are also secreted through the T1SS (Hommais *et al.*, 2008). Genome sequencing of *Pectobacterium* has revealed the presence of at least three T1SSs: a T1SS involved in protease secretion (Palacios *et al.*, 2001), a T1SS similar to the *HasA* system in *Serratia marcescens* that secretes siderophores (Bell *et al.*, 2004; Létoffé *et al.*, 1994), and a T1SS regulated by a diguanylate cyclase, which secretes a multi-repeat adhesin important for *Pectobacterium* virulence (Pérez-Mendoza *et al.*, 2011).

1.4.2.2 Type II secretion system (T2SS)

The T2SS, also known as the secretion-dependent (Sec) pathway or the general secretion system (GSP) constitutes a two-step process for secretion of proteins (Cianciotto, 2005;

Thomas *et al.*, 1997). Enzyme precursors secreted by the T2SS are synthesized with N-terminal signal sequences. Cleavage of these sequences leads to translocation of the enzymes from the cytoplasm to the periplasm by the *Sec* or *Tat* pathway (Cianciotto, 2005). The enzymes are then secreted across the outer membrane through the outer membrane proteins, encoded by the *out* gene cluster (Johnson *et al.*, 2006; Thomas *et al.*, 1997).

In *Pectobacterium*, the proteins secreted by the T2SS machinery are primarily involved in degrading the physical structure of potential hosts, and include acyltransferases, chitinases, pectinases, cellulases and lipases (Cianciotto, 2005). In fact, most of the PCWDEs involved in pathogenesis (except proteases), are secreted through the T2SS (Chapon *et al.*, 2001; Thomas *et al.*, 1997). The virulence factors, Nip and Svx, are also co-expressed and secreted by the T2SS (Charkowski *et al.*, 2012; Corbett *et al.*, 2005). Nip and Svx are associated with subversion of host defences, and are required for full virulence on potato stems and tubers (Corbett *et al*., 2005). Nip shows homology to necrosis- and ethylene-inducing elicitors from fungi and oomycetes (Mattinen *et al.*, 2004; Pemberton *et al.*, 2005) and Svx shows homology to an avirulence protein in *X*. *campestris* (Corbett *et al*., 2005). Proteins secreted via the T2SS also contribute to iron homeostasis, thereby controlling its acquisition. Iron assimilation is critical for *Pectobacterium* cell function (Douet *et al.*, 2009).

1.4.2.3 T3SS

The T3SSs are complex bacterial structures that enable Gram-negative pathogens to inject bacterial effector proteins directly into the host cell cytoplasm, where they can manipulate the host to achieve suppression of host defences (Coburn *et al.*, 2007; Collmer *et al.*, 2009; Grant *et al.*, 2006). There are two types of T3SSs: the translocation-associated T3SS (T3aSS) and the bacterial flagellum T3SS (T3bSS) (Devaux *et al*., 2006). The T3aSS is encoded by *hrp* (hypersensitive response and pathogenicity) and *hrc* (hypersensitive response and conserved) genes (Bogdanove *et al.*, 1996). These genes are often in a single cluster either on the bacterial chromosome or on a plasmid, or associated with mobile genetic elements (MGEs) (Alfano & Collmer, 2004) (see Section 1.6.2). The *hrp* genes encode a pilus that translocates effector proteins from the bacterial cytosol to the host cells. Upon delivery into the host, effectors assist in manipulating host defences and promote bacteria growth by targeting various

cellular processes that would otherwise elicit an immune response (Grant *et al.*, 2006). Hrc proteins are highly conserved amongst plant and animal pathogens and are often involved in secreting T3SS substrates across the bacterial envelope (Alfano and Collmer, 2004). Effectors can also elicit a hypersensitive response (HR) in non-host plants (Lahaye & Bonas, 2001), and therefore play a key role in defining pathogen host-specificity (Alfano and Collmer, 2004). Many Gram-negative bacteria secrete effectors via the T3aSS, including *Pseudomonas*, *Xanthomonas*, *Ralstonia*, *Erwinia* and *Pantoea* (Alfano & Collmer, 2004).

Expression of T3SS in bacteria is co-ordinately regulated by networks of transcription factors in response to environmental stimuli (Yang *et al.*, 2010). In *Pectobacterium*, the T3SS genes are activated by HrpL, a member of the extracytoplasmic factor family alternative sigma factor (Schechter *et al.*, 2006; Tang *et al.*, 2006). The T3SS does not appear to be essential for pathogenicity of *Pectobacterium* as several species (such as *P*. *wasabiae*) do not appear to encode a complete T3SS (Glasner *et al.*, 2008; Kim *et al.*, 2009; Pitman *et al.*, 2010). A functional T3SS is required, however, for expression of several genes in *P*. *carotovorum* subsp. *carotovorum*, including the virulence factor gene *pelB* (Hogan *et al*., 2013). The T3SS is also suggested to be important during the early stages of infection in *Pectobacterium* (Kim *et al.*, 2011). *Pectobacterium* have a restricted set of effector proteins, limited to only a small number of harpins or helper proteins (HrpN/HrpW) and the single-known cognate effector, DspA/E (Holeva *et al*., 2004). Hairpins play an important role in facilitating effector delivery to host cells because they possess many characters that indicate that they may interact with plant cell walls and plasma membranes (Alfano and Collmer, 2004). The number of T3 effectors encoded by *Pectobacterium* is much smaller than that produced by hemibiotrophic pathogens, however (Bell *et al.*, 2004; Charkowski *et al.*, 2012; Glasner *et al.*, 2008; Kim *et al.*, 2009). The effector, DspE, acts by initiating DspE-mediated host cell death functions during *P*. *carotovorum* pathogenesis in leaf tissue and promotes disease progression and maceration of plant tissue at the necrotrophic stage of infection (Hogan *et al.*, 2013; Kim *et al.*, 2011). The helper proteins facilitate the entry of DspE into host cells (Alfano and Collmer, 2004). *Pectobacterium atrosepticum*, *P*. *carotovorum* subsp. *carotovorum* and *P*. *carotovorum* subsp. *brasiliensis* encode homologous T3SSs at the same locus (Glasner *et al*., 2008). Both *P*. *carotovorum* subsp. *carotovorum* and *P*. *carotovorum* subsp. *brasiliensis*, but not *Pectobacterium atrosepticum*, encode HrpK and elicit HR response on tobacco plants (Glasner

et al., 2008). HrpK is a T3SS-secreted protein that aids in translocation of effectors across the plant cell wall in *P*. *syringae* (Petnicki-Ocwieja *et al.*, 2005). Mutation of *hrpK* in *P*. *carotovorum* subsp. *carotovorum* did not have any effect on the ability of *P*. *carotovorum* to elicit HR. Similarly, expression of *hrpK* of *P*. *carotovorum* subsp. *carotovorum* in *P*. *atrosepticum* did not confer HR elicitation onto *P*. *atrosepticum* suggesting that the lack of HR elicitation of *P*. *atrosepticum* was not due to lack of hrpK (Glasner *et al*., 2008).

1.4.2.4 T4SS

The T4SS translocates DNA and protein substrates across the bacterial cell envelope via direct cell-to-cell contact, by contact-independent protein export, and also by DNA release and uptake from the extracellular milieu (Cascales & Christie, 2003; Christie, 2004; Dillard & Seifert, 2001; Hofreuter *et al.*, 2001). The T4SSs were initially defined on the basis of homologies between components of three different macromolecular complexes: the *Agrobacterium tumefaciens* T-DNA transfer system required for exporting oncogenic T-DNA to susceptible plant cells; the conjugal transfer (Tra) system of the conjugative IncN plasmid pKM101; and the *Bordetella pertussis* toxin exporter, Ptl (Christie, 1997; Winans *et al.*, 1996). With the identification of additional systems involved in DNA and protein translocation, however, only some are composed of a complete set of proteins homologous to the *Agrobacterium* VirB proteins. Other transfer systems, such as the RP4 plasmid, seem to be chimeras of VirB protein homologs and Tra proteins of an unrelated ancestry. Protein delivery by T4SS is essential for the virulence of many plant and animal pathogens, including *Legionella pneumophila*, *Helicobacter pylori*, *Bartonella* spp., *Coxiella burnetti*, and *A*. *tumefaciens* (Alvarez-Martinez & Christie, 2009; Juhas *et al.*, 2008; Seubert *et al.*, 2003).

T4SSs with homology to the *A*. *tumefaciens* T-DNA transfer system (*virB1* – *virB11*) are distributed sporadically within *Pectobacterium*. It is present in *P*. *atrosepticum* SCRI1043, *P*. *wasabiae* strains SCC3193 and CFBP 3304, and *P*. *carotovorum* subsp. *brasiliensis* ATCC BAA 417 but is absent from *P*. *wasabiae* WPP163, *P*. *carotovorum* subsp. *carotovorum* WPP14 and *P*. *aroidearum* PC1 (previously classified as *P*. *carotovorum* subsp. *carotovorum* PC1) (Nykyri *et al*., 2012). A mutation in *virB4*, one of the genes in this cluster, in *P*. *atrosepticum* results in a small reduction in the virulence of this pathogen (Bell *et al*., 2004). The nature of the

material translocated through the T4SS in *P. atrosepticum* remains unknown (Davidsson *et al.*, 2013), however, and its role remains cryptic (Charkowski *et al*., 2012).

Pectobacterium atrosepticum SCRI1043 also harbours a second T4SS, with greater similarity to those involved in conjugal transfer of MGEs. This T4SS is encoded by a series of genes (*pilL* to *pilV*) on the MGE HAI2 and is predicted to be the pilus used for horizontal transfer of the island between donor and recipient strains (Vanga *et al.*, 2012). This T4SS is present in other enterobacteria and plant-associated bacteria such as *R. solanacearum*, where it is involved in adhesion to the host plant and twitching motility (Bell *et al.*, 2004; Pelicic, 2008). A similar T4SS cluster was found to be encoded on a MGE, GI_42 in the recently sequenced genome of *P*. *wasabiae* SCC3193 and also in the genome of *P*. *wasabiae* WPP163 (Nykyri *et al*., 2012). However, genes belonging to this cluster were not identified in other *Pectobacterium* strains or SRE (Nykyri *et al*., 2012). The role of type IV pili in transformation is well documented in both Gram-positive, such as *Streptococcus pneumonia* and Gram-negative bacteria, such as *Haemophilus influenza* (Dubnau, 1999). The type IV pili are also shown to be involved in conjugation of the MGE, PAPI-I between *Pseudomonas* species (Carter *et al.*, 2010). In addition, some type IV pili serve as receptors for bacteriophages and are involved in a number of bacterial processes, including cell-cell interaction and biofilm formation (Strom & Lory, 1993; Wall & Kaiser, 1999).

1.4.2.5 Type V secretion system (T5SS)

The T5SSs are simple secretion systems that can be divided into two sub-groups: autotransporters (AT) and two-partner secretion (TPS) systems that function to secrete larger proteins across the outer bacterial membrane. Secretion via the T5SS occurs in two steps, similar to T2SS. Proteins are secreted in a *sec*-dependent manner from the cytosol to the periplasm, and then across the outer membrane by a single protein (AT) or by a co-operating pair of proteins (TPS) (Jacob-Dubuisson *et al.*, 2004). In contrast to T2SS, however, T5SSs only secrete one type of protein (Jacob-Dubuisson *et al*., 2004).

The ATs are involved in the secretion of virulence factors in many pathogens including pathogenic *E. coli* and other Gram-negative bacteria (Dutta *et al.*, 2002; Henderson *et al.*,

2004), while the TPS is known to be associated with bacterial virulence in *Serratia marcescens* and *Proteus mirabilis* (Jacob-Dubuisson *et al*., 2004). In *Dickeya*, TPS contributes to bacterial adherence (Rojas *et al.*, 2002) and contact-dependent growth inhibition (Aoki *et al.*, 2010). A growth inhibition system known as Rhs (rearrangement hot-spot system), is also associated with the T5SS in *Dickeya* and *Pectobacterium*. This system was shown to be transferred by horizontal gene transfer (HGT), and therefore predicted to contribute to toxin diversity and intra-strain growth competition (Poole *et al.*, 2011). The function of the T5SS-related system in virulence of *Pectobacterium* and *Dickeya* remains unknown (Toth *et al*., 2006).

1.4.2.6 T6SS

The T6SS is the most recently described of all secretion systems. First identified in *Vibrio cholera* and *P. aeruginosa* (Mougous *et al.*, 2006), the T6SS encodes a cluster of 15 - 20 genes. Hcp (Hemolysin coagulated-like protein) and VgrG (Valine glycine rich) proteins are secreted substrates of this system. Structural analysis of these proteins has shown they bear resemblance to the tail spike protein of the T4 bacteriophage, and therefore, may function to penetrate host cells and directly deliver virulence factors into the host cells (Pukatzki *et al.*, 2007; Silverman *et al.*, 2012). The genome of *P*. *atrosepticum* has revealed the presence of *hcp* genes (Liu *et al*., 2008; Mattinen *et al*., 2007). Mutations in essential T6SS components of *P*. *atrosepticum* SCRI1043 showed significant reduction in virulence in potato tubers and stems (Liu *et al*., 2008). In SRE, the T6SS is related to microbial survival *in planta* and may be directly related to virulence (Nyryki *et al*., 2012).

1.4.3 Motility and adhesion

Flagella are an important characteristic of many bacterial plant pathogens, playing a role in motility and adhesion (Inglis *et al.*, 2003; Josenhans & Suerbaum, 2002). All three species of *Pectobacterium* encode flagellar gene clusters involved in motility. Motility may be important in contamination of tuber lenticels in wet soils; non-motile mutants of *P*. *carotovorum* subsp. *carotovorum* have reduced virulence on potatoes (Hossain *et al.*, 2005; Pérombelon, 2002; Pirhonen *et al.*, 1991). Flagella based motility is considered to be necessary for full virulence of *Pectobacterium* (Antunez-Lamas *et al.*, 2009; Hossain *et al.*, 2005; Mulholland *et al.*, 1993;

Pirhonen *et al.*, 1991). Mutations in the biosynthetic machinery responsible for flagellar synthesis cause a reduction in virulence of *P*. *atrosepticum* (Mulholland *et al*., 1993).

Pectobacterium species also produce strain-specific fimbriae for adhesion to their hosts (Wallace & Pérombelon, 1992). These fimbriae can either suppress or trigger host defences (Glasner *et al*., 2008). Genome sequencing of *P*. *atrosepticum* SCRI1043 revealed that the pathogen harbours genes homologous to two-partner secretion systems, such as *hecAB*, which play a role in host attachment (Buell *et al.*, 2003; Rojas *et al.*, 2002; Van Sluys *et al.*, 2002). Detailed molecular studies on adhesion are yet to be undertaken in *Pectobacterium*.

1.4.4 LPS and exopolysaacharide (EPS)

LPS and EPS are components of the bacterial cell surface that provide the initial interface between the bacteria and its plant host (Drigues *et al.*, 1985; Nikaido & Vaara, 1985). In enterobacteria, LPS is composed of three components: lipid A and the sugar moieties, the core region, and the so-called O-specific chain (Knirel & Zdorovenko, 1997; Rietschel & Brade, 1992). The role of LPS in bacterial-plant interactions is complex and is ill-defined. In addition to providing an indispensable barrier for the bacteria against toxic plant compounds (Nikaido and Vaara, 1985), the interaction of LPS with the host may promote pathogenesis (Dazzo *et al.*, 1991; Rudolph, 2001) and/or trigger defence-related responses (Silipo *et al.*, 2005). LPSs contribute to virulence of *P*. *atrosepticum* and *Dickeya* spp., where they appear to influence motility and secretion of PCWDEs (Toth *et al*., 1999).

The EPS is generally involved in binding nutrients, and is considered essential for bacterial growth and survival under stressful environmental conditions such as desiccation and UV damage (Condemine *et al.*, 1999; Lindow & Brandl, 2003). The EPS also plays a role in suppressing host defence responses (Malnoy *et al.*, 2005). Little is known about the impact EPS has on pathogenicity of *Pectobacterium*, but it is required for full virulence of *Dickeya* (Condemine *et al*., 1999). It also contributes to symptom development by the related species *E*. *amylovora* (Malnoy *et al*., 2005), and in occlusion of *R. solanacearum* in the xylem vessels of susceptible plants that results in wilting (Wallis *et al*., 1975). In *P*. *syringae*, EPS contributes

to virulence and epiphytic fitness by facilitating colonization and/or dissemination of the bacteria *in planta* (Yu *et al.*, 1999).

1.4.5 Iron Acquisition

Iron is a micronutrient, which acts as an important molecular and cellular signal that is required for basic biochemical functions in almost all organisms (Expert, 1999). In plants, iron is stored as non-toxic ferritins and is found in a scarce insoluble form (Fe^{3+}) in most organic tissues (Expert, 1999). As a result, the availability of iron maybe one of the limiting factors for the growth of pathogenic bacteria.

To successfully compete for Fe^{3+} in their environments and cause disease, many bacteria produce high affinity iron-chelating molecules known as siderophores (Expert, 1999). Species of *Pectobacterium* are known to produce several siderophores, most of which contain one of the two conserved iron-chelating groups, catechol or hydroxamate (Expert, 1999). However, iron acquisition mechanisms of *Pectobacterium* have not yet been studied extensively. Under iron limiting conditions, the closely related soft rot pathogen *D*. *dadantii* synthesizes two siderophores, chrysobactin and achromobactin, which are produced in a sequential manner in culture supernatants of bacterial cells (Franza *et al.*, 2005). Mutants impaired in the production of these siderophores are less virulent, suggesting iron acquisition as an important virulence determinant in *D*. *dadantii* (Expert, 1999).

Although *P*. *carotovorum* subsp. *carotovorum* strains produce chrysobactin and aerobactin siderophores, they do not require the production of these siderophores for successful infection (Ishimaru & Loper, 1992). This may be due to the rapid maceration of host tissues by these strains, which provides ample iron for bacterial metabolism. Alternatively, other siderophores may be involved in iron scavenging. For example, genome sequencing of *P*. *atrosepticum* SCRI1043 revealed the presence of genes for achromobactin uptake and transport, and enterobactin synthesis, which may be involved in iron acquisition (Bell *et al*., 2004).

1.4.6 Phytotoxins

Toxins are major components of pathogenesis in both plant and animal pathogens. The genome of *P*. *atrosepticum* SCRI1043 includes two biosynthetic clusters whose genes exhibit homology to genes of biosynthetic clusters encoding known phytotoxins (Bell *et al*., 2004). The first cluster harbours genes that encode non-ribosomal peptide synthetases (NRPS) with similarity to synthetases in *P*. *syringae*. These synthetases are involved in production of syringomycin, which induces necrosis in plant tissues by creating pores in the host's plasma membrane, leading to nutrient and ion leakage and consequently to cytolysis (Bell *et al.*, 2004; Bender *et al.*, 1999).

The second cluster has strong similarity with the biosynthetic cluster that encodes coronafacic acid (CFA). Coronatine (COR), a type I polyketide (Bender *et al*., 1999), is formed by the conjugation of CFA to coronamic acid (CMA), synthesized by *cma* genes. Conjugation is mediated by coronafacic ligase (*cfl*) (Bender *et al*., 1999). In *P*. *syringae*, COR mimics the action of methyl jasmonate (MeJA), a plant signalling molecule involved in the stress response associated with wounding and pathogen attack (Brooks *et al.*, 2005). It has also been shown that COR induces thickening of cell walls and shrinkage of chloroplasts (Palmer & Bender, 1995). *Pectobacterium atrosepticum* SCRI1043 also has a *cfl* homologue but lacks the *cma* genes, suggesting that a derivative of COR is synthesized by *P*. *atrosepticum* (Bell *et al*., 2004). Mutations in *cfa*6 and *cfa*7 genes, encoding the Cfa polyketide, show reduced blackleg symptoms on potato stems demonstrating that CFA is an important virulence factor in *P*. *atrosepticum* (Bell *et al*., 2004). *Pectobacterium carotovorum* subsp. *carotovorum* strains containing *cfa* and *cfl* genes have also been isolated from potato plants (Slawiak & Lojkowska, 2009). The role of Cfa in blackleg disease by *P*. *carotovorum* subsp. *carotovorum*, however, remains unclear.

Dickeya are also known to produce two phytotoxins, zeamine and zeamine II, both encoded by large polyketide synthetases. These phytotoxins act as potent antibiotics against both Gram-positive and Gram-negative bacteria pathogens and are involved in the inhibition of rice seed germination (Cheng *et al.*, 2013; Wu *et al.*, 2010; Zhou *et al.*, 2011).

1.4.7 Bacteriocins

Bacteriocins are bactericidal, extracellular toxins, produced by both Gram-positive and Gram-negative bacteria (Daw & Falkiner, 1996; Jack *et al.*, 1995) that kill closely related bacteria. Specific target receptors encoded on the bacterial membrane recognize the susceptible host cell and introduce lethality by expressing a cognate immune protein. *Pectobacterium* produces two types of bacteriocins: high molecular weight bacteriocin carotovoricin and low molecular weight bacteriocin carocin (Chuang *et al.*, 2007; Nguyen *et al.*, 1999). It is known that different types of carotovoricins have different host specificities due to their different tail fibre proteins (Nguyen *et al*., 1999). Carosin 2 exhibits ribonuclease activity and kills indicator cells by exhausting the supply of different kinds of RNA, leading to the inactivation of protein biosynthesis (Chan *et al.*, 2011).

The bacteriocins, pectocin M1 and M2, which exhibit cytotoxic activity depending on iron availability, have also been reported in *Pectobacterium* species (Grinter *et al.*, 2012). Furthermore, some strains of *Pectobacterium* produce the carbapenem antibiotic 1-carbapenem-2-em-3-carboxylic acid (*Car*), which is an indication that the *Pectobacterium* species is in competition with potato endophytes and secondary invaders for resources (Glasner *et al.*, 2008; Parker *et al.*, 1982). To date, the *car* genes have only been detected in *P*. *carotovorum* (Holden *et al.,* 1998). No bacteriocins have been identified in *P*. *atrosepticum* (Glasner *et al*., 2008).

1.5 Regulation of virulence

1.5.1 Regulation of PCWDE production via KdgR

PCWDEs are released into the host cells during the crucial stages of disease development, a process made possible by the complex regulatory and secretory networks in the pathogen (Toth *et al*., 2003). These networks are stimulated by several conditions including availability of oxygen and nitrogen, suitable temperature conditions, osmolarity, iron deprivation, growth of the bacteria, catabolite repression, plant degradation intermediates and exudates, and

DNA-destructive agents (Cuie *et al*., 2001; Toth *et al*., 2003). The regulatory networks can work on either one (targeted regulation) or several (global regulation) systems that are controlled by positive or negative feedback mechanisms in the pathogen to accelerate higher levels of exozyme production within the host plant (Collmer & Bateman, 1981; Hugouvieux-Cotte-Pattat & Robert-Baudouy, 1987).

The breakdown intermediates resulting from the degradation of pectins can be in the form of 5-keto-4-deoxyuronate (DKI), 2, 5-diketo-3-deoxygluconate (DKII) and 2-keto-3-deoxygluconate (KDG). These breakdown products interact with a transcriptional repressor, KdgR (Chatterjee *et al.*, 1985; Condemine *et al.*, 1986). In uninfected plants, the repressor binds to a 'KdgR box', which constitutes an operator for different genes that control production of pectate lyases, exozymes such as cellulases and proteases, and genes associated with type II and type III secretion systems implicated in pectinolysis (Condemine *et al.*, 1992; Condemine & Robert-Baudouy, 1987; Hugouvieux-Cotte-Pattat & Robert-Baudouy, 1989; Nasser *et al.*, 1994). During infection and as disease progresses, the breakdown intermediates activate the repressor, causing it to detach from the binding site. This results in the induction/de-repression of the pathogenicity factors. The coordinated production of enzymes and other pathogenicity determinants that occurs during precise stages in the infection process leads to the development of disease symptoms and the down-regulation of host plant defences (Hugouvieux-Cotte-Pattat *et al.*, 1996; Thomson *et al.*, 1999).

1.5.2 Regulation of virulence via quorum sensing

The production of PCWDEs and other virulence factors in *Pectobacterium* is also controlled through a mechanism known as quorum sensing (QS), which ensures their expression only occurs when the bacterial population is large enough to overwhelm the plant response (Jones *et al.*, 1993; Liu *et al.*, 2008; Pirhonen *et al.*, 1993). This population density-dependent regulation is thought to save energy, which may subsequently be required for functions in disease development and dissemination of the pathogen to other hosts. Quorum sensing may also prevent premature activation of plant defences as the cell wall fragments and exudates from the action of PCWDEs trigger defence responses in the host plant (Palva *et al.*, 1993; Salmond *et al.*, 1995).

An increase in the population size depends on the production and secretion of freely diffusible molecules known as autoinducers. In high concentrations, these autoinducers promote the expression of genes connected with virulence, stimulate conjugation, and induce the production of antimicrobial substances (Newton & Fray, 2004; Whitehead *et al.*, 2002). QS is activated by three transcriptional stimulators in *Pectobacterium*, which respond to small diffusible autoinducer molecules called N-acylhomosrine lactones (AHL), encoded by a synthase, *ExpI* (also known as *CarI*, *AhlI*, *HslI*, depending on the particular *Pectobacterium* strain) (Barnard & Salmond, 2007; Cubitt *et al.*, 2013; Fuqua *et al.*, 1994). To ensure successful infection, each species produces a unique AHL signalling molecule (Chatterjee *et al.*, 2005; Whitehead *et al.*, 2002). Once the AHL signalling molecules reach a threshold (at a cell density of about 10^6 cells per mL), three homologues of LuxR (CarR, ExpR/ExpR1 and ExpR2/VirR) are activated (Barnard and Salmond, 2007). The LuxR-type proteins are transcription factors that modulate transcription of their target genes, upon binding with the AHL molecule (Zhang *et al.*, 2002; Zhu & Winans, 2001). In *Pectobacterium*, VirR plays a central role in the repression of QS-regulated virulence factors (Burr *et al.*, 2006). The *expR* gene activates the transcription of the global repressor protein, RsmA (regulator of secondary metabolism, see below) in the absence of AHL, and is also involved in the production of exoenzymes (Cui *et al.*, 2005). The *carR* gene, unlike *virR* and *expR,* positively regulates the production of antibiotic carbapenem in carbapenem producing strains of *P*. *carotovorum* subsp. *carotovorum* (Coulthurst *et al.*, 2005; McGowan *et al.*, 1995). Disruption of *carI* and/or *carR* abolishes carbapenem production in *P*. *carotovorum* subsp. *carotovorum* (McGowan *et al*., 2005). *Pectobacterium carotovorum* subsp. *carotovorum* mutants in *expI* do not produce AHL or PCWDEs, resulting in non-pathogenic strains (Pirhonen *et al*., 1993). Inactivation of *expR*, however, does not affect AHL synthesis or extracellular pectinolytic enzyme production in *P*. *carotovorum* subsp. *carotovorum* SCC3193 (later re-classified as *P*. *wasabiae* by Nykyri *et al*., 2012) (Andersson *et al.*, 2000). Transcriptomic analysis of an *expI* mutant of *P*. *atrosepticum* SCRI1043 revealed approximately 26% of all *P*. *atrosepticum* genes (including genes and regulators of T6SS and other putative regulators demonstrated to control pathogenicity) to be QS-regulated or under QS regulation (Liu *et al*., 2008). Interestingly, genes of T1SS, T2SS and T3SS were found to be down-regulated in the *expI* mutant (Liu *et al*., 2008). This finding suggests that QS regulation of co-ordinated physical (PCWDE) and stealth (T3SS) attack is necessary for successful infection and disease development in *Pectobacterium* species. *Pectobacterium carotovorum*

subsp. *carotovorum* strains lacking the QS system (*expI expR* and *virR* triple mutant) were still able to grow and macerate plant tissue similarly to wild type cells under laboratory conditions (Põllumaa *et al.*, 2012). This indicates that the QS system is biologically relevant in the natural habitat, where densities and the composition of bacterial populations fluctuate in response to environmental cues (Põllumaa *et al.*, 2012).

The Rsm regulatory mechanism is inter-connected with the QS of PCWDEs and involves a small regulatory RNA gene, *rsmB*, in addition to RsmA (Chatterjee *et al.*, 2005; Cui *et al.*, 2005). RsmA binds to specific mRNA molecules of PCWDE genes preventing their translation and promoting their degradation by RNAses (Bejerano-Sagie & Xavier, 2007). RmsA functions as a repressor of virulence genes in unfavourable conditions (Charkowski *et al*., 2012). RsmB regulates gene expression by binding to RsmA and impeding its activity, in response to environmental stimuli, such as compounds from degraded plant tissue, as a signal to start the production of PCWDEs (Charkowski *et al*., 2012). In addition to PCWDEs, it has been recently shown that RsmA likely also regulates several other genes, including the T3SS, T6SS and genes used in butanediol fermentation (Kõiv *et al.*, 2013). Mutations in *rsmB* result in an avirulent phenotype due to the inhibited repression of virulence factors controlled by RsmA, suggesting both RsmA and *rsmB* are major virulence regulators in *Pectobacterium* (Cui *et al.*, 1995; Liu *et al.*, 1998). Further information on QS and other virulence regulators of *Pectobacterium* are detailed in the review by Põllumaa *et al*. (2012).

1.6 Evolution of virulence

Comparative analyses of bacterial, archaeal and eukaryotic genomes has indicated that a significant fraction of the genes in prokaryotic genomes have been acquired from other sources, including other prokaryotes, viruses, or even eukaryotes (Ochman *et al.*, 2000). This process is called HGT (Koonin *et al.*, 2001). In bacterial pathogens, virulence factors are often acquired through HGT leading to greater aggressiveness, the avoidance of host defences or even host switches (Keen, 2012).

1.6.1 Processes involved in HGT

HGT can occur by one of three mechanisms: transformation, conjugation or transduction.

1.6.1.1 Transformation

Transformation is the process by which bacteria uptake naked DNA from the environment (Griffith, 1928). This transfer method occurs naturally within various taxa including bacteria and archea (Lorenz & Wackernagel, 1994). Any cell that is able to uptake DNA is considered 'competent'; the induction of competency of a cell is generally related to the physiological stage of the cells and/or related to accumulation of an environmental competence factor (Lorenz and Wackernagel, 1994). The presence of large amounts of nutrients in growth media often hinders the development of competency in a bacterial cell (Lorenz & Wackernagel, 1994). The process of transformation starts with double stranded DNA binding to sites on the cell surface, which is then translocated in single strand form into the cell. Once in the cell, host-encoded recombination results in integration of the DNA into the genome of the bacterial cell. The proteins involved in this process are mainly related to the T2SSs and type IV pili (Chen & Dubnau, 2004).

1.6.1.2 Conjugation

During conjugation, DNA is transferred from a donor cell to a recipient cell via a sexual pilus (Lederberg & Tatum, 1946). Conjugative elements are divided into two types: (i) self-transmissible elements, those that encode the conjugation machinery, and (ii) mobilizable elements, those that depend on externally encoded conjugation systems. Circular, extrachromosomal elements known as plasmids and transposons are often transmitted by conjugation (Torrence & Isaacson, 2008). For plasmids, the conjugation process starts with the extension of the sex pilus from the donor cell to the recipient cell, which has recently been shown to occur at considerable distances (Babić *et al.*, 2008). The substrate, typically single stranded DNA from a replicating rolling circle, is then transferred by a type IV-like secretion system into the recipient cell (Christie, 2001). Because single stranded DNA is transferred, a

new double stranded plasmid copy is replicated in the recipient cell and a copy is retained in the donor cell through replication of the template plasmid (De La Cruz *et al.*, 2010).

1.6.1.3 Transduction

Transduction is the process by which DNA is moved by a virus that infects prokaryotic cells, known as a bacteriophage. It usually occurs between closely related species (Kokjohn, 1989), as the host range of the bacteriophage is limited by the specificity of the interaction between the bacteriophage and bacterial receptor sites. Transduction can either be generalized, specialized, or abortive. Generalized transduction takes place when random host DNA fragments become mistakenly packaged into a virulent phage particle, during the lytic cycle (the production of phage progeny) (Lwoff, 1953). Specialized transduction occurs when the host DNA flanking an integrated phage is replicated during phage induction and becomes integrated into the phage particle. In abortive transduction, host chromosomal sequences picked up by the bacteriophage are injected into the recipient and expressed transiently before being diluted out by cell division. Bacteriophages are used in genetic studies of bacteria as naturally occurring 'syringes' to pick-up, carry and inject DNA into new host (Mazodier & Davies, 1991). They may also be used to mobilise other linear DNA fragments or plasmids (Mazodier & Davies, 1991).

The following sections briefly outline the most important agents of HGT that contribute to virulence of pathogenic bacteria.

1.6.2 MGEs

MGEs are regions of DNA that are able to move throughout the genome of a single organism or between organisms via transformation, conjugation or transduction (Frost *et al.*, 2005). Movement of MGEs can result in the transfer, rearrangement, duplication, deletion or tandem accretions of genes harboured by these elements. Such genetic changes can lead to dramatic evolution of the host bacterium.

Self-replicating MGEs include plasmids, which are transferred to other prokaryotic cells through conjugation, and prophages that move through transduction (Frost *et al*., 2005). Non-replicative MGEs include transposons, integrons, genomic islands (GIs), and insertion sequence (IS) elements. Non-replicative elements all share three common hurdles for their proliferation: (i) the genetic element should excise from the host genome to form either an RNA or DNA molecule, (ii) the genetic element must be transmitted between organisms via HGT or within an organism and be ready for integration as a DNA molecule and, (iii) the element must be integrated into a replicon in a new location. Plasmids and bacteriophage must also be transmitted between cells, but only some integrate into the bacterial chromosome and therefore require mechanisms for chromosomal excision.

MGEs code for a variety of traits that can affect the fitness of both the host bacterium and a host's neighbours (Rankin *et al.*, 2010). For example, MGEs encode factors associated with symbiosis, metabolism, or antibiotic resistance. MGEs also contribute to bacterial pathogenesis. Of particular importance, the majority of bacterial toxins that cause toxin-mediated diseases (toxinoses) such as those responsible for diphtheria, anthrax, tetanus, botulism, cholera and toxic shock are encoded by MGEs (Novick, 2003). The impact of MGEs on neighbours can also be positive as the proteins encoded on these elements can have a beneficial effect on the degrading enzymes produced by the host's neighbours (Livermore, 1995). Nevertheless, they can also be detrimental, or even toxic, encoding products such as bacteriocins (Brown *et al.*, 2006; Dykes & Hastings, 1997; Riley & Wertz, 2002; van der Ploeg, 2005).

1.6.2.1 Plasmids

A plasmid is a collection of functional genetic modules, which is smaller than the cellular chromosome and usually does not contain genes required for essential cellular functions (Frost *et al*., 2005). Plasmids carry a diverse range of antimicrobial and biocide resistance genes and can carry toxin genes and virulence factors (Malachowa *et al*., 2010; Yamaguchi *et al*., 2001). Plasmids carry their own origin of replication and are maintained and propagated independently of the bacterial chromosome within bacterial cells (Mazodier & Davies, 1991).

A variety of plant and animal pathogens harbour plasmids contributing to virulence. Indeed, pathogenic animal pathogens such as *Yersinia*, *Shigella* and *Staphylococcus* species are considered phylogenetically separated from avirulent forms by the presence (or acquisition) of virulence-related genes encoded by large plasmids (Gemski *et al.*, 1980; McCarthy & Lindsay, 2012; Sasakawa *et al.*, 1988). In enterobacterial animal pathogens, these large 'virulence' plasmids usually encode T3SS proteins (Groisman & Ochman, 1996). In contrast to their enterobacterial plant pathogenic counterparts, T3SS proteins in enterobacterial phytopathogens appear to have been integrated into the bacterial genomes to form pathogenicity islands (Hacker *et al.*, 1997; Oh *et al.*, 2005; Toth *et al.*, 2006). In *R. solanacearum*, more than half of the T3SS effectors are found on a megaplasmid (Salanoubat *et al.*, 2002).

1.6.2.2 Prophage and integrons

Prophages are the integrated form of temperate bacteriophages, which account for a large proportion of variation seen between bacterial strains (Ohnishi *et al.*, 2001). In fact, they can represent as much as 10 - 20% of the genes in the bacterial genome (Casjens, 2000, 2003). Prophage regions usually carry genes associated with their movement, but they also carry genes encoding virulence factors. For example, the cholera toxin, shiga toxin and various staphylococcal toxins are all encoded by phages (Boyd & Brüssow, 2002; Wagner & Waldor, 2002; Waldor & Friedman, 2005). The role of prophage in disease symptoms has also been demonstrated in several plant pathogenic species including *Pectobacterium*, *Pseudomonas*, *Ralstonia*, and *Streptomyces* (Evans *et al.*, 2010; Fujiwara *et al.*, 2008; McKenna *et al.*, 2002; Varani *et al.*, 2013; Winstanley *et al.*, 2009).

Integrons are genetic elements that acquire and rearrange open reading frames embedded in gene cassette units and convert them to functional genes by ensuring their correct expression (Cambray *et al.*, 2010). They play an important role in the capture and spread of antibiotic resistance genes in Gram-negative bacteria by expressing multiple resistance phenotypes in synergy with transposons (Cambray *et al.*, 2010; Stokes & Hall, 1989). The impact of integrons related to bacterial virulence is less clear. Virulence factors such as heat-stable toxin (*sto*) and mannose-fucose-resistant hemagglutinin (*mrhA*), however, are

associated with *Vibrio chlorae* repeated sequence (VCR) islands, which are integron-like structures (Jackson *et al.*, 2011; Mazel *et al.*, 1998). T3SS effector genes, *avrPpiA1* and *avrPpiA2*, have also been identified in integron-like elements in phytopathogenic *P. syringae* (Arnold *et al*., 2001).

1.6.2.3 Transposons and IS elements

Transposons are DNA elements ranging from 100 - 65,000 bps that proliferate in the host genome and move from one location of the genome to another (Pray & Zhaurova, 2008). Transposons are divided into two groups. The first group requires transposition catalyzing enzymes called transposases that mediate autonomous transposition. The second group does not require such enzymes to transpose (non-autonomous transposition) but rely on counterparts that undertake autonomous transposition to transpose. Transposons are important elements for the spread of virulence factors among bacterial species as their mobilization can promote gene inactivation, modulate gene expression or induce illegitimate recombination (Muñoz-López & García-Pérez, 2010).

IS elements are similar to autonomous transposons, in that they encode a transposase, but they do not encode any genes contributing to the phenotype of the host and are typically much smaller (Adhya & Shapiro, 1969; Shapiro, 1969; Shapiro & Adhya, 1969). These have a simple genetic organization, and are capable of inserting at multiple sites in a target molecule (Mahillon & Chandler, 1998). IS elements are generally associated with 2 to 14-bp direct repeat sequences (DRs) of the target DNA flanking the IS. IS elements play an important role in disruption of virulence genes (Jackson *et al*., 2011). In bacterial plant pathogens, IS elements can disrupt virulence genes such as T3SS effector genes. Since, T3SS genes are generally organized in an operon, disruption of a single coding sequence may result in inactivation of effectors downstream. An IS element in *P. syringae* pv. *tomato*, the causative agent of bacterial speck disease of tomato, inserted into the T3SS effector gene *hopAG1*. Insertion into *hopAG1* interfered with expression of the genes *hopAH1* and *hopAI1* located downstream of *hopAG1* in the same operon (Greenberg & Vinatzer, 2003; Jackson *et al.*, 2011).

1.6.2.4 GIs and integrative and conjugative elements (ICEs)

GIs, sometimes referred to as horizontally acquired islands (HAIs), are large discrete segments of DNA (>10 Kb) that are acquired by HGT. They are generally integrated into the chromosome of the host bacterium, associated with atypical G+C content and delineated by direct repeats at either end of the island (Schmidt & Hensel, 2004). GIs are usually integrated at transfer RNA (tRNA) sites and carry IS elements, prophages, ICEs, or conjugative transposons (Buchrieser *et al.*, 1998). Based on the gene content of the islands, GIs are described as pathogenicity, symbiosis, metabolic, fitness or resistance islands (Figure 1.3) (Dobrindt *et al.*, 2004; Schmidt & Hensel, 2004).

GIs can generally excise from their chromosomal location, self-transfer and reintegrate into the chromosome of a new host cell (Figure 1.4). Transfer involves the excision of the GI from the host chromosome in a site-specific manner to form an extrachromosomal circular intermediate, followed by integration of the circular intermediate into the recipient chromosome via conjugative transfer, transformation or transduction (Lovell *et al.*, 2011; Qiu *et al.*, 2006).

Figure 1.3 General features of genomic islands (GIs). GIs are large segments of DNA characterized by atypical G+C% content, often flanked by direct repeats and integrases, and generally integrated at a tRNA site in a bacterial genome. GIs can be classified into pathogenicity islands, symbiosis islands, metabolic islands, resistance islands and fitness islands based on the gene functions encoded on the islands (Juhas *et al*., 2009).

GIs generally harbour mobility genes such as integrases, phage-related genes or transposases, which are required for their integration into the chromosome, their excision from the chromosome and/or transmission between cells (Dobrindt *et al*., 2004; Schmindt & Hensel, 2004). Integration and excision are generally facilitated by integrases or recombinases (*int*) (Burrus *et al.*, 2002; Burrus & Waldor, 2003; Ravatn *et al.*, 1998). Integrases catalyse site-specific recombination between short identical DNA sequences, allowing integration of a circular DNA element through a single-crossover event (Ramsay *et al.*, 2006). The integrases simultaneously bind short 'core' sequences at the site of recombination and flanking 'arm' sequences termed attachment sites (*att*). The core sequences within the attachment site of

the circular element are known as *att*P, the integration site on the host chromosome as *att*B, and the direct repeats flanking the integrated element/chromosome junctions are called *att*L and *att*R (Groth & Calos, 2004; Ramsay *et al.*, 2006). The excision function of *int* is stimulated by a class of proteins collectively known as excisionases or recombination directionality factor (Lewis & Hatfull, 2001) that bind and bend DNA within the attachment sites to promote excessive recombination (Sam *et al*., 2004).

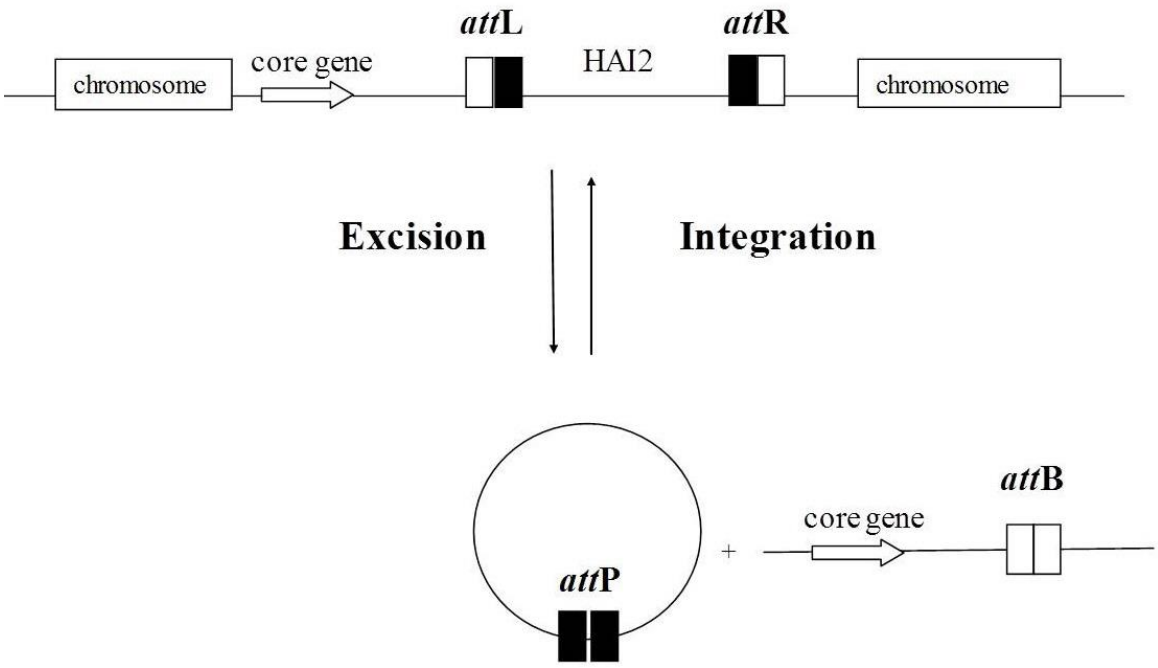


Figure 1.4 Excision and integration of genomic islands (GIs). Chromosomally located GIs are delineated by direct repeats flanking the island (*att*L and *att*R). Excision of the GI from the genome leads to the formation of a circular intermediate, resulting in regeneration of the *att*P site (phage attachment site). Excision also results in reconstitution of the GI integration site (the bacterial attachment site, *att*B) on the chromosome.

Although a number of GIs have been shown to be mobile, many GIs possess IS elements, indicating a transposition-like mechanism of transfer of at least portions of the island (Schmidt & Hensel, 2004). Similarly, many islands are considered to be remnants of integrated plasmids, where presumed loss of the associated mobility genes through reductive evolution has resulted in the anchoring of these islands in the genome of their host bacterium (Dobrindt *et al*., 2004; Hacker & Kaper, 2000; Schmidt & Hensel, 2004). One such example is the

permanently anchored 43 Kb GI, SGI1, which encodes multi-drug resistance in *Salmonella enterica* serovar Typhimurium DT104 (Boyd *et al.*, 2001; Boyd *et al.*, 2000).

Acquisition of a GI is a major step in the evolution of pathogenic bacteria as it can confer selective advantage over other less-flexible organisms in new ecological niches. Examples of major virulence factors encoded on GIs include adherence factors, toxins, iron uptake systems, effectors, and proteins associated with T3SS and T4SS (Hacker & Kaper, 2000). PAPI-1 of *P. aeruginosa* was the first of the GIs of Gram-negative bacteria for which inter-strain transfer was demonstrated (Qiu *et al*., 2006). Acquisition of GIs has subsequently been observed in plant pathogens such as *R*. *solanacearum* and *P*. *syringae* (Coupat *et al.*, 2008; Fall *et al.*, 2007; Lovell *et al.*, 2009).

ICEs (initially known as transposons) are GIs that are known to transfer by conjugation and to integrate into and replicate along with the host chromosome (Burrus *et al.*, 2002; Wozniak *et al.*, 2009). ICEs contribute to virulence of bacterial pathogens (Wozniak *et al.,* 2009). For example, the PAPI-1 ICE of *P*. *aeruginosa* PAPI-1 contributes to virulence in murine models of acute pneumonia and bacteraemia (Harrison *et al.*, 2010). The cargo genes of *ICEPm1*, conserved in *Proteus mirabilis*, *Providencia stuartii*, and *Morganella morganii* encode a yersiniabactin-related iron acquisition system and an adhesin/protease that both contribute to virulence in the mouse model of ascending urinary tract infection (Alamuri *et al.*, 2010; Flannery, 2011; Himpsl *et al.*, 2010; Nielubowicz *et al.*, 2008). SPI-7 of *Salmonella* species possess similarities to ICEs and encodes important virulence functions, including the major virulence antigen (Vi) and type IVB pili (Hashimoto *et al.*, 1991; Parkhill *et al.*, 2001). Distantly related SPI-7 family ICEs have also been identified in the *Enterobacteriaceae* family including various plant pathogenic bacteria such as *S. turgidiscabies*, *P. syringae and P*. *atrosepticum* (Bell *et al.*, 2004; Huguet-Tapia *et al.*, 2014; Lovell *et al.*, 2011; Pitman *et al.*, 2005).

Some GIs are smaller than 10 kb. These are referred to as genomic or gene islets (Hacker & Carniel, 2001; Hacker & Kaper, 2000). It is known that these horizontally acquired small DNA regions (often single genes) can influence virulence and are sometimes present within a GI exhibiting a mosaic structure (Marcus *et al.*, 2000). For example, the *mgtCB* genes of *Salmonella* are involved in the production of *Salmonella*-induced filaments (Stein *et al.*, 1996).

Given that these genes are the only virulence genes on the putative island, SPI3, it is considered that these virulence genes were likely acquired by *Salmonella* independently of other areas of SPI3 (Marcus *et al*., 2000). Another such example is the genomic islet of *P. putida* KT2440, which carries a lysophosholipase gene involved in haemolytic activity and two other genes encoding proteins related to insecticidal toxins (Vodovar *et al.*, 2006).

1.7 MGEs involved in virulence of *Pectobacterium*

Several MGEs have been described in strains of *Pectobacterium* that contribute to virulence of the bacteria. Bell *et al*. (2004) identified 17 putative GIs (HAI1 - HAI17) in the genome of *P. atrosepticum* SCRI103. Mutations in genes encoding the T4SS and a putative biosynthetic cluster encoding a polyketide phytotoxin (CFA), residing on HAI6 and HAI2, respectively, showed reduced blackleg symptoms on potato stems (Bell *et al*., 2004). Another GI, HAI5 of *P. atrosepticum* SCRI1043 contained the *rfb* cluster (biosynthetic cluster for the O-antigen of LPS). A mutation in *rfbI* caused altered LPS biosynthesis, reduced motility and decreased virulence (Evans *et al*., 2010). HAI9 and HAI17 of *P. atrosepticum* SCRI1043 are prophages (ECA29 and ECA41, respectively). These prophages also modulate virulence (Evans *et al*., 2010).

Recent genome sequencing also identified a variety of novel GIs in *P. wasabiae* SCC3193 (Nykyri *et al*., 2012). Inactivation of virulence cluster 2 (Vic2), located on one of these GIs (GI_7), reduced maceration of potato tissue by the pathogen in a potato slice assay (Nykyri *et al*., 2012). Genes belonging to Vic2 encode a putative lipoprotein transport system and a *hopL1*-like gene (characterized as a T3SS gene in *P. syringae*) (Nykyri *et al.*, 2012; Petnicki-Ocwieja *et al.*, 2002). Similarly, deletions in genes belonging to virulence cluster 1 (Vic1) (containing phage-related genes), partially overlapping with GI_6 of *P. wasabiae* SCC3193, exhibited a reduced capacity to macerate tubers (Nykyri *et al*., 2012). A double mutant in the T6SS also resulted in reduced maceration, demonstrating the importance of the T6SS in virulence of this *Pectobacterium* species (Nykyri *et al*., 2012). Genes of T6SS are encoded by GIs in this bacterium.

Finally, a mutation in *evr* (*Erwinia* virulence regulator), a two-component system regulator of *P*. *carotovorum* SCRI193, reduced virulence in potato tubers, bacterial growth in potato stems, and decreased swimming and delayed pellicle formation in liquid medium (Williamson *et al.*, 2010). *Evr* is involved in the synthesis of an orange pigment, which is itself encoded on a novel cryptic biosynthetic cluster known as the *Pectobacterium* orange pigment (pop) cluster. This cluster is flanked by two tRNA genes and is predicted to have been acquired by HGT (Wiliamson *et al*., 2010).

1.8 Aims of this study

Seventeen putative GIs were originally identified in *P*. *atrosepticum* SCRI1043 that might be involved in pathogenicity and/or ecological adaptation of the pathogen. A number of these were subsequently shown to be important for virulence on stems. More recently, several virulence factors encoded on putative GIs of other *Pectobacterium* genomes (such as *P*. *wasabiae* SCC3193) have been shown to be important for development of soft rot. There remains little understanding, however, of why some strains of *Pectobacterium* (especially *P*. *carotovorum* subsp. *carotovorum*) can invade the vascular tissue of potato and cause blackleg whereas others cannot. This is an important question to address given the economic impact of blackleg in New Zealand and overseas, and the lack of effective management strategies that presently exist to control the disease.

The hypothesis of this study was that *P*. *carotovorum* subsp. *carotovorum* and *P*. *atrosepticum* strains harbour similar GIs that support the invasion of the vascular tissue of potato by these pathogens, which result in blackleg. To test this hypothesis, a combination of genomic (comparative genomics) and post-genomic (functional genomics) approaches were used to identify GIs in the aggressive blackleg-causing strains such as *P*. *carotovorum* subsp. *carotovorum* ICMP19477 that encoded virulence factors, which were absent in those that only caused soft rot. In addition, the identification of virulence factors common to all strains might provide further insights into the mechanisms used by *Pectobacterium* to cause soft rot of potato tubers.

To complete this research, the project was divided into three specific objectives:

i) Evaluation of the species status of the blackleg causing strain, *P*. *carotovorum* subsp. *carotovorum* ICMP19477,

ii) Identification and characterization of GIs and other gene islets in ICMP19477 using genome sequencing and comparative genomics and,

iii) Functional analysis of the virulence factors and associated GIs in ICMP19477 by testing knockout mutants in pathogenicity assays on potato.

Chapter 2

Re-classification of *Pectobacterium carotovorum* subspecies *carotovorum* causing blackleg disease on potatoes in New Zealand

2.1 Abstract

The blackleg and soft rot disease complex on potato occurs sporadically in New Zealand, causing economic damage under optimal conditions of temperature and humidity for disease development. The latest outbreak of blackleg was observed in Canterbury in the 2013-2014 growing season, which coincided with a period of unseasonably heavy rainfall (Andrew Pitman, personal communication). Both *P. atrosepticum* and *P. carotovorum* subsp. *carotovorum* have previously been isolated from potato tubers with soft rot symptoms in New Zealand. *Pectobacterium atrosepticum* has also been associated with blackleg. In the 2004-2005 growing season, isolates from infected tubers and stems were collected in the Auckland, Waikato, Manawatu-Wanganui and Canterbury regions of New Zealand. The majority were identified as *P. carotovorum* subsp. *carotovorum* isolates by their growth at 37°C, carbon utilisation profiles and restriction fragment length polymorphisms. The *P. carotovorum* subsp. *carotovorum* isolates were largely unable to cause blackleg symptoms, although several were highly aggressive upon stem infection. Globally, *P. carotovorum* subsp. *carotovorum* isolates have rarely been associated with blackleg, and therefore these isolates were considered atypical.

To confirm the identity of the aggressive blackleg-causing *P. carotovorum* subsp. *carotovorum* isolates, polymerase chain reaction (PCR) assays were conducted using primers specific to the various species and sub-species of *Pectobacterium* known to cause blackleg and soft rot worldwide. Blastn analysis of 16S rRNA gene sequences and phylogenetic analyses of concatenated DNA sequences from eight housekeeping genes were also used to identify the isolates, by comparison with DNA sequences from previously characterised SRE. Of the 89 New Zealand isolates re-tested by PCR, 18 produced an amplicon specific to *P. carotovorum* subsp. *brasiliensis* but no PCR product characteristic of *P. carotovorum* subsp. *carotovorum*. Blastn analysis of the 16S rRNA gene sequence from one of these isolates, ICMP19477,

showed it had greatest sequence identity to *P*. *carotovorum* subsp. *brasiliensis* isolates from overseas. Multi locus sequence analysis (MLSA) also clustered ICMP19477 with other *P*. *carotovorum* subsp. *brasiliensis* isolates, including the tentative type strain ATCC BAA 417. Based on the current literature, this is the first report of *P*. *carotovorum* subsp. *brasiliensis* infecting potato in New Zealand. As *P*. *carotovorum* subsp. *carotovorum* is usually unable to cause blackleg and *P*. *atrosepticum* is only rarely isolated from tubers in New Zealand, *P*. *carotovorum* subsp. *brasiliensis* is likely to be an important component of the blackleg syndrome in New Zealand.

2.2 Introduction

A variety of SRE cause diseases on potato. *Pectobacterium carotovorum* is widespread in potato production areas worldwide: *P*. *carotovorum subsp*. *carotovorum* is typically associated with stem and tuber soft rot of potato, whereas *P*. *carotovorum* subsp. *brasiliensis* is a highly virulent subspecies of *P*. *carotovorum* associated with disease outbreaks of blackleg in Brazil (Duarte *et al.*, 2004) and South Africa (Duarte *et al.*, 2004). *Pectobacterium carotovorum* subsp. *brasiliensis* primarily infects potato in relatively cool temperatures during the growing season, but is principally found in humid subtropical regions (Duarte *et al*., 2004). *Pectobacterium atrosepticum* is considered to be almost exclusively a potato pathogen (Perombelon, 2002), causing seed tuber infection resulting in blackleg. *Pectobacterium wasabiae* has increasingly been isolated from potato with blackleg or soft rotting symptoms (Waleron *et al.*, 2013). Indeed, *P*. *wasabiae* has been detected in the USA (Ma *et al*., 2007), New Zealand (Pitman *et al*., 2008, 2010), Iran (Baghaee-Ravari *et al*., 2011), Syria (Nabhan *et al*., 2012a), South Africa and Zimbabwe (Moleleki *et al.*, 2013; Ngadze *et al.*, 2012), in Canada (De Boer *et al.*, 2012) and in Europe (Nabhan *et al*., 2012b). A recent study showed that vacuum-infiltration of tubers with *P*. *wasabiae* results in similar amounts of blackleg-like symptoms as tubers infected with *P*. *atrosepticum* (Pasanen *et al.*, 2013). *Dickeya solani* has also emerged as a significant economic threat to potato production, causing disease outbreaks under more temperate conditions than those normally required by other *Dickeya* spp. (Czajkowski *et al.*, 2009; Laurila *et al.*, 2008; Slawiak & Lojkowska, 2009; Toth *et al.*, 2011). *Dickeya solani* is predicted to have emerged by acquiring factors that enable it to out-compete other SRE on potato (Pédron *et al.*, 2014).

Given the diversity of SRE on potato, their different host ranges and variable severity of disease symptoms, accurate identification of SRE is essential for appropriate disease management. Carbon utilisation profiles have long been known to differentiate species or subspecies of *Pectobacterium* (Biolog, Haywood, CA, USA). PCR restriction fragment length polymorphisms (RFLPs) using the *pel* gene (Avrova *et al.*, 2002; Darrasse *et al.*, 1994; Hélias *et al.*, 1998; Toth *et al.*, 2001b) and genomic fingerprinting techniques such as AFLP have also been employed to study genetic diversity of *Pectobacterium* (Avrova *et al.*, 2002; Toth *et al.*, 2001a). More recently, MLSA and comparative genomics have re-defined taxonomic relationships amongst the SRE. These studies have led to the re-classification of numerous taxa. For example, De Haan *et al*. (2008) identified a subtype of *P*. *carotovorum* subsp. *carotovorum* that was associated with blackleg symptoms on potatoes grown in temperate climates. Comparison of the genomes of the blackleg causing *P*. *carotovorum* subsp. *carotovorum* with those of other SRE, however, led to their re-classification as *P*. *wasabiae* (Nabhan *et al.*, 2012a; Nykyri *et al.*, 2012). An understanding of the biochemical and genetic variation amongst SRE has supported the development of a variety of tools for the detection and differentiation of SRE (De Boer & McNaughton, 1987; De Boer & Ward, 1995; Kang *et al.*, 2003b; Lee & Yu, 2006; Singh *et al.*, 2000; Wells & Moline, 1991). PCR-based diagnostics have proven particularly useful for identification purposes, with specific tests available for all the major SRE on potato (De Boer & Ward, 1995; Duarte *et al.*, 2004; Kang *et al.*, 2003b; Nassar *et al.*, 1996).

The objective of the research in this chapter was to confirm the identity of isolates associated with an aggressive form of blackleg disease on potatoes in New Zealand, which had previously been designated as belonging to *P*. *carotovorum* subsp. *carotovorum* (Pitman *et al*., 2008). In New Zealand, both *P*. *atrosepticum* and *P*. *carotovorum* subsp. *carotovorum* were isolated from potato tubers with soft rot symptoms during the 1988 growing season (Crowhurst & Wright, 1988). In a second survey conducted in the 2004-2005 growing season, the majority of the pectolytic enterobacteria isolated from potato tubers from different growing regions (including Auckland, Central Plateau, Hawkes Bay, Waikato, Nelson, Northland, Otago, Southland, Wairarapa, Manawatu-Wanganui and Canterbury) were identified as *P*. *carotovorum* subsp. *carotovorum* by their growth at 37°C, carbon utilisation profiles and RFLP profiles (Pitman *et al*., 2008). These isolates were largely unable to cause blackleg symptoms,

although several were shown to be highly aggressive upon stem infection. Based on classical descriptions of disease produced by *P*. *carotovorum* subsp. *carotovorum*, this suggested that at least some *P*. *carotovorum* subsp. *carotovorum* isolates in New Zealand may have been misclassified or that a new genotype of *P*. *carotovorum* subsp. *carotovorum* may exist that can cause blackleg. Re-examination of these isolates was performed using a number of the tools available for detection and differentiation of SRE.

2.3 Materials and methods

2.3.1 Bacterial strains and culturing conditions

A collection of 89 pectolytic enterobacteria (Table 2.1) were isolated from stems and tubers of potato plants with typical disease symptoms (soft rot, wilting and external blackening), during the 2004-2005 growing season in New Zealand (Pitman *et al*., 2008). Samples were collected from diverse potato cultivars grown in numerous potato-growing regions: Auckland (10); Canterbury (34); Central Plateau (7); Hawkes Bay (5); Manawatu–Wanganui (12); Nelson (1); Northland (1); Otago (2); Southland (6); Waikato (10) and Wairarapa (1) (the total number of isolates from each geographical origin is given in parentheses). Type strains and other well characterised isolates of *Pectobacterium* were also used in this study for comparison with the New Zealand isolates. Type strains were obtained from the International Collection of Microorganisms from Plants (ICMP), Landcare Research, New Zealand or the American Type Culture Collection (ATCC) (Table 2.1). *Pectobacterium* strains used in this study were always grown at 28°C. For pathogenicity assays, bacterial cultures were revived from frozen stocks by streaking onto LB plates supplemented with 1.6% (w/v) agar to isolate individual colonies.

Table 2.1 *Pectobacterium* strains used in this study

Pectobacterium* strains and their geographical origins
Auckland (10)
ICMP19477 (NZEC1), **NZEC43**, NZEC68, NZEC78, NZEC110, NZEC181, NZEC118, NZE8974, NZEC210, NZEC211
Canterbury (54)
NZEC129, NZEC130, NZEC131, **NZEC132**, NZEC133, NZEC134, **NZEC137**, **NZEC138**, NZEC139, **NZEC140**, NZEC141, **NZEC142**, **NZEC143,** NZEC144, NZEC146, NZEC147, **NZEC149**, **NZEC150**, **NZEC151**, NZEC152, **NZEC153**, **NZEC154**, NZEC166, NZEC169, NZEC170, NZEC172, NZEC173, NZEC174, NZEC175, NZEC176, NZEC182, NZEC115, NZEC16, NZEC135
Central Plateau (7)
NZEC7, NZEC24, NZEC161, NZEC23, NZEC14, NZEC13, NZEC22
Hawkes Bay (5)
NZEC89, NZEC90, NZEC155, NZEC164, NZEC185
Manawatu-Wanganui (12)
NZEC4, NZEC5, **NZEC8**, NZEC38, NZEC119, **NZEC128**, NZEC157, NZEC158, NZEC184, NZEC188, NZEC9, NZEC10
Nelson (1)
NZEC179
Northland (1)
NZEC168
Otago (2)
NZEC165, NZEC177
Southland (6)
NZEC19, NZEC31, NZEC32, NZEC183, NZEC20, NZEC21
Waikato (10)
NZEC6, **NZEC25**, NZEC91, NZEC93, NZEC121, NZEC124, NZEC125, NZEC126, NZEC12, NZEC127
Wairarapa (1)
NZEC180
Collection strains
P. atrosepticum ICMP1526^T (United Kingdom), *P. atrosepticum* SCRI1043 (or ATCC BAA-672, United Kingdom), *P. carotovorum* subsp. *carotovorum* ICMP5702^T (Denmark), *P. carotovorum* subsp. *carotovorum* UGC32 (Peru), *P. carotovorum* subsp. *carotovorum* WPP14 (North America), *P. carotovorum* subsp. *brasiliensis* 1692 (or ATCC BAA 417, Brazil)

The total number of isolates from each geographical origin is shown in brackets. T, Type strain; *, Bacterial strain isolated from potato tubers or stems grown in the indicated region of New Zealand unless otherwise stated. The bacterial strains in bold letters are re-classified as *P. carotovorum* subsp. *brasiliensis*.

2.3.2 Subspecies-specific PCR

To identify different subspecies of *Pectobacterium*, PCR was carried out using subspecies-specific primers. In particular, PCR procedures to distinguish *P*. *atrosepticum*, *P*. *carotovorum* subsp. *carotovorum* and *P*. *carotovorum* subsp. *brasiliensis* were performed on purified DNA using primer pairs ECA1f/ECA2r (De Boer and Ward, 1995), EXPCCF/EXPCCR (Kang *et al.*, 2003a) and BR1f/L1r (Duarte *et al*., 2004), respectively. The nucleotide sequences and target loci for all the primers, and sizes of the respective amplicons used in these PCR assays are listed in Table 2.2. Each PCR reaction contained 2 µL (10-30 ng) genomic DNA as template, 2 µL each of 10 µM forward and reverse primers (Integrated DNA Technologies, USA), 4 µL of 2 mM dNTPs, 5 µL of 10× PCR buffer, 1.5 µL of 50mM $MgCl_2$ and 0.5 µL of DNA polymerase. The final volume was made up to 50 µL with sterile distilled water (SDW). PCR amplification was carried out using a GeneAmpR PCR System 9700 thermocycler (Applied Biosciences, USA). PCR amplification conditions consisted of an initial denaturation step at 95°C for 3 min, followed by 35 cycles of 94°C for 30 s, an annealing temperature specific to each primer pair (Table 3.2) for 30 s, and an extension step at 72°C for 30 s to 5 min depending on the size of the targeted fragment. A final extension step was carried out at 72°C for 7 min. Negative controls using water instead of DNA template, and positive controls using DNA of type strains of each bacterial species or subspecies, were included for each PCR reaction. Amplified DNA fragments from each sample were resolved on a 2% agarose gel containing Sybr Safe stain (Life Technologies, USA) in 1× TAE buffer at 100 volts, and visualized over UV light. The molecular size marker Hyperladder IV (Bioline, UK) was used to determine the size of amplified DNA fragments.

2.3.3 16S rRNA gene sequencing and analysis

The partial 16S rRNA gene from ICMP19477 was amplified by PCR using primer pair Yf1/Yr1 (Jensen *et al.*, 1993). ICMP19477 was used as a representative of the New Zealand strains from diseased stems of potato plants that were initially characterized as *P*. *carotovorum* subsp. *carotovorum*, but produced a *P*. *carotovorum* subsp. *brasiliensis*-specific amplicon upon PCR using primers BR1f and L1r. PCR amplification of the partial 16S rRNA gene sequence from this strain was carried out using the conditions for PCR described above. PCR

products were purified using a Qiaquick PCR purification kit (Qiagen, Netherlands), according to the manufacturer's protocol. Sanger sequencing of the purified DNA products was then performed in both directions by Macrogen Inc (Seoul, South Korea). DNA sequences obtained for the partial 16S rRNA sequence were assembled, trimmed and edited into a consensus sequence using Geneious v4.6 (Biomatters, available from http://www.geneious.com/), and deposited in Genbank under the accession number JQ771053. Finally, a blastn (Altschul *et al.*, 1990) similarity search was conducted on the consensus sequence generated from the Geneious assembler using the NCBI database.

2.3.4 MLSA

MLSA was performed by concatenating partial DNA sequences from ICMP19477 (JQ820114-JQ820120) for seven genes including the aconitate hydrase 1 (*acnA*), glyceraldehyde-3-phosphate dehydrogenase A (*gapA*), isocitrate dehydrogenase (*icdA*), malate dehydrogenase (*mdh*), mannitol-1-phosphate dehydrogenase (*mtlD*), glucose-6-phosphate isomerase (*pgi*), and gamma-glutamylphosphate reductase (*proA*) genes (Ma *et al.*, 2007). The concatenated sequence was then compared to those for 46 related taxa, using *Yersinia* species as an outgroup (Ma *et al*., 2007). The collection of concatenated DNA sequences was aligned in Geneious Pro v5.6.2 (Biomatters, available from http://www.geneious.com/). The alignment was performed using a cost-matrix of 65% with a pairwise alignment or multiple alignment gap opening penalty of 10, plus a gap extension penalty of 0.1. The resulting alignment was exported in nexus format and was used to generate a phylogenetic tree using the GTR substitution model in the MrBayes v. 3.0b4 program (Huelsenbeck & Ronquist, 2001). Each run was composed of four chains starting from random trees, and the analysis was run for 10,000,000 generations with trees sampled every 100 generations. The majority rule consensus tree was calculated after discarding the first 2500 trees corresponding to a burn-in period to create a final phylogenetic tree.

Table 2.2 Primers used in this study

Primer Name	Primer Sequence (5' to 3')	Amplicon size (bp)	Annealing Temperature	Source
16S rRNA sequencing primers				
Yf1	TGATGGAGGGGGATAACTACTGGA	1381	60	Jensen *et al*., 1993
Yr1	CCCCTACGGTTACCTTGTTACGAC			
Species-specific primers				
ECA1f	CGGCATCATAAAAACACG	690	60	De Boer and Ward, 1995
ECA2r	GCACACTTCATCCAGCGA			
EXPCCF	GAACTTCGCACCGCCGACCTTCTA	550	60	Kang *et al*., 2003
EXPCCR	GCCGTAATTGCCTACCTGCTTAAG			
BR1f	GCGTGCCGGGTTTATGACCT	322	51	Duarte *et al*., 2004
L1r	CARGGCATCCACCGT			
ADE1	GATCAGAAAGCCCGCAGCCAGAT	644	60	Naseer *et al*., 1996
ADE2	CTGTGGCCGATCAGGATGGTTTTGTCGTGC			
Primers for Multi Locus Sequence Analysis				
acnA3F	CMAGRGTRTTRATGCARGAYTTTAC	300	52	Ma *et al*., 2007
acnA3R	GATCATGGTGGTRTGSGARTCVGT			
gapA326F	ATCTTCCTGACCGACGAAACTGC	450	57	Ma *et al*., 2007
gapA845R	ACGTCATCTTCGGTGTAACCCAG			
icdA400F	GGTGGTATCCGTTCTCTGAACG	520	55	Ma *et al*., 2007
icdA977R	TAGTCGCCGTTCAGGTTCATACA			
mdh86F	CCCAGCTTCCTTCAGGTTCAGA	460	55	Ma *et al*., 2007
mdh628R	CTGCATTCTGAATACGTTTGGTCA			
mtlD146F	GGCCGGTAATATCGGCCGTGG	390	60	Ma *et al*., 2007
mtlD650R	CATTCGCTGAAGGTTTCCACCGT			
pgi815F	TGGGTCGGCGGCCGTTACTC	520	57	Ma *et al*., 2007
pgi1396R	TGCCTTCGAATACTTTGAACGGC			
proAF1	CGGYAATGCGGTGATTCTGCG	630	56	Ma *et al*., 2007
proAR1	GGGTACTGACCGCCACTTC			

2.3.5 Virulence assays

2.3.5.1 Preparation of bacterial inoculum

A 5 mL culture of each bacterium used in virulence assays was grown overnight in LB at 28°C with 180 rpm. The cells were then harvested by centrifugation at 4,000 rpm for 10 min, followed by washing with 5 mL of 10 mM $MgS0_4$ or 10 mM $MgCl_2$. Finally, the bacterial pellets were re-suspended in 10 mM $MgS0_4$ or 10 mM $MgCl_2$ to obtain 10^6 or 10^8 cfu per mL of bacterial cells (depending on the assay). Cfu per mL was estimated by absorbance at OD_{600} nm.

2.3.5.2 Blackleg assay

For blackleg assays, stems of fully-grown potato plants ('Ilam Hardy') with a height of approximately 20 cm were inoculated with 10 µL of bacterial suspension (equivalent to 10^4 cfu and 10^6 cfu in two separate experiments) under the second fully expanded leaf using a micropipette tip. Inoculation sites were covered with parafilm to avoid desiccation. Stems inoculated with 10 mM $MgSO_4$ were used as controls in this assay. The inoculated plants were then kept in an incubator at 22°C with a controlled alternate day and night cycle (16 h light and 8 h dark) and high humidity (>85%). Disease symptoms were assessed by measuring the length of the necrotic lesions on infected stems daily, over a period of six days. Twelve potato plants were used for each strain.

2.3.5.3 Statistical analysis

For both soft rot and blackleg assays, the least significant difference (LSD), i.e. the smallest difference between any two means that can be described as being significantly different, was calculated from the residual mean square analysis of variance from averaged means across all days using GENSTAT for Windows, version 6.1.0.200 (Lawes Agricultural Trust).

2.4 Results

2.4.1 PCR detection of *P. carotovorum* subsp. *brasiliensis*

A 322-bp PCR amplicon specific to *P. carotovorum* subsp. *brasiliensis* was amplified from DNA extracted from 18 of the 89 SRE tested including ICMP19477 (Table 2.1, Figure 2.1). A PCR product (550 bp) typical of *P. carotovorum* subsp. *carotovorum* was not amplified using the DNA from these 18 isolates (data not shown). In addition, no 690-bp PCR product specific to *P. atrosepticum* was amplified from the DNA of any of those isolates using primer pair ECA1f/ECA2r (data not shown).

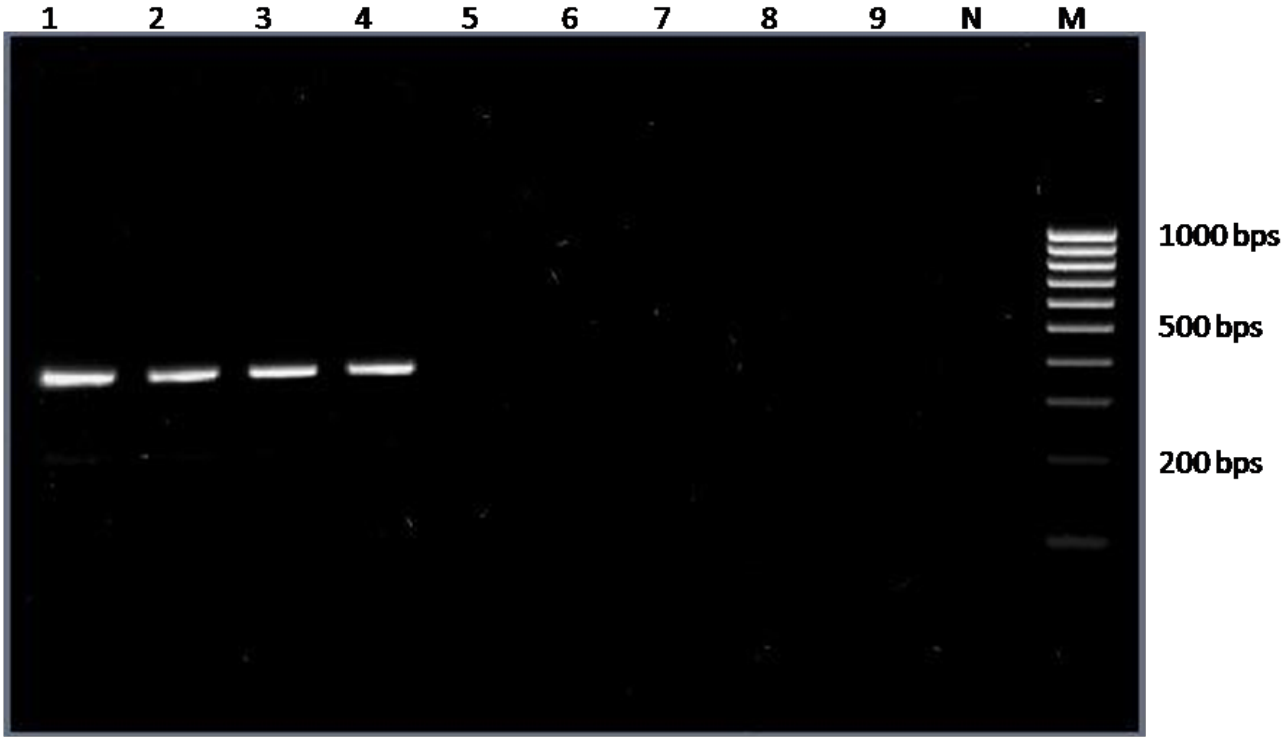


Figure 2.1 PCR detection of *Pectobacterium carotovorum* subsp. *brasiliensis* using the Br1f and Lr1 primers. A 322 bp PCR product was amplified using the Br1f and Lr1 primers from DNA samples extracted from *P. carotovorum* subsp. *brasiliensis* strains ATCC BAA 419, ATCC BAA 417, ICMP19477 and ICMP1947750. Lanes: 1, *P. carotovorum* subsp. *brasiliensis* ATCC BAA-419; 2, *P. carotovorum* subsp. *brasiliensis* ATCC BAA 417^{T}; 3, *P. carotovorum* subsp. *brasiliensis* ICMP19477; 4, *P. carotovorum* subsp. *brasiliensis* ICMP1947750; 5, *P. carotovorum* subsp. *carotovorum* ICMP5702^{T}; 6, *P. atrosepticum* ICMP1526^{T}; 7, *P. betavasculorum* ICMP4226^{T}; 8, *P. wasabiae* ICMP9121^{T}; 9, *P. carotovorum* subsp. *odoriferum* ICMP11533^{T}; N, No template control; M, Marker (1000-bp ladder); T, Type strain.

The 18 isolates from which a *P. carotovorum* subsp. *brasiliensis*-specific PCR product was amplified were collected from potatoes grown in a variety of regions in New Zealand, but in particular comprised a large proportion of the SRE from potato in Canterbury (13/23) (Figure 2.2).

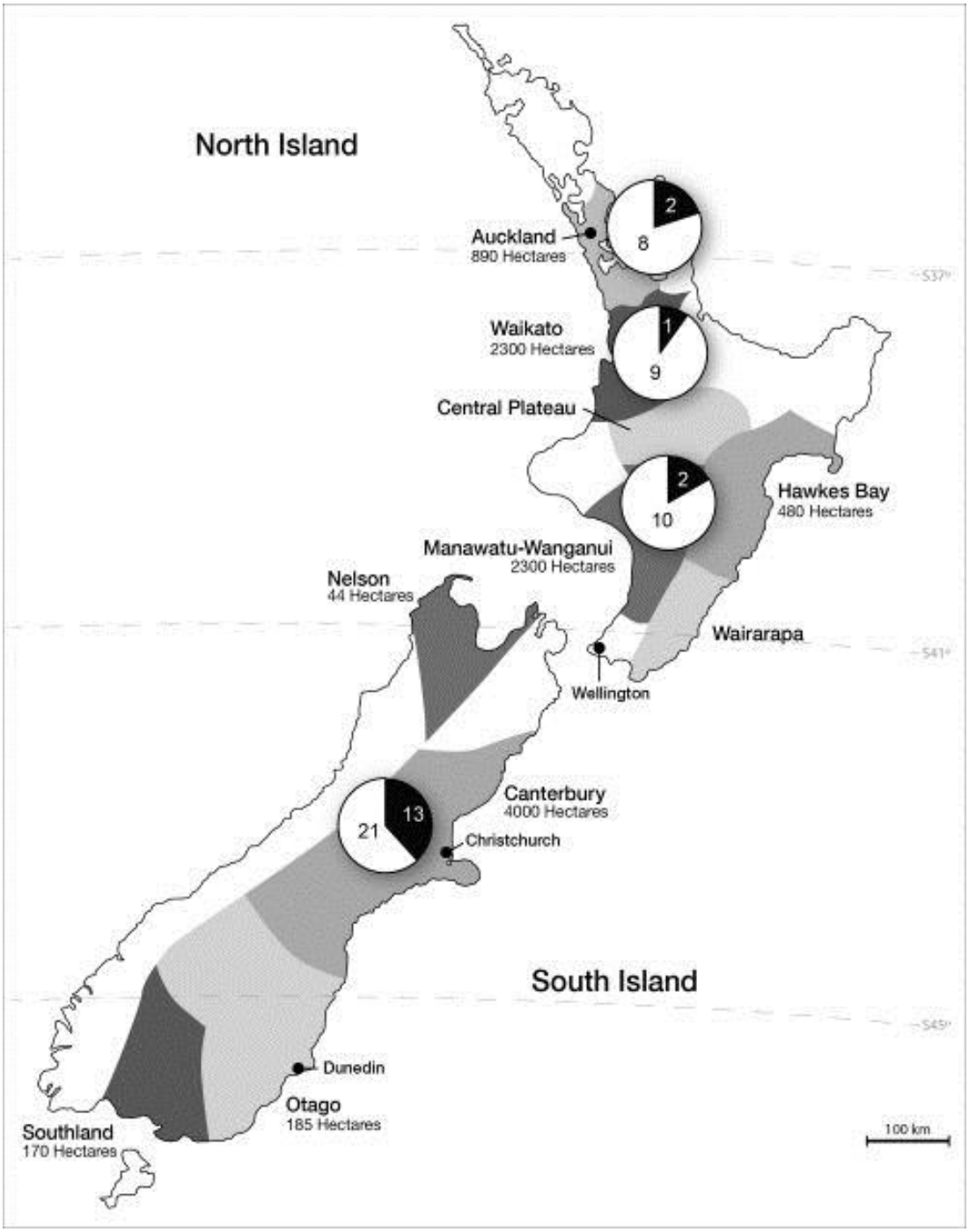
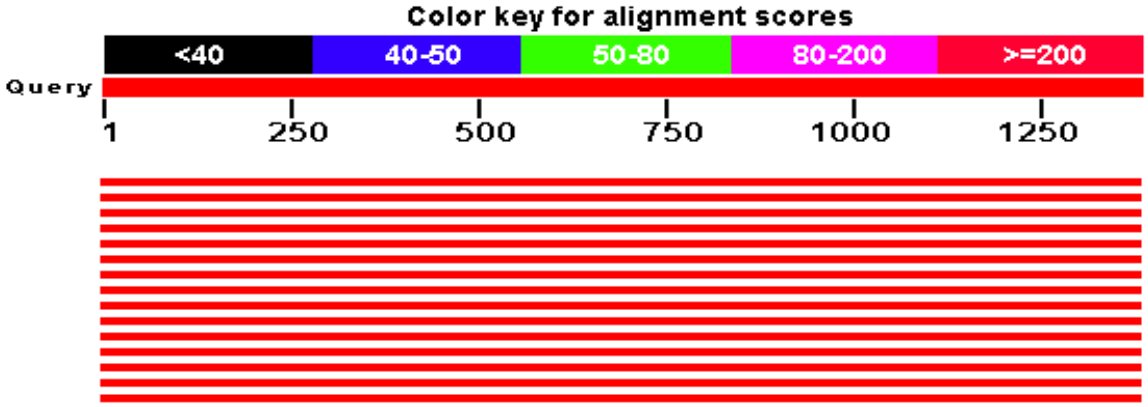


Figure 2.2 Locations in New Zealand (indicated by pie charts) from which potato tubers infected with *Pectobacterium carotovorum* subsp. *brasiliensis* were collected. Shading distinguishes major geographical localities where potato is cultivated. Total area (ha) of commercial potato cropping in each region is indicated. Each pie chart defines the proportion of enterobacterial isolates found to be *P*. *carotovorum* subsp. *brasiliensis* (black) or other *Pectobacterium* species (white).

2.4.2 Identification of *P. carotovorum* subsp. *brasiliensis* by 16S rRNA sequence analysis and MLSA

As a *P. carotovorum* subsp. *brasiliensis*-specific PCR product was amplified from ICMP19477, this strain was chosen to represent the collection of isolates tentatively identified as *P. carotovorum* subsp. *brasiliensis*. PCR amplification of the partial 16S rRNA gene from ICMP19477 with primers Yf1 and Yr1 (Jensen *et al*., 1993) generated a fragment of 1381 bp. Blastn analysis (carried out on 07 December, 2012) of the partial 16S rRNA gene sequence of ICMP19477 showed it had 99% identity (with an e-value of 0.0) to the 16S rRNA sequences of various *P. carotovorum* subsp. *brasiliensis* strains in the NCBI database (Figure 2.3).

Partial DNA sequences of *acnA*, *gapA*, *icdA*, *mdh*, *mtlD*, *pgi* and *proA* from ICMP19477 were successfully amplified by PCR using the primers described in Section 2.3.4. MLSA using the concatenated DNA sequences from these genes and 46 related taxa (Ma *et al*., 2007) resulted in a majority rule consensus tree that clustered ICMP19477 most closely with well characterized *P. carotovorum* subsp. *brasiliensis* isolates from overseas (Figure 2.4). These included the tentative type strain ATCC BAA 417. The *P. carotovorum* subsp. *brasiliensis* cluster (blue colour, cluster 3) is supported by a Bayesian confidence value or posterior probability of 100. MLSA also distinguished ICMP19477 from closely related isolates belonging to *P. carotovorum* subsp. *carotovorum* (cluster 2 and cluster 4). Cluster 1 also contained *P. carotovorum* subsp. *carotovorum* strains but the presence of *P. carotovorum* subsp. *odoriferum* SCRI482 in cluster 1 suggests that the other strains belonging to cluster 1 may actually belong to *P. carotovorum* subsp. *odoriferum*.



Description	Max score	Total score	Query cover	E value	Ident	Accession
Pectobacterium carotovorum subsp. brasiliensis NZEC1 16S ribosomal RNA gene, complete sequence	2551	2551	100%	0.0	100%	JQ771053.1
Pectobacterium carotovorum subsp. brasiliense strain 212 note sequence type I 16S ribosomal RNA gene, complete sequence	2540	2540	100%	0.0	99%	JF926716.1
Pectobacterium carotovorum subsp. brasiliense strain 212 note sequence type II 16S ribosomal RNA gene, complete sequence	2529	2529	100%	0.0	99%	JF926761.1
Pectobacterium carotovorum subsp. brasiliense strain 1001 note sequence type II 16S ribosomal RNA gene, complete sequence	2495	2495	100%	0.0	99%	JF926760.1
Pectobacterium carotovorum subsp. brasiliense strain 8 note sequence type I 16S ribosomal RNA gene, complete sequence	2495	2495	100%	0.0	99%	JF926723.1
Pectobacterium carotovorum subsp. brasiliense strain A17 note sequence type II 16S ribosomal RNA gene, complete sequence	2495	2495	100%	0.0	99%	JF926722.1
Pectobacterium carotovorum subsp. brasiliense strain 1033 16S ribosomal RNA gene, complete sequence	2495	2495	100%	0.0	99%	JF926720.1
Pectobacterium carotovorum subsp. brasiliense strain C18 note sequence type II 16S ribosomal RNA gene, complete sequence	2495	2495	100%	0.0	99%	JF926719.1
Pectobacterium carotovorum subsp. brasiliense strain 1001 note sequence type I 16S ribosomal RNA gene, complete sequence	2490	2490	100%	0.0	99%	JF926759.1
Pectobacterium carotovorum subsp. brasiliense strain C18 note sequence type I 16S ribosomal RNA gene, complete sequence	2490	2490	100%	0.0	99%	JF926718.1
Pectobacterium carotovorum subsp. brasiliense strain 1073 16S ribosomal RNA gene, complete sequence	2490	2490	100%	0.0	99%	JF926717.1
Pectobacterium carotovorum subsp. brasiliense strain 317 note sequence type I 16S ribosomal RNA gene, complete sequence	2484	2484	100%	0.0	99%	JF926725.1
Pectobacterium carotovorum subsp. brasiliense strain A17 note sequence type I 16S ribosomal RNA gene, complete sequence	2484	2484	100%	0.0	99%	JF926721.1
Pectobacterium carotovorum subsp. brasiliense strain 8 note sequence type II 16S ribosomal RNA gene, complete sequence	2477	2477	100%	0.0	99%	JF926724.1
Pectobacterium carotovorum subsp. brasiliense strain 371 note sequence type II 16S ribosomal RNA gene, complete sequence	2468	2468	100%	0.0	99%	JF926726.1

Figure 2.3 Blastn analysis of 16S rRNA gene sequence of ICMP19477 (NZEC1). Blastn analysis of 16S rRNA gene sequence of ICM19477 shows 99% similarity to 16S rRNA sequences of other *Pectobacterium carotovorum* subsp. *brasiliensis* strains in the NCBI database.

Pectobacterium atrosepticum (cluster 5) isolates were more closely related to *P*. *carotovorum* subsp. *betavasculorum* (cluster 6) isolates and a distinct group of *P*. *carotovorum* subsp. *carotovorum* strains (cluster 7) than to *P*. *carotovorum* subsp. *brasiliensis* and the closely related clusters representative of *P*. *carotovorum*. Further investigation identified that members of cluster 7 included several isolates whose taxonomic status had subsequently been revised, and that these represented the *P*. *wasabiae* isolates from potato (Kim *et al*., 2009; Nicola Perna and co-workers, unpublished; Pitman *et al*., 2010). The separation of *P*. *carotovorum* subsp. *brasiliensis*, *P*. *wasabiae* and *P*. *carotovorum* subsp. *betavasculorum* from *P*. *carotovorum* subsp. *carotovorum* was supported by a Bayesian confidence value or Bayesian posterior probability of 100%.

Dickeya (cluster 10), not unexpectedly, formed the most distantly related cluster. Since, *P*. *carotovorum* subsp. *carotovorum* is poorly understood taxonomically, the phylogenies also confirmed the existence of a novel, distantly related cluster of strains (cluster 8). *Pectobacterium carotovorum* subsp. *carotovorum* PC1 (of cluster 8) has recently been re-classified as belonging to a novel species of *Pectobacterium* known as *P*. *aroidearum*, which infects monocotyledonous plants (Nabhan *et al.*, 2012a).

2.4.3 Virulence of *P. carotovorum* subsp. *brasiliensis* ICMP19477

The capacity of ICMP19477 to cause blackleg disease on potato was confirmed by performing virulence assays using two concentrations of inoculum (10^4 or 10^6 cfu per mL). The assays were performed on potato stems as described in Section 2.3.5. The highly aggressive strain ATCC BAA 417 was used in assays for comparative purposes, as it was considered representative of the *P*. *carotovorum* subsp. *brasiliensis* isolates first identified in Brazil as causing blackleg. *Pectobacterium atrosepticum* strains ICMP1526 and SCRI1043 were also used in the experiments, as *P*. *atrosepticum* is well documented to cause blackleg in potato in more temperate climates.

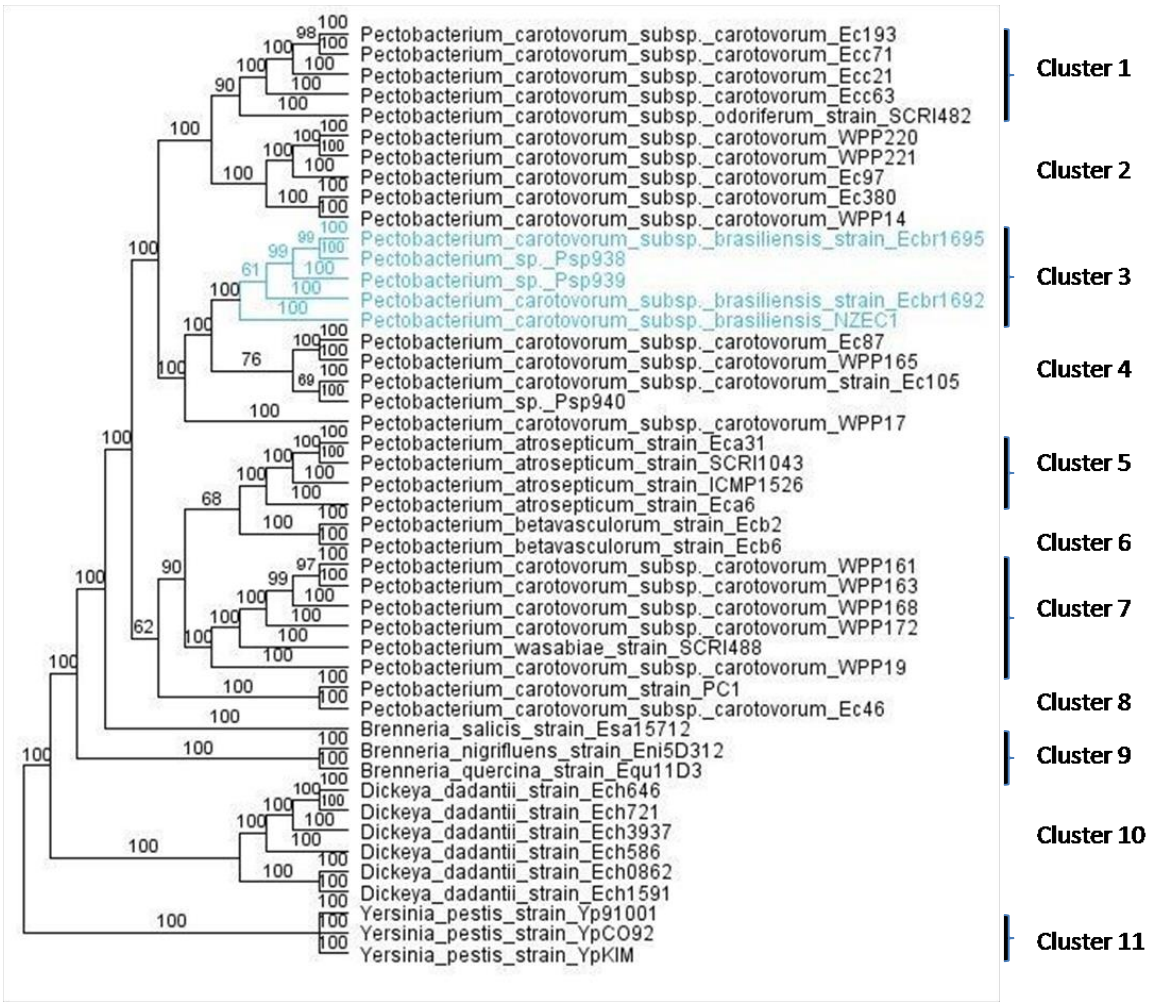


Figure 2.4 A Bayesian phylogenetic tree constructed using concatenated partial gene sequences of *acnA*, *gapA*, *icdA*, *mdh*, *mtlD*, *pgi* and *proA* for 46 *Enterobacteriaceae* taxa. The taxa are divided into 11 distinct and well-supported clusters. Numbers above nodes are posterior probabilities recovered by the Bayesian analysis. Sequences for *Yersinia pestis* were used as an outgroup.

Typical blackleg symptoms (i.e. typical blackening and decay of the stem) appeared on potato stems two days post inoculation (dpi) with the *P. carotovorum* subsp. *brasiliensis* isolates (Table 2.3; Figure 2.5). *Pectobacterium atrosepticum* strains SCRI1043 and *P. atrosepticum* ICMP1526, however, showed only small lesions after 2 days (Figure 2.4). Thus, for all plants that developed

lesions, lesion lengths on Day 2 varied between strains. At both concentrations, lesion lengths for SCRI1043 and ICMP1526 were substantially lower than for the other strains ($p<0.001$).

Table 2.3 Mean lesion lengths on potato stems two dpi with ICMP19477 and other *Pectobacterium* isolates. Plants were inoculated with either 10^4 or 10^6 cfu per mL of inoculum. 95% confidence limits are described in brackets.

Strain	10^4	10^6
ICMP19477	18.2 (17.8,18.5)	18.6 (18.2,18.9)
ATCC BAA 417	18.6 (17.9,19.2)	17.2 (16.7,17.8)
ICMP1526	2.9 (2.6,3.1)	3.1 (2.9,3.4)
SCRI1043	2.2 (2.1,2.3)	4.5 (4.3,4.7)

All 24 plants inoculated with either ICMP19477 or ATCC BAA 417 were dead by Day 3, and all for ICMP1526 and SCRI1043 by Day 6 (other than the one plant of ICMP1526 that showed no symptoms). Thus, at two dpi, the percentage of plants that developed symptoms (Table 2.4) did not vary significantly between the strains ($p>0.05$), or between concentrations of inoculum ($p=0.985$ for the strain by concentration interaction and $p=0.576$ for the overall concentration effect). Control plants inoculated with only $MgSO_4$ did not generate disease symptoms.

Table 2.4 Percentage of plants showing blackleg symptoms two dpi inoculated with either 10^4 or 10^6 cells of each *Pectobacterium* strain. 95% confidence limits are described in brackets.

Strain	10^4	10^6
ICMP19477	100.0 (73.5,100.0)	100.0 (73.5,100.0)
ATCC BAA 417	100.0 (73.5,100.0)	100.0 (73.5,100.0)
ICMP1526	91.7 (58.7,98.8)	100.0 (73.5,100.0)
SCRI1043	100.0 (73.5,100.0)	100.0 (73.5,100.0)

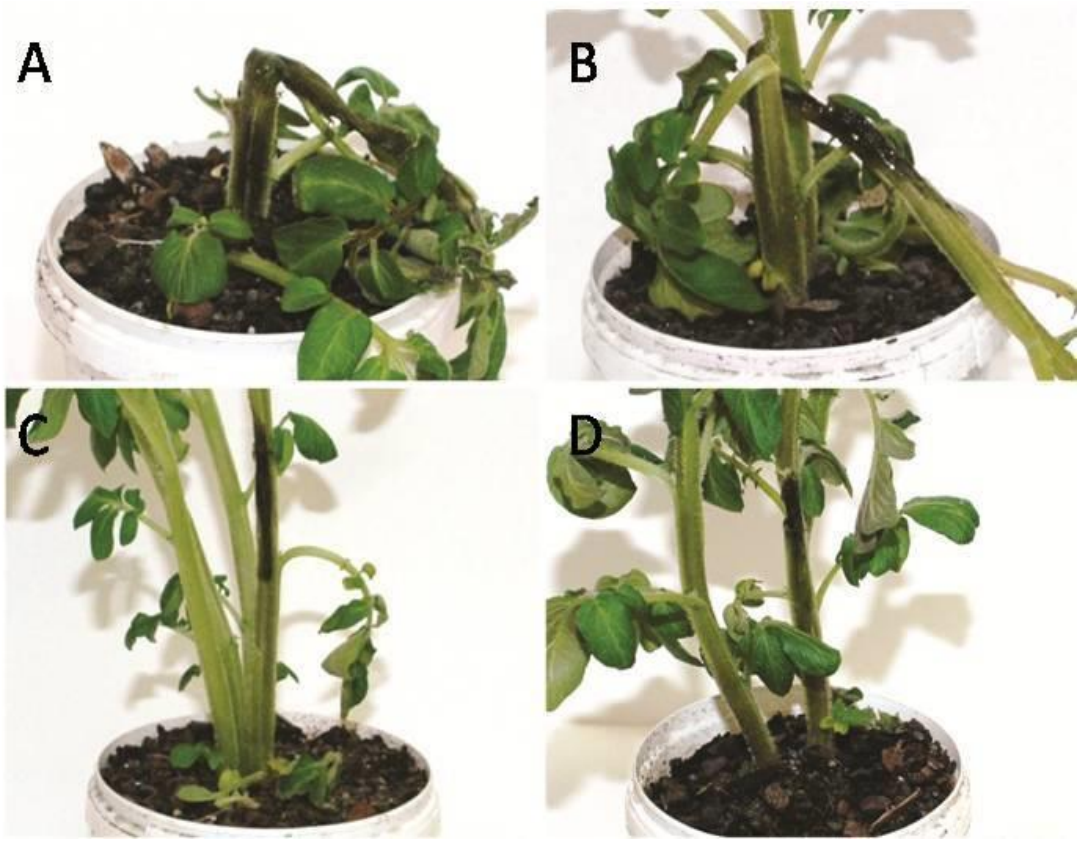


Figure 2.5 Disease lesions on stems of potato plants when inoculated with bacterial suspensions (A – D) two days post inoculation. Image A, *Pectobacterium carotovorum* subsp. *brasiliensis* ATCC BAA 417; B, *P*. *carotovorum* subsp. *brasiliensis* ICMP19477; C, *P*. *atrosepticum* SCRI1043 and D, *P*. *atrosepticum* ICMP1526. Lesion lengths for SCRI1043 and ICMP1526 were substantially lower than for the other *P*. *carotovorum* subsp. *brasiliensis* strains.

2.5 Discussion

In 2008, Pitman *et al*. identified aggressive isolates of *P*. *carotovorum* subsp. *carotovorum* in New Zealand that could elicit blackleg symptoms upon stem infection and showed greater levels of maceration of potato tubers. In this study, re-examination of the New Zealand isolates confirmed that 18 of these isolates were not *P*. *carotovorum* subsp. *carotovorum*, but belonged to the related subspecies *P*. *carotovorum* subsp. *brasiliensis*. *Pectobacterium carotovorum* subsp. *brasiliensis* was initially found only in subtropical climates in Brazil (Duarte *et al*., 2004). However, recent studies have reported this species to be present in temperate climates (de Haan *et al.*,

2008) and various other geographical locations including the United States, Israel (Ma *et al*., 2007), South Africa (van der Merwe *et al.*, 2010), Canada (de Boer *et al*., 2012), Peru, Germany, Syria and Japan (Nabhan *et al.*, 2012b). Thus, *P*. *carotovorum* subsp. *brasiliensis* is likely to be an important component of the soft rot and blackleg complex of potato in New Zealand and elsewhere globally.

An increase in the number of described species and subspecies of *Pectobacterium* infecting potato in the last decade has increased the challenges in their identification and differentiation by classical microbiological tests (De Boer *et al*., 2012). Pitman *et al*. (2008) used carbohydrate utilisation patterns to differentiate a collection of isolates from potato in New Zealand. As a result, the majority of them were classified as *P*. *carotovorum* subsp. *carotovorum*. Subsequent characterisation of a number of these *P*. *carotovorum* subsp. *carotovorum* isolates using phylogenetic analysis, however, re-classified them as members of *P*. *wasabiae* (Pitman *et al*., 2010). Another recent study also demonstrated that isolates of *P*. *carotovorum* subsp. *odoriferum* were initially misidentified as *P*. *carotovorum* subsp. *carotovorum* and that *P*. *carotovorum* subsp. *odoriferum* was present in Poland as far back as 1995 (Waleron *et al.*, 2014). Similarly, about 20% of *P*. *carotovorum* strains collected from diseased plants in Syria were subsequently identified as *P*. *carotovorum* subsp. *brasiliensis* (Nabhan *et al.*, 2012b), and a number of *P*. *carotovorum* subsp. *carotovorum* strains in a laboratory culture collection of *Pectobacterium* species isolated between 1970 and 1985 were *P*. *wasabiae* (De Boer *et al*., 2012).

Initial identification of *P*. *atrosepticum* and *P*. *carotovorum* subsp. *carotovorum* species was based on the ability of these species to produce acid from α-methylglucoside and reducing sugars from sucrose (De Boer *et al.*, 1979; Graham, 1972). However, De Boer *et al*. (2012) showed that carbon utilisation profiles, growth at 37°C and RFLPs are not sufficient to differentiate *P*. *carotovorum* subsp. *carotovorum* from *P*. *carotovorum* subsp. *brasiliensis*, because carbohydrate utilisation patterns and other phenotypic characteristics vary among strains of the same species or subspecies. For example, Canadian isolates of *P*. *carotovorum* subsp. *brasiliensis* differed from Brazilian isolates of *P*. *carotovorum* subsp. *brasiliensis* in terms of producing acid from α-

methylglucoside and reducing sugars from sucrose (De Boer *et al*., 2012). These results were consistent with those of Nabhan *et al*. (2012), who also showed that isolates of *P*. *carotovorum* subsp. *brasiliensis* belonging to different genotypic clades differ in carbohydrate utilisation patterns. Interestingly, Canadian isolates were far less aggressive than those from Brazil suggesting that different variants of *P*. *carotovorum* subsp. *brasiliensis* may be responsible for different disease symptoms.

Since identification methods based on biochemical and phenotypic characterization are neither rapid nor accurate, PCR protocols have been developed to effectively differentiate *Pectobacterium* at the species and subspecies level (Darrasse *et al.*, 1994; de Haan *et al.*, 2008; Duarte *et al.*, 2004; Frechon *et al.*, 1998; Nassar *et al.*, 1996). PCR on the DNA of the New Zealand isolates collected from diseased potato using species-specific primers identified 18 of them as *P*. *carotovorum* subsp. *brasiliensis*. These strains did not produce a band indicative of *P*. *carotovorum* subsp. *carotovorum* using the EXPCCF/ EXPCCF/EXPCCR primers (Kang *et al.*, 2003a), even though they were previously described as belonging to this subspecies. Pitman *et al*. (2008) initially identified these isolates using a different PCR protocol purportedly specific to *P*. *carotovorum* subsp. *carotovorum* (Darrasse *et al.*, 1994), which produced an amplicon from those re-classified as *P*. *carotovorum* subsp. *brasiliensis*. In another study, the EXPCCF/EXPCCR primers used for detecting *P*. *carotovorum* subsp. *carotovorum* did not amplify all strains of *P*. *carotovorum* subsp. *carotovorum*, but amplified some strains of *P*. *wasabiae* (De Boer *et al.*, 2012). Pitman *et al*. (2008) also showed that *P*. *carotovorum* subsp. *carotovorum* strains identified by carbohydrate utilisation profile, did not always produce PCR products with the primers described by Darrasse *et al*. (1994). These isolates clustered closely with *P*. *wasabiae* in phylogenetic analyses using the 16S rRNA gene and MLSA (Pitman *et al*. 2010). These ambiguous results demonstrate the limitations of using a PCR protocol designed to amplify a single specific product, suggesting that a polyphasic approach should be taken for diagnosis of SRE. The presence of multiple *Pectobacterium* species in fields, even in the same plant and/or tuber, also makes PCR diagnostics difficult (Kim *et al*., 2009).

Pectobacterium carotovorum subsp. *brasiliensis* strain ICMP19477 produced a PCR amplicon using the BR1f/L1 primers and was, therefore, tentatively used to represent the *P*. *carotovorum* subsp. *brasiliensis* isolates on potato in New Zealand. Comparison of the 16S rRNA gene sequence from ICMP19477 with those of enterobacteria in the GenBank database identified this strain as *P*. *carotovorum* subsp. *brasiliensis*. Hauben *et al*. (1998) previously defined the relationships between enterobacterial pathogens by 16S rRNA sequence analyses using the neighbour-joining (N-J) algorithm, which resulted in revision of the taxonomy of this group. Using a similar set of sequences and the same method, however, Sproer *et al*. (1999) reported *Brenneria* and *Pectobacterium* as a single clade. The differences in the taxonomic structure of SRE described by Hauben *et al*. (1998) and Sproer *et al*. (1999) indicate that analyses using this criterion might be fragile. Several other studies have demonstrated that the 16S rRNA gene provides only coarse resolution of enterobacterial phytopathogens and cannot guarantee species identity due to relatively conserved rates of mutation (Ibrahim *et al.*, 1993; Young & Park, 2007). Ma *et al*. (2007) also established that phylogenies built with single genes have been used to examine the relationships of the plant-pathogenic enterobacteria and, because these single genes do not have many informative characters, they may not accurately reflect interspecies taxonomic relatedness. 16S rRNA gene sequence also fails to differentiate strains of *P*. *carotovorum* subsp. *carotovorum* and *P*. *carotovorum* subsp. *brasiliensis* because of the presence of different homologs of the 16S rRNA gene in different strains (Naum *et al.*, 2008).

Given the limitations of 16S rRNA sequence analyses, MLSA was used to confirm the identity of ICMP19477. Stanley *et al*. (2006) showed that MLSA was more effective for bacterial speciation, as 16S rRNA gene lacks resolution below the genus level. According to Nabhan *et al*. (2012a), MLSA distinguishes strains with different adaptations at the infra-subspecies level. Kim *et al*. (2009) also stated that MLSA provides the most reliable classification of *P*. *carotovorum* strains. In a previous study by Ma *et al*. (2007), phylogenetic analyses based on sequences of concatenated housekeeping genes was successfully used to cluster *Pectobacterium* subspecies into distinct clades. In this study, comparison of the concatenated sequences of *acnA*, *gapA*, *icdA*, *mdh*, *mtlD*, *pgi*, and *proA* from ICMP19477 and other related enterobacteria clustered

ICMP19477 with *P*. *carotovorum* subsp. *brasiliensis* strains from overseas, readily distinguishing it from the closely related subspecies *P*. *carotovorum* subsp. *carotovorum*. Thus, the results of *P*. *carotovorum* subsp. *brasiliensis*-specific PCR, 16S rRNA sequence analyses and MLSA are consistent with ICMP19477 being closely related to other isolates belonging to *P*. *carotovorum* subsp. *brasiliensis*.

The topology of the Bayesian tree produced from MLSA analysis showed *P*. *carotovorum* subsp. *carotovorum* to be a highly divergent pathogen. MLSA does not always allow for the unambiguous identification of *P*. *carotovorum* subsp. *carotovorum*, however (Nabhan *et al*., 2012b). The limitation of MLSA lies in the fact that reliable results are produced only if multiple strains of the same subspecies representing genetically different taxa are included in building a phylogenetic tree. For example, Ma *et al*. (2007) showed that *P*. *carotovorum* subsp. *odoriferum* SCRI482 clustered with *P*. *carotovorum* subsp. *carotovorum* strains belonging to the same clade. With the inclusion of three other *P*. *carotovorum* subsp. *odoriferum* strains in a study by Nabhan *et al*. (2012a), the strain *P*. *carotovorum* subsp. *odoriferum* SCRI482 grouped with other *P*. *carotovorum* subsp. *odoriferum* strains to form a separate cluster, demonstrating the necessity of using more than one strain for phylogenetic analysis.

Pectobacterium carotovorum subsp. *carotovorum* is prevalent on potatoes worldwide (Costa *et al.*, 2006; Serfontein *et al.*, 1991; Yahiaoui-Zaidi *et al.*, 2003). It is considered the main causal agent of soft rotting of potato tubers. Numerous studies have also isolated *P*. *carotovorum* subsp. *carotovorum* from aerial stem rot lesions in fields. This has led to controversy over whether *P*. *carotovorum* subsp. *carotovorum* can also cause blackleg. For example, De Haan *et al*. (2008) showed that infiltration of potato tubers with several isolates of *P*. *carotovorum* subsp. *carotovorum* results in a substantial level of stem infection diagnosed as blackleg. Re-examination of these isolates using comparative genomics, however, re-classified them as belonging to *P*. *wasabiae* (Nykyri *et al*., 2012). This was consistent with the high degree of stem damage caused by isolates of *P*. *wasabiae* in New Zealand (Pitman *et al*. 2010). *Pectobacterium wasabiae* was later identified in South Africa, Canada and Finland; its capacity to cause blackleg

demonstrated by stem inoculation and propagation of diseased field-grown plants produced from infiltrated tubers. In Finland, *P*. *wasabiae* was able to cause similar amounts of blackleg-like symptoms as *P*. *atrosepticum* in field experiments. Similarly, Pitman *et al*. (2008) believed that several isolates of *P*. *carotovorum* subsp. *carotovorum* from New Zealand could cause blackleg. This study, however, re-classified these isolates as *P*. *carotovorum* subsp. *brasiliensis*. Given the recent re-examination and re-classification of a variety of putative *P*. *carotovorum* subsp. *carotovorum* isolates to other species and subspecies in this study and elsewhere, it seems that *P*. *atrosepticum*, *P*. *carotovorum* subsp. *brasiliensis* and *P*. *wasabiae* may be responsible for blackleg whilst *P*. *carotovorum* subsp. *carotovorum* isolates are in fact unable to cause this disease.

Pathogenicity tests on stems of potato plants infected with the *P*. *carotovorum* subsp. *brasiliensis* strain ICMP19477 confirmed the capacity of this isolate to cause blackleg disease upon stem infection and its relatively aggressive behaviour on potato stems compared to *P*. *atrosepticum*. The rapid development of blackleg caused by ICMP19477 is consistent with the highly aggressive behavior of *P*. *carotovorum* subsp. *brasiliensis* isolates from Brazil in experiments performed by del Pilar Marquez-Villavicencio *et al*., (2011). De Boer *et al*. (2012), however, showed that isolates of *P*. *carotovorum* subsp. *brasiliensis* in Canada were far less aggressive. Although these isolates conformed to the subspecies in PCR and MLSA, they differed from those in New Zealand, Brazil and other countries in diagnostic biochemical tests. This suggests that the virulence of *Pectobacterium* species attributed to blackleg disease is not restricted to species level but is strain dependent. It also suggests that the capacity to cause blackleg is associated with the presence of virulence genes that have been acquired by HGT. Furthermore, because all *Pectobacterium* species have overlapping symptomologies, it has been suggested that use of the term 'blackleg' disease caused by *P*. *carotovorum* subsp. *carotovorum*, *P*. *carotovorum* subsp. *brasiliensis* and *P*. *wasabiae* be avoided despite the similar symptomologies (De Boer *et al*., 2012). Instead, all stem rots caused by *Pectobacterium* and *Dickeya* species should be designated under a single name such as 'bacterial stem rot', since, in the past, blackleg has specially been referred to the disease caused by *P*. *atrosepticum* (De Boer *et al*., 2012).

The prevalence of *P*. *carotovorum* subsp. *brasiliensis* in Canterbury is of particular importance given the high proportion of seed production that takes place in this region. Globally, *P*. *atrosepticum* is usually associated with infections of seed tubers. *Pectobacterium carotovorum* subsp. *brasiliensis* appears to out-compete *P*. *atrosepticum*, however, as *P*. *atrosepticum* does not occur if *P*. *carotovorum* subsp. *brasiliensis* is present (De Boer, 2004). This may be why only four of 89 isolates collected from potato in New Zealand were *P*. *atrosepticum* whereas 18 were *P*. *carotovorum* subsp. *brasiliensis*, despite the climate being temperate and considered more habitable for *P*. *atrosepticum* (Pitman *et al*. 2008). Blackleg and soft rot assays conducted by del Pilar Marquez-Villavicencio *et al*. (2011) also showed *P*. *atrosepticum* to be less aggressive than *P*. *carotovorum* subsp. *brasiliensis* on stems and tubers. The mechanism associated with this dominance is not understood but some *P*. *carotovorum* subsp. *brasiliensis* strains such as ATCC BAA 417 produce carbapenem, an antimicrobial that is predicted to be involved in inhibiting growth of *P*. *atrosepticum* and *P*. *carotovorum* subsp. *carotovorum in vitro* (del Pilar Marquez-Villavicencio *et al.*, 2011). Interestingly, other *P*. *carotovorum* subsp. *brasiliensis* isolates used in the study did not inhibit other *Pectobacterium* species (both *P*. *atrosepticum* and *P*. *carotovorum* subsp. *carotovorum*), suggesting this *in vitro* phenotypic trait to be strain-specific rather than species-specific (del Pilar Marquez-Villavicencio *et al.*, 2011). The biological explanation for the antagonistic effect of ATCC BAA 417 against other *Pectobacterium* species *in vitro* is unknown since the same was not observed when ATCC BAA 417 was co-inoculated with *P*. *carotovorum* subsp. *carotovorum* in stems of potato plants (del Pilar Marquez-Villavicencio *et al.*, 2011). This suggests that the *Pectobacterium* strains are in competition as maceration progresses and nutrient availability decreases in other potato tissue.

In conclusion, the results of this study demonstrate that *P*. *carotovorum* subsp. *brasiliensis* should not be underestimated as a causal organism of potato blackleg in temperate regions such as New Zealand. Latent infection of seed tubers may result in significant disease levels in the following growing season and therefore, increase the economic risk of repetitive potato production especially when retaining tubers for the subsequent seasons. Given the climatic conditions in many growing regions of New Zealand are favourable for *P*. *carotovorum* subsp. *brasiliensis*

strains, this new subspecies adds additional complexity to the blackleg/soft rot complex on potatoes. Furthermore, given its wide distribution, it may in fact provide the greatest risk of all enterobacterial pathogens on potato production.

Since *P*. *carotovorum* subsp. *brasiliensis* strains are highly virulent, it is recommended to use seed potatoes free of virulent *P*. *carotovorum* subsp. *brasiliensis* strains able to cause blackleg. The difficulty distinguishing SRE by conventional biochemical and traditional phylogenetic means (i.e., 16S rRNA gene sequence analysis), however, that continued monitoring of the presence and spread of these pathogens using alternative molecular techniques is important. This requires the development of reliable methods able to distinguish virulent from non-virulent strains, which can be achieved by identifying and targeting bacterial genes through genome sequencing and comparative genomic analyses. Indeed, with the advent of new economical sequencing technologies, it is predicted that comparative studies of full genome sequences will become the standard method for defining bacterial species (Zeigler, 2003; Chun & Rainey, 2014). Whole genome studies of SRE could shed light on the bacterial genes required for blackleg symptoms and the distribution of these genes among *Pectobacterium* strains. In turn, these may be useful for diagnosis, and for understanding the relative risk of economic loss. In the following chapters, the genome of the newly classified *P*. *carotovorum* subsp. *brasiliensis* strain ICMP19477 was sequenced and annotated. The genome was then compared to that of other *Pectobacterium* strains to identify genes that may contribute to blackleg symptoms or the aggressive nature of *P*. *carotovorum* subsp. *brasiliensis* in New Zealand.

Chapter 3

Next Generation Sequencing (NGS) of *Pectobacterium* strains from potato

3.1 Abstract

MLSA revealed the taxonomic status of several highly aggressive isolates of *Pectobacterium* on potato in New Zealand as that of *P. carotovorum* subsp. *brasiliensis*. These isolates can cause blackleg disease. Here, we sequenced and annotated the genome of a representative of this group of *P. carotovorum* subsp. *brasiliensis* strains, ICMP19477, to identify the genetic factors that contribute to its colonisation of potato stems. For comparative purposes, the genome sequences were also obtained for the type strain for *P*. *carotovorum* subsp. *carotovorum* (ICMP5702), the type strain for *P*. *atrosepticum* (ICMP1526) and an atypical blackleg-causing potato isolate from Peru (*P*. *carotovorum* subsp. *carotovorum* UGC32). These isolates were considered representative of distinct lineages of the genus *Pectobacterium* that cause distinctive disease symptoms. Either 454 massively parallel pyrosequencing or HiSeq Illumina sequencing were used to obtain the genome sequences of these strains. Assembly of the raw sequence reads for each individual generated draft genome sequences that remained in a number of contigs. These sequences identified the gene composition, genome organisation for each genome and revealed the putative biological roles for many of the genes. The sequenced genomes will contribute to future comparative genomic studies using the complete genome of *P*. *atrosepticum* SCRI1043 and other previously sequenced *Pectobacterium* strains.

3.2 Introduction

In Chapter 2, the identities of several highly aggressive isolates of *P*. *carotovorum* subsp. *carotovorum* collected from potatoes in New Zealand were re-examined, as their capacity to cause blackleg was considered unusual amongst this group of pectolytic bacteria. Consequently, these isolates were re-classified as belonging to *P*. *carotovorum* subsp. *brasiliensis*, a highly

aggressive subspecies of *P*. *carotovorum* responsible for blackleg of potato in Brazil and South Africa (Duarte *et al.*, 2004; van der Merwe *et al.*, 2010). Their identities were defined primarily by MLSA. MLSA is sufficient to distinguish the various species and subspecies of *Pectobacterium* (Kim *et al*., 2009), but provides little information on the basis of differences in virulence. Thus, genome sequencing was applied as a more comprehensive method for taxonomic classification of the bacterium as well as for the identification of candidate genetic factors associated with the aggressiveness of *P. carotovorum* subsp. *brasiliensis*. The sequencing of the genomes of plant pathogenic bacteria has enhanced strain identification and typing, and has also helped to identify genetic patterns related to the virulence of pathogens (Vidaver & Lambrecht, 2004).

The sequencing of prokaryote genomes started in earnest in 1995, with the publication of the DNA sequences of *H. influenza* and *Mycoplasma genitalium* (Fleischmann *et al.*, 1995; Fraser *et al.*, 1995). It wasn't however, until the completion of the genome sequence of *Xylella fastidiosa*, the causative agent of citrus variegated chlorosis (Simpson *et al.*, 2000), that the genomics era finally reached plant pathology. The last decade has seen a proliferation in the sequencing of genomes from plant pathogenic bacteria. This has enabled various questions to be addressed including: What are the molecular determinants of pathogenicity? Are there distinctive features that will predict a bacterium's ability to survive on a host? How has the bacterium evolved? How does their sequence representation among related organisms shape phenotypic differences? And ultimately, how can information derived from sequence data be put to use for improved management of plant disease (Lindeberg, 2012)?

The major steps in a whole genome sequencing are: i) construction of small and large genomic libraries, ii) random selection of clones that cover the genome up to 10× fold coverage for next generation sequencing, iii) *in silico* sequence assembly of the raw sequence reads into contigs and scaffolds, iv) closure of the remaining gaps and, v) genome annotation including sequence analysis, structural gene finding and assignment or prediction of functions to gene products (Fraser & Fleischmann, 1997). The relatively small genome size of bacterial pathogens (0.5 to 10 Mb) makes them perfectly suited to the whole genome shotgun (WGS) approach (Fraser &

Fleischmann, 1997). A number of sequencing technologies have been developed for whole genome sequencing that emphasizes speed without too much loss of accuracy. Initially, these techniques expanded upon the so-called Sanger process that was first developed in the 1970s, gradually automating this process and increasing the number of samples that could be sequenced at one time. However, more recently, researchers have increasingly begun to rely on newer and even faster automated methods, including 454 sequencing and Illumina techniques. The automated Sanger method is considered a 'first-generation' technology, while newer methods are referred to as 'NGS' technologies. These newer technologies constitute various strategies that rely on a combination of template preparation, sequencing and imaging accompanied by genome assembly and alignment methods (Metzker, 2009). The major advance offered by NGS is the ability to produce an enormous volume of data cheaply; in some cases in excess of one billion short reads per instrument run (Horner *et al.*, 2009; Metzker, 2009).

NGS sequencing techniques are characterized by the production of hundreds of thousands or millions of short reads (25 – 500 bps) in a short span of time. Platforms enabling this include the Roche/454 pyrosequencing (Droege & Hill, 2008), the Illumina Genome Analyser (Bennett, 2004), Applied Biosystems SOLiD (Porreca *et al.*, 2006), the Helicos Heliscope sequencer (Harris *et al.*, 2008), Ion Torrent semiconductor sequencing (Merriman *et al.*, 2012) and more recently, the PacBio RS (English *et al.*, 2012). A detailed review of these methods is beyond the scope of this introduction and a brief comparison of these methods is detailed in (Table 3.1).

Table 3.1 Technical specifications of NGS platforms

Method	SMRT	Ion semiconductor	Pyrosequencing (454)	Sequencing by synthesis (Illumina)	Sequencing by ligation (Solid)	Sanger sequencing
Read length (in bps)	5,500 - 8,500	400	700	50 - 300	85 - 100	400 - 900
Accuracy (in %)	99.999	98	99.9	98	99.9	99.9
Reads per run (bases)	~400 megabases	~80 million	20 - 40 million	~3 billion	1.2 - 1.4 billion	N/A
Time per run	30 mins - 2 hrs	2 hrs	1 day	1 - 10 days depending upon sequencer and specified read length	1 - 2 weeks	20 mins - 3 hrs
Cost per 1 million bases (in USD)	$0.33 – $1	$1	$10	$0.05 - $0.15	$0.13	$2400
Merits	Longest read length, fast	Less expensive equipment, fast	Long read size, fast	Potential for high sequence yield	Low cost per base	Low individual reads
Demerits	Equipment expensive, moderate throughput	Homopolymer errors	Expensive runs, homopolymer errors	Expensive equipment, requires high concentrations of DNA	Slow, have issues sequencing palindromic s sequences	Very expensive
Company	Pacific Biosciences, CA	Ion Torrent, Life Technologies	454 sequencing, Roche	Illumia sequencing, Illumina	Solid, Life Technologies	
References	Eid *et al*., 2009	Rusk, 2011	Margulieus *et al*., 2005	Bentley *et al*., 2008	Shendure *et al*., 2005	Sanger *et al*., 1977

At the beginning of this PhD project (2010), sequencing a whole genome, together with the challenges of finishing and annotation, required resources that were beyond the scope of a typical laboratory. It was not yet practical and routine to sequence every individual genome using the traditional methods. It was only possible to produce better quality genome finishing with the combined advances of Sanger sequencing and pyrosequencing technologies. Today, many finished and unfinished genome sequences have been made publicly available using NGS technologies, and several databases have been established for bacterial genome sequence analyses. As a result of these advances, the rate at which genomes are being sequenced has increased exponentially leading to a reduction in time and the costs for sequencing a genome. From 2000 to 2010, the speed of DNA sequencing increased 500,000-fold (Metzker, 2009). This highlights the ever-increasing need for fast, accurate and readily available annotation tools in order to avoid a) a bottleneck at the step of annotations of these new genomes; and b) inaccurate or incomplete genome annotations adding misinformation to the databases upon which comparative genomic approaches are heavily reliant. The production of billions of NGS

reads has also challenged the infrastructure of existing information technology systems in terms of data transfer, storage and quality control, computational analysis to align or assembly read data, and laboratory information management systems for sample tracking and process management (Metzker, 2009). Advances in bioinformatics are ongoing, and improvements are needed if these systems are to keep pace with the continuing developments in NGS technologies (Metzker, 2009).

The following sections highlight the two NGS methods (454 pyrosequencing and Illumina sequencing) that have been used in this thesis to sequence the genomes of *Pectobacterium*. Pyrosequencing technology was the first alternative to the conventional Sanger method for *de novo* DNA sequencing, based on the sequence-by-sequence principle (Hyman, 1988; Melamede, 1989) and on the detection of released pyrophosphate (PPi) during DNA synthesis (Ronaghi, 2001). This method employed a series of four enzymes (DNA polymerase, ATP sulfurylase, luciferase and apyrase) and the substrates adenosine phosphosulfate and luciferin, to accurately detect nucleic acid sequences during synthesis (Gharizadeh *et al*., 2007). The sequencing primer was hybridized to a single-stranded DNA biotin-labeled template and mixed with the enzymes. The pyrosequencing method then relies on the luminometric detection of pyrophosphate that is released during primer-directed DNA polymerase catalysed nucleotide composition (Gharizadeh *et al.*, 2003). In addition to *de novo* genome sequencing, the pyrosequencing method was also used in the analyses of secondary structures, such as hairpin structures (Ronaghi *et al.*, 1999), analysis of single-nucleotide polymorphisms (Ahmadian *et al.*, 2000; Alderborn *et al.*, 2000; Nordström *et al.*, 2000) and mutation detection (Garcia *et al.*, 2000). Pyrosequencing had potential advantages of accuracy, flexibility, parallel processing and could be easily automated (Fakruddin & Chowdhury, 2012). Furthermore, this technique avoided the need of labelled primers, labelled nucleotides, and gel-electrophoresis (Fakruddin *et al*., 2012).

The first DNA pyrosequencing platform was developed by 454 Life Sciences Corporation (Branford, CT) which employed a highly multiplexed flow-through array capable of identifying 20 - 40 million bases per run. Sequencing was performed on randomly fragmented DNA using

microbead-based pyrosequencing chemistry which was hundreds of times faster than standard sequencing methods and was capable of sequencing 200,000 fragments per four-hour run (Margulies *et al.*, 2005). This increase in throughput came at the expense of read length, with pyrosequencing reads reaching about 700 bps in length on average. The major drawback of pyrosequencing technology was the incomplete extension through homopolymers and the assembly of pyrosequencing sequences from samples that contained large repetitive DNA sequences, which proved problematic for conventional fragment assembly programs (Margulies *et al*., 2005).

The Illumina sequencing method relies on sequencing by synthesis method that uses four fluorescently-labelled nucleotides to sequence the fragmented genomic DNA prepared in a sample library attached to a planar, optically, transparent surface on a flow cell surface *in parallel*. A single labelled deoxynucleitide triphosphate (dNTP) is added to the nucleic acid chain, during each sequencing cycle. The nucleotide label serves as a terminator for polymerization, so after each dNTP incorporation, the fluorescent dye is imaged to identify the base and then enzymatically cleaved to allow incorporation of the next nucleotide (Illumina Sequencing, Technology Spotlight, www.illumina.com/systems/nextseq-sequencer/technology.ilmn). Since, all four reversible terminator-bound dNTPs are present as separate single molecules, natural competition minimizes incorporation bias. Base calls are made directly from signal intensity measurements during each cycle and the images are compiled and processed to produce base sequences of each DNA sequence. This technique greatly reduces raw error rates compared to other technologies (Illumina Sequencing, Technology Spotlight, www.illumina.com/systems/nextseq-sequencer/technology.ilmn). This parallel approach generates close to 400 billion bases with high accuracy, with 1.3 billion reads per flow cell run as paired-end 250 bps reads. This method generates highly accurate base-by-base sequencing with minimal sequence-context specific errors, enabling robust base calling across the genome, including repetitive sequence regions (Illumina Sequencing, Technology Spotlight).

NGS technologies outperform traditional Sanger methods in read length throughput by a factor of 100 - 1000 compared to Sanger methods (Kircher & Kelso, 2010). Furthermore, refinements in

the current NGS technologies are resulting in a steady improvement in read length throughput and accuracy. However, upcoming technologies referred to as 'third generation' technologies may soon supersede NGS platforms in terms of increased throughput plus reduction in time to result and cost by eliminating the need for excessive reagents and harnessing the processivity of DNA polymerase (Schadt *et al.*, 2010). 'Third generation' sequencing methods under development include, labelling the DNA polymerase (Visigen Biotechnologies Inc., 2010), the rolling circle replication method (DNA nanoball sequencing, Complete Genomics), reading the sequence as a DNA strand transits through nanopore technology such as Oxford Nanopore's BASE platform (Clarke *et al.*, 2009) and IBM's proposed silicon-based nanopore technology (IBM Research, 2009), methods based on tunnelling electron microscopy (Tanaka & Kawai, 2009), and microscopy-based techniques, such as atomic force microscopy or transmission electron microscopy that are used to identify the positions of individual nucleotides within long DNA fragments (>5000 bps) by labelling nucleotides with heavier elements (e.g., halogen) for visual detection and recording (Xu *et al.*, 2009). Other methods include RNAP (RNA polymerase sequencing) (Greenleaf & Block, 2006) and *in vitro* virus high-throughput sequencing (IVV-HiTSeq) (Fujimori *et al.*, 2012). More recently, Pacific Biosciences have also released a powerful SMRT (single molecule real time) platform, producing read lengths of over 1 Mb and involving sequencing runs that can take as little as 30 minutes of instrument run-time (Korlach *et al.*, 2010). Ion Proton by Torrent (2012) promises to sequence a 'human exome' in a few hours and NIH/NHGRI have set up a much-publicised goal of a $1,000 genome (http://granths.nih.gov/grants/guide/rfa-files/rfa-hg-09-011.html).

The process of deciphering the sequence of a genome from the small constituent DNA fragments is called 'assembling' the genome, for which the extent of fragment read overlap is a crucial factor. The essential steps of genome assembly include: (i) quality analysis and trimming of sequence reads, (ii) detection of overlapping reads which are then grouped to form contigs, (iii) a multiple sequence alignment of the reads to form a scaffold and the construction of a consensus sequence for each contig layout, and (iv) identification of possible sites of misassembly by combining manual inspection with quality value validation (Huang & Madan, 1999). The first step usually

involves screening the base reads for vector contamination and low quality regions are often trimmed and removed (Myers *et al.*, 2000).

Various assemblers are used to assemble the reads into contigs and scaffolds. The placement of reads along a reference genome implicitly defines a set of contigs or contiguous regions of the assembly, as well as the relative order and orientation of these contigs in a structure, commonly known as a scaffold (Roach *et al.*, 1995). The majority of assemblers use algorithms by Needleman & Wunsch (1970), Smith & Waterman (1981) and Gotoh (1982). The construction of contigs and scaffolds consists of two major steps: (i) each overlap detection is evaluated based on the depth of coverage of two regions in the overlap by finding exact identical sub-sequences (often called k-mers), and (ii) reads are assembled into contigs based on unique overlaps. The sizes of the sub-sequences vary with different sequential platforms, and are dynamic in some assemblers. Contigs are then corrected and linked based on pairs. This process is called scaffolding, where genomic contigs are organized into larger frameworks and scaffolds are also corrected based on read pairs. The scaffold build is considered the true representation of the genome, although there will still be gaps between contigs, which are dealt with in the subsequent process of genome finishing (Scheibye-Alsing *et al.*, 2009). Finally, the generation of a consensus sequence for each contig is based on multiple alignments of reads in the contig. A series of rounds of multiple alignments is used for constructing a consensus sequence: the reads in a contig are sorted in an increasing order of their position in the contig, followed by repeatedly aligning the current read with the current alignment, and the resulting alignment is the current alignment for the iteration. The reads in the contig are considered one by one, in order. The current alignment is empty in iteration 1 and the current reads become the current alignment for iteration 2. It is assumed that the final consensus sequence corresponds to the original genomic sequence where the sequenced fragment originated.

Genome finishing involves determining the order and orientation of the consensus sequences of contigs obtained from the assemblies of random draft genomic sequences (Bastide & McCombie, 2007; Lee & Vega, 2004). For this, specific parts of the assembly are generally re-examined, due

to low quality of data, low coverage of the sequence, sites under suspicion due to mis-assembly, and numerous copies of insertion sequences or repeats. Thus, it is necessary to break these contigs apart at the repeated sequences and individually join the proper flanking regions using paired-end information, or using results of comparisons using a similar genome (referred to as 'reference' genome).

A typical genome assembly results in a number of contigs and scaffolds separated by gaps. The sequences of the genomic regions inter-spanning these gaps are generated by Sanger sequencing the amplified products generated from PCR techniques (primer pairs being designed across the contiguous segments). This is assisted greatly once one or two genomes of a given organism are sequenced to completion, such that the diversity of the species can be accessed through comparative genome hybridization using microarrays (Tettelin *et al.*, 2001), or by generating draft sequences of other stains of interest, and comparing them to a reference genome. Otherwise, a major drawback of using gap closure for genome finishing is that there is no prior knowledge of the estimated gap size, which is important to determine the presence of any missing genes in the gap regions. Also, manual closure of incomplete regions by Sanger sequencing is expensive. Hence, draft genome assemblies submitted to sequence databases often lack a significant number of nucleotides. Importantly, these gaps may contain essential functions for the organism studied. Once the genomes have been closed, low quality regions can be addressed by re-sequencing. Concurrently, the development of genome closure softwares such as GapFiller (Boetzer & Pirovano, 2012), will also overcome difficult genomic regions that cannot be covered by draft assemblies.

Following genome assembly and alignment the bacterial genome is annotated. Microbial genome annotation often consists of running an automatic annotation pipeline by combining the results of several gene prediction programs followed by manual curation of the results (MacLean *et al.*, 2009). Genome annotation comprises two distinct stages: (i) structural annotation that involves the correct identification and localization of distinct sequence elements such as genes, regulatory

elements, transposons, repetitive elements, etc., and (ii) functional annotation that attempts to predict the biological functions of the above mentioned elements.

In annotating bacterial genomes, gene prediction is a crucial step that identifies (long) open reading frames (ORFs). For a known sequence, ORFs are predicted using a variety of programs such as, GeneMarkS (Besemer *et al.*, 2001), GENESCAN (McEvoy *et al.*, 1998), Glimmer (Aggarwal & Ramaswamy, 2002), and Glimmer3 (Delcher *et al.*, 2007). Once the structural elements in a genome have been identified, the molecular functional and biological roles are predicted for each of these elements. Computational functional annotation can be divided into two parts; functional annotation based on homology or functional annotation based on protein signature. Homology-based functional annotations use sequence similarity-based gene finders or databases such as BLAST (Altschul *et al.*, 1997) or HMMer (Finn *et al.*, 2009; Sonnhammer *et al.*, 1997) to find protein coding regions that are similar to a query sequence, and which preferably have a known experimentally-defined function. BLAST compares the input or 'query' sequence to sequences in a database and identifies matches or 'subject sequences' that are similar to the given sequence. The limitation of homology based protein prediction is that the presence of structural domains and motifs that make up a protein are not properly analysed.

Protein-coding gene annotation or functional annotation is accompanied by the annotation of tRNA genes such as tRNAscan-SE (Lowe & Eddy, 1997) and programs to detect other non-coding RNA genes and other genomic features (Achaz *et al.*, 2007; Gardner *et al.*, 2009; Lagesen *et al.*, 2007; Langille *et al.*, 2008). Predictions of protein functions are inferred from homology to similar genes in other species with assigned functions. In protein-based annotation, motifs and domains can be discovered and extracted as profiles from multiple alignments of proteins with known similar functions using Hidden Markov Models (HMMs) (Gowri *et al.*, 2003). The results of these searches yield preliminary information for use in other analyses such as the function of proteins with a homologous sequence or cellular localisation (Dyrløv Bendtsen *et al.*, 2004). New protein sequences are analyzed using InterPro (Apweiler *et al.*, 2001; Mulder *et al.*, 2003), a database of

protein families, domains and functional sites in which identifiable features found in known proteins can be applied to unknown protein sequences (Quevillon *et al.*, 2005).

Homologous genes are not always available in databases, however, and other platforms are needed to provide a host of other information about protein-coding genes, such as chemical properties of the protein, cellular localisation and modular structure of the protein (Médigue & Moszer, 2007). Various prediction programs such as PSORTb (Gardy *et al.*, 2005), SignalP (Dyrløv Bendtsen *et al.*, 2004) and TMHMM (Krogh *et al.*, 2001) are available to detect the sub-cellular location of potential proteins. These programs use different methods of finding signal sequences or transmembrane segments to look at the amino acid content required to produce their predictions (Nakai & Horton, 1999). It is also becoming increasingly common to include software to reconstruct metabolic networks in the annotated genome (Maltsev *et al.*, 2006; Reed *et al.*, 2006; Schneider *et al.*, 2010). The last step is to employ a graphical user interface in order to present this information to the user and allow manual modification.

Automated pipelines are available that often integrate results from a number of protein-coding gene prediction programs to generate a results set that is as complete and accurate as possible. These pipelines use homology methods to transfer information from a closely related reference genome to the new sequence and can lead to the introduction and propagation of poor annotation and errors, however. Hence, manual curation is of utmost importance to catch and remove these errors. Dramatic advances in sequencing technologies, however, have opened possibilities to sequence multiple microbial genomes in a single day at low cost using a single sequencing machine so it is no longer feasible to manually curate the annotation of all sequenced genomes (Richardson & Watson, 2013). Introducing evidence tags/qualifiers stating how the annotation was assigned gives the user an idea of the reliability of the reference genome (Richardson & Watson, 2013). The concept of assigning a level of quality to annotation is not novel, but is seldom used (Gilks *et al.*, 2005; Janssen *et al.*, 2005).

Genome sequencing and annotation of phytopathogens such as *P. syringae* pv. *tomato*, *R. solanacearum*, *X. campestris* pv. *campestris* and *X. axonopodis* pv. *citri, X. fastidiosa*, *A. tumefaciens* and *P. atrosepticum* has yielded a wealth of information on novel and shared candidate phytopathogenicity determinants (Bell *et al.*, 2004; Buell *et al.*, 2003; da Silva *et al.*, 2002; Goodner *et al.*, 2001; Simpson *et al.*, 2000; Van Sluys *et al.*, 2002; Wood *et al.*, 2001). Many of these genomes were sequenced using Sanger DNA sequencing, however, (Sanger & Coulson, 1975) meaning the time and resources required to perform the tasks were prohibitive for most laboratories. In the wake of advances in DNA sequencing technologies, a plethora of bacterial genomes were sequenced and their sequences deposited into public databases (Mardis, 2008; Metzker, 2009; Shendure & Ji, 2008). In addition to new bacterial species, the genomes of multiple strains of the same bacterial species or subspecies were sequenced. This has led to a greater understanding of economically destructive variants of plant pathogenic bacteria and their emergence. The sequencing of isolates of *P. syringae* pv. *actinidae* (PSA) from outbreaks of kiwifruit canker in New Zealand and Italy, for example, revealed that these isolates represented a distinct lineage from those isolates of Japan and Korea (Mazzaglia *et al.*, 2012). Genomic studies suggest that the PSA isolates found in New Zealand and Italy are inhibited *in vitro* by copper-based compounds and antibiotics, unlike strains from Korea and Japan (Ferrante & Scortichini, 2010). Also within New Zealand, two biovars of PSA that differ in virulence have been identified (Vanneste *et al.*, 2012). One biovar (PSA-virulent, Psa-V) causes angular leaf spots and secondary symptoms on infected orchards, whereas the other biovar shows no symptoms beyond leaf spotting (known as Psa-LV, low virulence). Genome sequencing revealed that the Psa-V isolates had the greatest numbers of virulence genes such as effector proteins, phage integrases or transposases and genes orthologous to those in known vascular and woody pathogens (*P. sringae* pv. *lachrymans*, *P. syringae* pv. *pisi*, *P. syringae* pv. *aesculi* and *X. axonopodis*) (McCann *et al.*, 2013).

At the beginning of this thesis, only one 'complete' genome (*P. atrosepticum* SCRI1043) and three highly fragmented draft genome sequences (for *P. carotovorum* subsp. *carotovorum* WPP14, *P. carotovorum* subsp. *brasiliensis* ATCC BAA 417 and *P. wasabiae* WPP163) were available in the

public domain for *Pectobacterium* isolated from potatoes. Various virulence factors were identified in *Pectobacterium* that might contribute to virulence such as the PCWDEs, T3SS, flagella, coronafacic acid etc. (Bell *et al*., 2004). These virulence factors were thought to enable the pathogen to interact with its plant host in a 'stealth-like' manner, manipulating the responses of the host to invasion in an attempt to bypass defence systems (Toth & Birch, 2005). Genome sequencing revealed that the majority of virulence factors identified in these genomes are encoded on GIs. Forty-two 'variable regions' (VRs), which include, islands longer and shorter than 10 Kb were found in at least one of these *Pectobacterium* strains (Glasner *et al*., 2008). VRs correspond to a likely single evolutionary event, such as acquisition of a prophage, or a cluster of genes related to a common biological process (Glasner *et al*., 2008). These VRs encode for numerous secretion systems that contribute to virulence, polyketide and peptide synthetases, motility and taxis genes, phage remnants, various transporters and regulators and other gene islands such as galactonate, gluconate and sucrose isomerase islands (Glasner *et al*., 2008).

To confirm the taxonomic relationships of the isolates of *P. carotovorum* subsp. *brasiliensis* from New Zealand with other *Pectobacterium* and to identify the genetic factors responsible for pathogenicity and/or aggressiveness of the pathogen, the genome was sequenced for the representative strain *P. carotovorum* subsp. *brasiliensis* ICMP19477. For comparative purposes, the genomes of the type strains for *P. atrosepticum* (ICMP1526) and *P. carotovorum* subsp. *carotovorum* (ICMP5702) were also sequenced, as ICMP1526 causes blackleg whereas ICMP5702 does not. The type strain is usually the firstly isolated strain of the species, and exhibits all of the relevant phenotypic and genotypic properties cited in the species circumscriptions (Kim *et al.*, 2014). The genome of UGC32 was also sequenced as this was previously identified as a *P. carotovorum* subsp. *carotovorum* strain with the capacity to cause blackleg (Slawiak & Lojkowska, 2009). It was predicted that the assembly and annotation of these genomes and their comparison with the genome sequences for other *Pectobacterium* held in the public databases would aid in the identification of factors that support the invasion of the xylem, which results in blackleg development in potatoes.

3.3 Methods

3.3.1 Selection of *Pectobacterium* strains and DNA isolation

The bacterial strains used for genome sequencing are listed in Table 3.2. For DNA extraction, *Pectobacterium* strains were streaked onto LB agar and incubated overnight at 28°C. Individual isolated colonies were then incubated overnight in 5 mL of LB medium at 28°C with shaking at 180 rpm.

Table 3.2 Strains of *Pectobacterium* used for genome sequencing

Strain Designation	Species/Identity	Disease on potatoes	Origin	Genbank Accession no.
ICMP19477	*P. carotovorum* subsp. *brasiliensis*	Blackleg	New Zealand	ALIU00000000
ICMP1526	*P. atrosepticum*	Blackleg	Scotland	ALIV00000000
ICMP5702	*P. carotovorum* subsp. *carotovorum*	Soft rot	Denmark	AODT00000000
UGC32	*P. carotovorum* subsp. *carotovorum*	Blackleg	South Africa	AODU00000000

3.3.2 Genome Sequencing, Assembly and Alignment

For genome sequencing, genomic DNA was isolated from overnight cultures using the methods described previously in Chapter 5 (See Section 5.3.2.2).

3.3.2.1 454 pyrosequencing

454-platform pyrosequencing (Margulies *et al.*, 2005) of ICMP19477 and ICMP1526 was conducted using a FLX 454 sequencer by the Liverpool Advanced Genomics Facility (Liverpool University, UK) to generate DNA sequence reads for ICMP19477 and ICMP1526. Sequence data from the 454 sequencing platform was received in the standard flowgram format (SFF). Genome assembly of the SFF file was performed using the NewblerGS de novo assembler software (www.454.com) run on an UNIX platform (pgenome server, The New Zealand Institute for Plant and Food Research, New Zealand) to convert the sequence reads into large contigs. The output files were the contig files in .fasta format and .ace format. The .ace files were analyzed using

TABLET software (Milne *et al.*, 2010) to generate information regarding the length and coverage of the contigs. Contigs with a low coverage (< 250) were discarded.

3.3.2.2 HiSeq Illumina sequencing

A HiSeq 2000 system from Illumina sequencing services (http://www.illumina.com/systems/hiseq_2000_1000.ilmn) was used to generate paired-end reads for the genomes of *P. carotovorum* subsp. *carotovorum* UGC32 and *P. carotovorum* subsp. *carotovorum* ICMP5702. Initial processing of the Illumina sequence reads involved removal of adapter sequences and trimming of the sequence reads using FastQC software (Baraham Bioinformatics, UK). Duplicate reads were then eliminated using an in-house script; remove_duplicates.pl (Appendix A3.1) to ensure high accuracy of scaffold construction using paired-end reads. The SOAP *de novo* software v2.04 (Luo *et al.*, 2012) was used to assemble the trimmed pair-ends to obtain contigs and scaffolds. The default parameters (rank = 1, pair_num_cutoff = 3, asm_flags = 3, map_len = 32) were used for generating contigs except for the Kmer parameter. The –K parameter was optimised separately to obtain the best assembly for each genome by examining an assortment of values in the range 17 - 61. A K value of 37 was identified as the most effective value for assembly. Finally, assembly gaps for Illumina assembled reads were filled using the GapCloser v1.12 tool from SOAP (Luo *et al*., 2012). Assembly was performed iteratively with default parameters five times.

To obtain a draft sequence from both 454 and Illumina sequence data, the assembled contigs were scaffolded, ordered and oriented with respect to the reference genome sequences of SCRI1043 or PC1 with the aid of the MUMmer alignment tool (Delcher *et al.*, 1999b) and MAUVE 2.0 (Darling *et al.*, 2004). PC1 is a *P*. *carotovorum* subsp. *carotovorum* strain (later classified as *P*. *carotovorum* subsp. *aroidearum*) isolated from *Ornithogalum dubium* (Nabhan *et al*., 2012).

3.3.3 Assembly quality assessment

The quality of the contig assembly for each strain was assessed by calculating the N50 of the sequences assembled for each genome. The N50 is the length of the smallest contig in the set that contains the fewest (largest) contigs whose combined length represents at least 50% of the assembly (Miller *et al.*, 2010). Essentially, the N50 captures how much of the assembly is covered by relatively large contigs.

3.3.4 Gap filling and closure of assembled *Pectobacterium* genomes

Initial assembly of the sequence reads generated a number of contigs for each genome but also contained various miss-assemblies due to insertion sequence elements and repetitive regions (Mark Fiers, personal communication; Pop & Salzberg, 2008). To provide a more comprehensive genome annotation, gap closure was performed by PCR using primer pairs designed across un-linked contiguous sequences (known as gaps). Primer Express software version 3.0 (Applied Biosystems) was used to design the primers. PCR reactions were then performed with these primers using the genomic DNAs of the corresponding *Pectobacterium* strains as a template. Amplified PCR products generated from the PCR reactions were sequenced using the Sanger sequencing method (Macrogen, Korea). Based on the length of the PCR products, primer walking was carried out to obtain the full DNA sequence of the amplified fragment and consequently fill the gaps across the contig ends. Re-assembly of the contigs for each strain was then repeated manually to generate 'finished' draft sequences of the *Pectobacterium* genomes (ICMP19477 and ICMP1526). The schematic pipeline for whole genome sequencing and assembly is represented in Figure 3.1.

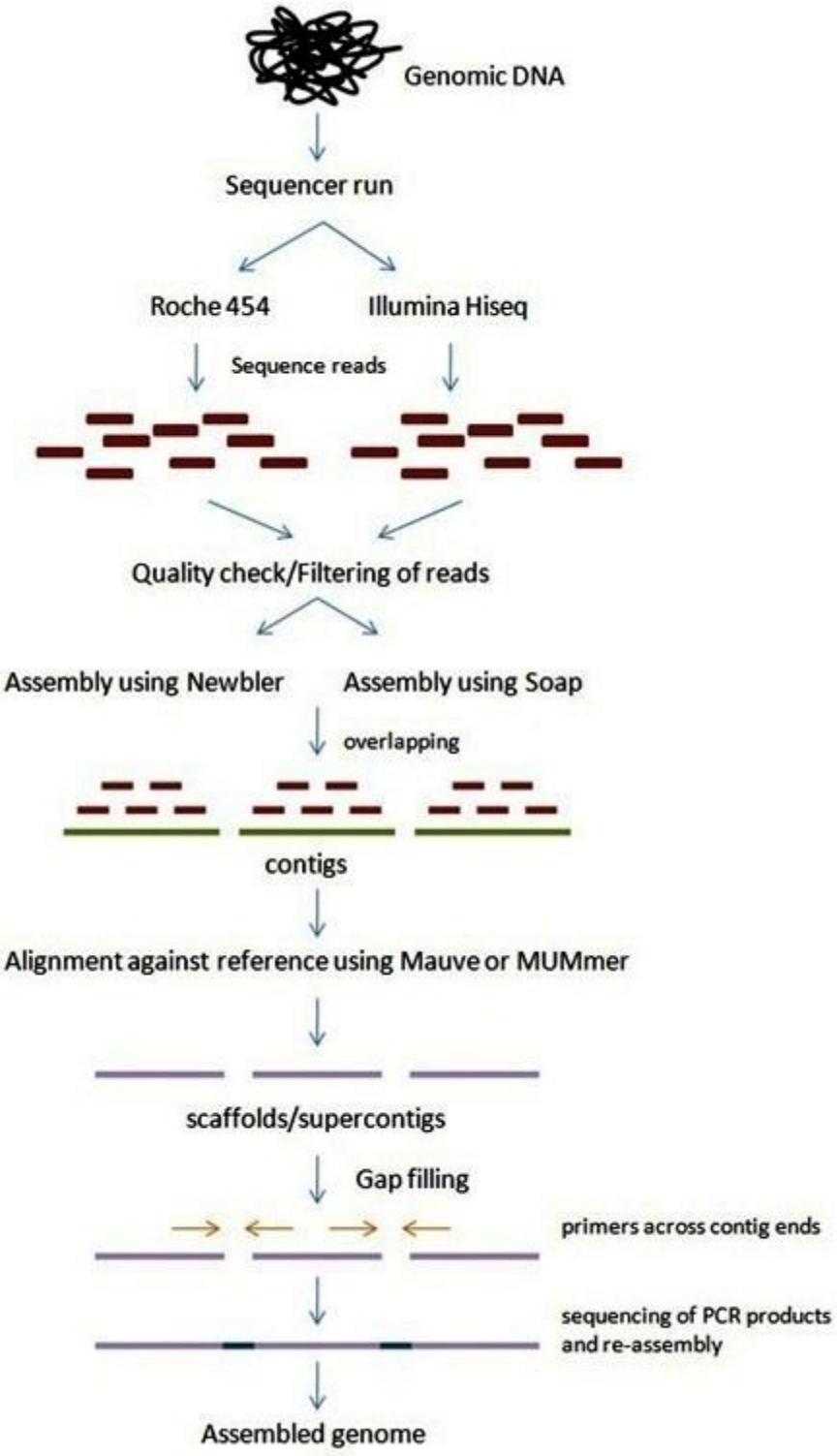


Figure 3.1 A road map of the bacterial genome sequencing.

3.3.5 Annotation of draft *Pectobacterium* genomes

The assembled genomes of *Pectobacterium* strains were processed using the NCBI prokaryotic Genomes Automatic Annotation Pipeline (PGAAP, http://www.ncbi.nlm.nih.gov/genome/annotation_prok/) and the Rapid Annotation Server (RAST, http://rast.nmpdr.org/) server (Aziz *et al.*, 2008b). The NCBI PGAAP combines HMM-based gene prediction methods with a sequence similarity-based approach, which combines comparison of the predicted gene products to the non-redundant protein database Entrez Protein Clusters , the Conserved Domain Database, and the COGs (Clusters of Orthologous Groups). A combination of gene prediction software tools: GenMark, GenMarkS and GLIMMER (Besemer *et al.*, 2001; Borodovsky & McIninch, 1993; Delcher *et al.*, 1999a, 1999b), was used to assign the coding regions in the genome of interest. Ribosomal RNAs were predicted by sequence similarity searching against an RNA sequence database using BLAST and/or using Infernal and Rfam models (Eddy, 2002; Griffiths-Jones *et al.*, 2003). Genes encoding tRNAs were identified using the tRNAscan-SE prediction software (Lowe & Eddy, 1997). A complete six-frame translation of the nucleotide sequence was carried out and the predicted protein sequences were masked (every nucleotide identified as a repeat is transformed to 'N' to aid sequence alignment and gene prediction) to identify missing genes. All predicted proteins were then searched using BLAST against all proteins from complete microbial genomes, which enabled conserved domains and other information to be added to the annotation for each protein. Frameshift errors were also detected and corrected by the annotation program to give an annotated genome. PGAAP required sequence files in .sqn and .agp format for generating annotated genomes. The .sqn files for each genome were generated using the tbl2asn program run on an UNIX terminal and .agp files were generated using an in-house bioinformatics script run from an UNIX platform (Appendix A3.2).

The RAST annotation server uses an approach similar to PGAAP to annotate a bacterial genome. The programmes tRNAscan-SE, 'search_for_rnas' developed by Niels Larsen (available from author) and GLIMMER2 were used to detect tRNAs, rRNAS and protein-coding regions,

respectively. The protein-coding genes were then searched against a small set of representative sequences from FIGfams (that have the property that they are universal or nearly universal in prokaryotes) to output a small set of genes (usually 8-15 genes). This set of genes was used to determine the closest phylogenetic neighbours of the newly sequenced genome. FIGfams is a collection of protein families utilized by RAST. Each FIGfam is assigned a set of similar and, presumably, homologous proteins, with all the members of the family sharing a common function. A decision procedure is associated with each protein (i.e. whether or not the protein is globally similar to the members and shares the common function). Two proteins will only be placed in the same family; 1) if both implement the same functional role and the region of similarity shared by the two sequences covers over 70% of each sequence, and 2) if they come from closely related genomes with >90% identity and the adjacent genes can be easily seen to correspond. Once the 'neighbouring genomes' have been determined, RAST forms a set of FIGfams that are present in these genomes and the new genome is searched for each of these FIGfams generated to accumulate the set of universal genes. The putative genes that are left out (of the universal gene set) are searched against the entire collection of FIGfams and also searched against a large non-redundant protein database embedded in the BLAST program to assign the function associated with each protein. Assembled sequenced reads in .fasta format were uploaded as separate jobs for each genome onto the RAST annotation server (http://rast.nmpdr.org/) to obtain annotated genome sequence files in Genbank (.gbk) format. Manual curation of the regions of interest was carried out, when necessary.

3.3.6 Genome visualization

The ICMP19477 genome annotated by PGAAP annotation pipeline (http://www.ncbi.nlm.nih.gov/genome/annotation_prok/) was visualized using Gview Server (https://server.gview.ca/). The genome of interest was uploaded to the server in a .gbk file to retrieve a circular image of the genome. The RAST annotation server (Aziz *et al.*, 2008a) was also used to view the different subsystem categories of the genome of interest.

3.3.7 Database submission

The annotated genome sequence for each of the four strains of *Pectobacterium* was deposited in GenBank under the accessions listed in Table 3.2.

3.4 Results

3.4.1 Whole genome sequencing and assembly of *Pectobacterium* strains

The genome sequence of at least one strain (Table 3.2) representative of the primary species or subspecies of *Pectobacterium* on potato in New Zealand was obtained by 454-platform massively parallel pyrosequencing or Hiseq Illumina sequencing technologies. They included *P. carotovorum* subsp. *brasiliensis* ICMP19477, the representative for blackleg causing strains in New Zealand as well as type strains *P. atrosepticum* ICMP1526 and *P. carotovorum* subsp. *carotovorum* ICMP5702. A summary of the statistics for each DNA sequencing dataset is listed in Table 3.3.

Assembly of the sequence reads generated by 454-platform pyrosequencing using Newbler assembler software (Margulies *et al*., 2005) generated preliminary draft genome sequences comprising 159 contigs and 94 contigs for *P. carotovorum* subsp. *brasiliensis* ICMP19477 and *P. atrosepticum* ICMP1526, respectively. Further assemblies and alignments of the *P. carotovorum* subsp. *brasiliensis* ICMP19477 contigs conducted with the aid of the MUMmer alignment tool (Kurtz *et al.*, 2004) and Mauve 2.0 (Darling, 2004) and using the published 'complete' *P. carotovorum* subsp. *aroidearum* PC1 genome as a reference resulted in 52 ordered and 107 unordered contigs. Upon alignment of these contigs with the published 'complete' *P. atrosepticum* SCRI1043 genome, however, Mauve detected additional contigs that could be aligned and ordered relative to homologous regions in the *P. atrosepticum* SCRI1043 genome. Hence, the contigs for *P. carotovorum* subsp. *brasiliensis* ICMP19477 were iteratively aligned and re-ordered using both *P. carotovorum* subsp. *aroidearum* PC1 and *P. atrosepticum* SCRI1043 until

no additional members of existing locally collinear blocks were detected. This iterative process produced 61 ordered and 98 unordered contigs for *P. carotovorum* subsp. *brasiliensis* ICMP19477. The initial alignments with *P. carotovorum* subsp. *aroidearum* PC1 and the re-ordered alignments with *P. atrosepticum* SCRI1043 for *P. carotovorum* subsp. *brasiliensis* ICMP19477 are compared using Mauve (Figure 3.2). The *P. atrosepticum* ICMP1526 contigs were also ordered and oriented with the aid of the MUMmer alignment tool and Mauve v2.0 (Darling, 2004) using the published 'complete' *P. atrosepticum* SCRI1043 genome as reference, resulting in 62 ordered and 32 unordered contigs.

The reads generated by Illumina sequencing were meticulously trimmed and corrected, and the lower quality reads discarded. Sequencing reads satisfying the quality control criteria were then assembled using the kmer custom parameter optimization strategy. kmer is the length of DNA that is used to construct the de Bruijn graph during assembly, and is the minimum number of identical bases required to join two overlapping reads (Zerbino & Birney, 2008). The Illumina sequence datasets provided nearly 10 million reads of 100 nucleotides or greater that passed the quality control check, for each strain. Assembly and alignment of the resulting pool of paired-end Illumina reads using the SOAP de novo software and MUMmer alignment tool using *P. carotovorum* subsp. *aroidearum* PC1 as the reference resulted in 249 and 125 contigs for *P. carotovorum* subsp. *carotovorum* ICMP5702 and *P. carotovorum* subsp. *carotovorum* UGC32, respectively.

Table 3.3 Summary statistics for the DNA sequence datasets generated for all *Pectobacterium* strains

Genome	**ICMP1977**	**ICMP1526**	**ICMP5702**	**UGC32**
Sequencing Method	454	454	HiSeq Illumina	HiSeq Illumina
Assembly Method	Newbler	Newbler	Soap	Soap
Coverage	50X	40X	600X	600X
Total no. of contigs	35	38	249	125
Longest contig size	1423741	1418603	954987	681885
N50	1120187	797800	448171	307291
Genome Finishing	Unclosed	Unclosed	Unclosed	Unclosed

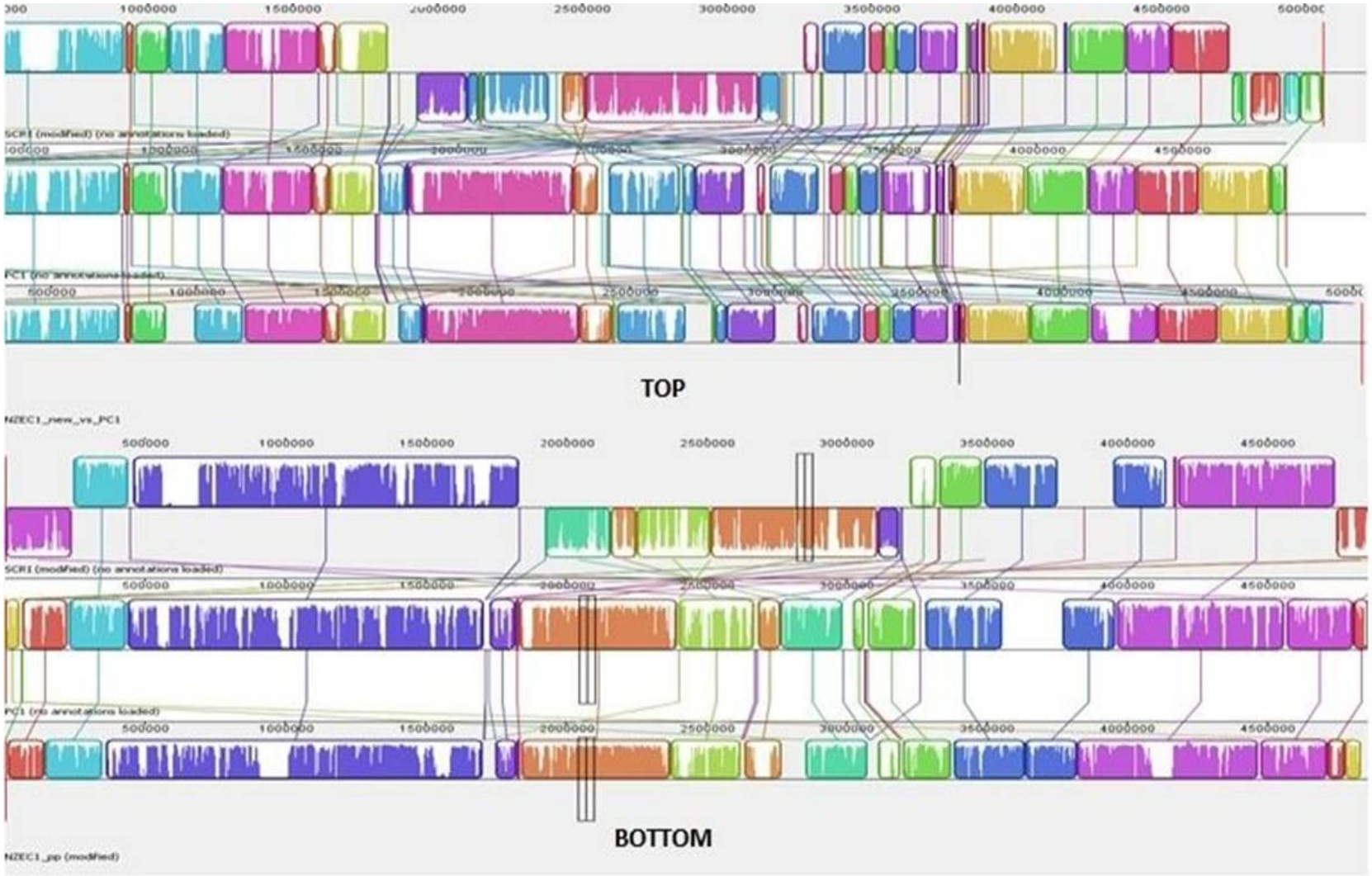

SCRI1043
PC1
ICMP19477
TOP
SCRI1043
PC1
ICMP19477
BOTTOM

Figure 3.2 Mauve alignment of the *Pectobacterium carotovorum* subsp. *brasiliensis* ICMP19477 genome before contig reordering with *P. atrosepticum* SCRI1043 (ordered using *P. carotovorum* subsp. *aroidearum* PC1 as reference genome) (top image) and after contig reordering with *P. atrosepticum* SCRI1043 (bottom image). Each alignment has three panels, one for each genome (*P. atrosepticum* SCRI1043, *P. carotovorum* subsp. *aroidearum* PC1 and *P. carotovorum* subsp. *brasiliensis* ICMP19477), composed of coloured segments (as blocks). Each of these blocks aligns to part of another genome and is presumably homologous to that particular region of the genome. The inverted regions are clearly depicted as blocks below a genome centre line. Regions outside blocks lack detectable homology among the input genomes. Inside each block, Mauve draws a similarity profile of the genome sequence. Areas that are completely white contain sequence elements specific to a particular genome. The reduction in the number of locally collinear blocks and an increase in average block length in the top alignment to the bottom alignment is clearly visible.

3.4.2 Gap Filling and closure of assembled *P. carotovorum* subsp. *brasiliensis* ICMP19477 and *P. atrosepticum* ICMP1526 genomes

It was anticipated that most of the gaps between contigs that were identified by preliminary assembly would be closed using primer walking, resulting in at least 90% coverage of each genome (Mark Fiers, personal communication). Indeed, re-assembly of the DNA sequences obtained from gap-filling PCR and Sanger sequencing of PCR amplicons using Newbler software resulted in the closure of at least 48 gaps in *P. carotovorum* subsp. *brasiliensis* ICMP19477 and approximately 35 gaps in *P. atrosepticum* ICMP1526. Improved 'finished' draft sequences for both *P. carotovorum* subsp. *brasiliensis* ICMP19477 and *P. atrosepticum* ICMP1526 were obtained in terms of coverage, reduction of gaps and reduction of poorly sequenced regions, resulting in 12 and 33 ordered contigs, respectively. The rest of the gaps remained unresolved possibly due to highly repetitive regions surrounding the gaps or more likely because of failure of PCR with the designed primer pairs across the contig ends, attributable to the length of the intervening gap. A higher quality genome assembly with fewer gaps was obtained for both the genomes than would have been produced by whole genome sequencing data alone.

The genome sequencing of *P. carotovorum* subsp. *brasiliensis* ICMP19477 and *P. atrosepticum* ICMP1526 was completed in March 2012 whilst the *P. carotovorum* subsp. *carotovorum* ICMP5702 and *P. carotovorum* subsp. *carotovorum* UGC32 genomes were still undergoing genome completion at the time of writing this thesis. Completion of these genome sequences would inevitably require some subsequent chromosome walking. Hence, no attempt to close the gaps of these genomes was made. Nevertheless, the high coverage draft genome sequences garnered for these strains provided sufficient information to carry out further studies.

3.4.3 General genome features of sequenced *Pectobacterium* strains

Pectobacterium carotovorum subsp. *brasiliensis* ICMP19477 had the largest genome size of approximately 4.98 Mb, whereas the genome of *P. carotovorum* subsp. *carotovorum* ICMP5702 was the smallest at about 4.79 Mb. The G+C content of the genomes ranged from 50.8 to 52.1%.

A summary of the genome features and contents of the *Pectobacterium* strains are presented in Table 3.4. The general features of the individual sequenced strains are described separately in the following sections, however.

Table 3.4 General genome features of all *Pectobacterium* strains used for comparison

Genome	ICMP19477	ICMP1526	ICMP5702	UGC32
Total base pairs	4979068	4873856	4788662	4804203
GC content	52.1	50.8	51.9	51.1
Number of CDSs	4435	4409	4338	4382
Number of tRNAs	67	68	78	75
Number of rRNAs	17	6	31	12

3.4.3.1 General features of the annotated *P. carotovorum* subsp. *brasiliensis* ICMP14977 genome

DNA sequencing determined that the genome of *P. carotovorum* subsp. *brasiliensis* ICMP19477 is 4,890,980 bps with an approximate G+C content of 52% (Table 3.4, Figure 3.3). *Pectobacterium carotovorum* subsp. *brasiliensis* ICMP19477 contains a total of 4645 predicted coding sequences (CDSs). Based on RAST annotation, biological roles were assigned to 2242 of the 4435 CDSs. The remaining CDSs comprise 902 conserved hypothetical proteins and 1132 CDSs of unknown function. Altogether, 67 genes coding for tRNAs representing all 20 amino acids were identified. Several rRNA operons organized in the order of 16S-23S-5S were also obtained. The number of CDSs identified in *P. carotovorum* subsp. *brasiliensis* ICMP19477 was similar to those of *P. carotovorum* subsp. *brasiliensis* ATCC BAA 417 (4836). The CDSs in *P. carotovorum* subsp. *brasiliensis* ICMP19477 were assigned to complete functional subsystems along with the subsystem categories (based on Gene Ontology: annotation based on the cellular component and molecular functions of the genes with their associated biological processes) (Figure 3.4). The majority was related to carbohydrate metabolism that include utilization and metabolism of sugar alcohols, amino acids, organic acids, di- and oligosachharides, monosaccharides and polysaccharides.

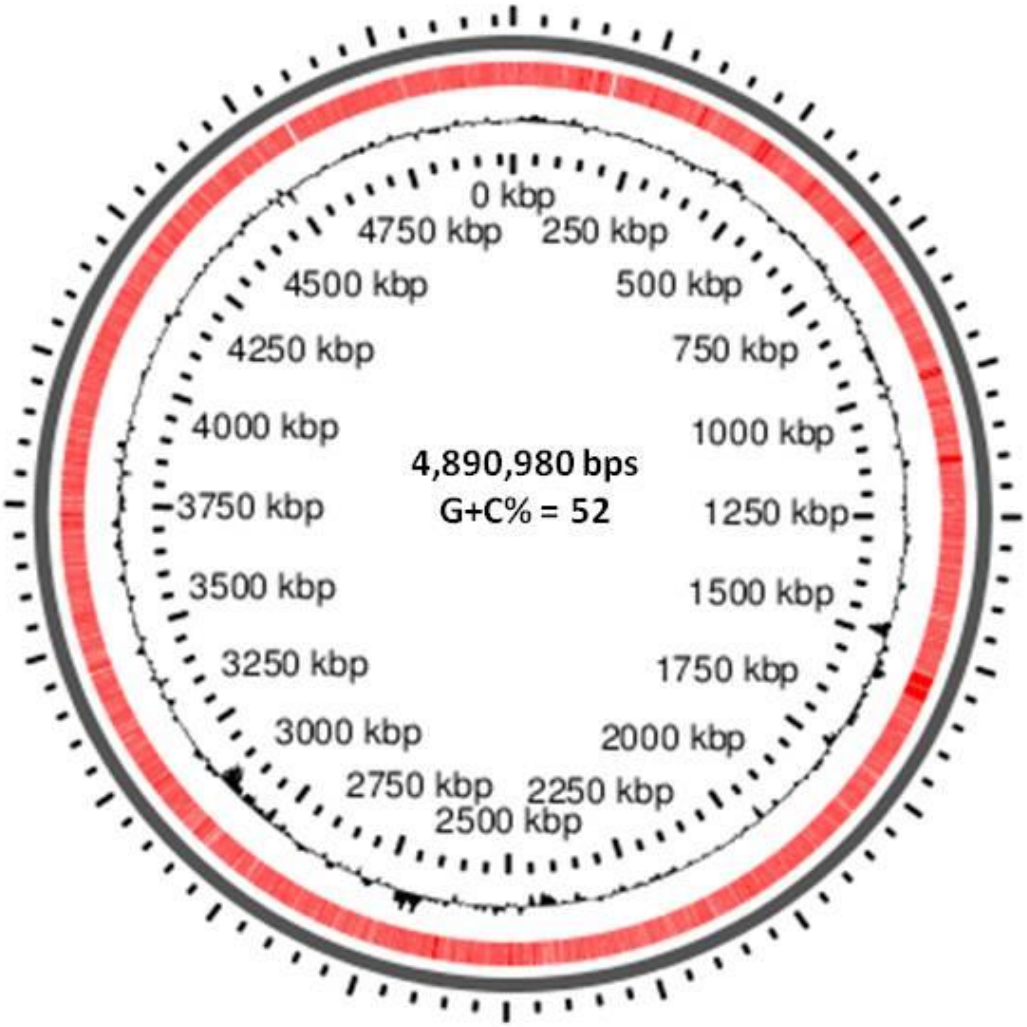


Figure 3.3 Schematic representation of the ICMP19477 genome. The grey outermost circle denotes the genome of ICMP19477. Inner to the grey circle is the red circle that corresponds to the total number of CDSs in that particular strain. The black circle (with irregular peaks) represents the GC content of ICMP19477 genome.

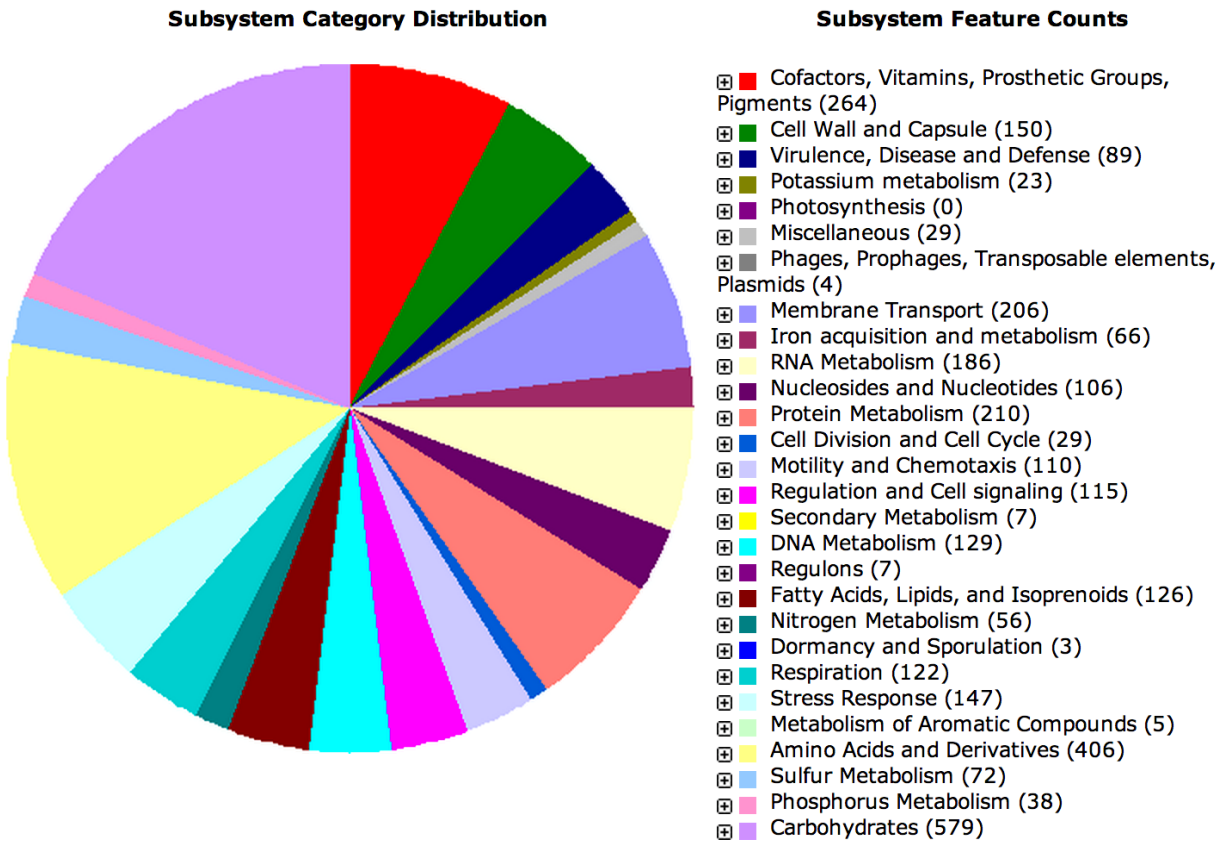


Figure 3.4 The subsystem category (Gene Ontology) for the genome of ICMP19477. The coloured pie chart represents the distribution of genes categorized under different gene ontologies for each genome (colour code for the GO categories are shown under subsystem feature counts). The subsystem feature counts represent the total number of CDSs of ICMP19477 belonging to a specific subsystem.

3.4.3.2 General features of the annotated *P. atrosepticum* ICMP1526 genome

The draft genome of *P. atrosepticum* ICMP1526 consists of 4,873,856 bps with an approximate G+C content of 50.8% (Table 3.3). A total of 4,409 CDSs and 67 genes coding for tRNAs are predicted to be present in the genome of *P. atrosepticum* ICMP1526. RAST annotation of *P. atrosepticum* ICMP1526 assigned biological roles to 2,348 CDSs with the remaining assigned as hypothetical proteins or proteins of unknown function. The number of CDSs identified in *P. atrosepticum* ICMP1526 was similar to the number of CDSs in *P. atrosepticum* SCRI1043 (4,472).

The distribution of the basic metabolic machinery of this strain into subsystems is represented in Appendix A3.3 and was similar to those of ICMP19477.

3.4.3.3 General features of *P. carotovorum* subsp. *carotovorum* ICMP5702 and *P. carotovorum* subsp. *carotovorum* UGC32 genomes

The genome of *P. carotovorum* subsp. *carotovorum* ICMP5702 was 4,788,662 bps and had a G+C content of 51.9% (Table 3.3). The genome of *P. carotovorum* subsp. *carotovorum* UGC32 contained 4,804,203 bps with a G+C% of 51.1 (Table 3.3). Annotation of their genomes identified 4,338 and 4,382 CDSs in *P. carotovorum* subsp. *carotovorum* ICMP5702 and *P. carotovorum* subsp. *carotovorum* UGC32, respectively. About 55% (2,350 predicted CDSs) or 56% (2,416 predicted CDSs) of the total identified CDSs in *P. carotovorum* subsp. *carotovorum* ICMP5702 and *P. carotovorum* subsp. *carotovorum* UGC32, respectively, were assigned biological functions after homology searching using the RAST annotation server. 45% (1,988) of the total CDSs of *P. carotovorum* subsp. *carotovorum* ICMP5702 were not found to be associated with any of the subsystem categories out of which 1,119 CDSs were assigned as non-hypothetical and the rest as hypothetical. In *P. carotovorum* subsp. *carotovorum* UGC32, 44% (1,966) of the total CDSs were not categorized into any subsystems as detected by the RAST server, of which 848 were hypothetical and the rest assigned as non-hypothetical. Seventy-eight and 75 tRNAs were predicted to be present in the genomes of *P. carotovorum* subsp. *carotovorum* ICMP5702 and *P. carotovorum* subsp. *carotovorum* UGC32, respectively.

The number of rRNAs in the genome of *P. carotovorum* subsp. *carotovorum* ICMP5702 was found to be higher (31) when compared to the 12 rRNAs in the genome of *P. carotovorum* subsp. *carotovorum* UGC32. This may be because of an annotation bias associated with the genome of *P. carotovorum* subsp. *carotovorum* ICMP5702 since the genome of *P. carotovorum* subsp. *carotovorum* ICMP5702 is present in a relatively high number of contigs.

3.5 Discussion

To date, the complete genome of only one *Pectobacterium* strain isolated from potato is available (*P. atrosepticum* SCRI1043) in the public domain. This genome (of *P. atrosepticum* SCR1043) was sequenced using the Sanger sequencing approach (Sanger & Coulson, 1975). By the beginning of the PhD project, the draft genomes of at least three other *Pectobacterium* strains, *P. carotovorum* subsp. *brasiliensis* 1692 (ATCC BAA 417) (Duarte *et al.*, 2004; Glasner *et al.*, 2008), *P. carotovorum* subsp. *carotovorum* WPP14 (Glasner *et al.*, 2008; Yap *et al.*, 2004) and *P*. *wasabiae* WPP163 (Ma *et al*., 2007; Nicole Perna and co-workers, and US DOE Joint Genome Institute, 2009) isolated from potatoes were available. These had been obtained using the high-throughput pyrosequencing method developed by 454 Life Sciences (Margulies *et al*., 2005). More recently, the complete genome of *P*. *wasabiae* SCC3193 (Koskinen *et al.*, 2012; Nykyri *et al*., 2012) was also generated using the same method. In this study, both pyrosequencing and Sanger sequencing were used to obtain the genomes of *P. carotovorum* subsp. *brasiliensis* ICMP19477 and *P. atrosepticum* ICMP1526. Pyrosequencing generates short reads of individual sequences, making assemblies of repeats problematic, and draft sequences may remain in several fragmentary contigs (Chaisson *et al.*, 2004; Glasner *et al.*, 2008). However, a significant number of the gaps were filled with the use of primer walking and long range PCR. The PCR products generated were sequenced using the Sanger sequencing method (Macrogen sequencing services). The iteration used in this study combined the pyrosequencing and Sanger data to produce the best possible high-quality draft genome assembly in the most timely and cost-effective manner for the two *Pectobacterium* genomes (*P. carotovorum* subsp. *brasiliensis* ICMP19477 and *P. atrosepticum* ICMP1526). The result was high-coverage 'finished' draft genomes for both strains; *P. carotovorum* subsp. *brasiliensis* ICMP19477 had 40× coverage and P. *atrosepticum* ICMP1526 had 30× coverage. Such coverage is ideally suited to further comparative analyses between strains and species (Schatz *et al.*, 2010). Coverage is defined as the ratio between the cumulative size of the set of reads and the size of the genome; thus, 40× coverage implies that the set of reads span fourty times as much DNA as the genome being sequenced (Pop, 2009).

The choice and justification of finishing a genome depends heavily on the scope of the project. A 'working' draft assembly can be produced more quickly than a 'finished' draft assembly and is very often sufficient to answer most biological questions about a phytopathogen and its interaction with a host (Glasner *et al.*, 2008; Meng *et al.*, 2014). The genomes of *P. carotovorum* subsp. *carotovorum* UGC32 and *P. carotovorum* subsp. *carotovorum* ICMP5702 were sequenced using Illumina sequencing technology, which offers a rapid and relatively inexpensive means of obtaining large amounts of highly informative genomic data (Bartram *et al.*, 2011). 'Working' draft assemblies of UGC32 and ICMP5702 genomes using the short reads generated by Illumina sequencing were produced more quickly than the 'finished' draft assemblies for the other two strains, which represented up to 97% of a genome's sequence. Draft genomes have been used extensively for comparative genomics to reveal the virulence gene component of plant pathogenic bacteria (Almeida *et al.*, 2009; Moreira *et al.*, 2010; Rodríguez-Palenzuela *et al.*, 2010). However, gaps and low quality regions may still remain in the data, decreasing the value of the working draft for studying features of the genome that span large regions or that requires high accuracy analyses.

The result of 'unsolvable' repeats is the fragmentation of the genome sequence into a (large) number of contiguous sequences separated by gaps for which there are no DNA sequence reads or where repeat sequences could not be resolved (Cahill *et al.*, 2010). This can result in the absence of important genes located in gaps between contigs. MGEs such as GIs are delineated by repeat sequences that may be hard to assemble. Such genetic elements make up a substantial proportion of the 'accessory' genome, which is important for the adaptation of a bacterium to its ecological niche (Dobrindt *et al.*, 2004; Malachowa & DeLeo, 2010). For example, 17 putative GIs have been predicted to be present in the genome of *P. atrosepticum* SCRI1043 that confer to nitrogen fixation, opine catabolism, antibiotic production, phytotoxin production and many more (Bell *et al*., 2004). A second drawback of a draft genome lies in the fact that it limits the study of genome organization as well as the function and transmission of the hereditary information of the organism (Mardis *et al.*, 2002). For instance, assembly and closure of the *Borrelia burgdoferi* genome revealed the presence of several previously unknown linear and circular plasmids

(Casjens *et al.*, 2000), while closure also demonstrated the presence of a second chromosome in *V. cholera* (Heidelberg *et al.*, 2000). Nevertheless, the generation of 'working' draft genome sequences of UGC32 and ICMP5702 garnered sufficient information to mine the gene content for biological meaning and to establish evolutionary relationships between these strains and other *Pectobacterium* (discussed in Chapter 4). The value of generating a 'complete' microbial genome has been reviewed by Fraser *et al*. (2002).

Due to the limitations of de novo assembly of bacterial genomes, a previously sequenced genome of a similar bacterium is often used as a 'reference' genome to construct contigs into a putative genome (Hunt *et al.*, 2014; Pop, 2009). The choice of reference, however, appears to have a major influence on the overall assembly of a genome and how one interprets the data for functional and evolutionary biology. For example, the genomes of ICMP5702 and UGC32 were obtained using PC1 as the 'reference' genome, which was essentially sufficient to align the set of contigs of ICMP5702 and UGC32 to the reference genome in order to characterize the genomes of ICMP5702 and UGC32. Similarly, ordering of contigs resulted in a draft assembly of ICMP1526 when the contigs of ICMP1526 were aligned against SCRI1043 (used as 'reference'). However, the same was not observed when the contigs of ICMP19477 were aligned against PC1. Though the majority of the contigs ordered and oriented against PC1, a number of large contigs were left out, which when re-aligned and ordered using SCRI1043 as the 'reference' genome resulted in an improvised draft genome. Hence, an iterative process using both genomes proved most useful in generating an accurate genome.

Once a genome is sequenced and assembled, it needs to be annotated to make sense of it. Annotation of the *Pectobacterium* genomes was carried out to gain insights into the location and context of a particular gene within the genome (e.g. if the gene is a part of a cluster and/or associated with a plasmid, phage or GI) and also whether a gene is expressed under defined conditions or life stages (such as during pathogenesis), is part of a known, novel or predicted regulatory network, or produces a protein product that is secreted or intracellular (Glasner *et al*., 2008). Annotation methods such as PGAAP and the RAST annotation server were used to

annotate the genomes of *Pectobacterium*. RAST was used to annotate the genomes of UGC32 and ICMP5702 instead of the PGAAP pipeline that was used to annotate the genomes of ICMP19477 and ICMP1526. This was because the pipeline was undergoing reconstruction at the time of submission, and the annotation of genomes was still in progress at the time of writing this chapter. These annotation tools take the ordered contig files as input, identify open reading frames that are likely to be genes, and compare these to a sophisticated database of genes and RNA sequences, providing a high quality annotation of the assembly. However, not every open reading frame is a gene, and it appears that many bacterial genomes are over-annotated. In addition, genes can be missed in the annotation (Wood *et al.*, 2012). These analyses showed that the genome size of *Pectobacterium* strains in this study is between 4.7 Mb to 4.9 Mb with the number of CDSs ranging between 4,338 and 4,435. The differences in the number of CDSs between the different strains are likely due to the 'fragmented' nature of the genomes, acquisition and loss of genes and/or also due to the application of different tools for their annotation. This has also been highlighted in studies by Ussery *et al*. (2004) and Nykyri *et al*. (2013). The genomic size difference can also be explained as a result of genomic expansion in some species (by HGT and duplication) and genomic reduction in others (e.g. IS mediated gene loss) (Bulach *et al.*, 2006). Genome expansion or reduction is related to the different lifestyles of pathogens, which may be adapted to different environments in and around the plant host. The CDSs of the annotated *Pectobacterium* strains used in this study were very similar to other *Pectobacterium* strains (SCRI1043, ATCC BAA 417 and WPP14) sequenced by Bell *et al*. (2004), Glasner *et al*. (2008) and Nykyri *et al*. (2012). The following chapter (Chapter 4) provides information on genes common to these sequenced *Pectobacterium* strains and also the presence of genes unique to each individual.

The number of rRNA genes in prokaryotic genomes can vary from one to as many as 15 copies (Pei *et al.*, 2010). Additional ribosomal operons can permit faster adaptation to new environmental conditions by increasing protein synthesis capacity (Klappenbach *et al.*, 2000). In particular, *Pectobacterium* is known to encode seven rRNA operons (*rrn*) (21 rRNA genes) (Kim *et al*., 2009). However, some strains of *P*. *wasabiae* and *P*. *carotovorum* subsp. *brasiliensis* WPP17

are known to contain only six *rrn*s due to deletion of an *rrn* operon (Kim *et al.*, 2009; Yap *et al.*, 2004). Spontaneous deletion of *rrn* operons has been reported for other bacteria too, such as *Yersinia* species (Deng *et al.*, 2002). Given that the newly annotated *Pectobaterium* genomes are incomplete, it is probable that the missing rRNA genes in ICMP19477, ICMP1526 and UGC32 are present in the gaps between contigs. In contrast, the presence of a high number of rRNA genes in ICMP5702 is likely due to artifacts in assembly.

The number of tRNA genes in bacterial genomes ranges from 126 in *Vibrio parahaemolyticus* to 29 in *Mycoplasma pulmonis* (Ussery *et al*., 2004). Since there are a maximum of 61 possible codons (and hence different tRNA genes), some genomes obviously have missing tRNAs, although all of the genomes can code for the use of all 20 amino acids (Ussery *et al.*, 2004). Differences in tRNAs in the newly sequenced *Pectobacterium* can be related to this or specifically because of the 'fragmented' nature of genomes. Subsequent gap closure and manual evaluation of the annotation process would shed light on the differences in the numbers of tRNAs in these *Pectobacterium* genomes.

More than 52% of the genes in each *Pectobacterium* genome were assigned to specific subsystem categories by RAST. Subsystems are groups of proteins with related functions, such as pathways of metabolism, complex structures, or phenotypes (Aziz *et al*., 2008). The rest of the genes that were not categorized to any of the subsystems were either non-hypothetical or hypothetical. The hypothetical proteins had no identifiable counterpart when searched against protein databases (conducted by RAST or PGAAP annotation tools), indicating the presence of unique genes (that have not yet been reported for other organisms) in the newly annotated *Pectobacterium* genomes. There are limitations to the annotation process, however. The fact that homology is the basis for these pipelines means that many genomes currently available may have been annotated using old, out of date genomes as a reference which in turn have been annotated based on even older more out of date genomes. Mis-annotations and errors may perpetuate throughout each new genome, ultimately propagating into secondary databases such as KEGG

and UNIPROT that use various annotation tools including RAST and PGAAP to blast the genes to assign their functional roles (Richardson & Watson, 2013)

In addition to essential housekeeping genes including genes required for DNA replication, DNA repair and cell division, transcription, translation and central carbohydrate metabolism such as tri-carboxylate cycle, glycolysis and gluconeogenesis, the genome of ICMP19477 also encoded essential genes required for purine and pyrimidine biosynthesis. Transport systems are central components of the host–pathogen relationship. A number of ion and ABC transporters and transporters for the uptake of carbohydrates, amino acids, peptides, nitrate/nitrite, sulphate and phosphate were also noted in the ICMP19477 genome. The genome annotation of ICMP19477 also revealed the presence of genes belonging to six different types of protein secretion systems, which are described here in a subsequent chapter (Chapter 4). Different protein secretion systems have been shown to be involved in the virulence of *Pectobacterium* species. Other virulence-related genes present in the ICMP19477 genome included genes involved in adhesion, resistance to antibiotics and toxic compounds, invasion and intracellular resistance and genes encoding for bacteriocins and ribosomally synthesized antibacterial peptides. The ICMP19477 genome also contained a number of genes coding for outer-membrane proteins or components of bacterial appendages (pili, fimbriae) implicated in the attachment of the bacterium to external surfaces. Genes involved in the biogenesis of type IV pilus responsible for movement of bacteria over epithelial surfaces without the use of flagella (Wall & Kaiser, 1999) were also found in ICMP19477 genome. Various chemotaxis genes were also identified, which may facilitate pathogenicity in specific environments. Chemotaxis is a complex behaviour used by many bacteria to sense specific chemicals or environmental conditions and move toward attractants and away from repellants (Adler, 1974; Blair, 1995). The presence of a number of adhesion/attachment proteins/factors suggest that these factors act as determinants for a wider host range in ICMP19477. Various genes encoding for siderophores and iron acquisition and metabolism were also observed in ICMP19477. The production of siderophores by pathogenic bacteria can greatly contribute to their virulence, because these molecules can remove iron from a wide variety of organic substrates (Ratledge & Dover, 2000).

MGEs such as prophages and transposons can have tremendous impacts on their hosts. Prophages can constitute as much as 10-20% of a bacterial genome and are major contributors to differences between individuals within species (Casjens, 2003). Bacteriophages can mediate the evolution and transfer of virulence factors and occasional acquisition of new traits by the bacterial host (Simpson *et al.*, 2000). Several components of phages and prophages were identified in the genome of ICMP19477 in this study by the web-based tool, RAST. Whether the phages are complete, incomplete or defective is unknown. Regions flanked by prophages, may contain genes essential for bacteria/plant interactions such as amino acid and iron/siderophore transporters, haemolysin and hemagglutinin proteins and other transporters that can also contribute important biological properties to their bacterial hosts.

Visualization of bacterial genome sequences is important for understanding gene structure, function and evolution. A multitude of tools are available for this purpose, including Artemis, Gview Server, Geneious software and many more. The Gview server was used to visualize the genome of ICMP19477. In addition to the visualization of the sequence features, this software shows the G+C content of this genome. The G+C content of bacteria has been used as a measure of relatedness. Differences in the G+C content in regions of a bacterial genome suggest that these regions were likely acquired via HGT. Regions acquired by HGT usually encode antibiotic resistance genes, virulence factors, phages and other beneficial traits. It has been observed that the ecological niche an organism occupies roughly correlates to both genome size and GC content (Foerstner *et al.*, 2005; Musto *et al.*, 2006). It would be interesting to investigate the functions of the genes encoded in these regions. GIs identified in the genome of ICMP19477 based on differences in the G+C content are listed in the following chapter (Chapter 4).

In summary, the genome sequences were sequenced, assembled and annotated for *P. carotovorum* subsp. *brasiliensis* ICMP19477 and other comparator *Pectobacterium* strains. These new genome sequences will enable comparative genomics of the SRE for the identification of GIs carrying novel virulence factors associated with pathogenicity and aggressiveness. It was particularly important to obtain the sequences for the type strains of *P. carotovorum* subsp.

carotovorum (ICMP5702) and *P*. *atrosepticum* (ICMP1526), as they provide important resources for *Pectobacterium* genomics, given their status as type strains. The comparative studies are described in the following chapter (Chapter 4).

3.6 Appendices

Appendix A3.1 Script to remove duplicate reads

```
#!/usr/bin/perl -w

use strict;
use warnings;

use Getopt::Long;
use Data::Dumper;
use PFR::FQ::Iterator;

{
   my $self = bless {
            'usage' => sub { exec 'perldoc', $0; },
            'step'  => 1e5,
            'gzipi' => 0,
            'gzipo' => 0,
            'bzipi' => 0,
            'bzipo' => 0,
   };

   GetOptions(
            'help' => $self->{'usage'},
            'f1|i1=s' => \$self->{'read1in'},
            'f2|i2=s' => \$self->{'read2in'},
            'o1=s'    => \$self->{'read1out'},
            'o2=s'    => \$self->{'read2out'},
            'step=s'  => \$self->{'step'},
            'gzipin'  => \$self->{'gzipi'},
            'gzipout' => \$self->{'gzipo'},
            'bzipin'  => \$self->{'bzipi'},
            'bzipout' => \$self->{'bzipo'},
            ) or $self->usage();

   exit( $self->main() );
}

sub main {
   my ($self) = @_;

   ## check command line
   $self->input_sanity() or return 1;

   ## get hold of some file handles
   my @input  = $self->input_filehandles();

   ## make a fastq file iterator
```

```
    $self->{'iterator'} = PFR::FQ::Iterator->new(\@input);

    ## actually remove duplicates
    $self->remove_duplicates();

    return 0;
}

## remove_duplicates - actually remove duplicates

sub remove_duplicates {
    my ($self) = @_;
    ## retrieve the iterator
    my $itr    = $self->{'iterator'};
    my @output = $self->output_filehandles();
    my $step   = $self->{'step'};
    my $total  = 0;
    my $filtered = 0;
    my %unique;                         # Memory!

    ## $record is an arrayref !
    while (my $record = $itr->get_next) {
            my ($read1, $read2) = @$record;

            my @test = grep { exists($unique{$_->{'seq'}}) } @$record;

            if (@test != @$record) {
               ## we have a non duplicate read...
               for(my $i = 0; $i < @$record && $record->[$i]; $i++) {
                       my $read  = $record->[$i];
                       my $write = $output[$i];
                       # $unique{ $read->{'seq'} } ||= [];
                       # push @{ $unique{ $read->{'seq'} } }, [ $read, $i ];
                       $unique{ $read->{'seq'} }++;
                       print $write join("\n", map { $read->{$_} } qw{id seq id2 qual}), "\n";
               }
            } else {
               $filtered++;
            }

            $total++;

            if ($total % $step == 0) {
               ## report where we've got to
               $self->report_status($total, $filtered);
            }
    }

    ## report again now we've finished
    $self->report_status($total, $filtered);

}
```

```
## report_status - simple routine to report percentages

sub report_status {
   my ($self, $total, $duplicates) = @_;
   my $unique = ($total - $duplicates);
   my $dups   = ($duplicates / $total) * 100;
   my $uniq   = ($unique     / $total) * 100;
   my $status = {
           Total     => sprintf("%9s", $total),
           Unique    => sprintf("%9s [%.2f%%]", $unique, $uniq),
           Duplicates => sprintf("%9s [%.2f%%]", $duplicates, $dups),
   };
   print STDERR join("\t", map { "$_ : $status->{$_}" } qw{Total Unique Duplicates}), "\n";
}

## input_filehandles - get an array of input filehandles. Either one or two.

sub input_filehandles {
   my ($self) = @_;
   my @filehandles;

   if($self->{'read1in'} && -f $self->{'read1in'}) {
           if($self->{'gzipi'}) {
              my @cmd = ('gzip', '-d', '-c', $self->{'read1in'});
              open($filehandles[0], '-|', "@cmd") or
                      die "Failed to open @cmd for reading.\n";
           } elsif($self->{'bzipi'}) {
              my @cmd = ('bzip2', '-d', '-c', $self->{'read1in'});
              open($filehandles[0], '-|', "@cmd") or
                      die "Failed to open @cmd for reading.\n";
           } else {
              open($filehandles[0], '<', $self->{'read1in'}) or
                      die "Failed to open ", $self->{'read1in'}, " for reading.\n";
           }
   }
   if($self->{'read2in'} && -f $self->{'read2in'}) {
           if($self->{'gzipi'}) {
              my @cmd = ('gzip', '-d', '-c', $self->{'read2in'});
              open($filehandles[0], '-|', "@cmd") or
                      die "Failed to open @cmd for reading.\n";
           } elsif($self->{'bzipi'}) {
              my @cmd = ('bzip2', '-d', '-c', $self->{'read2in'});
              open($filehandles[0], '-|', "@cmd") or
                      die "Failed to open @cmd for reading.\n";
           } else {
              open($filehandles[1], '<', $self->{'read2in'}) or
                      die "Failed to open ", $self->{'read2in'}, " for reading.\n";
           }
   }

   return @filehandles;
}
```

```
## output_filehandles - get an array of output filehandles. Either one or two

sub output_filehandles {
   my ($self) = @_;
   my @filehandles;

   if($self->{'read1out'}) {
           if($self->{'gzipo'}) {
              my @cmd = ('gzip', '-c', '>', $self->{'read1out'});
              open($filehandles[0], '|-', "@cmd") or
                      die "Failed to open @cmd for writing.\n";
           } elsif($self->{'bzipo'}) {
              my @cmd = ('bzip2', '-z', '-c', '>', $self->{'read1out'});
              open($filehandles[0], '|-', "@cmd") or
                      die "Failed to open @cmd for writing.\n";
           } else {
              open($filehandles[0], '>', $self->{'read1out'}) or
                      die "Failed to open ", $self->{'read1out'}, " for writing.\n";
           }
   }
   if($self->{'read2out'}) {
           if($self->{'gzipo'}) {
              my @cmd = ('gzip', '-c', '>', $self->{'read2out'});
              open($filehandles[0], '|-', "@cmd") or
                      die "Failed to open @cmd for writing.\n";
           } elsif($self->{'bzipo'}) {
              my @cmd = ('bzip2', '-z', '-c', '>', $self->{'read2out'});
              open($filehandles[0], '|-', "@cmd") or
                      die "Failed to open @cmd for writing.\n";
           } else {
              open($filehandles[1], '>', $self->{'read2out'}) or
                      die "Failed to open ", $self->{'read2out'}, " for writing.\n";
           }
   }

   return @filehandles;
}

## input_sanity - make sure noddy users have entered the required information

sub input_sanity {
   my ($self) = @_;

   my @messages;

   push @messages, "-i1 filename required" if (!$self->{'read1in'});
   push @messages, "-o1 filename required" if (!$self->{'read1out'});

   if($self->{'read1in'} && ! -f $self->{'read1in'}) {
           push @messages, sprintf(q{file '%s' does not exist}, $self->{'read1in'});
   }

   if($self->{'read2in'} && ! -f $self->{'read2in'}) {
```

```
        push @messages, sprintf(q{file '%s' does not exist}, $self->{'read2in'});
    }

    if($self->{'read2in'} && !$self->{'read2out'}) {
        push @messages, "-o2 filename required when -i2 specified";
    }

    if(@messages) {
        warn "Some errors have been encountered\n";
        warn "\n";
        map { warn "$_\n\n" } @messages;
        warn "Suggest $0 -help or perldoc $0 to view documentation\n";
    }

    return (@messages ? 0 : 1);
}

## usage - tell users what to do.

sub usage { (shift->{'usage'} || sub { exec 'perldoc', $0; })->(); }

=pod

=head1 NAME

./remove_duplicate_read_pairs.pl - Remove duplicate reads.

=head1 DESCRIPTION

Script to process a single end or paired end experiment set of fastq files.
The script removes duplicate reads.

=head1 USAGE

 ./remove_duplicate_read_pairs.pl [options]

Where options and [defaults] are:

 -i1 <read 1 input>   File to read from.
 -i2 <read 2 input>   File to read from.

 -o1 <read 1 output>  File to write unique reads to.
 -o2 <read 2 output>  File to write unique reads to.

 -step <step size>    When to display progress (every step reads) [1e5]

 -gzipi             Read compressed (gzip) input
 -gzipo              Write compressed (gzip) output

 -bzipi             Read compressed (bzip2) input
 -bzipo              Write compressed (bzip2) output

 -help              Display this help
```

=head1 EXAMPLE

./remove_duplicate_read_pairs.pl -i1 sequence_1_1.txt -o1 sequence_1_1.filtered.txt

=cut

__END__

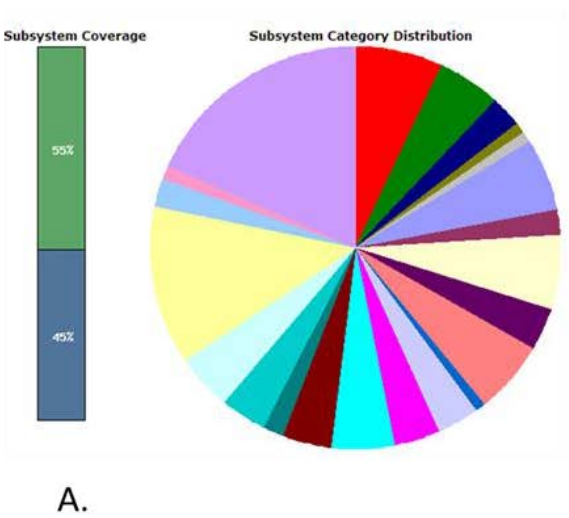
A3.2 Script to convert .sqn file to .agp file

```
#!/usr/bin/perl

use strict;
use warnings ;
use Bio::SeqIO ;
use Getopt::Long;
use File::Basename;

my $file;
my $outDir;
my $scaf_size_cutoff=0;
my $name_prefix="PGA";

GetOptions(
        'i=s' => \$file,    # scaf seq in fasta
        'size=i' => \$scaf_size_cutoff,
        'o=s'   => \$outDir,
        'name=s' => \$name_prefix);

die "Usage: $0 -i <sequence file> -o <out_dir> [-size <scaf_size_cutoff>] [-name <name>]\n" unless $file;
my ($file_name, $path, $suffix)=fileparse("$file", qr/\.[^.]*/);
my ($sample_name,$center)=split /_/,$file_name;

### Output file for contigs in Fasta format
### The various output file names can be tuned here.
my $contig_outfile = "$outDir/$name_prefix.contigs.fa";
my $scaffolds_outfile = "$outDir/$name_prefix.scaffolds.fa";
my $agp_outfile = "$outDir/$name_prefix.agp";


#open (FILE, ">$outdir/$contig_outfile") and
#   warn "Will write contigs to file '$contig_outfile' to $outDir Directory\n" or
#   die "Failed to write to file '$fasta_outfile'\n";

my $contig_out = Bio::SeqIO->new('-file' => ">$contig_outfile",
                                        '-format' => 'Fasta') and
   warn "Will write contigs to file '$contig_outfile' to $outDir Directory\n" or
   die "Failed to write to file '$contig_outfile'\n";

open (AGP_FILE, ">$agp_outfile") and
  warn "Will write contigs to file $agp_outfile to $outDir Directory\n" or
  die "Failed to write to file $agp_outfile\n";

#open (SCAFF_FILE, ">$outDir/$scaffolds_outfile") and
my $scaff_out = Bio::SeqIO->new('-file' => ">$scaffolds_outfile",
                                        '-format' => 'Fasta') and
   warn "Will write scaffolds to file $scaffolds_outfile to $outDir Directory\n" or
   die "Failed to write to file $scaffolds_outfile\n";
```

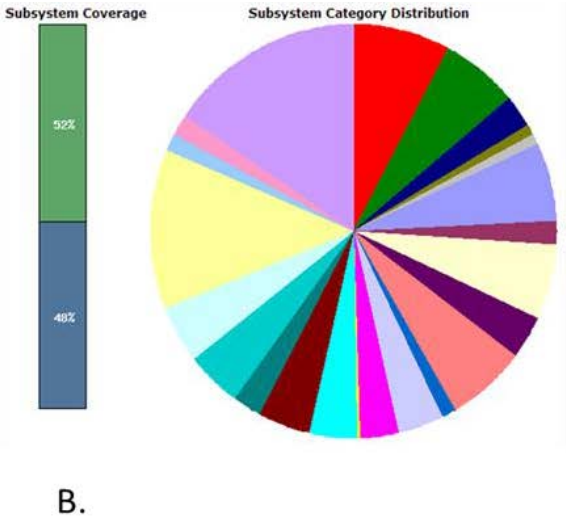
```
print AGP_FILE "# Generated from SOAPdenovo assembly file $file using script $0\n";

warn "Scaffold Sequence cutoff = $scaf_size_cutoff nt\n" if ($scaf_size_cutoff);


my $i = 0;# a counter, used for generating unique contig names

my $inseq = Bio::SeqIO->new('-file' => "<$file",
            '-format' => 'Fasta' ) ;


while (my $seq_obj = $inseq->next_seq ) {

  my $supercontig_id = $seq_obj->id ;
  my $supercontig_seq = $seq_obj->seq ;
  my $supercontig_desc = $seq_obj->description ;
  my $supercontig_length = length($supercontig_seq);

# only process the long supercontigs
  next if ($supercontig_length < $scaf_size_cutoff);



  if ($supercontig_id =~ m/NODE_(\d+)_length_\d+_cov_\d+/ or
     $supercontig_id =~ m/^(\d+)$/ or
     $supercontig_id =~ m/^scaffold(\d+)$/) {
     $supercontig_id = "${name_prefix}_scaffold_$1";
  }


#   print SCAFF_FILE ">$supercontig_id\n$supercontig_seq\n";
  my $scf = Bio::PrimarySeq->new(-seq => "$supercontig_seq",
                                 -id => "$supercontig_id");

  $scaff_out->write_seq($scf);

  my $start_pos = 1; # keep track of whereabouts in this supercontig we are
    my %substring_sequences;
  foreach my $substring_sequence ( split /(N+)/i, $supercontig_seq ) {
  #warn "\n$substring_sequence\n" if $supercontig_id eq '1160'; for #debugging only

  ### Define the AGP column contents
  my $object1 = $supercontig_id;
  my $object_beg2 = $start_pos;
  my $object_end3 = $start_pos + length($substring_sequence) - 1;
  my $part_number4 = $i;
  my $component_type5;
  my $component_id6a;
  my $gap_length6b;
  my $component_beg7a;
  my $gap_type7b;
  my $component_end8a;
  my $linkage8b;
```

```
  my $orientation9a;
  my $filler9b;
    if (  $substring_sequence =~ m/^N+$/i ) {
      ### This is poly-N gap between contigs
      $component_type5 = 'N';
      $gap_length6b = length($substring_sequence);
      $gap_type7b = 'fragment';
      $linkage8b = 'yes';
      $filler9b = '';
        } elsif ( $substring_sequence =~ m/^[ACGT]+$/i ) {
      ### This is a contig
      $i++; # a counter, used for generating unique contig names
      $component_type5 = 'W';
      $component_id6a = "${name_prefix}_$i";
      $component_beg7a = 1;
      $component_end8a = length($substring_sequence);
      $orientation9a = '+';
          ### Print FastA formatted contig
#       print FILE ">$component_id6a\n$substring_sequence\n";
      my $ctg = Bio::PrimarySeq->new(-seq => "$substring_sequence",
                                              -id => "$component_id6a");

      $contig_out->write_seq($ctg);


  } else {
      die "Illegal characters in sequence\n$substring_sequence\n";
  }
    $start_pos += length ($substring_sequence);
    if ($component_type5 eq 'N') {
      ### print AGP line for gap
      print                                                                                                                AGP_FILE
"$object1\t$object_beg2\t$object_end3\t$part_number4\t$component_type5\t$gap_length6b\t$gap_type7b\t$l
inkage8b\t$filler9b\n";
  } else {
      ### print AGP line for contig
      print                                                                                                                AGP_FILE
"$object1\t$object_beg2\t$object_end3\t$part_number4\t$component_type5\t$component_id6a\t$component
_beg7a\t$component_end8a\t$orientation9a\n";
       }
  }
}

$contig_out->close();
close AGP_FILE;
$scaff_out->close();
```

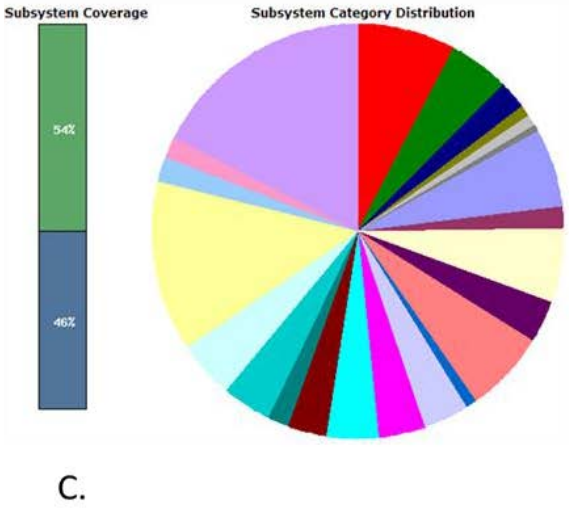
Appendix A3.3 The subsystem category (Gene Ontology) for the genomes of sequenced *Pectobacterium* strains: A, ICMP19477; B, ICMP1526; C, ICMP5702 and D, UGC32. Coloured pie chart represents the distribution of genes categorized under different gene ontologies for each genome (colour code for the GO categories are shown under subsystem feature counts). The subsystem feature counts represent the total number of CDSs for each of the sequenced *Pectobacterium* strain belonging to a specific subsystem. **E.** The total number of CDSs of each genome belonging to a specific subsystem.



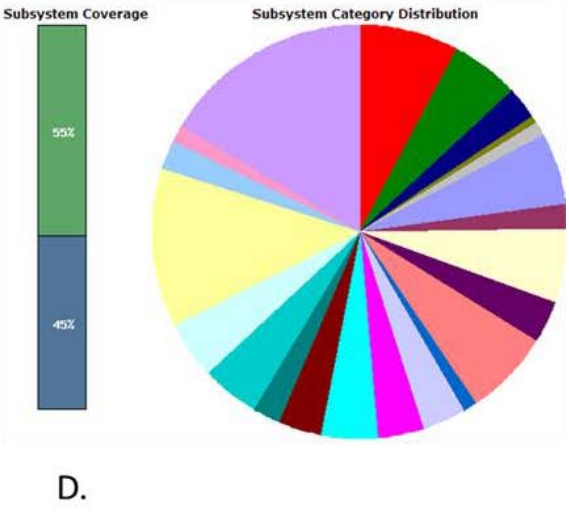
A.



B.



C.



D.

E.

Subsystem feature counts	A	B	C	D
Cofactors, Vitamins, Prosthetic Groups, Pigments	230	224	258	259
Cell Wall and Capsule	164	196	161	179
Virulence, Disease and Defense	85	81	75	84
Potassium metabolism	23	23	23	23
Photosynthesis	0	0	0	0
Miscellaneous	27	43	30	33
Phages, Prophages, Transposable elements, Plasmids	4	7	24	4
Membrane Transport	196	177	201	184
Iron acquisition and metabolism	64	56	53	68
RNA Metabolism	195	184	195	197
Nucleosides and Nucleotides	110	109	112	115
Protein Metabolism	192	202	210	223
Cell Division and Cell Cycle	28	39	27	36
Motility and Chemotaxis	112	111	122	112
Regulation and Cell signaling	112	107	112	121
Secondary Metabolism	7	7	8	7
DNA Metabolism	165	116	144	155
Regulons	0	0	7	2
Fatty Acids, Lipids, and Isoprenoids	128	107	97	111
Nitrogen Metabolism	53	66	58	70
Dormancy and Sporulation	3	3	5	3
Respiration	116	137	125	149
Stress Response	146	138	155	143
Metabolism of Aromatic Compounds	6	6	6	9
Amino Acids and Derivatives	423	402	436	427
Sulfur Metabolism	67	43	70	66
Phosphorus Metabolism	38	47	51	50
Carbohydrates	526	479	551	519

Chapter 4

Comparative genome analyses of the phytopathogen *P. carotovorum* subspecies *brasiliensis* ICMP19477 and other SRE

4.1 Abstract

To classify the group of unusually aggressive blackleg causing SRE isolated recently from potato in New Zealand and to define the genetic factors that might enable them to cause this disease, a comparative genomics analysis was performed using the genomes of the representative strain, ICMP19477, and of other sequenced SRE. This comparative analysis revealed that the SRE have a backbone of common housekeeping and virulence-associated genes, which enabled accurate taxonomic classification and probably provide the basis for pathogenicity on plant hosts. The genomes also carry a divergent set of gene modules, however, that could reflect adaptations to specific environments and enable them to infect specific plant organs or confer different degrees of aggressiveness.

Genomic approaches, including genome distance scores, average nucleotide identity and phylogenetic analysis confirmed that ICMP19477 is a member of *P*. *carotovorum* subsp. *brasiliensis*; ICMP19477 only differed to the type strain for *P*. *carotovorum* subsp. *brasiliensis* by 2% in pairwise comparisons and showed almost complete synteny. Analysis of the accessory genome of ICMP19477 revealed 69 GIs, many harbouring genes previously associated with pathogenicity in *Pectobacterium* or *Dickeya*. Putative gene islets were also identified, which encoded genes predicted to be involved in virulence of ICMP19477 and/ or other SRE. Of particular interest were the GIs predicted to encode a novel non-ribosomal peptide synthetase (NRPS) (PbNI_GI20) and the biosynthesis of CFA (PbN1_GI12), and a number of gene islets encoding genes predicted to be involved in the capture and utilization of sugars or the detoxification of phenolics. These GIs and gene islets were present in ICMP19477 and/or in other blackleg causing SRE, but were absent from those predicted to cause only soft rotting symptoms.

4.2 Introduction

Comparative genomics is a powerful tool for comparing the total information content or similarities and differences in the structure and organization of two or more genomes. As a result, comparative genomics can be used to describe and understand genome evolution (Wolfe and Li, 2003), which is closely linked to individual gene evolution. In other words, comparative genomics embodies an evolutionary approach to microbiology that can be far more elucidating than individual gene analysis.

The basis of comparative genomics is to identify and map conserved regions and to detect unique regions, which, in the case of pathogenic microbes, might be expected to define host range, mechanisms for survival in specific habitats and virulence mechanisms. Through the process of evolution, positive or negative purifying selection helps conserve regions of the genome, while areas under neutral selection are free to vary and thus expected to diverge more quickly (McDonald & Kreitman, 1991; Sawyer & Hartl, 1992). With this in mind, a major challenge of comparative genomics is to detect the series of events in a genome during evolution that result from nucleotide-level changes (such as single-nucleotide polymorphisms), gene duplications and losses, lateral gene transfers, and genomic recombinations.

Comparative genomic analyses are led by two approaches: (i) pairwise genome comparisons that identify syntenic blocks, i.e., genomic regions that are reasonably conserved in the two genomes and, (ii) multiple genome comparisons that carry out pairwise evaluations on multiple genomes compared to a fixed 'reference' genome. A variety of software has been developed to perform comparative genomics, such as OrthoMCL (Li *et al*., 2003) and Mauve (Darling *et al*., 2004). OrthoMCL uses BLAST to calculate the pairwise sequence similarity and a Markov Cluster algorithm to group orthologs and paralogs in a set of genomes. Mauve performs multiple alignments of closely related genomes to identify conserved syntenic regions.

Comparative genomics has revolutionized taxonomy and the understanding of the interplay between phenotype and genotype (Forst & Schulten, 2001). In particular, by associating the species-specific genes with the unique characteristics of that species, researchers have begun to dissect the relationship between genotype and phenotype. The repertoire of genes represented across all strains of a bacterial species is known as its 'pan genome'. The pan genome constitutes three parts: the genes common to all strains of the species known as the 'core genome', the 'dispensable or accessory genome' that consists of genes present in at least two but not all strains of a species, and the 'strain-specific' genome comprising genes unique or specific to a strain (Medini *et al*., 2005). Strain-specific genes are referred to as 'singletons' in this chapter. The core genome is the ancient part of the genus or species that shows the evolutionary origin of the bacterial group and contains traits that are necessary for the viability and fitness of the bacteria (Juhas *et al*., 2009). Genetic traits linked to variation in virulence, adaptation and antibiotic resistance, however, are more often governed by the dispensable or accessory genome (Tettelin *et al*., 2008). The accessory genome harbours GIs and other mobile genetic elements that are considered to have been acquired by HGT and to be beneficial to the recipient strain in specific environments and/ or in interactions with other competing microbes (Dobrindt *et al*., 2009; Juhas *et al*., 2009; Toth *et al*., 2001).

Since 2000, the blackleg-causing SRE *P. carotovorum* subsp. *brasiliensis* has been observed as an emerging pathogen of potato in Brazil, South Africa, Canada and Kenya (De Boer *et al*., 2012; Duarte *et al*., 2004; Onkendi *et al*., 2014; Van der Merwe *et al*., 2010). However, it is not clear whether *P*. *carotovorum* subsp. *brasiliensis* is an emerging pathogen in New Zealand due to the previous lack of accurate methodologies available to differentiate *P*. *carotovorum* subsp. *carotovorum* and *P*. *carotovorum* subsp. *brasiliensis*, and which resulted in the original inaccurate characterization of these strains (e.g. ICMP19477) as *P*. *carotovorum* subsp. *carotovorum*. Nevertheless, since genomic sequences are now readily available for various SRE, including ICMP19477 (Chapter 3), a comparative genomics analysis provides an opportunity to more accurately confirm the status of this blackleg-causing pathogen. Based on genomic information,

genetic factors can be identified that contribute to its success as a potato pathogen in New Zealand.

In a previous study by Glasner *et al*. (2008), the first comparative analysis of *Pectobacterium* was carried out using single strains of *P*. *atrosepticum*, *P*. *carotovorum* subsp. *carotovorum* and *P*. *carotovorum* subsp. *brasiliensis* in an attempt to explain the niche specificities of blackleg and soft-rot bacteria. No clear result was obtained as to why these bacteria have different niches, although variability in many traits such as toxin production, enzyme production, secretion system, taxis and motility genes and gene regulation and metabolism was observed (Glasner *et al*., 2008). This emphasized the need to sequence more genomes of SRE from different niches to identify or associate particular traits to niche specificity. Pan genome analyses of other bacterial species have also clearly shown that the sequencing of one or two genomes per species is not sufficient to understand within-species diversity and that sequencing of multiple strains is required to present a more consistent definition of the species itself (Mann *et al*., 2013; Medini *et al*., 2005; Tettelin *et al*., 2008).

In this chapter, whole genome phylogenetic analysis using the genomes of 14 selected SRE was carried out to determine the taxonomic position of the blackleg causing *P*. *carotovorum* subsp. *brasiliensis* strain ICMP19477. Three criteria were then used to distinguish more recently acquired regions of the ICMP19477 genome that may be associated with its relative aggressiveness on potato stems: i) estimation of a core genome shared amongst all strains plus strain-specific genes, ii) identification of putative GIs in ICMP19477, and iii) presence of any identified islands in other SRE that cause either blackleg or soft rotting on potato.

4.3 Methods

4.3.1 Bacterial genomes of SRE used for comparative genomics

The genomes of bacterial strains used in this study are listed in Table 4.1.

4.3.2 Genome comparisons using OrthoMCL, EDGAR and h-val analysis

A comparison of protein repertoires of ICMP19477 and 13 other selected SRE was conducted to calculate the pan genome, the core and singleton CDS numbers, and to identify virulence factors in ICMP19477 and/or in other blackleg-causing strains. Two programs were used to carry out these comparisons: (i) EDGAR (Blom *et al*., 2009) and, (ii) OrthoMCL (Li *et al*., 2003).

The Markov clustering algorithm OrthoMCL (Li *et al*., 2003) involves ortholog prediction based on pairwise reciprocal hits using an all-against-all blastp search of conceptual translations of a set of protein sequences from genomes of interest (Christian *et al*., 2003). OrthoMCL groups proteins into putative orthologous groups based on sequence similarity. A RBH (reciprocal best hit) is determined when the best match to a query sequence has the query sequence as its own best match in the reciprocal comparison. The blastp cutoff e-value and the Markov clustering inflation index are the two parameters used to predict the orthologous groups. A range of e- value cutoffs and clustering inflation indexes were tested in this study and it was concluded that the lower the e-value, the more stringent the ortholog detection, whereas a higher inflation index results in lower rates of false-positives. Hence, blastp alignments with an e-value of 1e-5 and an inflation index of 4 were chosen to achieve high sensitivity and specificity in ortholog detection. The results (orthologous groups or clusters) produced by the OrthoMCL program were then parsed using the R package (Maechler *et al*., 2013) and manually curated to categorize the orthologous pairs into strain-specific gene clusters and genes common to one or more genomes (the accessory genome).

Table 4.1 The genome sequences used in comparative analyses of SRE isolated from potato and other hosts

SRE	Strain	Bioproject/Genbank Accession no.	Total no. of contigs	No. of predicted CDSs	Description	Source
Pbr	ICMP19477	PRJNA72121	35	4436	Isolated from blackleg diseased stem in New Zealand	This study
Pbr	ATCC BAA 417^{T}	ABVX00000000	1371	4836	Type strain of *Pbr* causing blackleg disease isolated from potatoes in Brazil	Glasner *et al*., 2008
Pba	SCRI1043	BX950851.1	1	4492	Isolated from blackleg diseased stem in the United Kingdom	Bell *et al*., 2004
Pba	ICMP1526^{T}	PRJNA72123	38	4409	Type strain of *Pba* causing blackleg disease isolated from potatoes in the United Kingdom	This study
Pcc	ICMP5702^{T}	AODT00000000	249	4338	Type strain of *Pcc* isolated from soft rot of potato tubers in Denmark	This study
Pcc	WPP14	ABVY00000000	742	4540	Isolated from diseased tubers in the United States of America	Glasner *et al*., 2008
Pco	UGC32	AODU00000000	125	4382	Isoalted from blackleg diseased stems of potato plants in Peru	This study
Pcc	PC1	CP001657.1	1	4246	Soft rot causing *Po* isolated in the United States of America	US DOE Joint Genome Institute, 2009
Pcc	PCC21	CP003776.1	1	4263	Soft rot causing *Pcc* strain isolated from Chinese cabbage in Korea	Park *et al*., 2012
Pecwa	WPP163	CP001790.1	1	4437	Isolated from infected potato tubers in the United States of America	US DOE Joint Genome Institute, 2009
Pecwa	SCC3193	PRJNA122637	1	4705	Isolated from diseased potato stem in Finland	Koskinene *et al*., 2012
Dd	Dda3937	CP002038.1	1	4571	Isolated from African violet and causes soft rot and wilt of potato plant (n/k)	Glasner *et al*., 2011
Dd*	Ech586	CP001836.1	1	3970	Causes soft rot and wilt of potato plant (n/k)	US DOE Joint Genome Institute, 2009
Dd$^{\#}$	Ech703	CP001654.1	1	4144	Causes soft rot and wilt of potato plant (n/k)	US DOE Joint Genome Institute, 2009

Pbr – P. carotovorum subsp. *brasiliensis*; *Pba – P. atrosepticum*; *Pcc – P. carotovorum* subsp. *carotovorum*; *Po - P. aroidearum*; *Pecwa – P. wasabiae*; *Dd – D. dadantii*; * - reclassified as *D. zeae* # - reclassified as *D. paradisiaca*; T – type strain; n/k - geographical origin of isolation not known

The bidirectional best blast hit (BBH) method based on the well-known protein alignment tool BLAST (blastp) (Altschul, 1990) is employed by the EDGAR program to calculate 'orthologs' in determining the pan and core genomes. The proteins or protein clusters identified are then analyzed manually by conducting blast analysis using the NCBI blast server, to detect virulence determinants previously not identified by gene annotation methods.

H-val analysis is a useful method in determining true ORFs as it eliminates or avoids annotation bias (Fukiya *et al*., 2004). In this analysis, the nucleotide sequence for each ORF assigned in the genome of ICMP19477 is used as a query for a homology search. The homology search is conducted with blastn against the whole genome sequence of other comparator strains, with the default parameters used in the NCBI database. The h-val is calculated as: [(length of highest-score region) X (identities of hit shown in blastn)]/ (length of query sequence) (Fukiya *et al*., 2004). The h-val indicates how much the corresponding sequence (ORF) of the comparator genome resembled the query ORF in terms of length and sequence identity. An h-value >0.8 indicates that the query ORF should have an almost identical homologue in the comparator genome.

4.3.3 Phylogenetic analysis of the SRE using EDGAR

A phylogenetic tree was constructed using the core genomes of the 14 SRE described in Table 4.1 using EDGAR (Blom *et al*., 2009). The first step involved carrying out multiple alignments of the core genes for the collection of SRE using the gene alignment tool MUSCLE (Edgar, 2004). Non-matching parts of the alignments were masked using Gblocks (Talavera and Castresana, 2007) and removed. Gblocks is a computer program that eliminates poorly aligned positions and divergent regions of an alignment of DNA or protein sequences. The matching parts were then concatenated to build a single multiple alignment. An unrooted phylogenetic tree using maximum likelihood and 1000 bootstraps was generated from this alignment using the tree-building package PHYLIP (Felsenstein, 1995).

4.3.4 Genome content distances from multiple genome alignments using Mauve

The synteny in the genomes of ICMP19477, ATCC BAA 417, SCRI1043 and ICMP5702 was established by performing multiple genome alignments in MAUVE (Darling *et al*., 2004). A genome content distance matrix was produced as an output from these pairwise alignments. The distances were then correlated with average nucleotide identity (ANI) of the genomes using the ANIb (ANI blast) tool of the Jspecies package (Richter and Rossello-Mora, 2009) to measure the probability of two genomes belonging to the same species or not (a cut-off of 95-96% is used as a boundary for species circumscription).

4.3.5 GI predictions

Two bioinformatic approaches were used to identify GIs: sequence composition-based methods and a comparative genomics-based method. Sequence composition methods use a single genome for the prediction of GIs; whereas comparative genomics based methods require two or more genomes in comparisons.

Sequence comparison-based methods, SIGI-HMM (Waack *et al*., 2006), IslandPath-DIMOB (Hsiao *et al*., 2005), InDeGenius (Srivastava *et al*., 2010) and Alien Hunter (Vernikos and Parkhill, 2006) and comparative genomics based GI prediction tool, Islandpick (Langille *et al*., 2008) were used to predict GIs in ICMP19477. A brief summary of the sequence comparison-based methods is presented in Table 4.2. The automated predictions were manually curated. Wherever necessary, the borders of the predicted islands were adjusted such that genes associated with mobile genetic elements such as phages, transposes and integrases, when present adjacent to automatically predicted islands, were added to the initially detected islands.

Table 4.2 GI Prediction Programs

Program name	Method	Description	Availability	Precision (%)	Recall (%)	Accuracy (%)	Reference
SIGI-HMM	Sequence Composition-based	Uses codon usage and removes ribosomal regions	Command line	92.3	33.0	86.3	Waack *et al.*, 2006
IslandPath-Dimob	Sequence Composition-based	Uses dinucleotide bais with the presence of atleast one mobility gene	Command line	85.8	35.3	86.2	Hsiao *et al.*, 2005
InDeGenius	Sequence Composition-based	Uses variable length k-mers	Command line	61.3	27.6	82.4	Shrivastava *et al.*, 2010
Alien-Hunter	Sequence Composition-based	Uses variable length k-mers	Command line	38.0	77.0	70.8	Vernikos *et al.*, 2006
Island Pick	Comparative genomics based	Uses a data set of GIs reported in literature	Web based	n/d	n/d	n/d	Langille et al., 2008

n/d – not determined

4.3.5.1 Sequence composition-based methods

Sequence composition based methods rely on the different nucleotide pattern preferences (natural variation) exhibited by different organisms that constitute their genome signatures (Langille *et al*., 2008). These methods generally depend on sequence and annotated features of a genome such as difference in G+C content, dinucleotide bias, codon bias, flanking direct repeats, tRNA and other mobility genes (e.g. integrases, transposes, etc.) to characterize GIs (Dobrindt *et al*., 2004; Karlin, 2001; Vernikos & Parkhill, 2008). Various computational tools (see below) have been developed to aid the identification of islands in genomic sequences that are based on sequence composition methods (Langille *et al*., 2008). These tools, in general, calculate the frequency of nucleotide sequences of a certain length (generally between 1 and 9), referred to as a k-mer where k is the length of the sequence, for a sub-region of a genome and compare these results with the expected frequencies of a genome (Langille *et al*., 2008). GIs are then identified based on any deviation from the genome frequencies.

4.3.5.2 Comparative genomics prediction methods

Comparative genomics approaches use multiple genomes to define clusters of genes present in one genome that are not present in closely related genomes. They are based on the observation that GIs are sporadically distributed among closely related species or in strains that can be found

in distantly related species (based to sequence divergence in 16S rRNAs or other orthologs) (Ragan, 2011). Various gene alignment programs such as Mauve (Darling *et al*., 2004) and MUMmer (Delcher *et al*., 2002), detect conserved regions (that are unlikely to have horizontal origins) by aligning multiple genomes; regions that are unique to a genome can be considered putative GIs for the genome they reside in. This was achieved here by using the IslandPick program that uses a combination of methods including Mauve (Darling et al., 2004), MUMmer (Delcher et al., 2002) and blastn (Altschul, et al., 1997) similarity searches, and comparing the genome of ICMP19477 (query genome) with other closely related sequenced *Pectobacterium* genomes (comparison genomes).

4.4 Results and discussion

4.4.1 The pan genome of the SRE

The genomes of ICMP19477, ICMP1526, UGC32 and ICMP5702, which were sequenced in this study (Chapter 3), along with the genomes of SCRI1043, ATCC BAA 417, WPP14, PC1, PCC21, SCC3193, WPP163, Ech586, Ech703, and Dda3937 which were sequenced elsewhere, were compared using EDGAR (Blom *et al*., 2009) and OrthoMCL (Li *et al*., 2003) to generate the pan genome, core genome and strain- or species-specific genome.

Comparison of the genomes of the fourteen SRE predicted that they varied from 4.6 Mb to 5.1 Mb in size. The average size of the SRE genome was approximately 4.8 Mb and had an average G+C content of 52.5%. The genomes were composed of between 3,970 (Ech586) and 4,836 (ATCC BAA 417) CDSs, which divided the genomes into the core and the accessory components.

4.4.1.1 Pan genome analysis using OrthoMCL

The pan genome of the 14 SRE comprised a total of 6,711 groups of proteins, encoded by a total of 57,224 CDSs. 2,124 of these protein families (encoded by a total of 30,204 CDSs in the 14

genomes analysed) were common to all genomes and constituted the core genome (Figure 4.1). A variety of accessory gene clusters (or orthologs) were also identified, which were present in two or more strains but not in all (Figure 4.1 and Appendix A4.1).

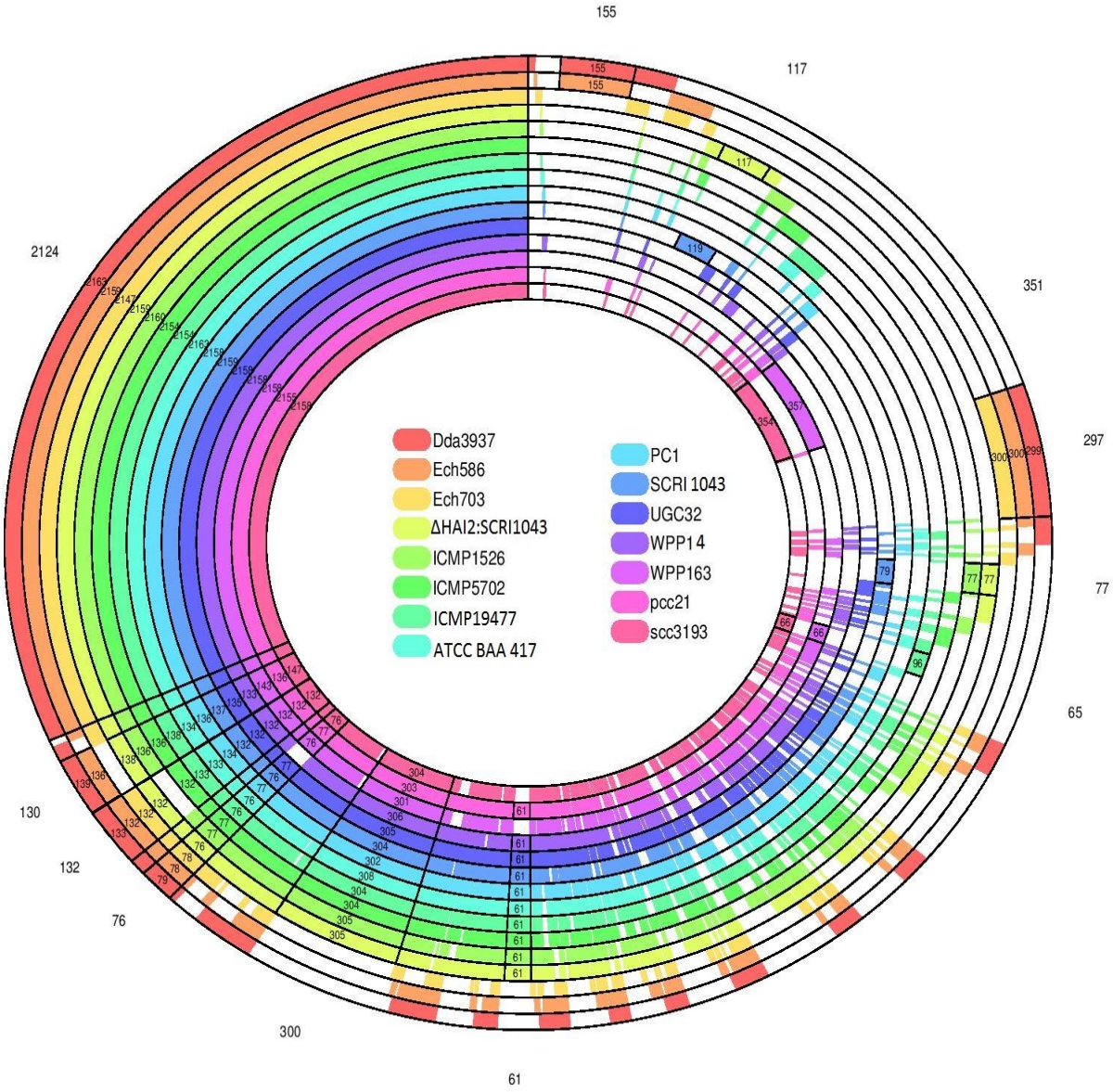


Figure 4.1 Core and shared gene orthologs between strains of SRE genomes. The numbers outside the rainbow plot show the number of gene ortholog groups present in at least two SRE genomes used in this study.

4.4.1.2 Pan genome analysis using EDGAR

The pan genome of the 14 SRE was also predicted to consist of 57,224 CDSs by EDGAR (Appendix A4.2). Among these genes, a core genome was identified that comprised of 28,658 CDSs.

4.4.1.3 Comparison of the pan genome predictions using OrthoMCL and EDGAR

OrthoMCL and EDGAR produced similar results, the core genome of the SRE making up at least 50% of the pan genome. The percentage of the pan genome represented by the core genome was comparable to that of other plant pathogens such as *Ralstonia* (Remenant *et al*., 2010), *Xanthomonas* (Blom *et al*., 2009) and *Pseudomonas* (Baltrus *et al*., 2011), which comprised 48%, 63% and 64% of the pan genomes, respectively (Table 4.3). The high degree of variation in the pan genomes of these pathogens could be related to their broad ecological niches, but may also reflect a higher inherent level of genetic recombination that enables them to evolve rapidly to habitat changes.

Table 4.3 Pan genome type of each selected bacterial species

Species	Core CDS (%)	No. of genomes	Pan genome type	Reference
E. amylovora	89	12	Open	Mann *et al*., 2013
E. coli	44	17	Open	Rasko *et al*., 2008
L. monocytogenes	80	26	Open	Deng *et al*., 2010
M. tuberculosis	98	9	Closed	Wozniak *et al*., 2011
P. syringae	64	19	Unknown	Baltrus *et al*., 2011
R. solanacearum	48	6	Unknown	Remenant *et al*., 2010
X. oryzae	63	4	Open	Blom *et al*., 2009

OrthoMCL is a sophisticated program, which uses RBHs to predict orthologs in genome comparisons. But it requires in-house scripts to compute the pan and core genome and to determine singletons for each genome being analyzed. OrthoMCL also has special requirements including hardware necessities such as high memory and disk space, and installation of relational databases MySQL and MCL programs for its efficient and smooth running. This program works on an UNIX-compatible server and can only be installed by a bioinformatician for a real setup, as special permissions are a prerequisite. Generally, the availability of a server is limited and is its

use is often time-consuming. On the other hand, EDGAR is an easy-to-use automated program that uses BBHs as orthology criterion for genomic calculations. The only requirement of this program is to open a 'project' account by registering online for using the web-based comparative genomics services. A drawback of BBHs is that only one-to-one orthologous pairs are found, and therefore for duplicated genes or paralogs only a single hit will be found (Blom *et al*., 2009). However, in a recent study, the comparison of the performance of BBHs with 11 orthology estimation methods concluded that BBHs gave comparable results to the more sophisticated methods such as OrthoMCL (Altenhoff & Dessimoz, 2009). Since, EDGAR was easily available and was sufficiently fast to handle the large amounts of sequence data generated in the comparative genomics analyses, in-depth analysis of the singletons and other genomic analyses (e.g. phylogenetic tree building) were carried out using this program.

4.4.1.4 The accessory genome of the SRE

The accessory genome of each strain was variable, with the initial number of strain-specific genes (singletons) for each genome (predicted by OrthoMCL and EDGAR) representing approximately one third of the predicted pan genome (Table 4.4). The majority of the singletons were not assigned function. The remainder of the CDSs was, however, enriched for those encoding proteins involved in DNA replication, mostly of phage origin, and regulation.

From comparative analysis using both OrthoMCL and EDGAR, a large number of predicted singletons were noted for draft genomes such as ATCC BAA 417 and WPP14. Since these genomes were highly fragmented, there was a high probability that the majority of the gene fragments were mis-annotated due to sequencing errors or inconsistent genome assemblies (repetitive regions associated with short pyrosequencing reads). Glasner *et al*. (2008) reported that gene fragments resulting from likely sequence errors in the draft genomes lead to gene count differences between species even within orthologous regions. Also, the use of different gene prediction methods for the annotation of different strains may have proven a barrier for classification of clusters due to their high error rates. This possibility was highlighted in a recent

comparison of the proteomes of SCC3193 and other closely related SRE (Nykyri *et al*., 2013). This problem was over-come by unifying the gene predictions for each genome before clustering the proteomes for their comparative analyses.

In this study, a different approach was taken for determining the accuracy of the core genome and predicted strain-specific genes, as it was believed that there was a strong likelihood that unifying the gene predictions would not remove the annotation biases caused by the various levels of genome fragmentation (which ranged from the highly fragmented genomes to those near to completion). Hence, blastn analysis of the singletons of each genome (predicted by EDGAR) was iteratively carried out against the whole genome sequence of the comparator genomes. Some of the shorter predicted CDSs from the reference genome used in the comparison, initially assigned as singletons, showed high homology to the sequence of a comparator genome, where the sequence of the comparator genome was not assigned as a CDS. The homology score was related to the h-val and, therefore, the CDS with an h-val score of >0.8 was removed from the singleton list and placed in the accessory genome list. Accordingly, the number of genes predicted to be singletons for a particular genome was re-calculated and a clear reduction in the number of singletons from the initial analysis was observed for the highly fragmented genomes (Table 4.4). Similar approaches in eliminating false strain-specific genes have been used in other recent studies by Mann *et al*. (2013) and Pedron *et al*. (2014).

Table 4.4 Singleton summary of each strain

Strain	No. of genes predicted to be singletons by OrthoMCL	No. of genes predicted to be singletons by EDGAR	Final no. of genes predicted to be singletons after manual curation
Pectobacterium carotovorum* subsp. *brasiliensis			
ICMP19477	160	157	105
ATCC BAA 417	724	681	481
Pectobacterium atrosepticum			
SCRI1043	96	193	72
ICMP1526	316	341	254
UGC32	240	257	187
Pectobacterium carotovorum* subsp. *carotovorum			
WPP14	518	431	170
ICMP5702	262	272	203
PC1	180	192	119
PCC21	235	241	189
***Pectobacterium* wasabiae**			
SCC3193	317	229	161
WPP163	149	143	105
Erwinia crysanthemi* or *Dickeya			
Dda3937	559	615	436
Ech703	448	313	238
Ech586	297	481	187

4.4.1.5 Does the pan genome predicted in this study represent the pan genome of the SRE?

The number of genomes required to estimate the size of a species pan genome has been mathematically modelled by several groups (Hogg *et al*., 2007; Tettelin *et al*., 2005; Tettelin *et al*., 2008) leading to the concept of 'open' and 'closed' pan genomes (Medini *et al*., 2005). Depending on the sequencing of new strains, the addition of new genes to the gene repertoire of a species would be considered as finite (closed pan genome) or unlimited (open pan genome) (Tettelin *et al*., 2008). Species with an open pan genome usually span different ecological niches. For example, both *E*. *coli* and *S. agalactiae* have open pan genomes (Table 4.4), where, with the addition of each newly sequenced genome, there is a predicted gain of approximately 300 and 33 novel genes, respectively (Medini *et al*., 2005; Rasko *et al*., 2008). In contrast, the gene repertoire of a closed pan genome reaches a saturation point where no further addition of genes is predicted. One such example is the genome of *Buchnera aphidicola* that showed almost no gene rearrangements (lateral exchange) in the past 50-70 million years and therefore, its pan genome was denominated as closed (Tamas *et al*., 2002). Another such example is the closed pan genome of *Bacillus anthracis*, where the number of specific genes added to the gene repertoire dropped to zero after the sequencing of only four genomes (Tettelin *et al*., 2005). A recent

example of a closed pan genome is that of *S. aureus*, which was predicted to eventually reach a genetic saturation with a defined ecological niche (Boissy *et al*., 2011). Species with closed pan genomes usually have limited access to the global microbial gene pool with a high degree of specialization to a narrow ecological niche and mostly have a low capacity to acquire foreign genes (Medini *et al*., 2005).

Although the SRE consist of multiple species and genera, due to the vageries of SRE taxonomy, the status of the SRE pan genome was examined. Singleton development analysis estimated that about 314 novel CDSs would be added to the pan genome with each additional SRE genome sequenced (Figure 4.2). This estimate of strain-specific genes was smaller than earlier estimates (Glasner *et al*., 2008), which was not surprising as the number of unique genes was expected to drop with the addition of new genomes in the comparison. Nevertheless, the data suggested that the pan genome of the SRE is 'open'.

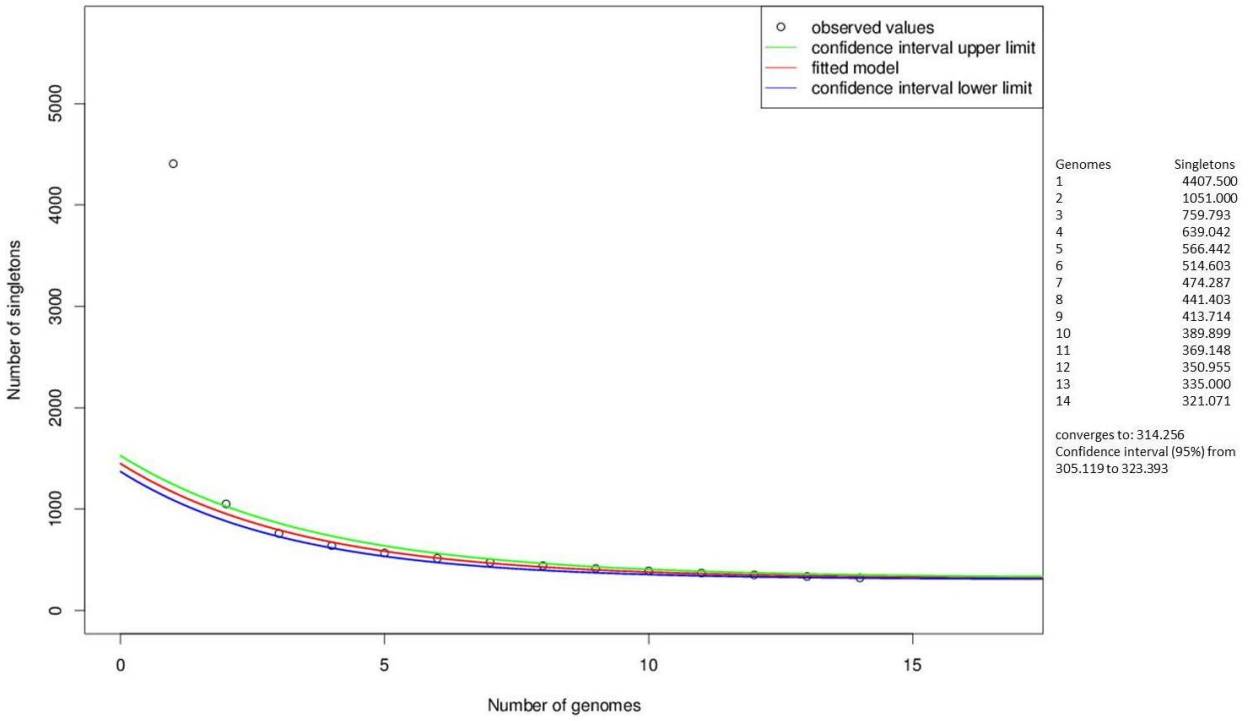


Figure 4.2 Singleton development plot for the soft rot erwiniae (SRE) genomes. The number of singletons (314 ± 9) almost reaches a saturation point with an increasing number of SRE genomes. The red line shows the number of singleton genes as a function of the number of compared genomes (number of genomes = 14, Table 4.1). The mean values for all the possible strain combinations of each respective gene count are taken. The green and dark-blue lines show the upper and lower limit of a 95% confidence interval of the curve fit. The fitted model converges to the predicted number of singleton genes, as calculated by the EDGAR program.

The core genome development plot showed that the core genome for the SRE is likely to comprise 2047 CDSs (or genes), only slightly lower than the 2,124 predicted by OrthoMCL (Figure 4.3). Nevertheless, more genomes will need to be sequenced before the pan genome is closed and the core genome of the SRE is confirmed.

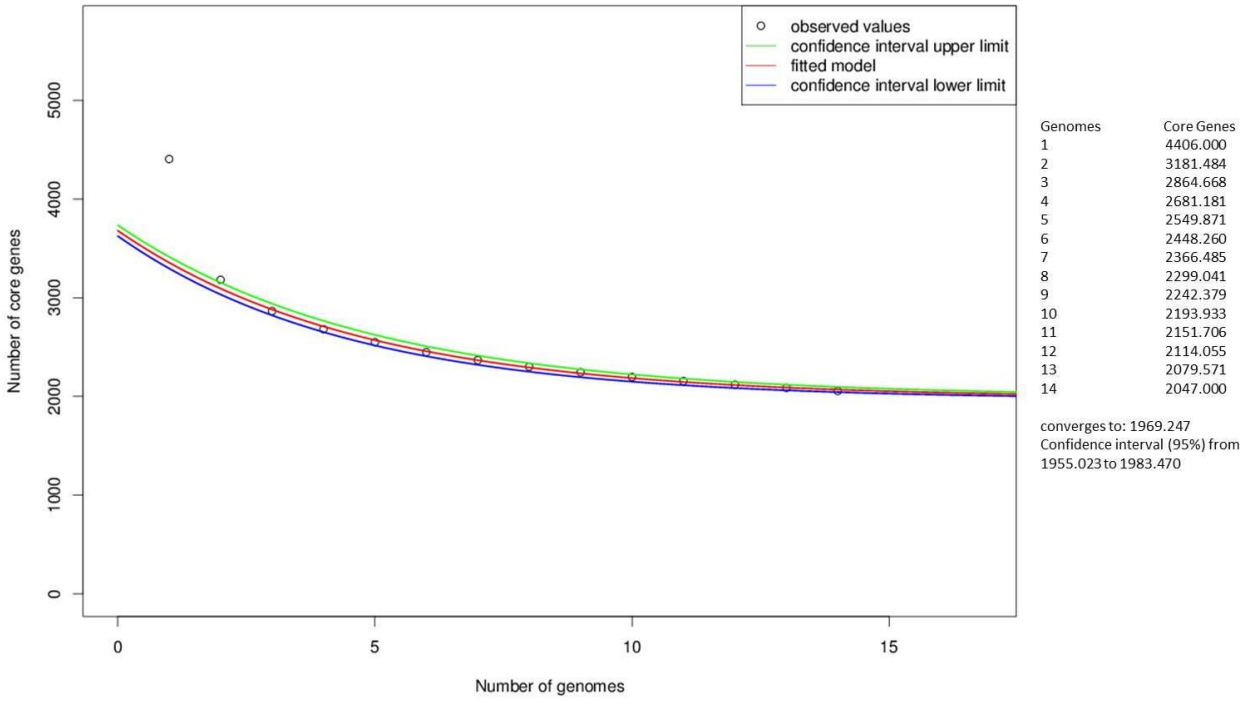


Figure 4.3 Core development plot for the eight complete and six draft soft rot erwiniae (SRE) genomes. A reduction in the core genome is observed with the inclusion of every added genome sequence, indicating that the pan genome of *Pectobacterium* is 'open'. The red line shows the number of core genes as a function of the number of compared genomes (number of genomes = 14, Table 4.1). The mean values for all the possible strain combination of each respective gene count are taken. The green and dark-blue lines show the upper and lower limit of a 95% confidence interval of the curve fit. The fitted model converges to the predicted number of core genes, as calculated by the EDGAR program.

4.4.2 Functional annotation of the genes in the core and accessory genome of the SRE

As expected, most of the genes in the core genomes of the SRE were proteins associated with fundamental biological functions such as genes involved in housekeeping functions including cellular processes, DNA and RNA metabolism, protein processing and secretion, and energetic and intermediary metabolism. Functional annotation of the remaining core genes showed they were probably involved in metabolic processes, especially energy production and conversion, carbohydrate transport and metabolism and amino acid transport and metabolism, replication, transcription, translation, the tricarboxylic acid cycle, and nucleotide, fatty acid and phospholipid metabolism (Appendix A4.3).

Genes associated with housekeeping functions were also found within the dispensable genome. The majority of the CDSs in the accessory genome, however, could not be assigned to one of the COG classifications and the predicted proteins were assigned unknown functions. In fact, between 73 and 84% of proteins of unknown function (or conserved proteins of unknown function) were encoded by the strain-specific genes (Appendix A4.6). These genes may hold clues to the traits and mechanisms underlying the biological diversity of the SRE, and rigorous gene-by-gene mutagenesis and phenotypic analysis would determine their specific roles in disease development. These genes were probably acquired after the radiation of each genus/species and, as such, were likely acquired by HGT and not by vertical gene transfer. Genes associated with mobile elements were abundant in the dispensable genome, supporting the hypothesis that the majority of specific traits depend on lateral transfer events.

4.4.3 Confirmation of the species status of ICMP19477 using genomic information

The taxonomic status of ICMP19477 was confirmed using several genome-based analyses described below.

4.4.3.1 Phylogenetic analysis using the core SRE genome identifies ICMP19477 as *P. carotovorum* subsp. *brasiliensis*

Phylogenetic analysis of the SRE in this study using their predicted core genomes, grouped ICMP19477 into a clade with *P. carotovorum* subsp. *brasiliensis* ATCC BAA 417 (Figure 4.4). This clade was distinct from a second clade containing other *P. carotovorum* subsp. *carotovorum* strains including the type strain ICMP5702, suggesting that *P. carotovorum* subsp. *brasiliensis* represents a group of SRE that are genetically distinct from other *P. carotovorum* strains.

UGC32, which was previously identified as a strain of *P. carotovorum* subsp. *carotovorum*, was placed in a clade closer to strains of *P. atrosepticum* in this tree (Figure 4.4). *Pectobacterium atrosepticum* and *P. wasabiae*, both associated with blackleg of potato, were themselves

grouped together in a lineage distinct from the *P. carotovorum* subspecies. This may explain why UGC32 was previously shown to cause blackleg and to encode CFA (Slawiak and Lojkowska, 2009), as CFA is a known virulence factor in *P. atrosepticum* (Bell *et al.,* 2004). Misidentification of *Pectobacterium* species or difficulties in *Pectobacterium* taxonomy has already been documented in Chapter 2 as well as by recent studies performed by other reseachers. For example, Gross *et al*. (1991) were unable to classify over 50% of *Pectobacterium* isolates obtained from potato and Pitman *et al*. (2008) were unable to type 13% of their isolates.

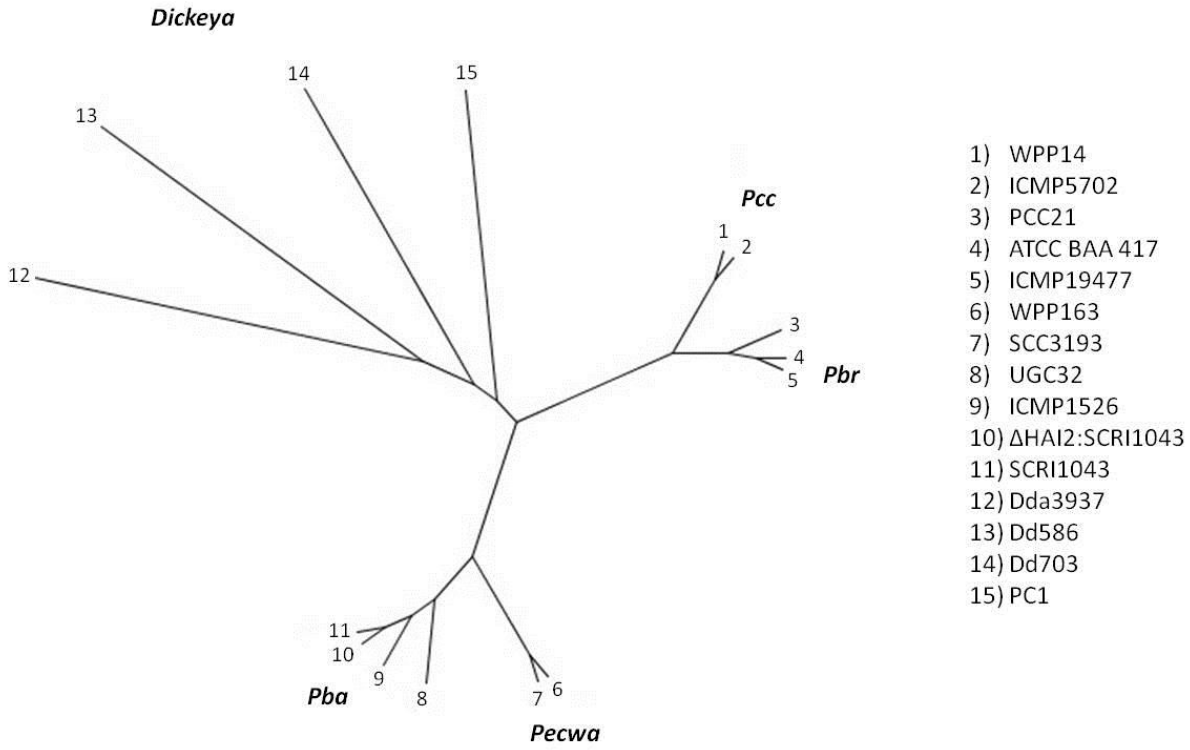


Figure 4.4 A maximum likelihood phylogenetic tree constructed from a multiple alignment using concatenated sequences of the core genome of each soft rot erwiniae (SRE). The maximum likelihood phylogenetic tree was constructed using the LG substitution model. Gamma distribution was used to model evolutionary rate differences among sites. The phylogenetic tree was generated using 1,000 bootstrap replications, with all branches supported by a bootstrap value of 1 (%). The genome of ΔHAI2: SCRI1043 (HAI2 removed from the genome, Dy *et al*., 2013) was used as an internal control to check the confidence of this tree. *Pbr – P. carotovorum* subsp. *brasiliensis*; *Pba – P. atrosepticum*; *Pcc – P. carotovorum* subsp. *carotovorum*; *Pecwa – P. wasabiae*; *Dd/Dda – D. Dadantii*

4.3.2 Pairwise alignment of ICMP19477 with other *Pectobacterium* strains shows nearly complete synteny between ICMP19477 and *P. carotovorum* subsp. *brasiliensis* ATCC BAA 417

Genome conservation is largely indicated by the location, the order and orientation of a gene (Darling *et al*., 2010). Comparison of multiple genomes based on their gene orders reveals segments of homologous gene content and also evolutionary relationships of organisms (Zhang *et al*., 2010). In this context, synteny refers to multi-gene regions where DNA sequences and gene order are conserved between genomes (Barloy-Hubler *et al*., 2001; Bentley and Parkhill, 2004). Whole genome alignments of several strains of *Pectobacterium* used in this study were conducted to establish their degree of synteny, using the MAUVE aligner program (Darling *et al*., 2004). Alignment of the *Pectobacterium* genomes for ICMP5702, SCRI1043 and ATCC BAA 417 using ICMP19477 as the reference revealed that the majority of the genomic regions existed as collinear blocks (Figure 4.5). Some of the homologous genomic regions, however, appeared to be present as inverted regions while other regions were missing entirely (Figure 4.5). This may be linked to the intra- and extracellular movement of DNA in bacteria, likely arising due to recombination events that lead to gene acquisition and loss, duplication, inversion, transposition, and genome rearrangement.

ICMP19477 showed almost complete synteny with ATCC BAA 417, differing only by ~2% using ANIb. In contrast, ICMP19477 differed from ICMP5702 and SCRI1043 by 7% and 11% respectively. This suggests that ICMP19477 and ATCC BAA 417 belong to the same species, as the threshold used to distinguish bacterial species is 95% (Richter and Rossello-Mora, 2009). Furthermore, *P. carotovorum* subsp. *brasiliensis* may be a distinct species rather than just a subspecies of *P.* carotovorum. These results were consistent with the MLSA data collected in Chapter 2, confirming the identity of ICMP19477 as *P. carotovorum* subsp. *brasiliensis*.

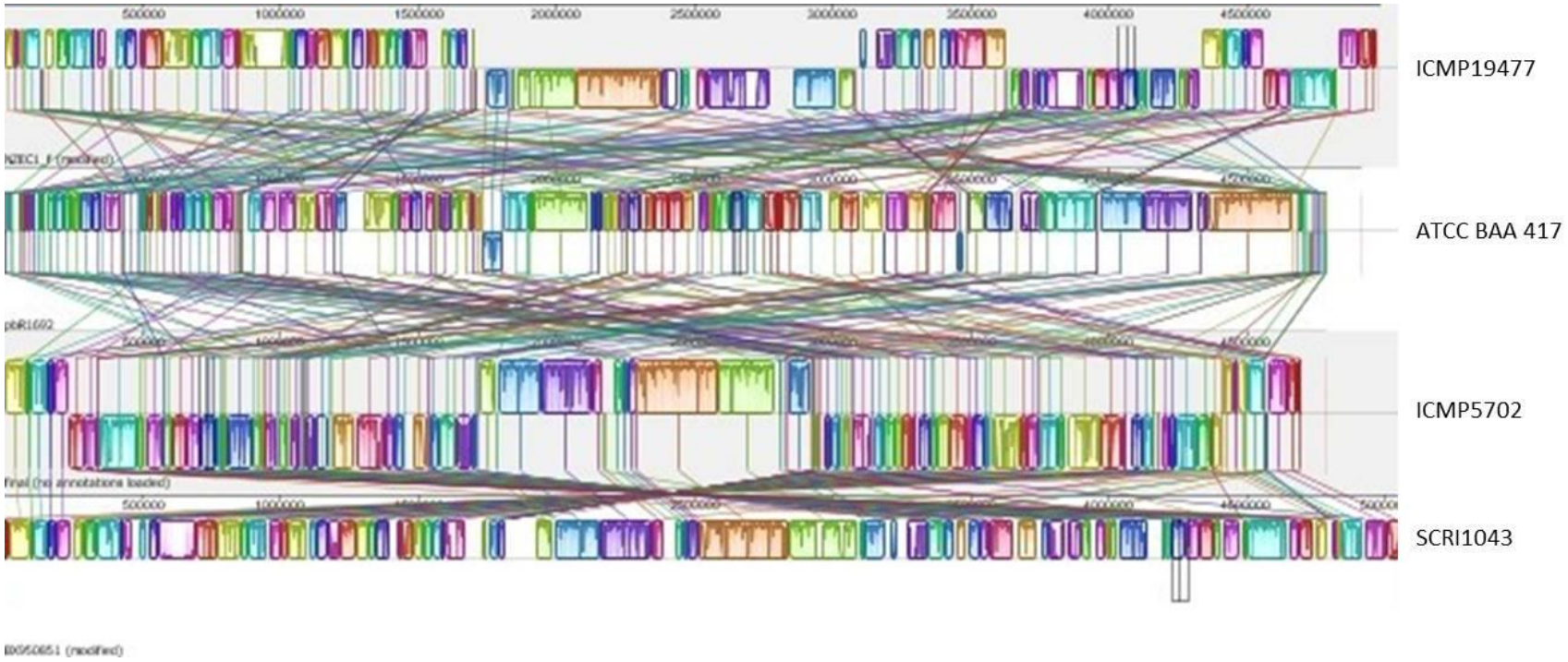


Figure 4.5 Pairwise alignment between the genomes of ICMP19477, ATCC BAA 417, ICMP5702 and SCRI1043 using MAUVE. Coloured blocks outline genome sequence that aligns to part of another genome, and is presumably homologous and internally free from genomic rearrangement (Locally Collinear Blocks or LCBs). Areas that are completely white were not aligned and probably contain sequence elements specific to a particular genome or may be due to gaps between contigs. Blocks below the centre line indicate regions that align in the reverse complement (inverse) orientation. A profile is drawn within each LCB with the height of the colour corresponding to the average degree of sequence conservation.

4.4.3.3 H-val analysis of ICMP19477 with other *Pectobacterium* strains showed high similarity to *P. carotovorum* subsp. *brasiliensis* ATCC BAA 417

The h-val of each ORF was calculated and a score between 0 and 1 was assigned to each ORF of ICMP19477 used in the comparison (Appendix A4.4). An h-val of 0.1 suggested the absence of the ORF in the other comparison genome. In contrast h-vals >0.8 indicated that the ICMP19477 ORF should have an almost identical homologue in the comparison genome. However, it was difficult to evaluate the presence of a counterpart for an ORF that had an h-value in the middle range (Fukiya *et al*., 2004). Hence, the h-vals were colour coded based on the similarity level and were plotted in a circular diagram using the CIRCOS package (Figure 4.6). This provided a better representation of the genomes used in the genome comparisons (Mark Fiers, personal communication).

Visual diagrammatic representation of the colour-coded h-vals showed that ICMP19477 was a member of *P*. *carotovorum* subsp. *brasiliensis* due to the higher level of identity of the ORFs in ICMP19477 and ATCC BAA 417. H-val analysis is a useful approach to conduct comparative analysis, as there is negligible annotation bias during comparison of the genomes under study.

4.4.4 GIs of ICMP19477

Alignment of the ICMP19477 genome with the genomes of other SRE revealed the presence of unaligned regions, which were likely to be strain-specific (or isolate-specific) sequences, or rearranged regions. Strain-specific (or isolate-specific) sequences or rearranged regions, known as GIs, are predicted to have been acquired by HGT, and to encode many niche specific traits such as virulence determinants (Dobrindt *et al*., 2004). Therefore, two different bioinformatic approaches were applied here to detect putative GIs in the genome of ICMP19477 and to identify genes that might be related to the virulence of this pathogen.

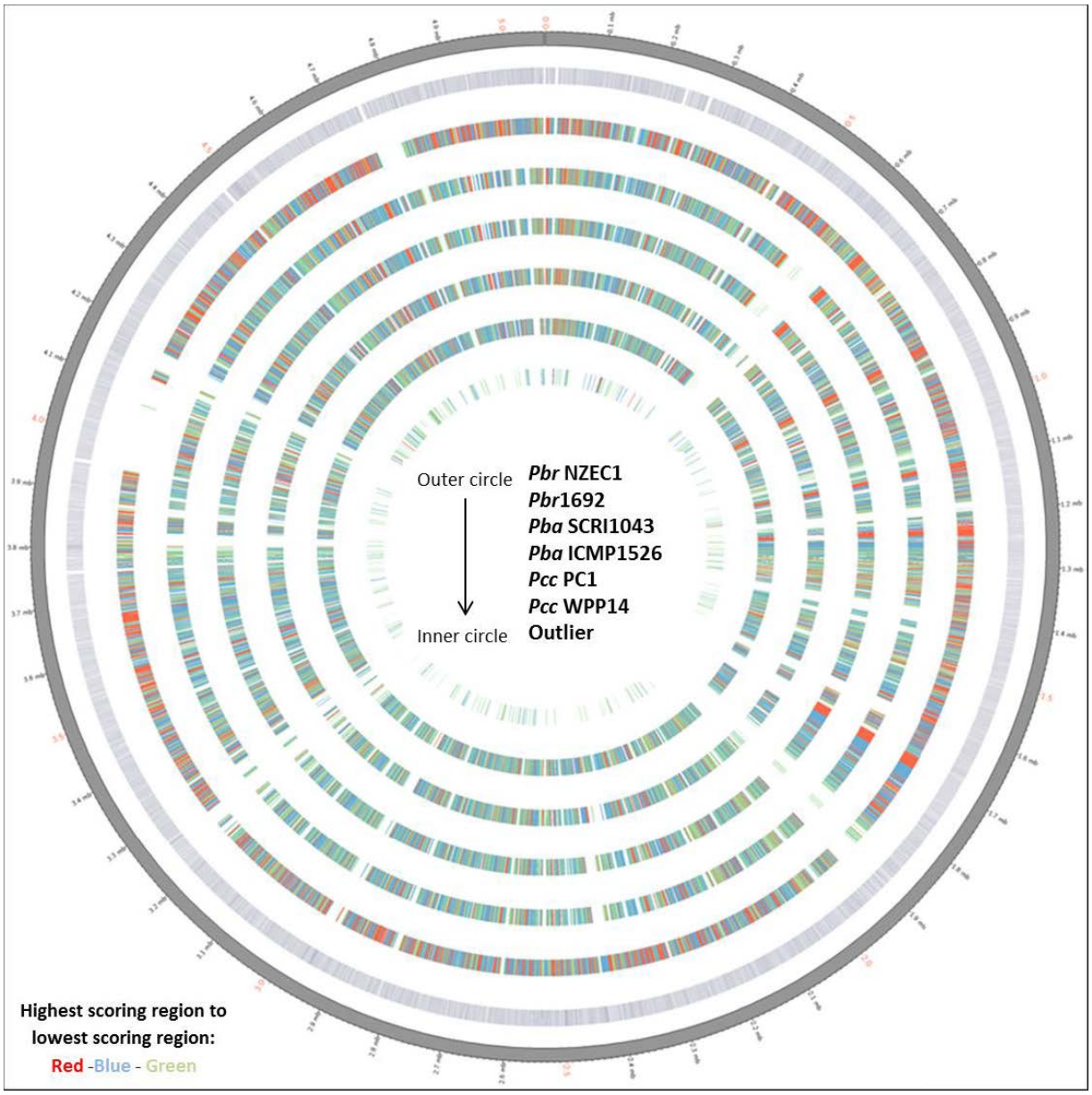


Figure 4.6 A circos plot of the h-vals resulting from blastn analysis of all the coding sequences in *Pectobacterium carotovorum* subsp. *brasiliensis* ATCC BAA 417 (*Pbr*1692), *P*. *atrosepticum* SCRI1043, *P*. *atrosepticum* ICMP1526, *P*. *carotovorum* subsp. *carotovorum* PC1 and *P*. *carotovorum* subsp. *carotovorum* WPP14 using *P*. *carotovorum* subsp. *brasiliensis* ICMP19477 (*Pbr*NZEC1) as the reference. *Erwinia amylovora* strain ATCC BAA 2158 was used as an outlier to generate this CG plot. H-values between 0.8 and 1 is colour coded as red, values between 0.4 and 0.6 is coded as blue and scores less than 0.4 are colour coded as green. Regions of similarity are colour coded based on the h-vals and the white regions in this plot depict unique regions in the strain being compared. ICMP19477 was most similar to ATCC BAA 417 from this colour-coded h-val analysis.

Firstly, a sequence composition-based method was used. Such methods have been shown to be capable and versatile tools for detecting GIs and straightforward to apply, as they require only a single genome (i.e., the target genome). However, they are limited in that they may lead to false positive GI predictions due to factors associated with bias in sequence composition, such as clusters of highly expressed genes (Karlin *et al*., 1998), and missing ameriolated islands due to ancient acquisition events (Lawrence & Ochman, 1997). Nevertheless, the SIGI-HMM and IslandPath-Dimob methods used here were found to have the highest overall accuracy of various detection methods tested in a recent bioinformatics study (Langille *et al*., 2008).

Secondly, a comparative genomics-based approach was used which, although more robust, is often less accessible as it requires availability of several other closely related genomes to esentially look for clusters of genes present in one genome and absent from others. Genome alignment methods use pairwise alignment of the query genome against the comparison genomes to identify unique regions in the query genome based on insertions, deletions, inversions and other rearrangements in all of the genomes. In other words, GIs were predicted by searching for breaks in synteny between the ICMP19477 genome (target genome) and a set of closely related bacteria. However, there are certain limitations to this approach also: (i) inclusion of very distant genomes (with extensive rearrangements) which may be the only ones available for comparison may result in false positive predictions of GIs due to difficult genome alignments, (ii) GIs that have inserted before the divergence of the genomes being compared will not be predicted in the subsequent very closely related genomes and, (iv) that comparative genomics approach relies on the manual selection of the query genome and the available comparator genomes, which may result in inconsistent selection criteria due to the unfamiliarity of different phylogenetic distances within genera (Langille *et al*., 2008).

The features of GIs that can be appropriately detected by each of these methods have been listed in Table 4.5.

During the GI analyses, it quickly became evident that the prediction of GIs was ambiguous. Different GI prediction methods estimated different numbers of individual and contiguous ORFs in the genome of ICMP19477, based on evidence such as divergent base composition and the presence of flanking insertion sequences, tRNAs and/or mobility genes. Thus, the different computational methods for predicting islands were pooled and successive ORFs were manually evaluated and clustered as/into a single region. A total of 79 GIs (Figure 4.7) were identified in the genome of ICMP19477. Further curation of the automated predictions by spanning the islands across multiple ordered contigs (since, the genome of ICMP19477 is not complete), resolved these PbN1_GIs into a total of 69 (i.e. predicted islands consisting of >5 ORFs) and clusters of shorter islands termed gene islets (islands consisting of <5 ORFs) (Appendix A4.5). These GIs were numbered as PbN1_GI1 to PbN1_GI69 in accordance with the increasing genomic coordinates of ICMP19477. The presence or absence of these identified ICMP19477 islands in the other comparator genomes was also determined by conducting blastn searches of the GI nucleotide sequences in the other genomes. Since, the majority of the genomes were draft sequences, the presence of GIs were estimated based on the query coverage and nucleotide identities (calculated as the h-val) of the particular genome region being compared. H-val scores between 0.8 and 1.0 inferred the same GI was present, a score between 0.6 and 0.8 inferred that a similar GI (insertion or deletion event may have occurred) was present, and a score of less than 0.4 suggested that the GI was absent or a different or a new GI had been inserted in that location of the comparator genome. A summary of these is listed in Appendix A4.5.

Various proteins were encoded by the GIs identified in ICMP19477, including those involved in pathogenicity, adaptation, regulation, mobility and cell processes (Appendices A4.2, A4.5). Many of the determinants identified on the GIs were already known to influence the virulence of *Pectobacterium*. For example, genes related to type II, III and IV secretion systems, phytotoxin production and host cell wall degrading enzymes were found on GIs by Bell *et al*. (2004) and Glasner *et al*. (2008). GIs or gene islets specific to only *P*. *carotovorum* subsp. *brasiliensis* were also detected that might provide clues to the aggressive nature of this pathogen. Specific GIs and

their candidate virulence determinants are discussed in more detail in the remainder of this chapter.

Table 4.5 A summary of the features of GIs restricted to each detection method

Features of GIs	Detection method	Pitfalls of detection method
Usually large segments > 5 Kb	Comparative genomics to identify large insertions	Single HGT events are harder to detect over large horizontally acquired regions
Inserted at a tRNA site	Using tools such as BLAST and tRNAscan-SE	Not all GIs are inserted at a tRNA site
Flanked by direct repeats (DRs)	Repeat finder tool such as REPuter, direct repeat finder	Not all GIs are flanked by DRs and identification is difficult due to variances in the length of DRs
Unusual G+C% content and sequence composition bias	Various sequence composition based detection methods	Predictions are biased due to positive predictions (highly expressed genes) and negative predictions (gene amelioration)
Generally unstable and can excise spontaneously	Comparative genomics	Need of closely related sequenced genomes for comparison
Encode mobility genes or elements such as integrases, transposases and insertion elements	Similarity search using BLAST	Not all GIs are associated with mobility genes, tools not available for detection
Associated with virulence factors, phage-related genes, and genes of unknown functions	Using functional databases for comparison such as COGs	Tools not available for detection

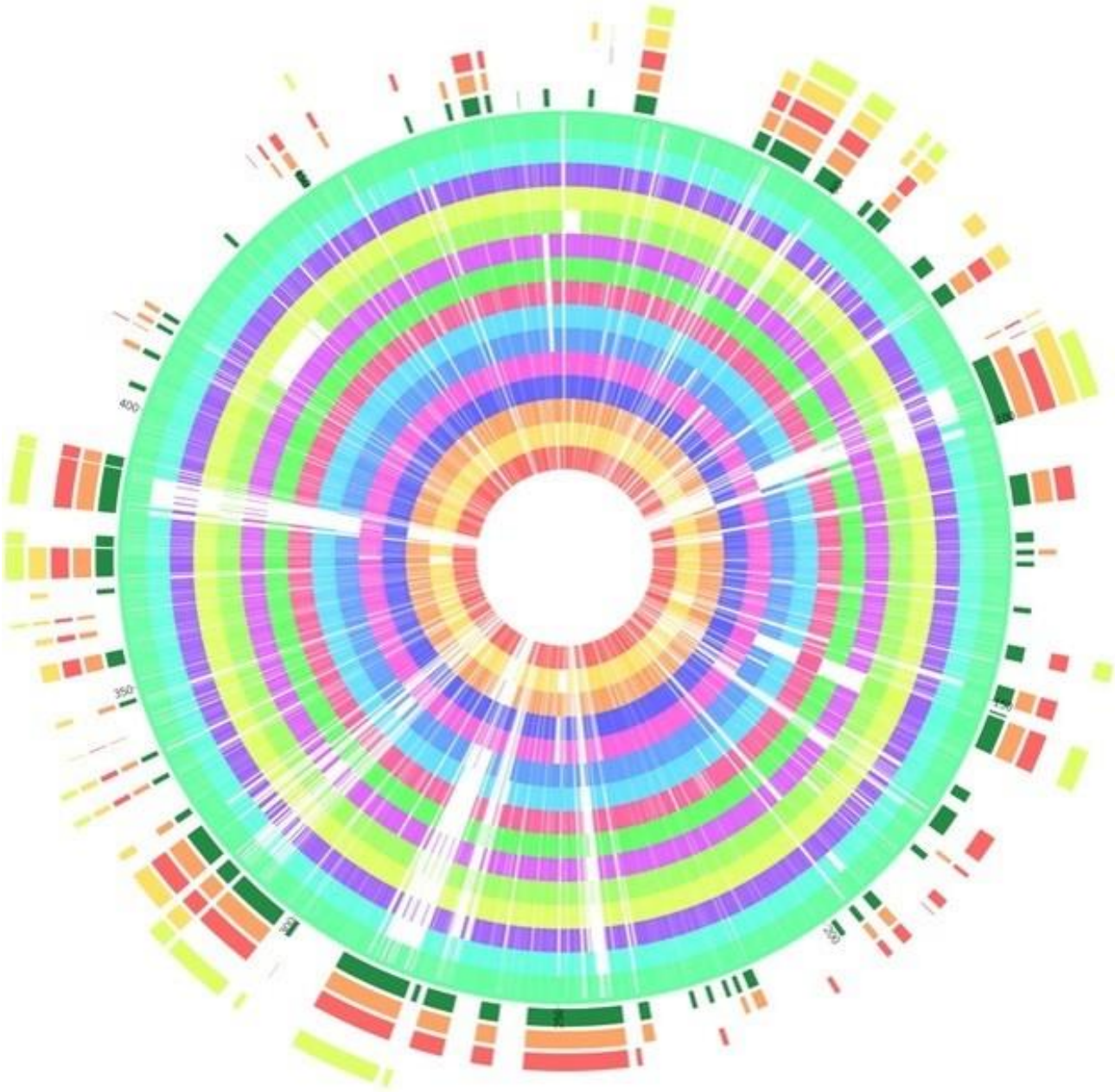


Figure 4.7 A circos plot showing a comparison of the genome of ICMP19477 with the genomes of 15 other soft rot erwiniae (SRE) and the location of computationally predicted genomic islands (GIs). The circles from outside to inside represent: islands associated with tRNAs, islands detected by islandpath-DIMOB, SIGI-HMM detected islands, alien hunter detected islands and indegenius detected islands. Inside the circle representing the islands detected by Indegenius (in green) are the genomes of the 15 SRE. From outside to inside, the circles represent: ICMP19477, ATCC BAA 417, SCRI1043, SCRI1043ΔHAI2, ICMP1526, UGC32, ICMP5702, WPP14, PC1, PCC21, WPP163, SCC3193, Dd586, Dd703 and Dda3937. The genome of SCRI1043ΔHAI2 was used to confirm the diagrammatic representation was correct, as it has HAI2 removed from the genome (Dy *et al*., 2013; Vercoe *et al*., 2013).

4.4.5 Previously characterized virulence determinants of *Pectobacterium* in the genome of ICMP19477

4.4.5.1 PCWDEs

A total of 64 PCWDEs including pectinases, polygalacturonases, cellulases, xylanases and proteases were identified in the genome of ICMP19477 (Table 4.6). PCWDEs are the major virulence determinants of SRE. These enzymes degrade the structural components of the plant cell wall, enabling infections that lead to soft rot and wilt symptoms.

Most of the genes encoding PCWDEs in ICMP19477 were detected in all the sequenced strains (Table 4.6). In addition, a novel M20 family peptidase, PepV, was encoded on the putative GI, PbN1_GI51. This island also carries two copies of the hydantoinase B/oxoprolinase gene and other ABC transporter genes. The peptidase and the hydantoinase genes are predicted to encode proteins that exhibit hydrolase and catalytic activity, respectively. *PepV* and the hydantoinase B/oxoprolinase genes were also present in ATCC BAA 417, but were absent from other SRE. ICMP19477 also encoded an additional putative rhamnogalactouronase lyase, which was produced by ICMP5702 but not all other comparator strains.

The known regulators of PCWDEs, KdgR and ExpI, were encoded by all strains of SRE used in this study. Genes encoding DgA, RdgB, RexZ, were only encoded in the genomes of the *Pectobacterium* strains.

Table 4.6 PCWDEs encoded in the genomes of the SRE used in this study and the accession numbers of their CDSs

Enzyme	ICMP19477	ATCC BAA 417	ICMP5702	WPP14	SCRI1043	ICMP1526	UGC32	PC1	PCC21	Dda3937	Dd586	Dd703	SCC3193	WPP163
Pectate lyases														
pelL	A_15177	B_03243	C_2191	D_19342	ECA2553	E_17512	F_1755	G_1773	H_17800	I_03201	J_2245	K_2204	L_1991	M_2045
pelW	A_14522	B_01100	C_2057		ECA2402	E_16807	F_1612	G_1908	H_19680	I_03361	J_1962	K_2033	L_2143	M_2201
	A_13102	B_19083	C_1806	D_02609	ECA2135	E_04231	F_1340	G_2170	H_22490				L_2484	M_2511
pelI	A_07805	B_10277	C_2969	D_07878	ECA1094	E_13965	F_2729	G_0988	H_10230	I_00058	J_2937		L_3319	M_3322
PelA	A_04992	B_04132	C_4261	D_05524	ECA4067	E_20159	F_4074	G_3858	H_38500				L_4192	M_4035
pelB	A_04987	B_04137	C_4260	D_05529	ECA4068	E_20164	F_4073	G_3859	H_38510				L_4193	M_4036
pelC	KCQ_04977		C_4259	D_05534	ECA4069	E_20174	F_4072	G_3860	H_38520	I_04192	J_3789	K_0352	L_4194	M_4037
pelZ	A_04972	B_04147	C_4258	D_05539	ECA4070	E_20179	F_4071	G_3861	H_38530	I_04191	J_3790	K_0351	L_4195	M_4038
pelX	A_05908	B_13194	C_4049	D_11198	ECA4510	E_09116	F_10	G_4259	H_42370	I_00143	J_4161	K_3996	L_4748	M_4562
	A_18257	B_14269	C_1155	D_11368	ECA3112	E_00905	F_643	G_2863	H_29360					
hrpW	A_13062	B_14583	C_1793	D_22506	ECA2112	E_04171	F_1328	G_2182	H_22630	I_02943	J_2002			
Pectin lyases														
Pnl	A_17207	B_19663	C_2591	D_07668	ECA1499	E_19692	F_2093	G_1371	H_14030	I_03551		K_3378	L_1691	M_1777
Pectin methlyesterases														
pemA	A_17502	B_07154	C_978		ECA3253	E_00115	F_479	G_3047	H_30850	I_03374	J_3087	K_2766	L_1138	M_1228
pemB	4448275-4449479	B_16043	C_102	D_12090	ECA0107	E_06747	F_2525	G_4146	H_00960	I_03435			L_4605	M_4427
Pectin acetylesterase														
paeX	A_14552	B_01130	C_2063	D_06799	ECA2408	E_16837	F_1618	G_1902	H_19620	I_03363	J_1956	K_2039	L_2136	M_2195
Polygalactouronases														
pehX	A_18262	B_14264	C_1156	D_11363	ECA3111	E_00910	F_644	G_2862	H_29350	I_00261	J_3904	K_0200	L_1293	M_1410
pehN	A_07310	B_15783	C_2862	D_21038	ECA1190	E_14435	F_2418	G_1092	H_11240	I_04155			L_3221	M_3223
pehA	A_07800	B_10282	C_2968	D_07883	ECA1095	E_13970	F_2728	G_0989	H_10240				L_3318	M_3321
pehK	A_02907	B_14024	C_600	D_21293	ECA3552		F_215	G_3371 G_2125	H_33770	I_00206	J_3319	K_0972		
Oligogalacturonide lyase														
Ogl	A_14642	B_01220	C_2082	D_06889	ECA2426	E_16927	F_1637	G_1884	H_19440	I_03686	J_1940	K_2056	L_2117	M_2176
Rhamnogalactouronate lyase														
rhiE	A_09675	B_11037	C_3276	D_02994	ECA0804	E_11860	F_3053	G_0682	H_07150	I_01465	J_2097		L_0801	M_0910
	A_02032		C_426											
	A_17507	B_07149	C_979	D_12945	ECA3252	E_00120	F_480	G_3046	H_30840	I_03373	J_3086	K_2765	L_1139	M_1229
Cellulases														
	A_00545	B_12934	C_3941	D_13080	ECA4370	E_08396	F_3773	G_0076	H_41310	I_01990	J_4061	K_3895	L_0074	M_0073
	A_00550	B_12939	C_3942	D_13085	ECA4372	E_08401	F_3774	G_0075	H_41320				L_0073	M_0072
bcsZ	A_00555	B_12944	C_3943	D_13090	ECA4373	E_08406	F_3775	G_0074	H_41330				L_0072	M_0071
	A_00560	B_12949	C_3944		ECA4374	E_08411	F_3776	G_0073	H_41340				L_0071	M_0070
	A_00530	B_12919	C_3938	D_13065	ECA4366	E_08381	F_3770	G_0079	H_41280				L_0077	M_0076
celV	A_12447	B_02737	C_1654	D_03174	ECA1981	E_03495	F_1187	G_2337	H_24390				L_2582	M_2612
celS	A_16467	B_08523	C_2457	D_16303	ECA2827	E_18727	F_1968	G_1503	H_15250				L_2949	M_2966
Proteases and peptidases														
	A_11255	B_16708	C_3624	D_07531	ECA0388	E_10390	F_3459	G_0378	H_03960				L_0441	M_0429
	A_10690	B_21075	C_3508					G_0472			J_3669	K_0479	L_0586	M_0567
	A_09680	B_11032	C_3277	D_02989	ECA0803	E_11855	F_3054	G_0681	H_07140				L_0800	M_0909

Enzyme	ICMP19477	ATCC BAA 417	ICMP5702	WPP14	SCRI1043	ICMP1526	UGC32	PC1	PCC21	Dda3937	Dd586	Dd703	SCC3193	WPP163
	A_09250	B_16293	C_3203	D_13320	ECA0879	E_12395	F_2966	G_0761	H_07830	I_01405	J_0785		L_0891	M_0998
Prt1	A_17722	B_06934	C_1022	D_15546	ECA3211	E_00325	F_522	G_3006	H_30420	I_03712	J_3049	K_2731	L_1176	M_1266
	A_16772	B_16503	C_2488	D_08563	ECA3126	E_00835	F_632	G_1464	H_14840	I_04145	J_1841	K_1778	L_1276	M_1367
					ECA2515	E_17312							L_2036	M_2090
	A_12872	B_14418	C_1756	D_14599	ECA2074	E_03981	F_1289	G_2217	H_22960	I_01925	J_0978	K_2929	L_2503	M_2529
	A_12507	B_02797	C_1671	D_03234	ECA1988	E_03535	F_1200	G_2325	H_24270	I_03123	J_1719	K_1728	L_2575	M_2605
	A_16057	B_20230	C_2367	D_00839	ECA2771	E_18437	F_1890	G_1581	H_16040	I_02891	J_2439	K_2396	L_2864	M_2881
	A_16202	B_08768	C_2400	D_06364	ECA2785	E_18507	F_1922	G_1553	H_15770	I_03861	J_2059		L_2894	M_2909
	A_16302	B_08673	C_2422	D_06474	ECA2802	E_18602	F_1939	G_1530	H_15570	I_02906	J_2453	K_2412	L_2916	M_2931
	A_06220	B_09390	C_2639	D_05054	ECA1450	E_15735	F_2143	G_1326	H_13560				L_2993	M_3008
	A_06990	B_15648	C_2797	D_21263	ECA1290	E_14945	F_2307	G_1168	H_11920	I_02500	J_1123	K_1079	L_3161	M_3163
	A_02037	B_13354	C_428	D_18622	ECA3785	E_21330	F_4371	G_3562	H_35540	I_03095	J_3505	K_0677	L_3888	M_3747
													L_4075	M_3929
			C_3078		ECA0980	E_13375			H_09100	I_03393	J_0892			
	A_12602	B_02892	C_1691	D_03334	ECA2007	E_03635	F_1224	G_2287	H_23550					
				D_02779	ECA2163	E_04371	F_1371	G_2143	H_22200					
	A_19197	B_04667	C_1348	D_09035			F_843	G_2675	H_27410	I_03579	J_1383			
		B_15588	C_2802				F_2319	G_1156	H_11870					
								G_1154						
	A_14652	B_01225	C_2083	D_06894	ECA2427	E_16932	F_1638	G_1883	H_19430	I_03688	J_1937	K_2058	L_2116	M_2175
	A_19092	B_04777												
	A_14747	B_01330	C_2104	D_21991	ECA2448	E_17027	F_1658	G_1863	H_19240	I_02091	J_1925	K_2069	L_2093	M_2148
	A_17852	B_06804	C_1056	D_17940	ECA3193	E_00495	F_558	G_2976	H_30140	I_02814	J_3009	K_2709	L_1210	M_1304
Other PCWDEs														
Deacetylase	A_18122	B_07684	C_1112	D_19754	ECA3143	E_00750	F_613	G_2925	H_29630	I_01077	J_4199	K_4018	L_1259	M_1353
Deacetylase	A_09800	B_17980	C_3305	D_20813				G_0659	H_06900	I_00907	J_3274	K_0901		
Racemase	A_19082	B_04787	C_1334	D_09085	ECA2965	E_01775	F_809	G_2688	H_27510				L_1438	M_1554
Esterase	A_19087	B_04782	C_1335	D_09080			F_810	G_2687	H_27500					
							F_1348						L_2478	M_2504
Xylosidase	A_01712	B_14723	C_361	D_10441	ECA3847	E_21640	F_4304	G_3622	H_36210				L_3951	M_3811
Hydrolase	A_20487	B_02222	C_1550	D_17965	ECA1873	E_02950	F_1041	G_2433	H_25430	I_03836	J_1661	K_1670	L_2718	M_2738

A - KCO; B_ - PcarbP_0102000; C - final; D_ - PcarcW_0102000; E - KCQ; F - UGC32; G - PC1; H - PCC21; I - Dda3937; J - Dd586; K - Dd703; L - W5S; M - Pecwa

4.4.5.2 Secretion systems associated with pathogenicity or virulence

SRE encode six secretion systems, T1SS to T6SS, to transport molecules such as enzymes, effector proteins, pilus proteins and heavy metals from the bacterial cytosol to the environment or directly into the host cells (Tseng *et al*., 2009). All six secretion systems were identified here in the genome of ICMP19477, which are described below.

4.4.5.2.1 T1SS and T2SS

T1SSs and T2SSs are involved in the export of virulence factors proximal to host cells. Numerous proteases of the SRE are secreted by the T1SS (Delepelaire and Wandersman, 1991; Wandersman *et al*., 1990) that contribute to plant cell wall degradation (Martis *et al*., 1999). Previous studies on SRE have shown that the T1SS-secreted protease of SCC3193 and T1SS-secreted adhesin of SCRI1043 are required for virulence on potato (Martis et al., 1999; Perez-Mendoza et al., 2011a). The presence of the protease and adhesin genes in both ICMP19477 and the type strain of *P. carotovorum* subsp. *brasiliensis* suggest that these genes also probably contribute to the virulence of *P. carotovorum* subsp. *brasiliensis* on potato (Martis *et al*., 1999; Perez-Mendoza *et al*., 2011a).

The PCWDEs in SRE are secreted via T2SS. The *out* cluster of the T2SS, comprising of genes *outB-outO* (KCQ_18252, KCO_18262, KCO_18267, KCO_18272, KCO_18277, KCO_18282, KCO_18287, KCO_18292, KCO_18297, KCO_18302, KCO_18307, KCO_18312, KCO_18317, KCO_18322) and *outS* (KCO_18247), is present just before the carbapenem antibiotic cluster encoded on a GI, PbNI_GI48 in the genome of ICMP19477. These genes were conserved in all strains of SRE used here with the exception of *outN* that was absent from all of the *Dickeya* and *P. wasabiae* strains. *Out* mutants of *Pectobacterium* strains are impaired in the production of pectic enzymes and cellulase, making them avirulent, suggesting the importance of T2SS in virulence of SRE (Pirhonen, 1991; Reeves *et al*., 1993).

4.4.5.2.2 T3SS

The T3SS is an effector protein translocator machinery involved in suppressing the basal defence system of the host and is required for virulence in many biotrophic and hemibiotrophic plant pathogens (Buttner and He, 2009). In ICMP19477, 26 putative T3SS proteins were encoded on PbNI_GI37. Comparative analysis together with the other SRE revealed the presence of homologous T3SS genes in the genomes of all the *Pectobacterium* and *Dickeya* species, with the exception of *P. wasabiae* strains SCC3193 and WPP163 and *Dickeya* strain Dd703 (Table 4.7). Genes *hrpV*, *hrpP* and *hrcQ*, components of the T3SS machinery, were also absent from all the three-comparator strains of *Dickeya* species.

The composition and organization of the T3SS genes are largely conserved in SRE including *P. carotovorum* subsp. *brasiliensis*. With the exception of *P. atrosepticum* and *P. wasabiae* strains and *Dickeya* strains Dd586 and Dd703, all other strains of SRE used in the study encode HrpK. HrpK is a T3SS protein that aids in translocation of effectors across the plant cell wall in *P. syringae* (Glasner *et al*, 2008; Petnicki-Ocwieja *et al*., 2005). The absence of *hrpK* in *P. atrosepticum* initially suggested that HrpK might be involved in host specificity, especially as only *P. carotovorum* subsp. *brasiliensis* and *P. carotovorum* subsp. *carotovorum* strains have been shown to be able to elicit a HR response on tobacco (Glasner *et al*., 2008). However, in disagreement with this, mutation of the *hrpK* gene in *P. carotovorum* subsp. *carotovorum* WPP14 did not affect the ability of this strain to elicit HR. Also, the expression of the *hrpK* gene from *P. carotovorum* subsp. *carotovorum* WPP14 from a plasmid in *P. atrosepticum* did not confer HR elicitation to *P. atrosepticum* (Glasner *et al*., 2008), suggesting that the lack of HR elicitation by *P. atrosepticum* is not due to the absence of *hrpK* gene. All told, this gene in unlikely to play a role in host specificity.

In addition to the main secretory components of the T3SS, the genome of ICMP19477 also contained putative T3SS effector proteins, resembling the *srfABC* T3SS-associated gene cluster of *Salmonella* species. These effector proteins were encoded on a different GI, PbN1_GI36 in ICMP19477. Comparative analyses of the SRE genomes revealed that a similar gene cluster was

also present in all of the other strains used here. The *srfABC* operon is encoded on a GI, SPI1, in *Salmonella* species and contributes to virulence during the intenstinal phase of infection (Lostroh and Lee, 2001). This putative operon has been observed in other *Enterobactericiae* species, including *Enterobacter* and *Yersinia* species (Clara *et al*., 2010). Orthologous srfC, secreted through a T3SS was also detected in plant pathogen *P*. *syringae* (Petnicki-Ocwieja *et al*., 2002), suggesting that the *srf* gene operon may be involved in virulence of certain species. Hence, it is been hypothesized that the T3SS in the SRE may have an important role under certain conditions, and that these species use alternate and different virulence strategies to modify host defences during different stages of infection.

As *P*. *wasabiae* can still cause disease on potaoes without a T3SS apparatus, it seems that the T3SS is not essential for pathogenicity of SRE. Indeed, it was recently shown that there was no clear correlation between virulence and the presence of T3SS in *Pectobacterium* species (Kim *et al*., 2009). However, the T3SS has been suggested to be important during the early stages of infection (Holeva *et al*., 2004; Kim *et al*., 2011) and does contribute to virulence in some species (Holeva *et al*., 2004; Rantakari *et al*., 2001). For example, mutations in genes required for the secretion apparatus in *P*. *atrosepticum* 1039 led to a reduction in virulence on potato (Glasner *et al*., 2008; Holeva *et al*., 2004). Furthermore, the *hrpL* gene is required for the expression of effector protein DspE/F from *P*. *carotovorum* subsp. *carotovorum*, which is involved in evoking cell death of host plants (Hogan *et al*., 2013; Kim *et al*., 2011). The non-essential role of the T3SS in SRE is in contrast to the hemibiotrophic pathogens such as *P*. *syringae*, where the T3SS is of utmost importance to pathogenicity (Collmer *et al*., 2009).

Table 4.7 T3SS genes in the genomes of SRE and the accession numbers of their CDSs

T3SS genes	ICMP19477	ATCC BAA 417	ICMP5702	WPP14	SCRI1043	ICMP1526	UGC32	PC1	PCC21	Dda3937	Dd586	Dd703	SCC3193	WPP163
hrpN	A_13022	B_14563	C_1784	D_14754	ECA2103	E_04121	F_1320	G_2189	H_022670	I_04152	J_1909			
hrpW	A_13062	B_14583	C_1793	D_22506	ECA2112	E_04171	F_1328	G_2182	H_022630	I_02943	J_2002			
dspE	A_13067	B_14588	C_1794	D_19719	ECA2113	E_04176	F_1329	G_2181	H_022620	I_02945	J_2004			
dspF	A_13072	B_14593	C_1795	D_19724	ECA2114	E_04181	F_1330	G_2180	H_022610	I_02946	J_2005			
srfC	A_13652	B_00225	C_1882	D_03414	ECA2217	E_04641	F_1419	G_2095	H_021390	I_01645	J_2048	K_1880	L_2361	M_2383
srfB	A_13657	B_00230	C_1883	D_03419	ECA2218	E_04646	F_1420	G_2094	H_021380	I_01646	J_2049	K_1881	L_2360	M_2382
srfA	A_13662	B_00235	C_1884	D_03424	ECA2219	E_04651	F_1421	G_2093	H_021370	I_01647	J_2050	K_1882	L_2359	M_2381
hrpW **chaperone**	A_13057	B_14578	C_1792	D_22501	ECA2111	E_04166	F_1327	G_2183	H_022640	I_02942	J_2001			
hrpV	A_13017	B_14558	C_1783	D_14749	ECA2102	E_04116	F_1319	G_2190	H_022680					
hrcC	A_13007	B_14548	C_1781	D_14739	ECA2100	E_04106	F_1317	G_2192	H_022700	I_04208	J_1906			
hrpG	A_13002	B_14543	C_1780	D_14734	ECA2099	E_04101	F_1316	G_2193	H_022710	I_04209	J_1905			
hrpF	A_12997	B_14538	C_1779	D_14729	ECA2098	E_04096	F_1315	G_2194	H_022720	I_04210	J_1904			
hrpE	A_12992	B_14533	C_1778	D_14724	ECA2097	E_04091	F_1314	G_2195	H_022730	I_03340	J_1898			
hrpD	A_12987	B_14528	C_1777	D_14719	ECA2096	E_04086	F_1313	G_2196	H_022740	I_03341				
hrcJ	A_12982	B_14523	C_1776	D_14714	ECA2095	E_04081	F_1312	G_2197	H_022750	I_03342	J_1896			
hrpB	A_12977	B_14518	C_1775	D_14709	ECA2094	E_04076	F_1311	G_2198	H_022760	I_03343	J_1895			
hrp **pili**	A_12972	B_14513	C_1774	D_14704	ECA2093	E_04071	F_1310	G_2199	H_022770	I_03344	J_1894			
hrcV	A_12932	B_14473	C_1767	D_14664	ECA2085	E_04036	F_1303	G_2206	H_022840	I_00609	J_1888			
hrpQ	A_12927	B_14468	C_1766	D_14659	ECA2084	E_04031	F_1302	G_2207	H_022850	I_00610	J_1887			
hrcN	A_12922	B_14463	C_1765		ECA2083	E_04026	F_1301	G_2208	H_022860	I_00611	J_1886			
hrpP	A_12917	B_14453	C_1763	D_14639	ECA2081	E_04016	F_1299	G_2210	H_022890					
hrcQ	A_12912	B_14448	C_1762	D_14634	ECA2080	E_04011	F_1298	G_2211	H_022900					
hrpK	A_13117	B_19098	C_1810	D_02624			F_1343	G_2167	H_022460	I_02940				
hrpS	A_12962	B_14503	C_1772	D_14694	ECA2090	E_04061	F_1308	G_2201	H_022790	I_03345	J_1893			
hrpO	A_12897	B_14438	C_1760	D_14624	ECA2078	E_04001	F_1296	G_2213	H_022920	I_00616	J_1881			
hrpT	A_13012	B_14553	C_1782	D_14744	ECA2101	E_04111		G_2191	H_022690	I_04207	J_1907			

A - KCO; B_ - PcarbP_0102000; C - final; D_ - PcarcW_0102000; E - KCQ; F - UGC32; G - PC1; H - PCC21; I - Dda3937; J - Dd586; K - Dd703; L - W5S; M – Pecwa

4.4.5.2.3 T4SS

The T4SS is required for plasmid conjugation and is unique in its ability to transport nucleic acids, in addition to proteins, into host cells (Christie and Cascales, 2005). In 2004, Bell *et al*. identified a T4SS in the genome of SCRI1043, encoded on a GI called HAI7. This T4SS included a large contiguous operon constituting genes from *virB1* to *virB11*. Studies have revealed, however, that the conjugation machinery is involved in delivering virulence proteins into the host cells (Seubert *et al*., 2003). Mutations in the *virB4* gene in SCRI1043 also caused significant reductions in the lesion lengths on potato plants infected with this pathogen (Bell *et al*., 2004). Homologous *vir* genes were encoded on a GI, PbN1_GI26, in ICMP19477. Very similar genes were also present in the genomes of ATCC BAA 417, Dda3937 and Dd586, and in SCC3193 (Table 4.8). However, this secretion system was absent from the other SRE: ICMP5702, WPP14, PC1, UGC32, WPP163 and Dd703, suggesting that this operon has been acquired by HGT in other strains.

A second T4SS operon, previously undescribed in SRE, is present in ICMP19477. This is encoded on a putative GI, PbN1_GI44, and comprises a conjugal transfer protein cluster constituting genes from *trbB* to *trbL* (Table 4.8). The *trb* genes are involved in transfer of DNA between cells (Li *et al*., 1998). Genes of the trb system are related to those of several other bacterial conjugation or protein secretion systems (Li *et al*., 1998), including the VirB system of *A*. *tumefaciens* (Shirasu *et al*., 1990), the Ptl system of *Bordetella pertussis* (Farizo et al., 1996) and the Tra system of plasmid F (Frost *et al*., 1994). A conjugal transfer protein, traG was also present in this locus. The *trb* gene cluster was conserved in both *P*. *carotovorum* subsp. *brasiliensis* strains, but absent from all other comparator genomes used in this study and is proposed here to be specific to *P*. *carotovorum* subsp. *brasiliensis*. Given that a mutation in one of the core genes of the other T4SS cluster of SCRI1043 resulted in reduced blackleg symptoms in potato plants (Bell *et al*., 2004), the presence of an additional T4SS cluster in *P*. *carotovorum* subsp. *brasiliensis* suggests that this cluster may indeed contribute to the aggressive nature of this species in causing blackleg disease. However, similar genes (~72%) were also identified in the genome of Dda3937 suggesting this cluster has been acquired by HGT in *P*. *carotovorum* subsp. *brasiliensis* strains and *Dickeya* strain Dda3937.

PbN1_GI44 also encoded a putative VirD2 relaxase, in addition to a conjugal transfer coupling protein, traG. Relaxases are key enzymes in bacterial conjugative transfer systems, which together with accessory proteins, form a nucleoprotein complex, known as a relaxosome (Francia *et al*., 2004). Conjugal coupling proteins are involved in the coupling of the relaxosome with the mating pair formation system, resulting in the active mediation of DNA transport (Llosa *et al*., 2002, 2003). The conjugal and relaxase genes were probably involved in the acquisition of PbN1_GI44 by HGT. The benefit to *P. carotovorum* subsp. *brasiliensis* (if there was one) of acquiring the GI remains unknown.

Table 4.8 T4SS genes of SRE used in this study and the accession numbers of their CDSs

T4SS genes	ICMP19477	ATCC BAA 417	SCRI1043	ICMP1526	Dda3937	Dd586	SCC3193
T4SS cluster I							
Hypothetical	A_12262	B_07799					
Conjugal transfer protein		B_21760	ECA1612	C_19142	D_02744	E_1451	F_1615
virB11	A_16672	B_21870	ECA1613	C_19137	D_02745	E_1452	F_1616
virB10	A_16667	B_21220	ECA1614	C_19132	D_02746	E_1453	F_1617
virB9	A_16662	B_21225	ECA1615	C_19127	D_02747	E_1454	F_1618
virB8	A_16657	B_21230	ECA1616	C_19122	D_02748	E_1455	F_1619
virB7	A_16652	B_21235	ECA1616A		D_02749		F_1620
virB6	A_16647	B_21240	ECA1617	C_19117	D_02750	E_1456	F_1621
Putative plasmid protein	A_16642	B_21245		C_19112	D_02751	E_1457	F_1622
virB5	A_16637	B_21250	ECA1619	C_19107	D_02752	E_1458	F_1623
virB4	A_16632	B_21685	ECA1620	C_19102	D_02753	E_1459	F_1624
virB2	A_16627	B_21690	ECA1621	C_19097	D_02754	E_1460	F_1625
virB1	A_16622	B_21695	ECA1622	C_19092	D_02755	E_1461	F_1626
T4SS cluster II							
TrbL/virB10/	A_12257	B_07804			D_01746		
TrbG/virB9	A_12252	B_07809			D_01745		
TrbF	A_12247	B_07814			D_01744		
TrbB	A_17277	B_07849			D_01738		
TrbC/virB2	A_17282	B_07844			D_01739		
TrbD	A_17287	B_07839			D_01740		
TrbE/virB4	A_17292	B_07834			D_01741		
TrbJ	A_17297	B_07829			D_01742		
TrbJ signal peptide	A_17302	B_07824					
TrbL/virB10/	A_17307	B_07819			D_01743		
*virD2/*relaxase	A_17312	B_07909			D_01717		
traG	A_17382	B_07859			D_01736		

A - KCO; B_ - PcarbP_0102000; C - KCQ; D - Dda3937; E - Dd586; F- W5S

4.4.5.2.4 T5SS

The T5SS is the simplest of the secretion systems and includes an auto-transporter and a two-partner secretion system (hecAB) (Henderson *et al*., 2004). In SRE, this system is little explored. Comparative genomic analysis of ICMP19477 with other strains of SRE showed that homologous *hecA* (KCO-11235) and *hecB* (KCO_11240) genes of ICMP19477 were present in ATCC BAA 417, SCRI1043, ICMP1526, SCC3193 and WPP163 and absent from rest of the strains (Appendix A4.2). *HecB* gene was present in the genome of Dd586, however. Genomic analysis by Rojas *et al*. (2002) and Bell *et al*. (2004) implied that the genes encoded by the T5SS were acquired by HGT. Mutants in the *hecA* gene of *Dickeya* species impaired attachment, aggregation and subsequent killing of leaf epidermal cells of tobacco seedlings, suggesting an important role in pathogenicity (Rojas *et al*., 2002). However, other than this gene, T5SSs have not been found to contribute to virulence of *Pectobacterium* species. Hence, it is hypothesized that the *hec* genes in *P*. *carotovorum* subsp. *brasiliensis*, *P*. *atroseptium* and *P*. *wasabiae* (blackleg causing *Pectobacterium* species) provide a common mechanism of host attachment that promotes disease.

4.4.5.2.5 T6SS

The T6SS has been discovered in many pathogenic species, including *V. cholera* (Puktzki *et al*., 2006), *Burkholderia mallei* (Schell *et al*., 2007), *E*. *coli* (Dudley *et al*., 2006), and *P*. *aeruginosa* (Mougous *et al*., 2006). To date, Hcp and VgrG proteins have been identified as the major proteins secreted through this system (Pukatzki *et al*., 2006; Schell *et al*., 2007). This apparatus resembles a tailed bacteriophage assembly and is composed of a filamentous tube of Hcp promoters capped by VgrG proteins (Koskiniemi *et al*., 2013; Siverman *et al*., 2012). Hcp and VgrG have been associated with virulence in both plant and animal pathogens and are potential effector proteins delivered through T6SS (Pukatzki *et al*., 2006). Accordingly, the number and type of *hcp* and *vgr*G encoding genes varies among bacterial species and strains.

ICMP19477 harbours a T6SS cluster encoded on a putative GI, PbN1_GI54, which is also present in ATCC BAA 417 and all other strains of SRE used in this study. PbN1_GI61 is similar (~88%) to the T6SS locus of *P*. *atrosepticum* SCRI1043. In addition, ICMP19477 has 38 genes that code for Hcp or VgrG, which may be related to the function of the T6SS (Appendix A4.2). At least 44 genes encoding for Hcp proteins are present in the genome of ATCC BAA 417 (Appendix A4.2).

The function of the secreted proteins appears to be diverse in *Pectobacterium* species. In SRE, the T6SS is related to microbial survival *in planta* and may be directly related to virulence (Nyryki *et al*., 2012). For instance, a mutant deficient in hcp1 of *P*. *atrosepticum* was unaffected for virulence but overexpression of this protein resulted in a hyper-virulent phenotype, suggesting that the proposed T6SS in *P*. *atrosepticum* is important for virulence (Liu *et al*., 2008). Interestingly, mutations in essential T6SS components, *vas*E and *vip*B, of *P*. *atrosepticum* SCRI1043 resulted in significant reduction in virulence in potato tubers and stems (Liu *et al*., 2008). *VasE* is similar to *V*. *cholera* gene VCA0114, which is necessary for a functional T6SS and VipB is analogous to the VipB protein of *V*. *cholera* that is involved in T6SS-dependent secretion of *V*. *cholera* V52 (Bonemann *et al*., 2009; Zheng *et al*., 2011). However, it was later found that the phenotype of the *Pectobacterium* T6SS mutants was dependent on the virulence assay and the concentration of bacterial cells used in conducting experiments (Liu *et al*., 2008; Nykyri *et al*., 2012). In another study, addition of potato tuber and stem extracts to the *P*. *atrosepticum* bacterial cultures *in vivo* resulted in the induction of T6-secreted vgrG and hcp proteins (Mattinen *et al*., 2007). More recently it was shown that mutagenesis of both T6SS clusters in *P*. *wasabiae* resulted in delayed symptoms of disease development, indicating the importance of T6SS in pathogenicity of *Pectobacterium* (Nykyri *et al*., 2012). Though the T6SS does not secrete major virulence factors, it may be related to fitness *in planta* and its role in pathogenesis may be in conjuction with PCWDEs, the T3SS and other virulence determinants. The components of T6SS have not been studied so far in *P*. *carotovorum* subsp. *brasiliensis* species and are thought to be promising candidates in the future for characterization of effectors secreted by this system.

4.4.5.3 Iron acquisition

Genome sequencing of SCRI1043 has shown the presence of genes for the achromobactin uptake and transport system and enterobactin synthesis (Bell *et al*., 2004). Genes encoding both achromobactin and enterobactin were also identified in the genome of ICMP19477, ATCC BAA 417 and all other comparator strains used in this study.

In addition to achromobactin and enterobactin, ICMP19477 and ATCC BAA 417 carried a locus (present on a GI, PbNI_GI13) encoding a putative siderophore-independent system similar to ferric citrate uptake in *E*. *coli*. This locus had already been identified in the genome of SCRI1043 (Bell *et al*., 2004) (located in the same genomic region as in ICMP19477). It was subsequently detected in the genomes of ICMP1526, UGC32, PC1 and PCC21, but not in the sequence for the other strains.

Iron is a necessary cofactor for enzymes involved in important cellular functions. Thus, the capacity to obtain iron from the host is essential for survival of pathogenic microorganisms. Microbial siderophores are well characterized in human pathogens, acting as pathogenicity determinants in species such as *E*. *coli* and *Shigella flexneri*, where the clusters encoding the siderophores are thought to be on GIs acquired by HGT (Ishimaru & Looper, 1992; Ratledge & Dover, 2000). In plants iron is acquired from the environment by the roots using mechanisms based on reduction or chelation (Morrissey & Guerinot, 2009). Iron is then mobilized from the root tissues to the xylem vessels by citrate and is distributed throughout the plant via phloem by the non-proteogenic amino acid nicotianamine that is able to chelate iron and other metals (Curie *et al*., 2009). The storage reservoirs of iron in plants are chloroplasts, plastids and vacuoles (Lanquar *et al*., 2005). Bacteria associated with plants produce siderophores under iron-depleted conditions that bind the iron and actively take it up into the cell (Gulick, 2009). It has been hypothesized that even during wounding of a host plant, the bacteria at the colonization sites use some iron substrates released in damaged tissues for infection of the host (Expert & Brian, 2012).

In *Dickeya* and *E*. *amylovora*, the production of siderophores, chrysobactin and achromobactin, plays an important role in plant infection and full pathogenicity (Dellagi *et al*., 1998, 1999; Franza *et al*., 2005). In contrast, even though there is an apparent similarity in symptoms elicited by *Dickeya* and *Pectobacterium*, *P*. *carotovorum* does not require the production of siderophores for successful infection (Ishimaru & Loper, 1992). The role of siderophores in virulence of *P*. *carotovorum* subsp. *brasiliensis*, however, is yet to be explored.

4.4.5.4 Toxins

4.4.5.4.1 Production of toxins involved in competitive fitness or bacterium-bacterium interactions

SRE use various traits to recognize and outcompete other bacteria in the same niche. Many produce a range of toxins to eliminate competing bacterial species and/ or strains.

4.4.5.4.1.1 Phenazine

Phenazines are versatile secondary metabolites of bacterial origin that function in biological control of plant pathogens and contribute to the ecological fitness and pathogenicity of the producing strains (Mavrodi *et al*., 2010). These secondary metabolites exhibit broad-spectrum antibiotic activity against bacteria, fungi and eukaryotes (Fitzpatrick, 2009).

Genome sequencing of ICMP19477 identified a gene cluster (*phzA*-*phzG*) predicted to be involved in the production of phenazine, which was located on a GI, PbN1_GI31. The genes involved in phenazine biosynthesis in ICMP19477 showed similarity to those involved in phenazine antibiotic production in the closely related plant pathogen *Pantoea agglomerans*, which contributes to suppression of the fire blight disease-causing phytopathogen *E*. *amylovora* (Bell *et al*., 2004; Giddens *et al*., 2003).

To date, the phenazine biosynthesis cluster has only been identified in SCRI1043 (Bell *et al*., 2004). Comparative analyses of the SRE genomes here revealed that a similar gene cluster was also present in the genome of *P. atrosepticum* ICMP1526. This cluster was absent, however, from ATCC BAA 417 and the other comparator strains. Comparisons of the islands encoding the phenazine biosynthesis clusters in the three SRE genomes (ICMP19477, SCRI1043, ICMP1526) revealed high levels of similarity (~93%) in the overall organization of the core biosynthesis genes (*phzA*-phzG) (Figure 4.8). These islands contained four additional genes (KCO_15707, KCO_15712, KCO_15722, KCO_14727) adjacent to the phenazine gene cluster (one of them encoding a putative polyketide synthetase) that were also absent from the other genomes. A second *phzF* gene (KCO_18812) was nevertheless identified in all the genomes used in this study. The function of this gene is unknown.

The presence of the phenazine biosynthesis cluster in taxonomically distinct SRE suggests that it has been acquired horizontally in these strains, providing a competitive advantage to its recipient in soil or on its plant host.

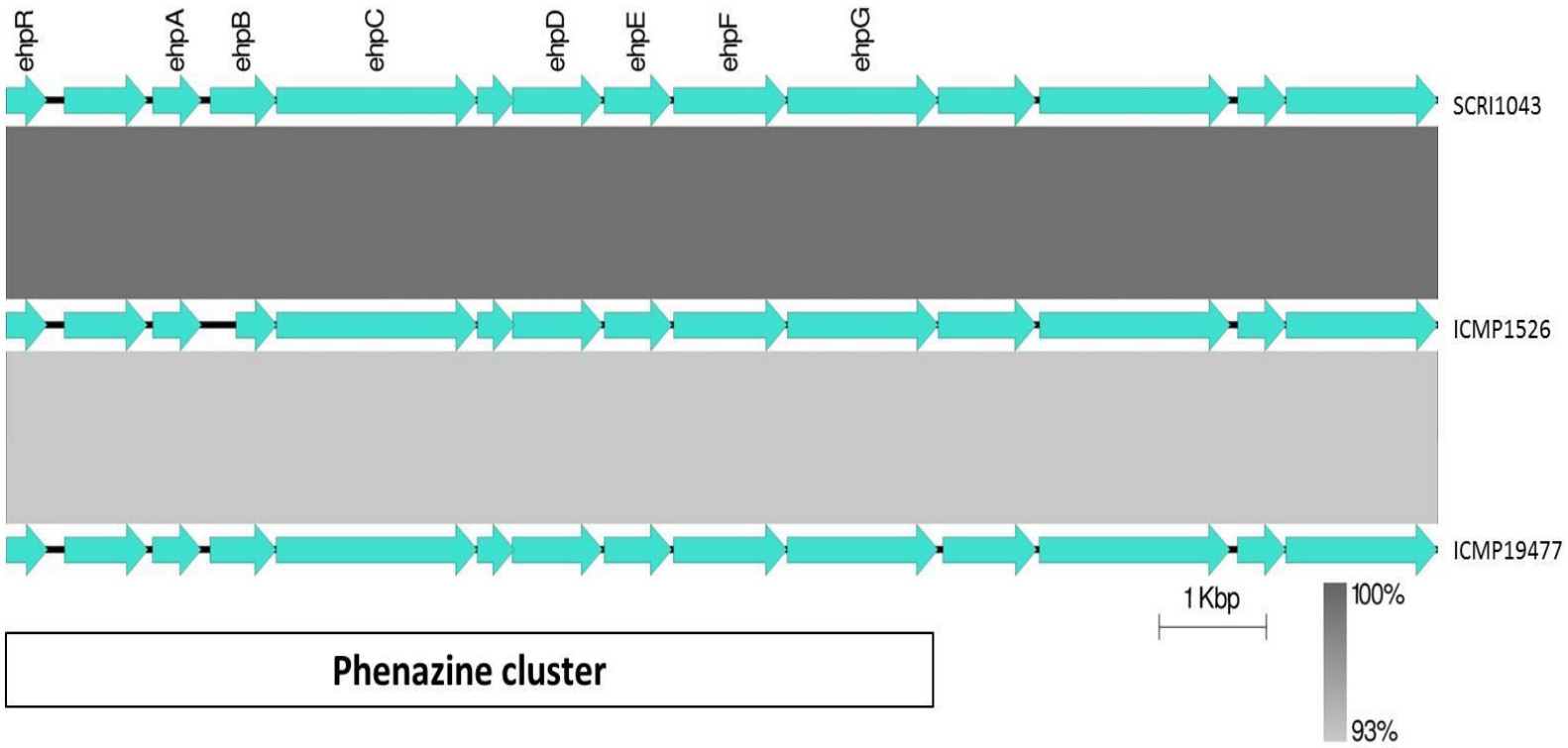


Figure 4.8 A schematic diagram showing a comparison of the phenazine cluster (ephR – ephG) from *Pectobacterium atrosepticum* SCRI1043, *P. atrosepticum* ICMP1526 and *P. carotovorum* subsp. *brasilienis* ICMP19477. The phenazine clusters from *P. atrosepticum* strains show ~93% similarity to that found in ICMP19477. The four extra genes present outside the phenazine cluster belonging to the same GI, are conserved in the three strains (SCRI1403, ICMP1526 and ICMP19477) but absent from the rest of the comparator genomes used in this study.

4.4.5.4.1.2 Carbapenem

A cluster of genes predicted to be responsible for the synthesis of carbapenem is present on a GI (PbN1_GI75) in ICMP19477. This cluster is composed of nine genes: a quorum-sensing transcriptional activator of carbapenem gene expression (*carR*), five enzymes required for the construction of the carbapenem molecule (carABCDE), two proteins responsible for a novel β-lactam resistance mechanism conferring carbapenem immunity in the producing host (carFG), and a protein of unknown function (carH) (Derzelle *et al*., 2010). The presence of a QS regulator is not suprising, as other SRE produce the carbapenem antibiotic 1-carbapen-2-em-carboxylic acid (Car) in response to the quorum sensing of bacterial cells (Barnard & Salmond, 2007). In *P*. *carotovorum*, *carR* responds to the primary molecule involved in quorum sensing, N - (3-oxohexanoyl)-L-homoserine lactone (OHHL), which is produced by the product of the unlinked *carI* gene. This OHHL-dependent transcriptional activation allows cells to coordinate the expression of carbapenem with cell density (McGowan *et al*., 1995). Mutations in *carR* or *carI* inhibit the production of carbapenem antibiotic and mutations in *carI* alone affect exoenzyme production (Bainton *et al*., 1992; Bodman *et al*., 2003).

Comparative genome analysis of SCRI1043, WPP14 and ATCC BAA 417 by Glasner *et al*. (2008) indicated that only ATCC BAA 417 encoded a carbapenem antibiotic-producing gene cluster. In this study, PbN1_GI75 (or an equivalent island) was identified in the genomes of ICMP19477, ATCC BAA 417 as well as ICMP5702 and absent from the rest of the comparator strains. The regions surrounding the carbapenem cluster were highly conserved in all the strains, however, suggesting that PbN1_GI75 and the related islands were acquired by HGT. This finding suggests that the presence of toxins in enterobacterial genomes is strain-dependent and not species-dependent. With this in mind, del Pillar Marquez-Villavicencio *et al*. (2011) suggested (not confirmed) that the production of carbapenem by *P*. *carotovorum* subsp. *brasiliensis* might provide the pathogen with a competitive advantage over other potato endophytes, secondary invaders and pathogenic SRE.

4.4.5.4.1.3 Bacteriocins

In addition to phenazine and carbapenem, ICMP19477 also produces potent narrow-spectrum protein antibiotics known as bacteriocins that may kill closely related strains and species of SRE (Table 4.9). Genes encoding the phage tail-like bacteriocin, carotovoricin (Er), were detected in all SRE. Er is produced in multiple forms (ErA and ErB) at high frequency by *P. carotovorum* strains (Nguyen *et al*., 2001; Tovkach & Mukvich, 2003). The expression and production of Er is expressed and co-regulated with the expression of the tissue-macerating enzyme pectate lyase (Nguyen *et al*., 2002; Zink *et al*., 1985). Since multiple strains of *P. carotovorum* are usually found in a field with diseased plants, and sometimes even in a single diseased plant (Yap *et al*., 2004), and also because of the strong correlation observed between pectinolytic and carotovoricinogenic abilities of *P. carotovorum* (Gorb & Tovkach, 1997), it has been hypothesized that this bacteriocin plays an important role in the ecology of these phytopathogens. However, the role of these toxins in the structure of *Pectobacterium* populations in fields or in the plant is unknown.

Recently, two novel colicin-like bacteriocins (pectocin M1 and pectocin M2) were identified in *P. carotovorum* subsp. *carotovorum* PC1 and ATCC BAA 417, respectively that target other *Pectobacterium* species (Grinter *et al*., 2012). Genes encoding pectocin were also present in the genomes of ICMP19477 (KCO_12557) and *P. carotovorum* subsp*. carotovorum* PCC21 but were absent from the genomes of the remaining strains (Table 4.9). Furthermore, a cluster of genes encoding a novel bacteriocin, colicin D, was identified in the genomes of ICMP19477 and ATCC BAA 417 (Table 4.9). The colicin D biosynthetic gene was adjacent to those encoding ferrodoxin (2Fe-2S) and an S-type pyocin-domain containing protein on a putative GI (PbN1_GI43) in the genome of ICMP19477.

To date, no apparent bacteriocins have been identified in the *P. atrosepticum* genome. In this study, however, a candidate colicin and its potential immunity gene were identified, which were specific to the *P. atrosepticum* genomes (SCRI1043 and ICMP1526) (Table 4.9).

Colicins and S-type pycoins have been intensively studied in Gram-negative bacteria. These bacteriocins consist of two proteins: one that is responsible for antimicrobial activity (the killing protein) and the other for immunity (the immunity protein). The killing proteins have functional domains that confer receptor-binding, translocation and DNAse (RNAse) activity (Michel-Brand and Baysse, 2002; Riley, 1998). The interaction of these bacteriocins with specific membrane receptors on target cells determines the spectrum of target cells, which is generally narrow. The host strain is protected from its own bacteriocin through interaction with the immunity protein that is co-produced with the bacteriocin (Roh *et al*., 2010). It has also been shown that some colicins such as carosin S1 are induced in the presence of glucose and lactose (Chuang *et al*., 2007). Various bacteriocins have been shown to enter target cells in a TonB-dependent manner (Cascales *et al*., 2007; Mora *et al*., 2008), although the exact mechanism is yet to be elucidated. Various tonB-dependent receptors for iron compounds and colicins have been identified in the *Pectobacterium* genomes in this study. For example, colicin V was found in all strains of SRE used in this study and tonB receptor for colicin was found only in some strains (Table 4.9). However, the role of these receptors is unclear.

The presence of different bacteriocins in a single strain such as ICMP19477 suggests that a combination of several different bacteriocins plays a role in exhibiting antimicrobial activity against other competing SRE. It has been proposed that bacteriocins may have a key role in bacterial population dynamics (Riley & Wertz, 2002). The search for new bacteriocins that are effective against plant pathogens that cause serious diseases in important crops will be helpful in developing new and environmentally friendly methods for disease control in the field (Roh *et al*., 2010).

Table 4.9 Bacteriocins of SRE identified in this comparative study and the accession numbers of their CDSs

Bacteriocin	ICMP19477	ATCC BAA 417	ICMP5702	WPP14	PC1	PCC21	ICMP1526	SCRI1043	UGC32	SCC3193	WPP163	Dda3937	Ech586	ECh703
Carocin S1 or S-type pyocin							G_05151	ECA1669		I_1594	J_1762			
Yersiniabactin	A_16907	B_18938	C_2522	D_08388	E_1436	F_014530			H_2027					
Pesticin				D_03289										
Bacteriocin immunity	A_22184	B_018439	C_739	D_22551	E_3222	F_032110	G_07485	ECA3378						
tonB-dependent receptor	A_18827	B_05047		D_02028	E_2751	F_028190				I_1403	J_1518			
Colicin V	A_18607	B_17545	C_1238	D_07993	E_2794	F_028640	G_01290	ECA3053	H_706	I_1354	J_1470	K_02100	L_2768	M_2571
Colicin							G_03580	KCQ_03580						
Colicin immunity							G_03585	ECA1997						
Bacteriocin immunity			C_3161			F_008260				I_2824	J_2843			
Colicin-like bacteriocins	A_12557	B_02847			E_2303	F_023640								
Colicin D	**A_12552**	**B_02842**												
Carocin D S-type pyocin	A_12547	B_02837				F_023870								

A - KCO; B_ - PcarbP_0102000; C - final; D_ - PcarcW_0102000; E - KCQ; F - UGC32; G - PC1; H - PCC21; I - Dda3937; J - Dd586; K - Dd703; L - W5S; M – Pecwa

4.4.5.4.2 Production of toxins involved in bacterium-host interactions

SRE are known to produce secondary metabolites including polyketide synthetases and NRPSs. These molecules play an important role in manipulating the host defence system, enabling survival and disease development. Several gene clusters were detected in ICMP19477 that are likely to be involved in phytotoxicity.

4.4.5.4.2.1 Presence of a *cfa* cluster in the genome of ICMP19477 absent from ATCC BAA 417

A putative secondary metabolite biosynthetic cluster predicted to synthesize CFA was identified for the first time in *P*. *carotovorum* subsp. *brasiliensis*. In ICMP19477, the *cfa* cluster is encoded on a GI, PbN1_GI15, which is approximately 100 kb in size, delineated by DRs (5′ AGTGGTGCCCGGACTCGGAATCGAACCAAGGACACGAGGATTTTCAATC 3′) and adjacent to a tRNA. Comparative analysis of this cluster with other comparator strains revealed that the structure and organization of the *cfa* biosynthetic cluster in ICMP19477 is highly conserved when compared to that in SCRI1043 (present on the GI, HAI2), ICMP1526 and UGC32 (Figure 4.9). Bi-directional comparative analyses showed that the remainder of the GI surrounding the *cfa* cluster in ICMP19477 was different to HAI2 of SCRI1043 and the islands in ICMP1526 and UGC32 (Figure 4.9). The distinct structure of PbN1_GI15 suggests that a discrete GI encoding CFA may have been acquired independently during evolution of ICMP19477. Interestingly, the *cfa* cluster was not identified in the genomes of ATCC BAA 417 nor the other SRE in the comparison.

CFA is a component of COR, a phytotoxin (type I polyketide) produced by several pathovars of *P*. *syringae* (Bender *et al*., 1999). COR is formed by the conjugation of Cfa, synthesized by the *cfa* gene cluster, to CMA, synthesized by *cma* genes. Conjugation is mediated by *cfl* (Bender *et al*., 1999). COR mimics the action of MeJA, a plant signaling molecule involved in stress responses associated with wounding and pathogen attack (Brooks *et al*., 2004). Genome analysis of SCRI1043 has shown the presence of a *cfa* operon, similar in structure and organization to that found in *P. syringae*, encoded by the GI, HAI2. SCRI1043 also has a *cfl* homologue but lacks *cma*

genes suggesting that a derivative of COR is synthesized by this strain (Bell *et al*., 2004). More recently, a highly similar *cfa* biosynthetic cluster was identified in a Gram-positive bacterium, *S. scabies* (Bignell *et al*., 2009). *Streptomyces scabies* infects potato tubers. The presence of similar islands in distantly related species suggests that this cluster is highly mobile and is important for interactions between bacteria and potato (Pickard *et al*., 2003).

Mutants in the *cfa*6 and *cfa*7 genes of SCRI1043, encoding the Cfa polyketide, showed reduced blackleg symptoms on potatoes, demonstrating that Cfa is an important virulence factor in *P*. *atrosepticum* (Bell *et al*., 2004). In a different study, mutations in the *cfa*-like biosynthetic cluster of *S*. *scabies* resulted in reduced potato root disease symptoms. However, the strain was still able to cause excessive stunting of roots and shoots and chlorosis, eventually leading to the death of the tobacco plant host (Bignell *et al*., 2009). These results are consistent with the suggestion that CFA and CFA-related compounds are important for host-pathogen interactions on potato. However, the presence of the *cfa* biosynthetic cluster in only specific strains of *P*. *carotovorum* subsp. *brasiliensis* suggests this cluster may not be essential for pathogenicity but does contribute to disease severity. Consequently, the role of Cfa in blackleg disease by ICMP19477 remains unclear.

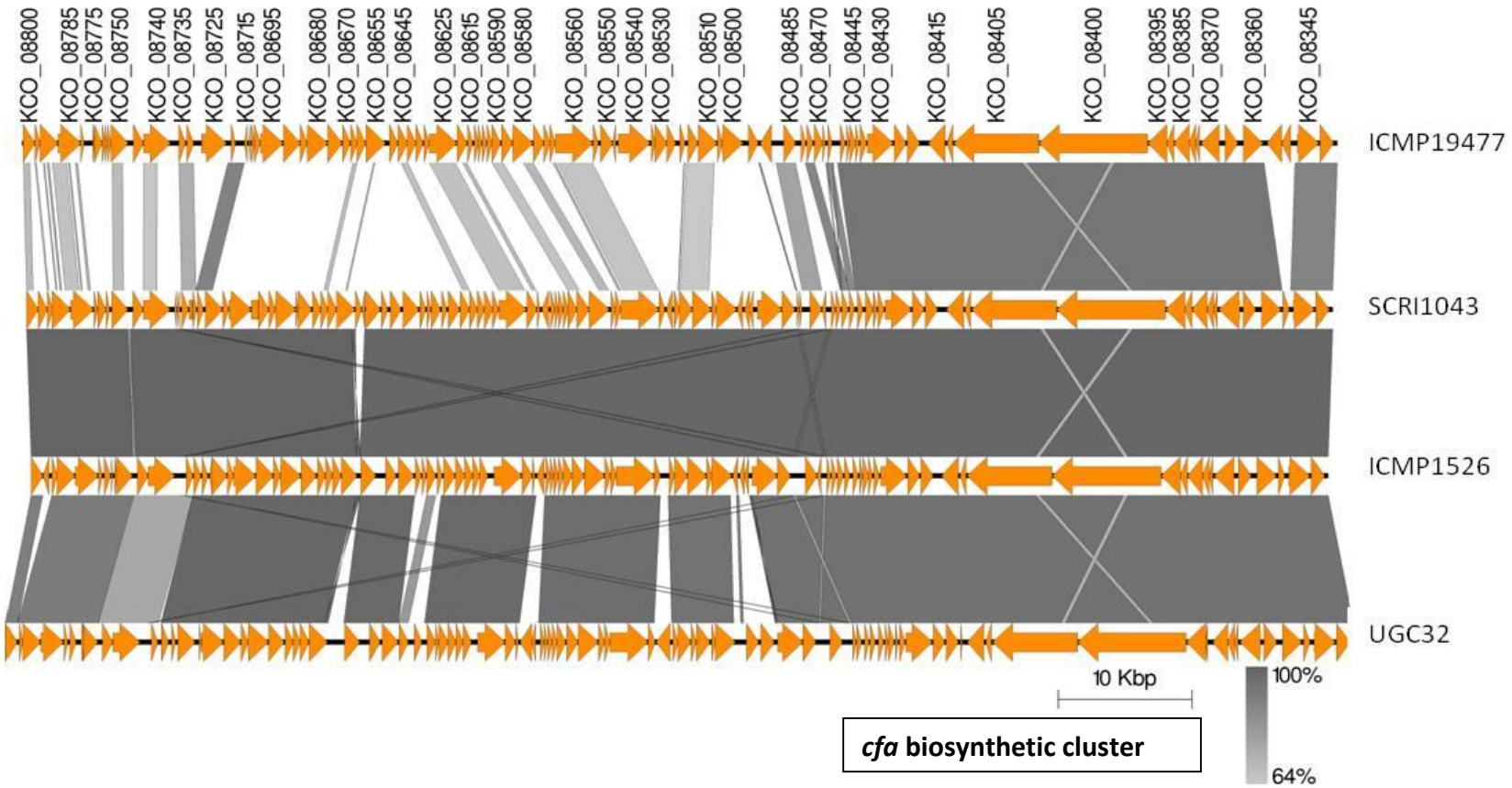


Figure 4.9 A schematic diagram showing PbN1_GI15 of ICMP19477 and a comparison of its structure with HAI2 of SCRI1043 and similar islands from ICMP1526 and UGC32. The *cfa* biosynthetic cluster is highly conserved in all four strains. However, the region surrounding the *cfa* cluster (delineated by direct repeats) is different in ICMP19477.

4.4.5.4.2.2 Presence of a NRPS cluster in ICMP19477 and its similarity to other *Pectobacterium* strains

A NRPS biosynthetic cluster was identified on a GI, PbN1_GI24 in ICMP19477 and also on similar islands in the other blackleg-causing strains ATCC BAA 417, SCRI1043, ICMP1526 and WPP14. The biosynthetic cluster contained two large CDSs (~21 Kb) (partial CDS NRPS1, KCO_17262 and full CDS NRPS2, KCO_06050) separated by a third smaller CDS (KCO_06045). This third CDS was subsequently found to belong to the first larger CDS (KCO_17262). This false annotation was due to the highly repetitive nature of the NRPS sequences. Unfortunately, since NRPS genes are generally associated with highly repetitive sequences, despite repeated attempts (gap filling), the gap between KCO_17262 and KCO_06045 was not resolved. A CDS encoding an ABC transporter, which is an accessory protein required for NRP synthesis, was located adjacent to the NRPS genes in the same cluster. The presence of an ABC transporter protein implies that the NRPS product might be secreted. A putative transcriptional regulator (KCO_06065) shared the same intergenic region with the NRPS gene cluster and was present between two flavin reductase genes (KCO_06070 and KCO_06060) transcribed in the opposite direction.

The *nrps* genes from ICMP19477 has similarity to the syringomycin (*syr*) and syringopeptin (*syp*) genes of *P*. *syringae* pv. *syringae*. The *nrps* genes have also been identified in SCRI1043 (encoded on GI HAI6), but their function in pathogenicity has not been studied. Nevertheless, these toxins appear to be major virulence determinants in *P*. *syringae* pv. *syringae* (Zhang *et al*., 1995), where a reduction in virulence was observed when immature cherry fruits were inoculated with *syr* and *syp* mutants (Xu & Gross, 1988). Syringomycin and syringopeptin are involved in inducing necrosis by forming transmembrane pores causing electrolyte leakage (Bender *et al*., 1999).

Comparative analysis showed similar NRPS sequences were present in the genomes of ATCC BAA 417 and WPP14 (as fragmented genes in the draft genomes). Interestingly, *syr* and *syp* homologues were not found in the genomes of PC1 or other SRE that cause soft rot as opposed to blackleg symptoms, as shown in a CG diagram of PbN1_GI24 from ICMP19477, HAI6 of

SCRI1043 and a similar region from the genome of PC1 (Figure 4.10) (other incomplete draft genomes were not included in this figure). The presence of the NRPS homologues in the blackleg causing strains of *P. atrosepticum* and *P*. *carotovorum* subsp. *brasiliensis* and in WPP14 (isolated from a stem lesion), and their absence from soft rotting SRE, suggested that the production of this peptide phytotoxin might have an important role in blackleg disease development. Such variability amongst strains in secondary metabolite production also suggests differences in how SRE interact with either their plant hosts or competing microbes, and calls for further analysis of the distribution of these systems amongst SRE.

Surprisingly, despite their unusually large size, these genes in SRE remain uncharacterized. In addition to the production of phytotoxins, however, NRPSs are also known to be involved in the production of antibiotics, siderophores or other secondary metabolites (Neilan *et al*., 2008). More recently, it has been hypothesized that the genes belonging to the NRPS/PKS hybrid clusters of *D*. *solani* strain Ds0432-1 may be related to high virulence and the competitive replacement of common *Pectobacterium* species with *D*. *solani* isolates in many European countries (Garlant *et al*., 2013).

The association of ABC transporters with the gene clusters of non-ribosomally synthesized peptides is not uncommon (Pearson *et al*., 2004). ABC transporters are known to influence many cellular processes including antibiotic resistance, nutrient acquisition, adhesion, protein secretion, environmental sensing, spore formation, conjugation, and growth under stress conditions (Sutcliffe & Russell, 1995). This has been well studied in Gram-positive bacteria, where putative transporters have been associated with several NRP toxins such as rapamycin (Schwecke *et al*., 1995), bleomycin (Calcutt *et al*., 1994) and many more (Pearson *et al*., 2004). Mutations in the ABC transporter gene of *Microcystis aeruginosa* PCC7806 resulted in complete abolition of the production of NRP toxin, microcystin, under laboratory growth conditions (Pearson *et al*., 2005). In another study, ABC transporters have been shown to play diverse roles in both fungicide resistance and pathogenesis of *Fusarium graminearum* (Ammar *et al*., 2013). To date, the role of the ABC transporter protein in association with the NRP toxins remains unknown in SRE.

Additional CDSs harboured on PbN1_GI24 included those that encode a putative zeta toxin family protein (KCO_06120) and a 4 Kb novel β-glucoside operon (consisting of 3 ORFS; KCO_06085, KCO_06090 and KCO_06095) for sugar utilization. From the pan genome analysis, it was observed that the gene encoding the zeta toxin family protein of ICMP19477 was also present in WPP163 and SCC3193, but absent from rest of the comparator strains including ATCC BAA 417. Zeta toxins are known to stabilize MGEs, which transmit resistance in many other pathogens. They are also thought to be a part of a post-segregational killing system that secures stable inheritance of low and medium copy number plasmids during cell division and induces plasmid-deprived offspring cells (Meinhart *et al*., 2003). The bacteriotoxic zeta toxin protein is inhibited by its cognate antitoxin epsilon, in its inactive state. Upon degradation of epsilon, the zeta toxin is released, which allows this enzyme to trigger autolysis of the bacterial cells. A putative hypothetical protein (KCO_06115) was found adjacent to the zeta toxin in ICMP19477. This protein was also present in the genomes of SCC3193 and WPP163 but absent from the rest. Whether this protein is a homologue of an antitoxin is currently unknown. Zeta toxins of *Streptococcus pyogenes* (PezT) are also encoded on HAIs that provide the host with virulence factors and resistance to different antibiotics (Brown *et al*., 2001; Khoo *et al*., 2007). Given that this toxin is associated with a GI in ICMP19477, it is hypothesized that this toxin is involved in stabilising this MGE.

The 4 Kb β-glucoside operon of ICMP19477 consists of a transcriptional anti-terminator *bgl*G (KCO_06095), 6-phosp-beta-glucosidase (KCO_06090) and phosphoenolpyruvate-dependent phosphotransferase (PTS) system β-glucoside-specific transporter subunit IIABC (KCO_06085). This operon was also present in the genomes of PCC21 and PC1. Though the same operon was not seen in other *Pectobacterium* genomes, similar genes were present at different locations in other *Pectobacterium* genomes. The PTS is important for carbohydrate acquisition in many bacteria (Postma *et al*., 1993) and consists of several proteins that transport PTS-dependent carbohydrates in both Gram-negative and Gram-positive bacteria (Cote *et al*., 2000). The components of PTS include: a non-specific enzyme I (EI), a sugar-specific enzyme II (EII) and, a heat-stable histidine protein (HPr) (Postma *et al*., 1993). The functional domains of EII protein

participate in the translocation of the specific carbohydrate, which is readily utilized by the bacterium. In *E*. *crysanthemi* and *P*. *carotovorum* subsp. *carotovorum*, β-glucosidases are involved in the degradation of the plant cell wall polysaccharide matrix and convert cellobiose and cellodextrin to glucose (An *et al*., 2004). Cellobiose and cellodextrin are produced as a result of the degradation of cellulose by extracellular endoglucanases and exoglucanases. Elsewhere, it has been suggested that the operons associated with β-glucoside utilization are repressed in the presence of a more readily metabolized carbon source, such as glucose (Rutberg, 1997). Such a mechanism, known as catabolite repression, is important for competition in natural environments and plays a crucial role in the expression of virulence genes that enable bacteria to access new sources of nutrients (Gorke and Stulke, 2008). The exact role of this operon in ICMP19477 is yet to be determined.

Though the *nrps* genes of *P*. *carotovorum* subsp. *brasiliensis* strains show similarity to those of *P*. *atrosepticum* strains, the region adjacent to these genes (delineated as an island in ICMP19477) is unique (Figure 4.10), suggesting the presence of a distinct GI, potentially acquired by HGT.

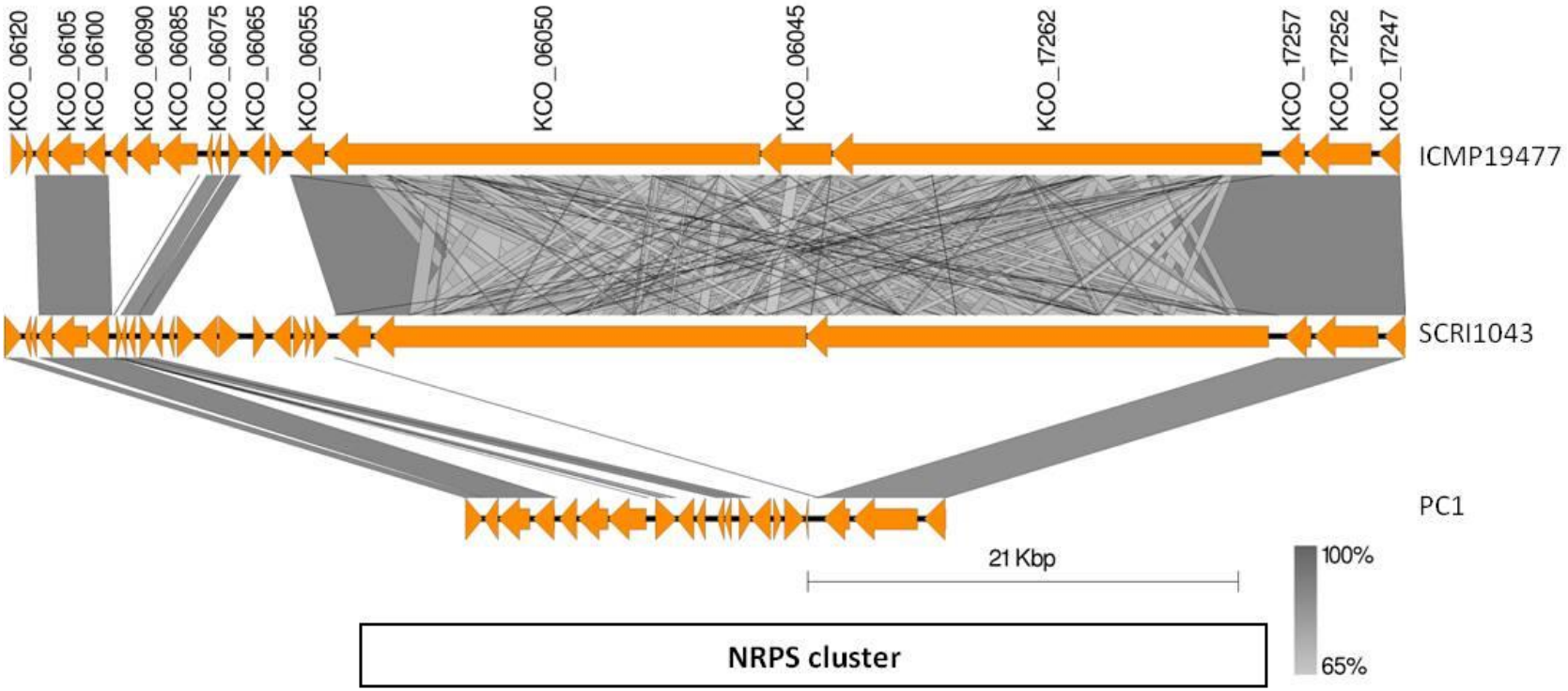


Figure 4.10 A schematic diagram showing a comparison of the genetic structure and organization of PbN1_GI24 from ICMP19477, HAI6 from SCRI1043 and a similar region from PC1. The NRPS cluster is highly conserved in both ICMP19477 and SCRI1043. Due to the repetitive sequences associated with the NRPS genes, the blast analysis is not clearly depicted in those genes. The region outside the NRPS gene cluster in ICMP19477 is different to that of SCRI1043, suggesting that the NRPS clusters are harboured on distinct islands in these SRE. Regions outside the islands are conserved in all three strains.

4.4.6 Identification of the presence and absence of other pathogenicity/gene islets

In addition to previously identified virulence determinants in *Pectobacterium*, comparative genome analysis led to the identification of other potential virulence factors encoded on small genomic regions or virulence loci (termed pathogenicity/gene islets) in ICMP19477. Gene islets have made similar contributions to the evolution of bacterial virulence as GIs, but do not fulfill the criteria of islands due to their small size (generally 5-7 Kb) and/or lack of genetic stability or mobility (Schmidt and Hensel, 2006). Examples of pathogenicity islets involved in HGT and/or virulence have been reported in various bacteria such as *Hemophilus*, *Pseudomonas* and *Salmonella* (Marcus *et al*., 2000; Morales *et al*., 2004; Smoot *et al*., 2002).

4.4.6.1 Sucrose isomerase (SIM) gene islet

A gene islet of ~5.7 Kb, consisting of 4 CDSs encoding for 6-phospho-alpha-glucosidase, PTS system alpha-glucoside-specific transporter subunit IIBC, a transcriptional regulator and sucrose isomerase, was identified only in the genomes of *P*. *carotovorum* subsp. *brasiliensis* and *P*. *atrosepticum* strains (Figure 4.11). The regions surrounding these four genes were conserved in all other *Pectobacterium* strains, suggesting an insertion or deletion event that could have occurred during the progression of speciation.

The potato shoot system uses photosynthesis to produce sugars and starch. Sugars, in plants, derived from photosynthesis, act as substrates for energy metabolism and the biosynthesis of complex carbohydrates, providing sink tissues with the necessary resources to grow and to develop. Sucrose and starch are the final photosynthates in leaves, and sucrose translocates to sink organs through the stem in plants (Blankenship, 2002). The reason why sucrose is transported rather than glucose is probably because sucrose is harder for bacteria to metabolize and so transporting sucrose reduces the risk of infection. In addition to being an important storage and transport sugar in plants, sucrose is an energy source for many phytopathogenic bacteria.

SIM catalyzes the isomerization of sucrose into isomaltulose, trehalulose and other glucose and fructose byproducts (Salvucci, 2003). These sucrose isomers cannot be metabolized by plant cells and are therefore, likely to be advantageous to the infecting pathogen (Bornke *et al*., 2001). The PTS is found only in bacteria, where it catalyzes the transport and phosphorylation of numerous monosaccharides, amino acids, sugars, polyoligosaccharides and other sugar derivatives. To carry out its catalytic function in sugar transport and phosphorylation, the PTS uses PEP as an energy source (Roseman, 1964). To date, at least three different PTS systems responsible for the uptake of glucose, fructose and sucrose in bacteria have been identified using classical genetic and biochemical analysis (Dominguez *et al*., 1998; Dominguez & Lindley, 1996; Malin & Bourd, 1991). Studies have also shown that the sugar:PTS system is a key factor in the pathogenicity of *Spiroplasma citri*, a plant-pathogenic mollicute (Andre *et al*., 2003). Futhermore, *sim* mutants of *E*. *amylovora* did not produce significant fire blight symptoms on apple seedlings, stating the importance of sucrose metabolism for colonization of host plants in that system (Bogs & Geider, 2000). It has been proposed that reduced virulence of *sim* mutants could be due to the low level of nutrients in xylem vessels, requiring access to sucrose and sorbitol for colonization of the host (Bogs & Geider, 2000).

SRE vary in the utilization of carbon sources (Yap *et al*., 2004), numerous indels involved in metabolite uptake and degradation accounting for these genetic differences (Glasner *et al*., 2008). To date, the role of sucrose isomerase in SRE is unknown. The presence of islets carrying the sucrose isomerase gene in *P*. *carotovorum* subsp. *brasiliensis* and *P*. *atrosepticum*, however, suggests that these species are likely to metabolize more and different sugars from plant cells than other SRE. Furthermore, since *P*. *carotovorum* subsp. *brasiliensis* and *P*. *atrosepticum* are best known for inducing blackleg, these strains may be better adapted to capture the limited sugars available in plant stems.

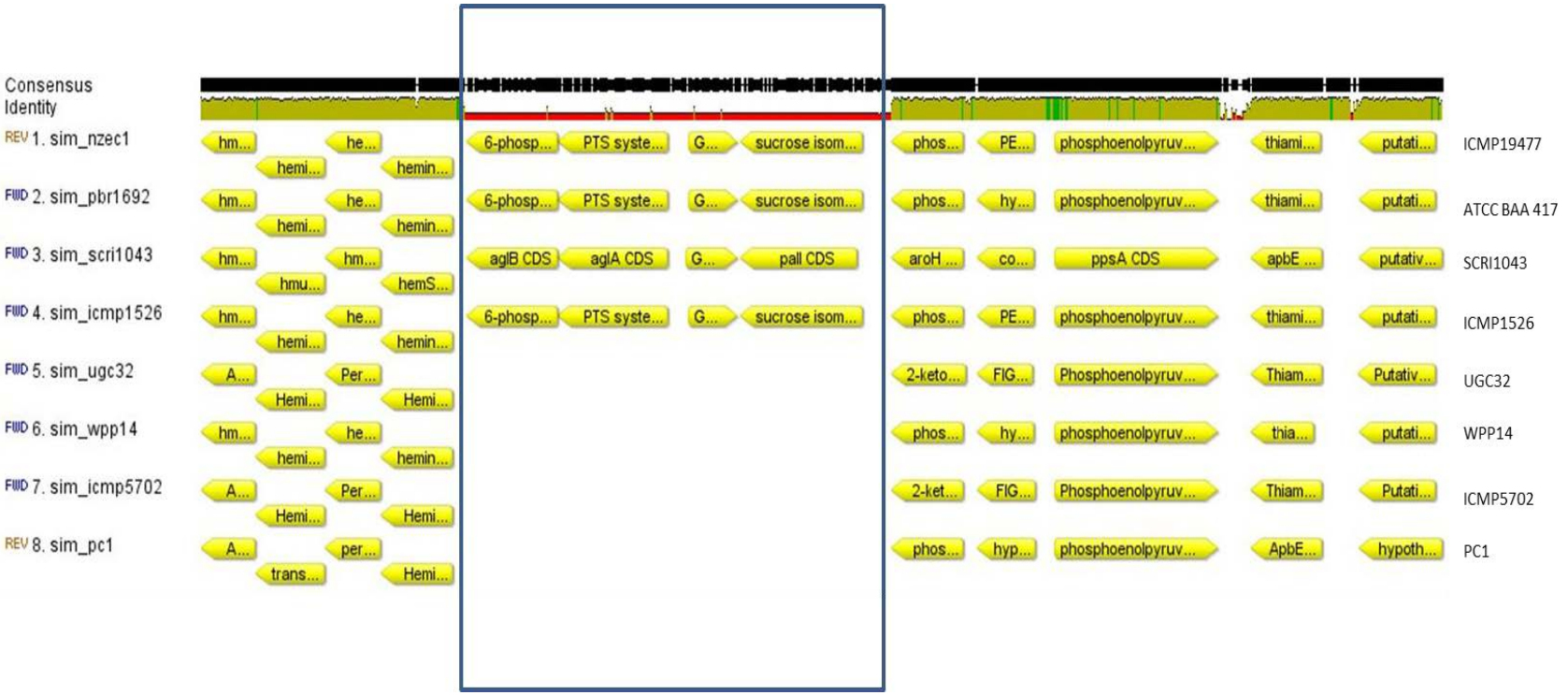


Figure 4.11 A schematic representation of the sucrose isomerase (SIM) gene islet in *Pectobacterium carotovorum* subsp. *brasiliensis* and *P*. *atrosepticum* strains. The four genes belonging to the SIM gene islet are absent from the remainder of the *Pectobacterium* strains used in this comparative analysis. The regions outside this islet are conserved in all strains. (nzec1 = ICMP19477)

4.4.6.2 Phenolic acid decarboxylase (PAD) gene islet

Comparative analysis led to the identification of the *pad* gene (KCO_03377) along with its transcriptional regulator (KCO_03372) in the genomes of ICMP19477 as well as those of *P. carotovorum* subsp. *brasiliensis*, *P. atrosepticum*, *Dickeya* and *P. wasabiae* strains. Both the genes were absent from the genomes of ICMP5702, WPP14 and PC1 (Figure 4.12).

Phenolic acids, important lignin-related aromatic acids, are the natural constituents of plant cell walls produced by plants to defend themselves against pathogens (Lynd *et al*., 2002). These acids (particularly ferulic, p-coumaric, and caffeic acids) bind the complex lignin polymer to the hemicellulose and cellulose in plants and have been suggested to be involved in cell wall extensibility, which plays a key role in cell defence (Bokern, 1991). Microorganisms living in a plant-soil ecosystem survive phenolic acid stress by the expression of a *padA* gene (Calvin *et al*., 1997). The *padA* is transcriptionally regulated by p-coumaric, ferulic, or caffeic acid; these three acids are the substrates of PAD.

In response to microbial attack, plants activate defence responses that lead to induction of a broad spectrum of antimicrobial compounds some of which may be species specific. These induced defence mechanisms are expressed at the site of attack (i.e., in the HR) as well as at a distance (signaled by methyl salicylate) to the site of primary infection to protect the plant from the spread of infection and future attack (van Loon, 2000). Plants defend themselves against bacterial pathogens by recognizing either the T3SS effectors or their actions and initiating a cascade of defence responses that often results in programmed cell death of the plant cell being attacked. It has been shown recently that a plant phenolic compound, p-coumaric acid (PCA), represses the expression of T3SS genes of the plant pathogen *D. dadantii*, suggesting that plants can also defend themselves against bacterial pathogens by manipulating the expression of the T3SS (Yan *et al*., 2009).

Some microorganisms have evolved ways of suppressing plant defence systems and other microorganisms are able to tolerate these plant defence chemicals by eliminating plant-derived toxins (Federici *et al*., 2004) or by detoxification of the phenolic acids (Asante *et al*., 2008). Decarboxylation by microorganisms is one of the important processes of detoxification and degradation of phenolic acids (Asante *et al*., 2008). Certain bacteria such as *Lactobacillus plantarum*, *Pediococcus pentosaceus* and *Bacillus subtilis* are resistant to the toxicity of phenolic acids (Nguyen *et al*., 2011). This resistance is due to the rapid induction of the *pad*A or *pad*C gene, which encodes a phenolic acid decarboxylase that can rapidly degrade these antimicrobial acids into less toxic vinyl derivatives. This resistance mechanism is termed the phenolic acid stress response (PASR) (Tran *et al*., 2008). The phenolic acid decarboxylase gene *pad*A is involved in the PASR in Gram-positive bacteria. The *pad*A gene (named padC in *B*. *subtilis*) encodes the PadA enzyme and *padR* encodes the PadR transcriptional repressor. Deletion of *pad*A leads to growth inhibition in the presence of phenolic acids, especially at low pH (Barthelmebs *et al*., 2000), while deletion of *pad*R leads to constitutive overexpression of padA (Gury *et al*., 2004) and, consequently, to high resistance to phenolic acids. Recently it was shown that in *E*. *crysanthemi*, efflux pump gene expression is upregulated in the presence of phenolic acid signaling molecules (Ravirala *et al*., 2006). Multidrug efflux proteins, which together with bacterial outer membrane proteins form a continuous channel across the periplasm and both membranes, are essential for bacterial resistance to antimicrobial chemicals and survival in plant tissue (Barabote *et al*., 2003).

The function of PAD in Gram-negative bacteria is unknown. The presence of PAD and its associated transcriptional regulator gene, as small gene islets in only a subset of SRE suggests, however, that these bacteria have evolved strategies to co-opt defence-related plant signaling molecules to enhance their resistance to toxic plant chemicals and consequently, to survive in a hostile host environment. It is hypothesized that PAD manipulates the plant defence system by degrading the phenolic compounds produced by the plants when attacked by pathogens.

A 705 bp CDS encoding for a protein involved in β-1, 3-glucan synthesis was also located in the PAD locus in the genomes of both *P*. *carotovorum* subsp. *brasiliensis* and in PC1, but was absent

from the remainder (Figure 4.12). Blastx analysis of this gene showed 48% similarity/identity to β-1, 3-glucan-synthesis protein of *P*. *syringae*. A variety of bacterial and fungal cell wall components including LPS, peptidoglycan, and β-1, 3-glucan are collectively referred to as pathogen-associated molecular patterns (PAMPs) (Janeway, 1989). Cyclic glucans are produced by a number of bacteria, including *X*. campestris pv. *campestris*, *R*. *solanacearum*, *A*. *tumefaciens*, *Bradyrhizobium japonicum*, *Rhizobium meliloti*, and *Brucella abortus* (Arellano-Reynoso *et al*., 2005; Bhagwat *et al*., 1993, 1999; Breedveld and Miller, 1994; Castro *et al*., 1996; Dunlap *et al*., 1996; Ingram-Smith & Miller, 1998; Talaga *et al*., 1996; Vojnov *et al*., 2001) and have been shown to be important for a number of symbiotic and pathogenic plant–microbe interactions (Breedveld & Miller, 1994; Mithofer *et al*., 2001). Bacterial cyclic β-1,2-glucan has been shown to be associated with systemic suppression of plant defences by the black rot pathogen *X*. *campestris* pv. *campestris* (Rigano *et al*., 2007). Systemic suppression is a novel counterdefencive strategy that may facilitate pathogen spread in plants and may have important implications for the understanding of plant–pathogen coevolution and for the development of phytoprotection measures (Rigano *et al*., 2007). Previous studies by Talaga *et al*. (1994) have demonstrated the role of periplasmic cyclic glucans involved in pathogenicity and elicitation of the HR by *P*. *syringae* pv. *syringae*, suggesting the importance of cyclic glucans in plant infection. The actual role role of cyclic glucans in *Pectobacterium* species is speculative and it would be interesting to know whether these species employ similar mechanisms (systemic suppression) to *X*. *campestris* pv. *campestris* for colonizing host plants.

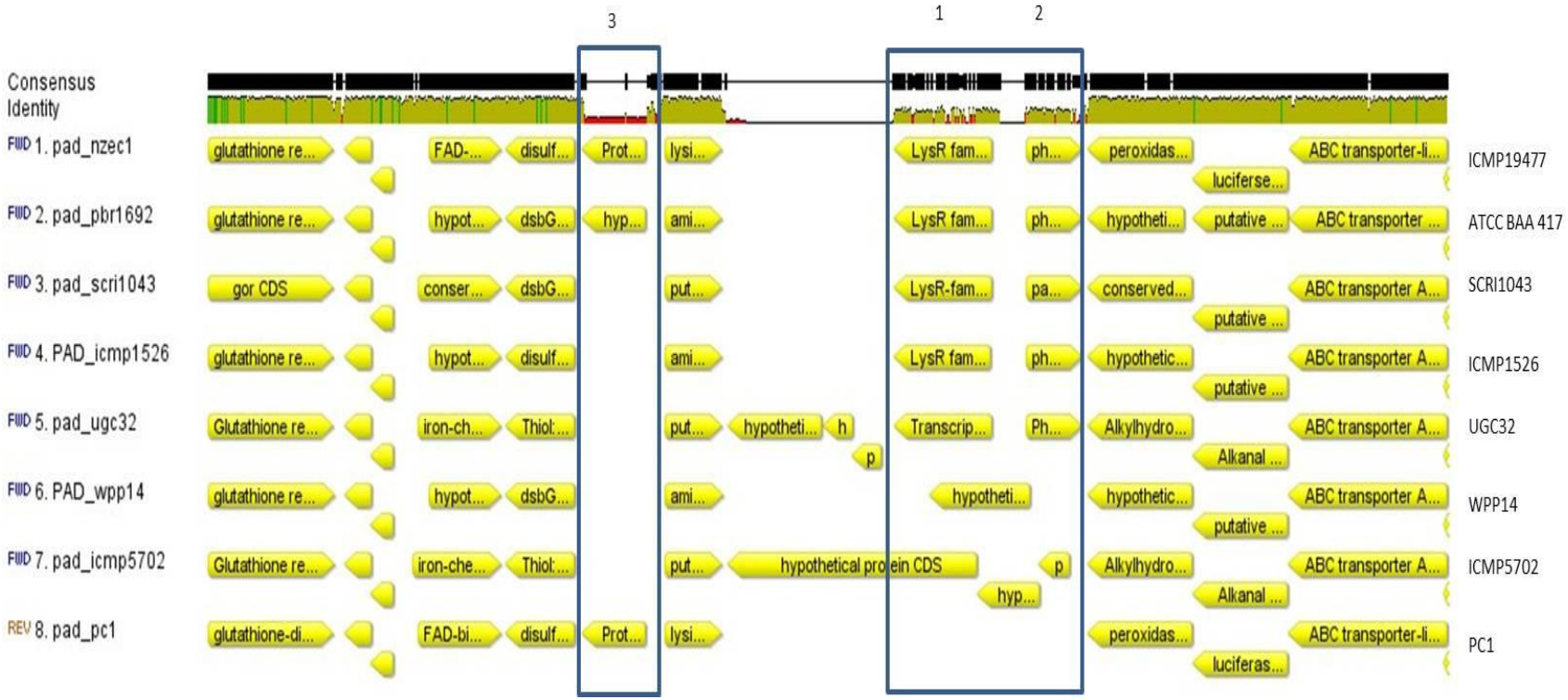


Figure 4.12 A schematic representation of the presence of the phenolic acid decarboxylase (PAD) gene islet in soft rot erwiniae (SRE). The two genes (2,1 - phenolic acid decarboxylase and adjacent transcriptional regulator, respectively) belonging to PAD gene islet are present in blackleg causing *Pectobacterium atrosepticum* and *P. carotovorum* subsp. *brasiliensis* strains and UGC32, but absent from the rest of the *Pectobacterium* strains (with the exception of *P. wasabiae* strains, data not shown) used in this comparative analysis. The regions outside this islet are conserved in all strains. The β-1,3-glucan (3) was present only in the ICMP19477, ATCC BAA 417 and PC1 strains, however. (nzec1 = ICMP19477)

4.4.7 Genes unique/specific to *P*. *carotovorum* subsp. *brasiliensis* strains

In addition to the gene islets described above, this study led to the identification of genes that were only present in the genomes of the *P*. *carotovorum* subsp. *brasiliensis* strains (Table 4.10). The majority of these *P*. *carotovorum* subsp. *brasiliensis*-specific genes were annotated as encoding hypothetical proteins. Some genes were annotated, however, as encoding proteins of known function. For example, a tripartite ATP independent periplasmic (TRAP) transporter was encoded by the *dct* locus (*dctM*, *dctQ*, *dctP*). A second gene cluster was present at this locus, which encoded two copies of the hydantoinase B/oxoprolinase gene and an ABC transporter.

TRAP transporters are known to be involved in the transport of C_4-dicarboxylates such as malate and succinate (Valentini *et al*., 2011). These transporters are encoded by the *dct* locus. The *dctP* gene encodes the peripasmic C_4-dicarboxylate-binding protein (Shaw *et al*., 1991) and is very distantly related to the citrate-binding protein of the citrate transport system in *Salmonella typhimurium* (Tam & Saier, Jr., 1993). The *dctQ* and *dctM* genes act as integral membrane proteins, essential for C_4-dicarboxylate transport activity (Forward *et al*., 1997). Homologous transport systems are widely distributed among Gram-negative bacteria and are present in *E*. *coli*, *S*. *typhimurium*, *B*. *pertussis*, *H*. *influenza* and many others (Forward *et al*., 1997).

Nutrient acquisition from host tissue is considered to be a critical process. In *Arabidopsis*, a highly efficient sugar transport protein is specifically synthesized upon wounding or elicitor treatment (Truerniet *et al*., 1996). Several genes also encoding for transport proteins were specifically upregulated in the course of infection of the host tissue by *D*. *dadantii* (Okinaka *et al*., 2002). Mutations in a different *dct* gene, *dctA1*, that encodes for a dicarboxylate transporter in *P*. *syringae* pv. tomato DC3000, has been shown to impair growth of these mutants *in planta*, suggesting the importance of transport and utilization of dicarboxylates in the virulence of DC3000 (Mellgren *et al*., 2009). Transporters enable the acquisition of organic nutrients and inorganic elements, efflux of toxin compounds, ion homeostasis, signal sensing, energy production, and various other important cellular functions (Paulsen *et al*.,

2004). It has also been recently shown that citrate uptake by a citrate carrier protein (Cit1) from *P*. *atrosepticum* is critical for full bacterial virulence (Urbany & Neuhaus, 2008). These data suggest that bacterial plant pathogens differ in the strategies they use to get necessary nutrients during infection.

Given the impact the *dct* cluster has in other systems, the presence of the *dct* genes in *P*. *carotovorum* subsp. *brasiliensis* strains should not be under estimated, as these genes might provide clues to the aggressive nature of this pathogen. Certainly, differences in nutritional requirements or access to nutrients could provide a competitive advantage against other competing strains inside the host.

Table 4.10 Genes unique to *P. carotovorum* subsp. *brasiliensis*

Gene Product	ICMP19477	ATCC BAA 417
Phosphate acetyltransferae	KCO_02947	PcarbP_010200013984
Hypothetical	KCO_05438	PcarbP_010200008905
Hypothetical	KCO_05593	PcarbP_010200009085
Hypothetical	KCO_06970	PcarbP_010200015668
Integrase/transposase	KCO_07905	PcarbP_010200019398
Hypothetical protein	KCO_07910	PcarbP_010200019393
Hypothetical protein	KCO_07915	PcarbP_010200019388
Putative aminotransferase	KCO_07920	PcarbP_010200019383
GCN5-like N-acetyltransferase	KCO_07925	PcarbP_010200019378
TRAP dicarboxylate transporter subunit DctM	KCO_09015	PcarbP_010200011749
TRAP dicarboxylate transporter subunit DctQ	KCO_09020	PcarbP_010200011744
TRAP dicarboxylate transporter subunit DctP	KCO_09025	PcarbP_010200011739
Hypothetical	KCO_09290	PcarbP_010200016253
Hypothetical	KCO_10105	PcarbP_010200006487
Hypothetical	KCO_11525	PcarbP_010200007464
Colicin D	KCO_12552	PcarbP_010200002842
Nucleotide excision repair endonuclease	KCO_12767	PcarbP_010200003057
Hypothetical	KCO_12837	PcarbP_010200014378
Hypothetical	KCO_12942	PcarbP_010200014483
Hypothetical	KCO_13077	PcarbP_010200014598
Putative DNA-binding transcriptional regulator	KCO_13337	PcarbP_010200000055
Hypothetical	KCO_13342	PcarbP_010200000060
Hypothetical	KCO_13347	PcarbP_010200000065
Hypothetical	KCO_13352	PcarbP_010200000070
Regulatory protein	KCO_13357	PcarbP_010200000075
Hypothetical	KCO_13527	PcarbP_010200000100
Amine oxidase	KCO_13642	PcarbP_010200000215
Glyoxalase/bleomycin resistanceprotein/dioxygenase	KCO_15997	PcarbP_010200020170
Hypothetical	KCO_16087	PcarbP_010200021120
Hypothetical	KCO_16267	PcarbP_010200008708
Hypothetical	KCO_17402	PcarbP_010200007949
Cobyrinic acid a,c-diamide synthase	KCO_17412	PcarbP_010200007934
M20 family peptidase PepV	KCO_17422	PcarbP_010200007929
Hypothetical	KCO_17427	PcarbP_010200007924
Conjugal transfer protein	KCO_17432	PcarbP_010200007919
M20 family peptidase PepV	KCO_19092	PcarbP_010200004777
Hydantoinase B/oxoprolinase	KCO_19097	PcarbP_010200004772
Hydantoinase B/oxoprolinase	KCO_19102	PcarbP_010200004767
ABC transporter-binding protein	KCO_19107	PcarbP_010200004762
ABC transporter-binding protein	KCO_19117	PcarbP_010200004752
Transport system inner membrane protein	KCO_19122	PcarbP_010200004747
Tyrosine-based site-specific recombinase	KCO_21222	PcarbP_010200008019
Nuclease	KCO_21282	PcarbP_010200007994
Lipoprotein	KCO_21302	PcarbP_010200007959
Hypothetical	KCO_21307	PcarbP_010200007954
Hypothetical	KCO_22119	PcarbP_010200018499
Putative hemolysin-like protein	KCO_22458	PcarbP_010200021920

4.5 Conclusions

The phylogenetic analysis carried out using the genome information collected for various SRE confirmed ICMP19477 as a strain of *P*. *carotovorum* subsp. *brasiliensis* (as described in Chapter 2). This strain had previously been identified as belonging to *P*. *carotovorum* subsp. *carotovorum* (Pitman *et al*., 2008). It is possible that other *Pectobacterium* isolates from potato have been mistakenly mis-classified and that *P*. *carotovorum* subsp. *brasiliensis* is more common than previously thought. Certainly, recent research has described the widespread distribution of this pathogen on potato in other countries (De Boer *et al*., 2012; Duarte *et al*., 2004; Onkendi *et al*., 2014; Van der Merwe *et al*., 2010).

Comparative genomic analysis using ICMP19477 and a number of other SRE also enabled the identification of genetic factors that might contribute to the pathogenicity or aggressiveness of *P*. *carotovorum* subsp. *brasiliensis* as well as the GIs and gene islets whose horizontal transfer may have lead to their acquisition. A total of 69 GIs and 10 gene islets were identified that comprised 38.2% of the total CDSs (1694 out of 4435 CDSs) and 37.4% of the assembled genome for ICMP19477. The size of the GIs ranged from 5,415 bps to 165,181 bps. There were 27 GIs larger than 10 Kb, eight GIs > 50 Kb and two GIs > 100 Kb. At least 17 GIs harboured known virulence genes. For example, PbN1_GI12, PbN1_GI22 and PbN1_GI23 encode *cfa*, *virB* and T6SS clusters, respectively. These gene clusters have already been shown to be involved in virulence of *P*. *atrosepticum*. Other islands or gene islets encoded clusters with no association with virulence of SRE. One such example was the gene islet carrying an auxin efflux carrier protein along with its transcriptional regulator. These genes were present only in the *P*. *carotovorum* subsp. *brasiliensis* and *P*. *carotovorum* subsp. *carotovorum* strains. It would be interesting to inactivate these genes and compare the epiphytic fitness of the resulting mutants to the wild type strains to determine whether these genes are responsible for the aggressive nature of these subspecies on tubers, when compared to *P*. *atrosepticum*.

Although many putative GIs and possible virulence factors were identified, a common set of genes that could be related to the lifestyle of a particular species of SRE or to specific disease symptoms in the host was not determined. For example, the *cfa* biosynthetic cluster involved in virulence of *P*. *atrosepticum* was found in ICMP19477, but was absent from the other *P*. *carotovorum* subsp. *brasiliensis* strain. Both cause blackleg. Another such example was the presence of an NRPS cluster in blackleg causing *Pectobacterium* strains (*P*. *atrosepticum* and *P*. *carotovorum* subsp. *brasiliensis*) isolated from potatoes. This cluster was also present in WPP14 but was absent from other *P*. *carotovorum* subsp. *carotovorum* strains isolated from potatoes such as ICM5702 as well as UGC32 and strains of *P*. *wasabiae*.

Additional genes involved in the environmental fitness and adaptation of the bacterium *in planta* were also identified such as the novel bacteriocins of *P*. *carotovorum* subsp. *brasiliensis*. *P*. *carotovorum* subsp. *brasiliensis* is known to have activity against *P*. *atrosepticum* (del Pilar Marquez-Villavicencio *et al*., 2011). The presence of novel bacteriocin, colicin D, in both the strains of *P*. *cartovorum* subsp. *brasiliensis* and its absence from the remaining SRE suggests that this bacteriocin may be involved in providing a competitive advantage to this subspecies. Further characterization of this toxin by construction of NRPS mutants, and their use in competition assays, would shed some light on the competitive nature of *P*. *carotovorum* subsp. *brasiliensis*.

In summary, comparative genomics provided a selection of GIs and gene islets that harbour genes that may be involved in the virulence or ecological fitness of *P*. *cartovorum* subsp. *brasiliensis* and/or other blackleg causing SRE. In the future, however, key questions that need to be addressed include: i) are the putative virulence genes identified in this study functionally active and what roles do they play in host pathogen interactions or fitness, ii) which *in vitro* and *in planta* conditions activate these genes and, iii) what are the events leading to the transfer of these genes?

4.6 Appendices

Appendices A4.1, A4.2, A4.3, A4.4, A4.5 and A4.6 are on the enclosed CD.

Chapter 5

Construction of knockout mutants of *P. carotovorum* subspecies *brasiliensis* ICMP19477

5.1 Abstract

Pectobacterium carotovorum subsp. *brasiliensis* ICMP19477 harbours various virulence genes encoded on putative GIs and small gene islets. Five genes of interest, *sim* (KCO_20372), *pad* (KCO_03377), *nrps1* (KCO_06050), *abc* (KCO_06055) and *cfa7* (KCO_08405) were selected for functional studies on the basis of their presence in blackleg causing *Pectobacterium* (*P*. *atrosepticum* and *P*. *carotovorum* subsp. *brasiliensis*) and evidence that they are involved in virulence of other bacterial plant pathogens. Non-functional copies of the five genes were cloned into either the suicide vector pKNG101 or the counter-selectable plasmid pK18mobsacB. Allelic exchange mutagenesis of *P. carotovorum* subsp. *brasiliensis* ICMP19477 was then performed to create mutants in which these genes were inactivated. Allelic exchange mutagenesis initially involved selection of 'single crossover' events, in which plasmid marker genes were incorporated along with the non-functional copy of the target gene. 'Double crossover' events were then induced to produce mutants that had lost the plasmid markers and were consequently sucrose resistant. The virulence of resulting mutants was compared to that of wild type *P. carotovorum* subsp. *brasiliensis* ICMP19477 in virulence assays. The assays were carried out on stems and tubers of potato plants. No effect on virulence was observed when the NRPS and CFA clusters were disrupted. Single crossover mutants in the islands encoding *sim* and *pad* significantly reduced both lesion length on stems of potato plants and maceration of potato tubers. This data suggested that the *sim* and *pad* loci might be important for virulence of *P. carotovorum* subsp. *brasiliensis* ICMP19477.

5.2 Introduction

In the previous chapter, comparative genome analysis led to the identification of candidate virulence factors in the genome of ICMP19477. Many of the candidate virulence factors identified were encoded on GIs and others were present on gene islets. To understand the roles of these genes in virulence of ICMP19477, attempts were made to generate knockout mutants of ICMP19477 in five genes of interest (*cfa7*, *nrps1*, *abc*, *sim* and *pad*). The *cfa7*, *nrps1* and *abc* genes are encoded on GIs, PbN1_GI15, PbN1_GI24 and PbN1_GI24, respectively. The *sim* and *pad* genes are encoded on gene islets. Allelic exchange was used to create knockout mutants for ICMP19477. The virulence of these mutants was then compared to the virulence of wild type ICMP19477.

Reverse genetic analysis through allelic exchange has been extensively used to introduce recombinant or mutated alleles into genomes of both Gram-negative and Gram-positive prokaryotes to decipher the unknown function of genes (Prentki & Krisch, 1984). In general, gene inactivation by allelic exchange mutagenesis is carried out by conjugation. Conjugation is usually performed by triparental matings using a strain transformed with a mutagenic plasmid (carrying the inactivated gene construct) as a donor, a helper strain containing a plasmid with transfer (*tra*) genes, and the recipient strain.

Suicide vectors carrying the inactivated gene construct are generally desirable for this type of gene inactivation process. However, the suicide vector must have several criteria: it must i) be conditional for replication to allow selection for integration into the chromosome, ii) carry a selectable marker (i.e. a counter-selectable gene) for subsequent selection and, iii) be transferable to a variety of other bacteria for efficient gene inactivation (Zhou *et al*., 2002). The most commonly used counter-selectable markers include genes that confer sucrose (*sacB*), streptomycin (*rpsL*), or fusaric acid (*tetAR*) sensitivity (Dean, 1981; Gay *et al*., 1985; Maloy & Nunn, 1981).

The first step in allelic exchange mutagenesis involves the integration of a plasmid carrying an inactivated gene construct within the reciprocal target sequence by homologous recombination, producing a chromosomal duplication. Following integration, the integrated plasmid can be excised via a second crossover event, resulting in allelic exchange (Appendix A5.1). Allelic exchange ultimately leaves one copy of the gene on the chromosome, either the wild type copy or the mutant copy. If the plasmid is still integrated in the chromosome, allelic exchange is initiated with the use of the counter-selectable markers as the cell will die in the presence of the counter-selective compound (Reyrat *et al*., 1998).

5.3 Methods

5.3.1 Bacterial strains

Pectobacterium and *E*. *coli* strains were cultured in LB medium at 28°C for 24 h and at 37°C for 16 h, respectively. When appropriate, cultures were grown with antibiotics at the following concentrations: 50 µg/mL kanamycin (Kn), 170 µg/mL chloramphenicol (Chl), 100 µg/mL ampicillin (Amp) and 50 µg/mL streptomycin (Str). Bacterial cells were harvested by centrifugation at 7,500 rpm for 10 min, the supernatant discarded and the pellet used for subsequent DNA preparation. Alternatively, for long term storage, equal volumes of an overnight culture were mixed with 40% glycerol and stored at -80°C. The bacterial strains and plasmid vectors used in this study are listed in Table 5.1.

Table 5.1 Bacterial strains and plasmid vectors used in this study

Bacterial strain/plasmid	Description/Genotype	Source/Reference	Antibiotic resistance
P. carotovorum* subsp. *brasiliensis			
ICMP19477	Wild type	Pitman *et al*., 2008	
ICMP19477Δ*nrps1*[2]	ICMP19477 carrying a double crossover mutation in *nrps1*	This study	Chl
ICMP19477Δ*sim*[1]	ICMP19477 carrying a single crossover mutation in *sim*	This study	Chl
ICMP19477Δ*pad*[1]	ICMP19477 carrying a single crossover mutation in *pad*	This study	Chl
ICMP19477Δ*cfa7*[1]	ICMP19477 carrying a single crossover mutation in *cfa7*	This study	Chl
ICMP19477Δ*nrps1*[1]	ICMP19477 carrying a single crossover mutation in *nrps1*	This study	Chl
ICMP19477Δ*abc*[1]	ICMP19477 carrying a single crossover mutation in *abc*	This study	Chl
E. coli			
DH5α	F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK- mK+), λ–	Invitrogen	
S17-λpir	TpR SmR recA, thi, pro, hsdR-M+RP4: 2-Tc:Mu: Km Tn7 λpir	Simon *et al*., 1983	Str
CC118-λpir	Δ(ara-leu) araD ΔlacX74 galE galK phoA20 thi-1 rpsE rpoB argE (Am) recA1 λpir	Herroro *et al*., 1990	
HH26 pNJ5000	Mobilizing strain for conjugal transfer	Grinter, 1983	Tet
Plasmids			
pACYC184	Plasmid carrying the chloramphenicol cassette	Rose, 1988	Chl, Tet
pK18mobsacB	Vector used for delivery of knockout gene constructs	Schafer *et al*., 1994	Kn
pKNG101	Vector used for delivery of knockout gene constructs	Kaniga *et al*., 1991	Str
pK18mobsacB-*sim*::*chl*	pK18mobsacB carrying the *sim* knockout gene construct containing the chloramphenicol cassette	This study	Kn, Chl
pK18mobsacB-*pad*::*chl*	pK18mobsacB carrying the *pad* knockout gene construct containing the chloramphenicol cassette	This study	Kn, Chl
pK18mobsacB-*cfa7*::*chl*	pK18mobsacB carrying the *cfa7* knockout gene construct containing the chloramphenicol cassette	This study	Kn, Chl
pK18mobsacB-*nrps1*::*chl*	pK18mobsacB carrying the *nrps1* knockout gene construct containing the chloramphenicol cassette	This study	Kn, Chl
pK18mobsacB-*abc*::*chl*	pK18mobsacB carrying the *abc* knockout gene construct containing the chloramphenicol cassette	This study	Kn, Chl
pKNG101-*sim*::*chl*	pKNG101 carrying the *sim* knockout gene construct containing the chloramphenicol cassette	This study	Str, Chl
pKNG101-*pad*::*chl*	pKNG101 carrying the *pad* knockout gene construct containing the chloramphenicol cassette	This study	Str
pKNG101-*cfa7*::*chl*	pKNG101 carrying the *cfa7* knockout gene construct containing the chloramphenicol cassette	This study	Str
pKNG101-*nrps1*::*chl*	pKNG101 carrying the *nrps1* knockout gene construct containing the chloramphenicol cassette	This study	Str
pKNG101-*abc*::*chl*	pKNG101 carrying the *abc* knockout gene construct containing the chloramphenicol cassette	This study	Str

The antibiotic resistances are represented as follows: Kn, kanamycin; Chl, chloramphenicol; Tet, tetracycline; Str, streptomycin.

5.3.2 Standard DNA manipulations

5.3.2.1 Preparation of plasmid DNA

Overnight bacterial cultures were grown in LB medium supplemented with appropriate antibiotics in a shaker at 37°C. Bacterial cells were harvested by centrifugation at 7,500 rpm for 10 min. A QIAprep Spin Miniprep kit (Qiagen, Netherlands) was used for the preparation of plasmid DNA following the manufacturer's instructions. Briefly, the supernatant was discarded and the pelleted bacterial cells were resuspended in 250 µL of buffer P1 containing RNAase in a microcentrifuge tube. 250 µL of buffer P2 was added and mixed gently by inverting the tube 4-6 times. The lysate was neutralized with 350 µL of buffer N3 and the cell debris and chromosomal DNA were removed by centrifugation for 10 min at 13,000 rpm. The supernatant was transferred to a spin column, which was centrifuged for another min and the flow through was discarded. DNA bound to the column was washed with 500 µL of buffer PB and centrifuged as above. The flow-through was discarded and the spin column washed again with 750 µL of buffer PE and centrifuged for 1 min. The spin column was then centrifuged for an additional min to ensure any residual wash buffer was removed before final elution of plasmid DNA by addition of 50 µL of elution buffer and centrifugation as above. The eluted DNA was quantified as described in section 5.3.2.5, and then stored at -20°C prior to further use.

5.3.2.2 Isolation of genomic DNA

A DNeasy blood and tissue kit (Qiagen, Netherlands) was used for the isolation of chromosomal DNA of bacterial cells. The isolation procedure was carried out as recommended by the kit's manufacturer. LB broth containing overnight bacterial cultures (1-1.5 mL) was centrifuged at 7,500 rpm for 10 min and the supernatant discarded. The pellet was resuspended in 180 µL of resuspension solution (ATL buffer) and 20 µL of Proteinase K. The suspension was then mixed thoroughly by vortexing and the cells lysed by incubating at 56°C for 1 h. 20 µL of 10 mg/mL RNase A was added to the suspension mixture and incubated at room temperature for 30 min to avoid

contamination with RNA. An equal volume of ethanol-AL buffer (200 µL each) was added and mixed thoroughly by gentle vortexing. The clear lysate was transferred to a DNeasy mini spin column and centrifuged at 8,000 rpm for 1 min to bind DNA to the column, followed by washing of the column with 500 µL of buffer AW1 and 500 µL of AW2 buffer. For the elution of bound DNA, 200 µL of elution buffer was added to the spin column and the column centrifuged at 8,000 rpm for 1 min. The supernatant containing the genomic DNA was then stored at -20°C for further use. The eluted DNA was quantified as described in section 5.3.2.5.

5.3.2.3 DNA purification of PCR products and restriction digests

PCR amplicon and restriction DNA fragments were purified using the MiniElute PCR Purification Kit (Qiagen, Netherlands) following the manufacturer's instructions. Briefly, the PCR product or restriction digest was mixed with five volumes of PBI buffer and transferred onto a spin column. The column was then centrifuged at 13,000 rpm for 1 min and material bound to the column washed with 750 µL of PE buffer. The column was centrifuged again and the flow through discarded. The empty column was spun for an additional minute to remove residual ethanol from the column. The column was then placed in a fresh 1.5 mL centrifuge tube and 10 µL of elution buffer was added to the center of the membrane avoiding any contact with the membrane. The column was allowed to stand for 2 min at room temperature and then centrifuged at maximum speed for 1 min to elute the DNA. The eluted DNA was quantified as described in Section 5.3.2.5.

5.3.2.4 DNA purification of gel extracted products

Purification of gel extracted products was carried out using a MinElute Gel Extraction Kit (Qiagen, Netherlands) as described by the manufacturer. For gel extraction, DNA fragments were electrophoresed on a low-melt agarose gel (0.8 to 1.2%). The band was visualized using a UV-illuminator (GelDoc, Bio-Rad, USA) and the gel was excised using a sterile scapel. The sliced gel was then added to the binding buffer and

incubated at 65°C until the agarose melted, followed by washing using the final wash buffer. The purified DNA was eluted using sterile water.

5.3.2.5 Determination of DNA concentration

The concentration of genomic DNA, plasmid DNA and the purified DNA products were measured spectrophotometrically using a NanoDrop Spectrophotometer (Thermo Fisher, USA) with a light absorbance of 260 nm wavelength.

5.3.3 Cloning

5.3.3.1 Restriction digestion

Restriction digestions were carried out to clone genes of interest into a suitable vector. Digests of DNA using restriction endonuclease enzymes were carried out in suitable buffers according to the manufacturer's recommendations (New England Biolabs; Roche). A restriction digestion was generally performed in a single reaction tube: 10 µL (approximately 1 µg) of DNA was mixed with 1 µL (2-3 units/µg DNA) of appropriate restriction enzyme and 2 µL of the reaction buffer (variable according to the enzyme used in the reaction). Where appropriate, double digestions using 1 µL each of two different compatible restriction enzymes was also used in a single reaction tube. The final volume was made up to 20 µL by adding double distilled water. The tube containing the reaction mixture was mixed gently and incubated at an appropriate temperature (generally 37°C) for 1 h followed by incubation at 85°C for 10 min for the deactivation of the enzyme. The digest was then purified as described above and stored at -20°C for further use.

5.3.3.2 Plasmid dephosphorylation

To prevent recircularisation of linearized plasmids, these were dephosphorylated using alkaline phosphatase (NEB, Germany). For this purpose, 10 units of CIP (Alkaline Phosphatase, Calf Intenstinal, NEB, Germany) were added to the restriction digestion

sample after heat inactivation of the restriction digests. Subsequently, the resulting mixture was incubated for 1 h at 37°C before it was inactivated at 85°C for 10 min.

5.3.3.3 Ligation

DNA ligation reactions were carried out to insert the digested DNA fragments into plasmids using T4 DNA Ligase (Invitrogen, California). The procedure was carried out as per the manufacturer's instructions Specifically, reaction mixtures consisted of 4 ul 5X ligase reaction buffer, insert:vector molar ratio's of 3:1, either 0.01-0.1 µg DNA plus 0.1 unit T4 DNA ligase (cohesive end) or 0.1-1.0 ug DNA plus 1.0 unit T4 DNA ligase (blunt end), to a total volume of 20 µL with autoclaved distilled water .

The amount or concentration of insert and vector DNA to be used in a ligation reaction was calculated as:

(ng of vector x size of insert in Kb x 3) / size of vector in Kb = ng of insert

The reactions were incubated at room temperature (depending on the enzyme used) for 5 min for cohesive ends or at 14°C for 16-24 h for blunt ends. The ligation mix was then ready to use for transformation.

5.3.3.4 Transformation of bacterial cells

Transformation of *E. coli* or *Pectobacterium* was carried out either by heat-shock transformation or electroporation of competent cells. Competent cells were constructed in the lab or were purchased commercially.

5.3.3.4.1 Preparation of chemically competent *E. coli* cells

Cells of *E. coli* were grown overnight in 5 mL of LB medium. After 16 h, a 50 µL aliquot of the overnight grown culture was added to 5 mL of LB broth. The culture was then grown to an OD_{600} 0.4–0.6 at 37°C (for approximately 2 h) followed by centrifugation

at 4,500 rpm for 5 min at 4°C. The pellet was resuspended in 700 µL 50 mM ice cold $CaCl_2$ and was incubated on ice for 30 min. The final steps involved, centrifugation of the cells at 4,500 rpm for 1-2 min, removal of the supernatant and subsequent resuspension of the cells in 100 µL of 50 mM $CaCl_2$.

5.3.3.4.2 Heat-shock transformation

1-5 µL of DNA (plasmid DNA or ligation reaction) was added to 50 µL of chemically competent *E. coli* cells (on ice) and mixed gently. The cells were incubated on ice for 30 min followed by heat-shock treatment for 30 s at 42°C (in a water bath without shaking), and immediate transfer to ice for 1-2 min. 450 µL of LB was added and the contents were incubated at 37°C with shaking (250 rpm) for 1 h. 50-200 µL of each transformation reaction was then spread on a pre-warmed plate containing the appropriate antibiotics and was incubated overnight at 37°C for the selection of transformants.

5.3.3.4.3 Preparation of electrocompetent *Pectobacterium* cells

A single colony of *Pectobacterium* strain was inoculated in 10 mL of LB broth and incubated overnight at 28°C with moderate shaking at 250 rpm. A small volume of the overnight culture was added to 1 L of LB media and the culture grown at 37°C with shaking to an OD_{600} of 0.5 to 0.9 (log phase growth). The cells were chilled in an ice water bath for 10 to 15 min followed by splitting the 1 L culture into 250 mL aliquots, each into 300 mL pre-chilled bottles. The cells were then centrifuged at 4,000 rpm for 25 min at 4°C, discarding the supernatant, resuspending the pellet in 200 mL of ice-cold distilled water and repeating the centrifugation process. The pellet was resuspended in 100 mL of ice-cold distilled water and the cells centrifuged again at 4,000 rpm for 25 min at 4°C. The pellet was resuspended in 20 mL of ice-cold 10% glycerol. Two 20 mL suspensions were pooled and transferred to a single pre-chilled 50 mL tube, providing two pre-chilled 50 mL tubes each containing 40 mL of cells in 10% glycerol**.** Centrifugation was carried out at 4,000 rpm for 10 min at 4°C followed by resuspending each pellet in 1 mL of ice-cold 10% glycerol to generate

electrocompetent cells. 50 µL aliquots of these electrocompetent cells were transferred to pre-chilled Eppendorf tubes followed by flash freezing in liquid nitrogen and storage at -80°C.

5.3.3.4.4 Electroporation

1-5 µL of DNA (plasmid DNA or ligation reaction ~ 10 µg DNA) was added to 100 µL of electro-competent *Pectobacterium* cells and mixed gently. The electroporation mixture was transferred to a pre-chilled 0.1 cm cuvette. Samples were electroporated with 12.5 kV/cm voltage using a Gene Pulser (Bio-Rad Lab) electroporator. 250 µL of room temperature LB was added immediately after electroporation. The solution was then transferred to a 15 mL tube and the tube shaken for at least 1 h at 37°C for the expression of antibiotic resistance genes. 10-50 µL of each transformation mix was spread on to selective medium as above (Section 5.3.3.3.2).

5.3.4 Preparation of knockout gene constructs

Approximately 1.4 Kb each of the upstream and downstream regions of the target gene were amplified from the genomic DNA of ICMP19477 using primer pairs LF/LR and RF/RR (Figure 5.1), respectively. The primer pairs targeting individual genes used for creating knockout gene constructs are listed in Table 5.2. The chloramphenicol cassette was amplified using primer pair chlF and chlR (Table 5.2) and the pACYC184 vector as a template. The LF and RR primers had *Xba*I and *Bam*HI restriction sites at the 5′ ends, respectively. The LR and RF primers (Figure 5.1) contained, at their 5′ ends, an extension of approximately 20 nucleotides homologous to the chlF and chlR primers, respectively. The three PCR products obtained in the first step were mixed at equimolar concentrations and subjected to an overlap-extension PCR with LF and RR primers to generate a final gene knockout construct containing a chloramphenicol resistance cassette flanked by both the upstream and the downstream gene homologous regions. The overlap-extension PCR was carried out with Expand High Fidelity Taq Polymerase (Roche), according to the manufacturer's recommendations; the conditions used were 94°C for 30 s, 94°C for 30 s, 60°C for 30 s, 68°C for 5 min (30

cycles), and a final extension at 68°C for 7 min. The final PCR products were purified and the DNA concentration determined.

5.3.5 Construction of recombinant plasmids for allelic exchange mutagenesis

Restriction digestions were carried out on plasmids pKNG101 and pK18mobsacB using restriction enzymes *Xba*I and *Bam*HI. The plasmid maps of pKNG101 and pK18mobsacB are shown in Figure 5.2. The digested products were analyzed on agarose gels, purified and then the linear plasmids were dephosphorylated prior to ligations. Ligations were carried out as described in the previous section (Section 5.3.3.3) to generate recombinant plasmids containing the allelic exchange cassette for each target gene.

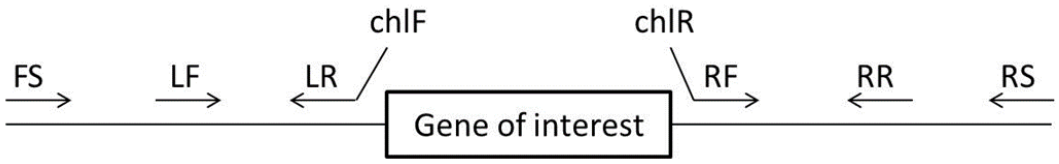


Figure 5.1 Preparation of a knockout gene construct. Approximately 1.4 Kb regions upstream and downstream of the gene of interest were amplified using primer pairs LF/LR and RF/RR, respectively. The LR and RF primers at their 5′ ends contain complementary sequences of chlF and chlR primers, respectively. FS and FR are primers designed externally to the regions of the knockout gene construct to validate the authenticity of the knockout construct after crossover events in mutant *Pectobacterium* strains. chlF and chlR primers were used to amplify the chloramphenicol cassette from the plasmid pACYC184.

Table 5.2 List of primers used in this study

Primer Name	Primer Sequence (5' to 3')	Amplicon size (bp)	Annealing Temperature (°C)
Primers to amplify upstream gene regions			
PAD_LF	GGCCTCTAGAGCCGGGTTTGGACACGGCAT		
PAD_LR	CGAAGTGATCTTCCGTCACAGCGCTTCATCGCAGCCAGTCTT	1399	60
SIM_LF	GGCCTCTAGAGCCAGACTCCCAATTACCCCTGCC		
SIM_LR	CGAAGTGATCTTCCGTCACACGCCAGCATTGCACTGAGCGT	1399	60
CFA_LF	GGCCTCTAGAATGGCTCCCCGGCGTGATGA		
CFA_LR	CGAAGTGATCTTCCGTCACACCGCCAACGCCTTTTTGGATGC	1399	60
NRPS_LF	GGCCTCTAGATGCGCAGCGGCAGTGTGTTA		
NRPS_LR	CGAAGTGATCTTCCGTCACACGCCGATCTGCGTCAGCAGT	1399	51
ABC_LF	GGCCTCTAGACAGCTACTGGCGGTGATGCGT		
ABC_LR	CGAAGTGATCTTCCGTCACACAGACATCTTCCGCAAGGCGGATT	1399	57
Primers to amplify downstream gene regions			
PAD_RF	CGCTTATGTCTATTGCTGGTCGCCTAGCGTGAGCATCAGTCG		
PAD_RR	GGCCGGATCCATCGTTCCGCCGCCCGTTC	1399	60
SIM_RF	CGCTTATGTCTATTGCTGGTCCTCGTAGCAACCGGCAATCTCAA		
SIM_RR	GGCCGGATCCGATCAGTAAGGTCAACCAGGCGGC	1399	52
CFA_RF	CGCTTATGTCTATTGCTGGTAATCGGTACGATCCAGCAGCCG		
CFA_RR	GGCCGGATCCTATCGAGCGGCGCAACCCC	1399	57
NRPS_RF	CGCTTATGTCTATTGCTGGTGGCAGTCAGTTCGCGGAAGGT		
NRPS_RR	GGCCGGATCCAGACGCTGGAAGCGCTGCTG	1399	55
ABC_RF	CGCTTATGTCTATTGCTGGTGAAATGCCATTGGTGACGCTGGC		
ABC_RR	GGCCGGATCCTCACATGACGACGGCCGATGAAC	1399	55
Primers to amplify regions outside the gene knockout construct			
PADSF	GCGAGAAAATCATCCGGCAC		
PADSR	GTACCACGGCAAACAGCTTG	4864	60
SIMSF	TGTGTGCTCCATCACTTCCC		
SIMSR	TGTGTGCTCCATCACTTCCC	5244	60
CFASF	GCAGTTCAAAATGGCTCCCC		
CFASR	TTTCTGGAGGATGCAGCAGG	5913	60
NRPSNF	TGTTAGCTGCACTGTGCTGA		
NRPSNR	ACCACGCTGTTTATGACCGT	16545	60
ABCSF	CAATTAGGCCCGGCTCAAGA		
ABCSR	AAGGGCTTAAGCGACGAACA	4726	60
Primers to amplify chloramphenicol cassette and sequence regions outside chloramphenicol cassette			
ChlF	TGTGACGGAAGATCACTTCG		
ChlR	ACCAGCAATAGACATAAGCG	860	58
ChlF_seq	GAAAGCAAATTCGACCCGGT		
ChlR_seq	ACGTCTCATTTTCGCCAAAAGT	n/d[1]	n/d
Other primers			
M13F	CGCCAGGGTTTTCCCAGTCACGAC		
M13R	AGCGGATAACAATTTCACACAGGA	n/d[2]	60
sacBF	ATGATTGAACAAGATGGATTG		
sacBR	TTGTTCAAGGATGCTGTCT	2683	51

The blue and purple coloured letters represent the restriction sites *Xba*I and *Bam*HI, respectively. The red and green colored letters represent the complementary sequences of chlF and chlR primers, respectively. n/d[1] – not determined, used for sequencing purposes only, n/d[2] – primers to confirm the gene constructs are correctly inserted in the parent plasmid.

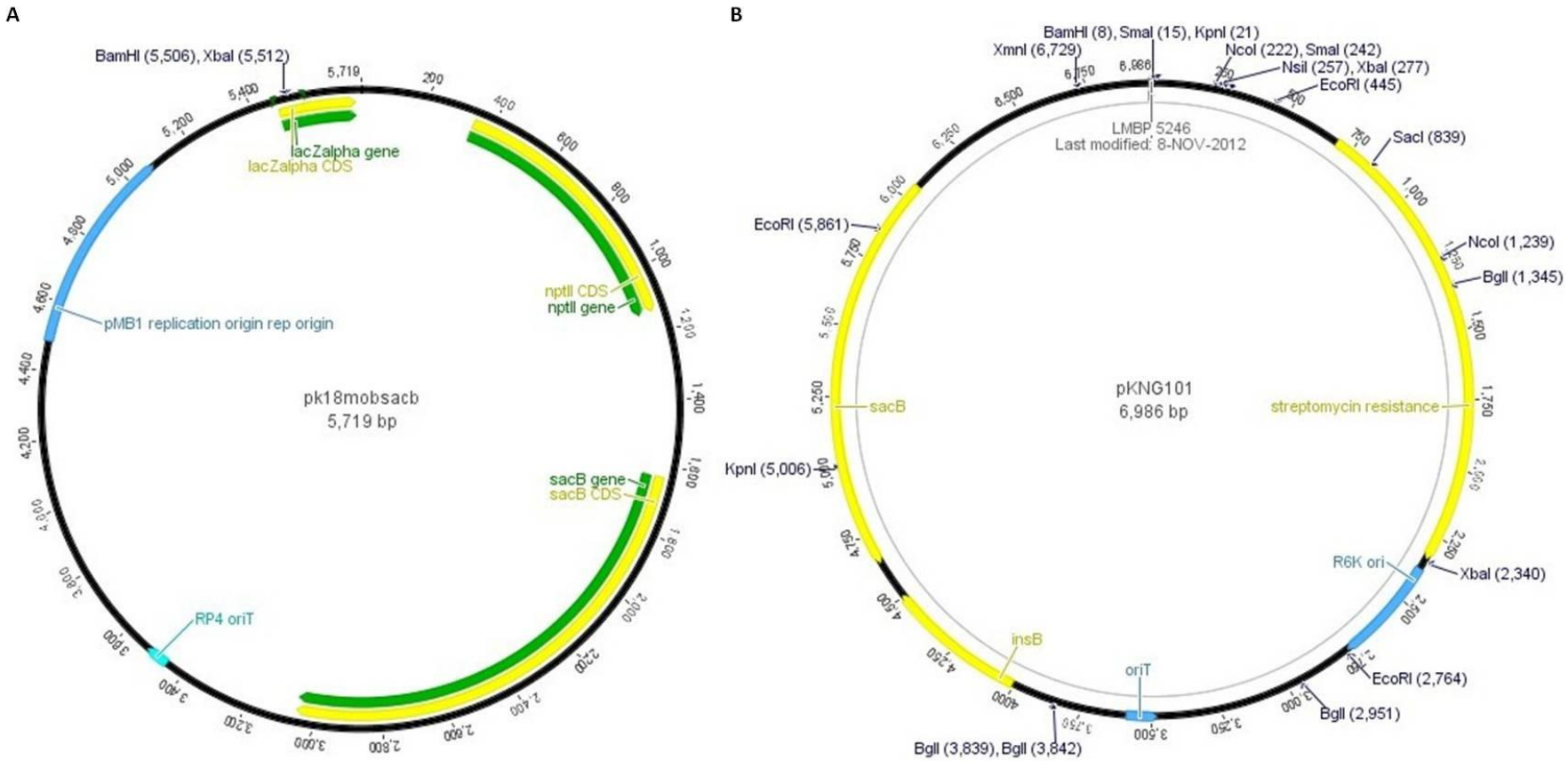


Figure 5.2 Plasmid maps of A, pK18mobsacB and B, pKNG101. Abbreviations: *nptII*, gene coding for kanamycin resistance; *sacb*, gene coding for levansucrase; *lacZalpha*, encodes for blue-white colour screening of colonies for ones with plasmids carrying inserts; *oriT*, origin of tranfer; pMB1, origin of replication. Expression of the *sacB* is lethal to Gram-negative bacteria when grown in the presence of 10% sucrose, providing a direct selection for loss of the plasmid.

5.3.6 Screening, selection and validation of positive clones

To check whether *E. coli* transformants carried recombinant plasmids with the correct inserts, PCR was performed with the appropriate primer pairs for each target gene, as shown in Table 5.2. Genomic DNA of ICMP19477 was used as a positive control. The gene construct in each recombinant plasmid was also confirmed by carrying out PCR using M13F/M13R primer pair (Table 5.2). Once the plasmids (constructs with correct inserts) were identified, they were transformed into selected *E. coli* strains; C118-λpir if the plasmid backbone was pKNG101 and S17-λpir for pK18mobsacB. The resulting recombinant *E. coli* strains were used as donor strains in the conjugation experiments to transfer the constructs to *Pectobacterium*.

5.3.7 Allelic exchange mutagenesis by conjugation

5.3.7.1 Transconjugation

Three strains were used to carry out a triparental mating to generate knockout mutants of ICMP19477. These included: i) an *E. coli* strain carrying the mutagenic plasmid (Kn^R or Str^R, depending on the plasmid used and Chl^R) as donor, ii) *E. coli* HH26 + pNJ5000 helper strain (Tet^R) and, (iii) wild type ICMP19477 as the recipient strain. Overnight cultures of each were grown in 5 mL LB including the appropriate antibiotics. 10 µL of each culture were mixed together and then the 30 µL mixture was spotted onto the surface of an LBA plate containing no antibiotics. The agar plate was incubated at 25°C for 24 h. After incubation, the mixture spot was scraped with a wire loop and streaked out onto minimal glucose agar plates supplemented with chloramphenicol and incubated at 25°C for 3 - 4 days or until visible colonies appeared. The individual colonies were then re-streaked onto the same media and incubated for an additional 3 - 4 days at 25°C to obtain individual colonies and to counter-select the *E. coli* donor strains. At this stage, the colonies were transconjugants (Chl^R).

Transconjugants were then screened for Kn^R or Str^R depending on the plasmid used, to identify whether the vector was still present (indicative of a first crossover event

had been mediated by incorporating the plasmid with the knockout gene construct into the recipient chromosome. Colony PCR using primers specific to *sac*B (sacBF/sacBR, Table 5.2) were performed to confirm the presence of the mutagenic plasmid in the recipient strains.

5.3.7.2 Creating mutants with a second crossover

Sucrose sensitive derivatives of ICMP19477 (first crossover mutants) were grown overnight in 5 mL LB without any antibiotics. The cultures were then diluted 100-fold in 1× Phosphate buffer. 100 µL aliquots of the diluted cultures were spread on to minimal sucrose agar plates supplemented with chloramphenicol and incubated at 25°C for 3-4 days. Colonies that were sucrose resistant and chloramphenicol resistant were then patched on to agar plates supplemented with kanamycin (pK18mobsacB) or streptomycin (pKNG101) to confirm the loss of the plasmid from the knockout mutant strain. Finally, sucrose and chloramphenicol resistant, kanamycin (pK18mobsacB) or streptomycin (pKNG101) sensitive colonies were screened for deletion of the plasmid by colony PCR using sacBF/sacBR primers.

5.3.8 Screening and selection of knockout mutant strain

Deletion of target genes in ICMP19477 was confirmed by PCR using gene specific primers (Table 5.2), amplifying from the external region of the deleted genes. The gene-deleted mutants were expected to produce different PCR products (~3.8 Kb) than the wild type, although the sizes of the PCR products were dependent on the design of the individual gene knockout constructs (shown in Table 5.2).

5.3.9 Pathogenicity assays

5.3.9.1 Blackleg assays

Blackleg assays were carried out as described previously in Chapter 2 (Section 2.3.5.2). For each strain, eleven plants were inoculated with bacterial cells, with either a

concentration of 10^2 (1st experiment) or 10^4 cfu. Plants inoculated with $MgSO_4$ were used as controls. Two dpi, each plant was scored for infection, using a severity scale that included disease symptoms: 0 = no symptoms at site of inoculation, 1 = symptoms at site of inoculation, 2 = intermediate lesions (extending upwards and downwards from the site of inoculation) and, 3 = wilting and collapse of stems (Figure 5.3). The scores were analyzed with analysis of variance. Data for the two experiments was combined into a single analysis, including bacterial strain and experiment (=concentration), and their interaction. The experiments were run concurrently so that any observed differences were likely to be associated with the bacterial concentrations rather than due to the experiments having been carried out on separate occasions.

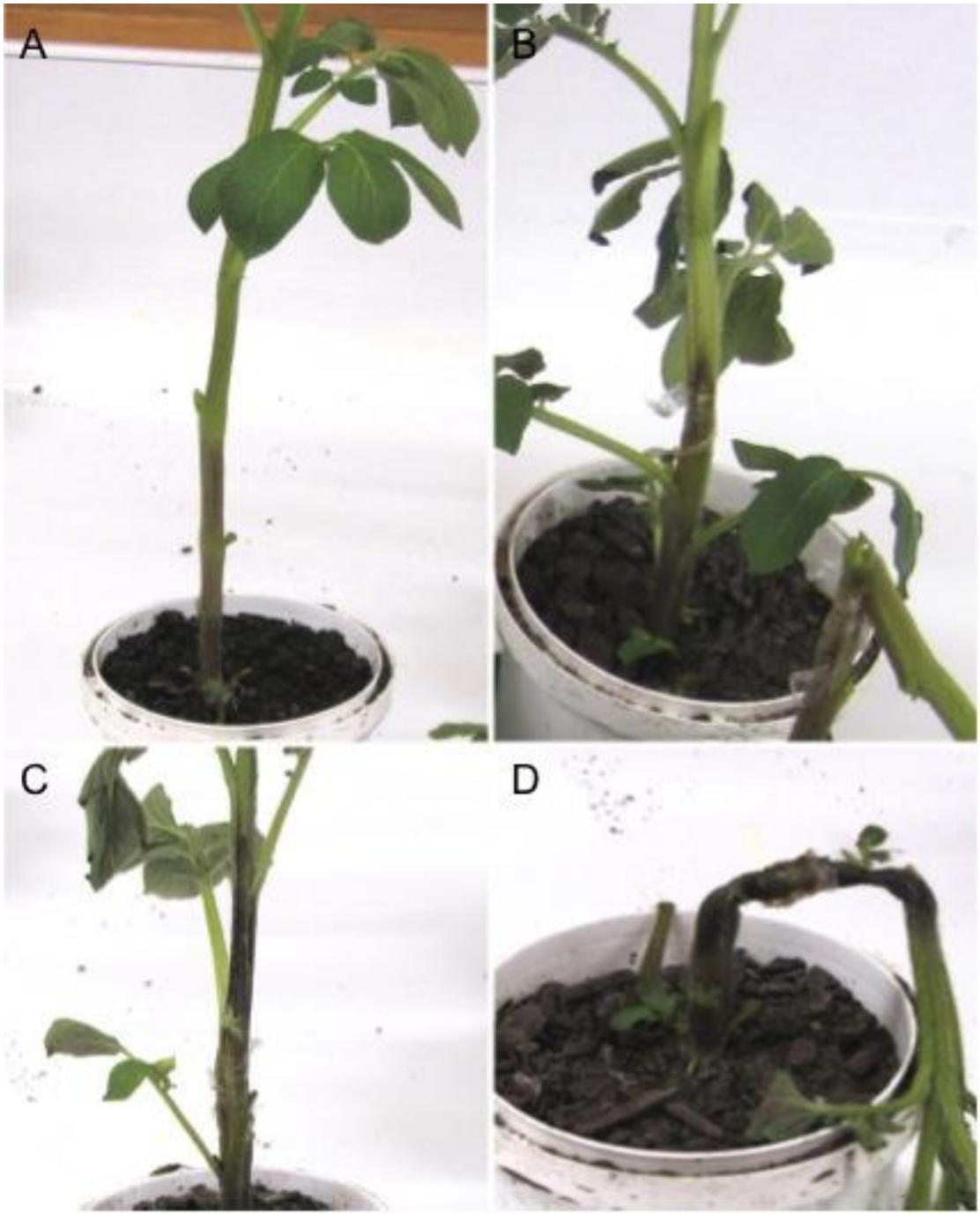


Figure 5.3 Disease severity scale for potato plants infected with wild type and/or knockout gene mutants of ICMP19477 two days post inoculation. A, 0 = no symptoms at site of inoculation; B, 1 = blackleg symptom at site of inoculation; C, 2 = intermediate lesions (extending upwards and downwards from the site of inoculation) and D, 3 = wilting and collapse of stems of potato plants.

5.3.9.2 Soft rot assays

To establish the capacity of bacterial strains to macerate potato tubers, potato 'Ilam hardy' tubers were surface-sterilised by submersion in 10% sodium hypochlorite for 10 min, followed by rinsing the tubers with sterile distilled water The tubers were air-dried and a uniform bore was made in the tubers using a cork borer (~ 3mm diameter by 10 mm deep). A 10 µL aliquot of the prepared bacterial suspension (equivalent to 10^4 cfu) was inoculated into each bore hole using a sterile pipette tip. The tuber plugs were then placed back in the holes and sealed with petroleum jelly to maintain anaerobic conditions and to reduce dessication. Tubers inoculated with 10 mM $MgSO_4$ were used as negative control treatments in this assay. Each inoculated tuber was placed on a layer of moist tissue paper in a plastic box, wrapped in a black plastic bag and subsequently incubated at 25°C in a dark room for 48 hr. The weight of each tuber was then recorded after the incubation period, before macerated tissue was removed by scraping. Each tuber was washed and re-weighed and the severity of soft rotting was recorded by calculating the weight of tissue loss after removal of the rotten tissue. Twelve potato tubers were used for each strain. The percentage of weight loss (i.e. 100*(initial wt – final wt)/final wt)) was analysed using a binomial generalized linear model approach (McCullagh & Nelder, 1989), with a logit link.

5.4 Results

5.4.1 The authenticity of the parent plasmids for allelic exchange

The authenticity of plasmids pK18mobsacB and pKNG101 were confirmed by maintaining them in *E. coli* on kanamycin and streptomycin antibiotics, respectively. The colonies that grew on these antibiotics were also streaked onto agar plates containing sucrose (10% w/v). None of the colonies were able to grow on agar plates containing sucrose, indicating that the *sacB* cassette was expressed and that the sucrose sensitive phenotype was stable. To confirm the sizes of the two plasmids, purified DNA of each plasmid was digested with *Xba*I and *Bam*HI, as restriction sites for these enzymes were located on multiple cloning sites. Restriction digestion

showed the resulting linearized plasmids were ~ 5.7 Kb and 7.0 Kb for pK18mobsacB and pKNG101, respectively (data not shown), which was in accordance with the sizes shown in the maps for each plasmid.

5.4.2 Construction of knockout constructs for functional studies in ICMP19477

Overlap PCR successfully generated PCR products (consisting of flanking sequences of the target gene (Figure 5.4) interrupted by a chloramphenicol resistance cassette (Figure 5.4) of approximately 3.8 Kb for all the target loci (*cfa7*, *nrps1*, *abc*, *sim* and *pad*) (Figure 5.5). Each of these products was then successfully cloned into both pK18mobsacB and pKNG101 using *Xba*I and *Bam*HI to generate the mutagenic plasmids (see Table 5.1) required for functional analysis in ICMP19477.

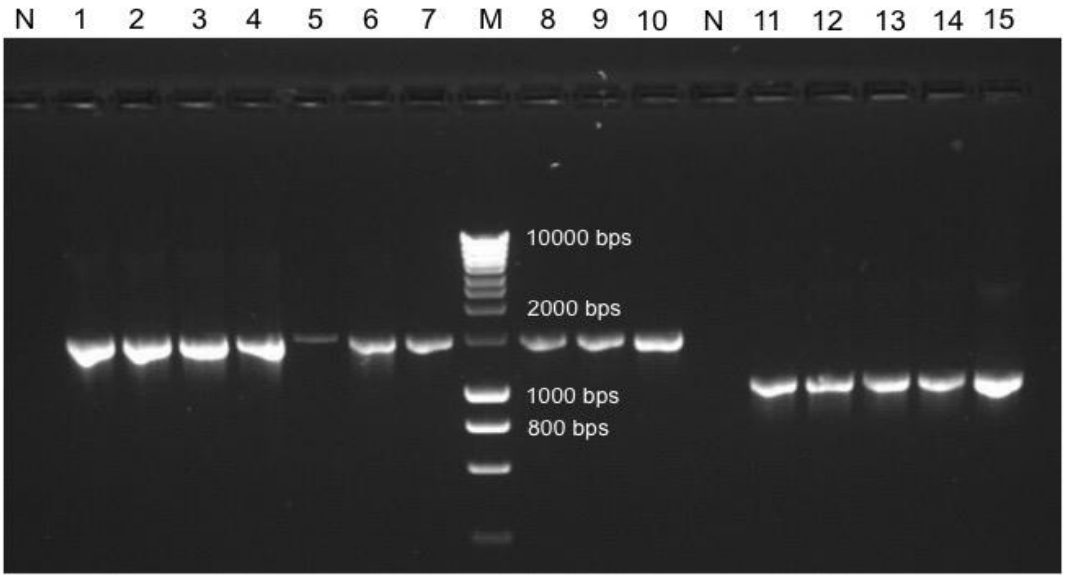


Figure 5.4 Gel image of the approximately 1.4 Kb PCR products (upstream and downstream) amplified for the target loci using gene specific LF/LR primers (Lanes 1-5) and RF/RR primers (Lanes 6,7,8-10), and the amplified chloramphenicol cassette (~ 1.0 Kb) from each knockout gene construct (using chlF/chlR primers, Lanes 11-15). Lanes: N, no template control; 1,6,11, *sim* knockout gene construct; 2,7,12, *pad* knockout gene construct; 3,8,13, *cfa7* knockout gene construct; 4,9,14, *nrps1* knockout gene construct; 5,10,15, *abc* knockout gene construct; M, Marker (1 Kb ladder)

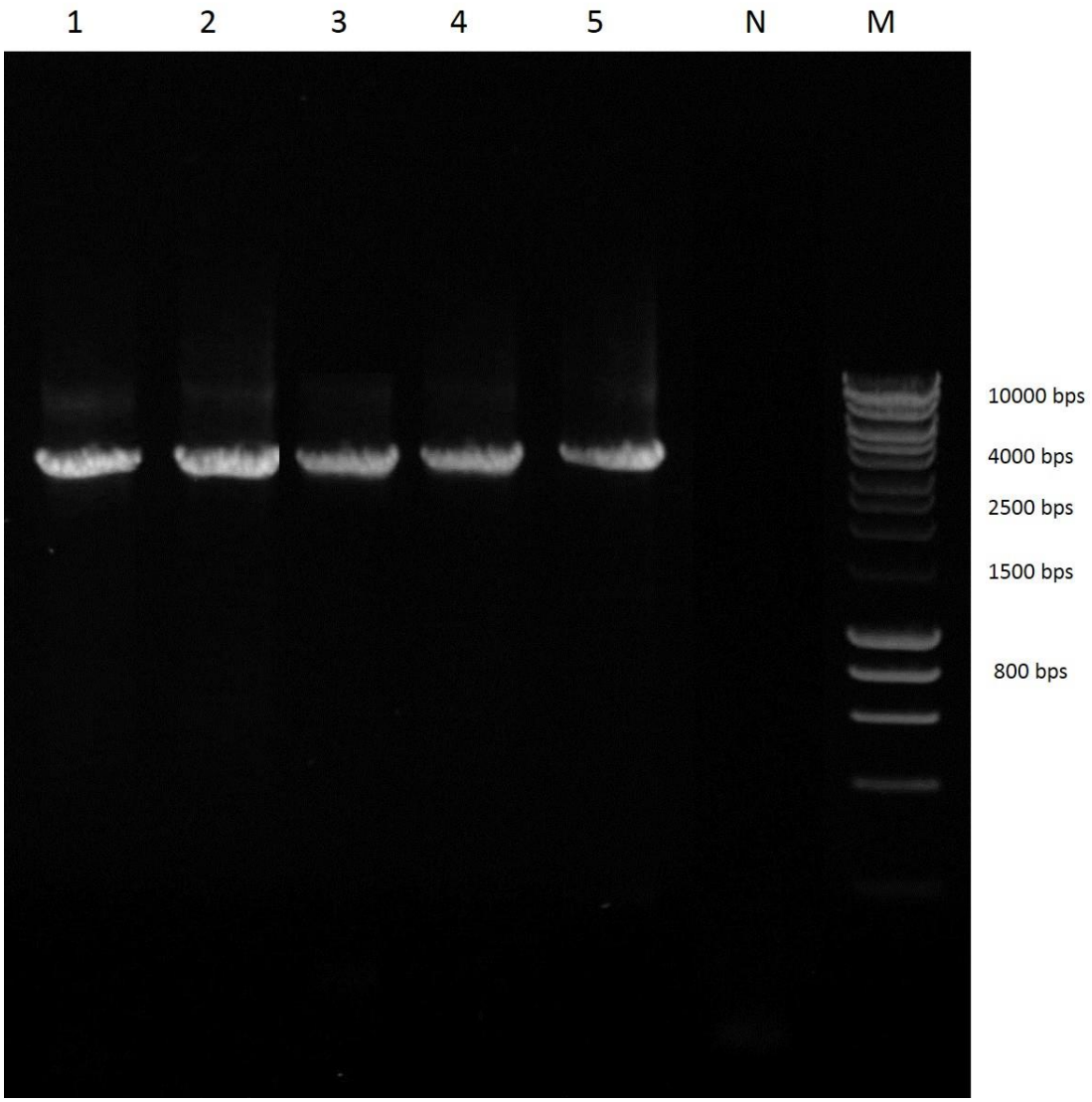


Figure 5.5 Gel image of the approximately 3.8 Kb PCR products (consisting of flanking sequences of the target gene interrupted by a chloramphenicol resistance cassette) for all the target loci, which were amplified using the LF/RR primers for each gene knockout construct. Lanes: 1, *sim* knockout gene construct; 2, *pad* knockout gene construct; 3, *cfa7* knockout gene construct; 4, *nrps1* knockout gene construct; 5, *abc* knockout gene construct; N, no template control; M, Marker (1 Kb ladder)

5.4.3 Validation of gene knockout constructs

To validate the gene knockout plasmids constructed to target *pad*, *sim*, *cfa7*, *nrps1* and *abc*, PCR was carried out with the primer pairs LF and RR for each construct (Figure 5.5). PCR was also carried out using primer pairs LF/chlR and chlF/RR to confirm the expected sizes of each flanking sequence containing the chloramphenicol cassette (~ 2.4 Kb) used for homologous recombination (Figure 5.6) and with the M13F and M13R

primer pair to confirm the fragments had been cloned into the plasmids appropriately (data not shown). All PCRs produced appropriately sized products indicative that the gene knockout constructs had been cloned into the two suicide vectors. After the validation of the gene knockout constructs, the mutagenic plasmids were used for conjugation.

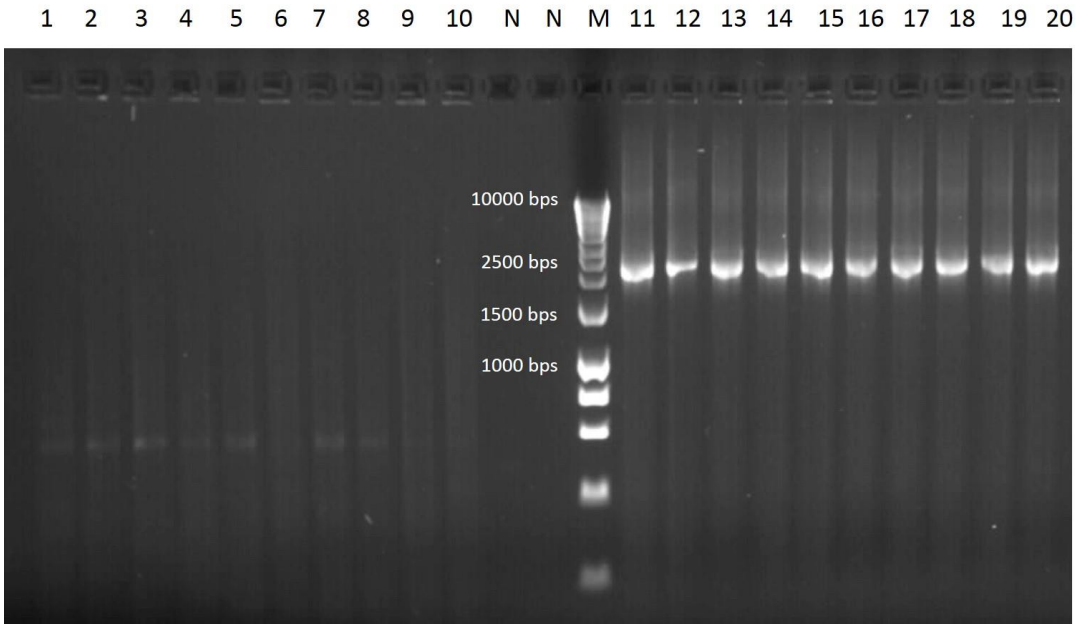


Figure 5.6 Gel image of the approximately 2.4 Kb PCR products for each knockout gene construct, which was amplified using the LF/chlR and chlF/RR primer pairs to confirm the integrity/authenticity of each flanking sequence containing the chloramphenicol cassette (~ 1.4 Kb). Lanes: 1-10, no template control; N, negative control using DNA of wild type ICMP19477 as template; 11,12, *sim* knockout gene construct; 13, 14; *pad* knockout gene construct; 15, 16, *cfa7* knockout gene construct; 17,18, *nrps1* knockout gene construct; 19,20, *abc* knockout gene construct; M, Marker (1 Kb ladder)

5.4.4 Transconjugation

5.4.4.1 Using mutagenic pKNG101 plasmids

Despite repeated attempts, not a single colony was observed during the transconjugation process using recombinant pKNG101 derivatives containing the allelic exchange cassettes for pad, *sim*, *cfa7*, *nrps1* or *abc* and the recipient ICMP19477,

with the exception of one colony that was produced using the plasmid carrying the *nrps1* gene construct. The single colony produced (Str^R, Chl^R) using the *nrps1* gene construct was expected to carry a single or double crossover event given the inability of pKNG101 to replicate in ICMP19477.

5.4.4.2 Using mutagenic pK18mobsacB plasmids

At least 100 chloramphenciol resistant colonies were observed for each set of gene deletion experiments grown on minimal glucose media supplemented with chloramphenicol using the mutagenic pK18mobsacB plasmids. These colonies were also resistant to kanamycin and were expected to have an allelic exchange cassette embedded into the chromosome at the target site due to homologous recombination (through a single).

5.4.4.3 Screening transconjugants for sucrose resistance or sensitivity

The transconjugants obtained using the mutagenic pKNG101 or pK18mobsacB plasmids were subsequently screened for sensitivity or resistance to sucrose to identify whether the allelic exchange cassettes had recombined into the chromosome via a single or double crossover event. All colonies were sensitive to sucrose, consistent with the occurrence of single crossover events. Furthermore, the presence of the vector in colonies generated using pK18mobsacB was confirmed by PCR using the *sacB* primers (Figure 5.7).

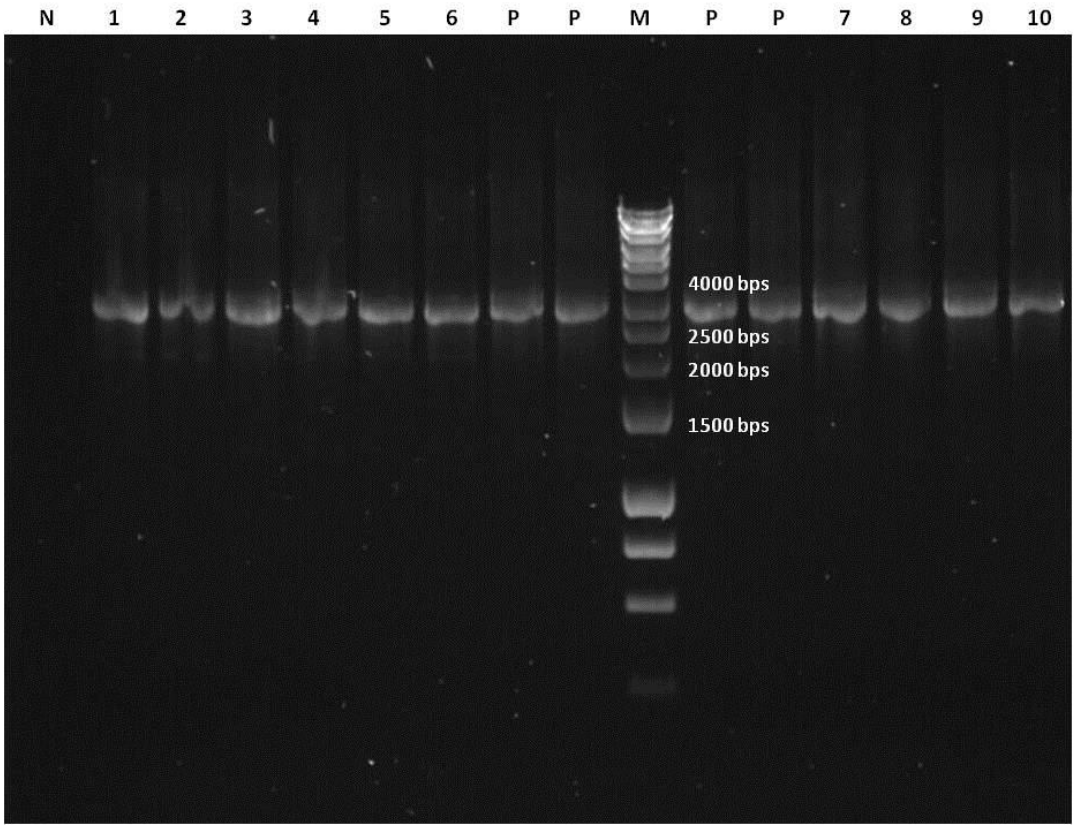


Figure 5.7 Amplification of the approximately 2.7 Kb *sacB* cassettes from transconjugants carrying different gene knockout constructs. Lanes 1,2, ICMP19477 transconjugant carrying pK18mobsacB-*sim*::*chl;* 3,4, ICMP19477 transconjugant carrying pK18mobsacB-*pad*::*chl*; 5,6, ICMP19477 transconjugant carrying pK18mobsacB-*cfa7*::*chl*; 7,8, ICMP19477 transconjugant carrying pK18mobsacB-*nrps1*::*chl*; 9,10 ICMP19477 transconjugant carrying pK18mobsacB-*abc*::*chl*; N, negative control; M, Marker (1 Kb ladder)

5.4.5 Gene knockout mutant strains

5.4.5.1 Using mutagenic pKNG101 plasmids

To generate a mutant carrying an inactivated *nrps1* gene construct, the chloramphenicol resistant and sucrose sensitive colony obtained during the transconjugation process was grown overnight on non-selective LB media. Cells were then spread onto minimal sucrose media supplemented with chloramphenicol to select sucrose resistant colonies. Only one colony was observed after 4 days of incubation at 28°C, which was chloramphenicol and sucrose resistant but

streptomycin sensitive, suggesting that the mutagenic plasmid was lost after the counter-selection process. To confirm whether the colony was an authentic *nrps1* gene knockout mutant of ICMP19477, primer pair NRPSNF/NRPSNR was used to amplify the external region (upstream and downstream) of the deleted gene. As a result, a shorter PCR product (~ 4.0 Kb) was amplified from the gene-deleted colony. In contrast, PCR with external primers using genomic DNA of ICMP19477 as template did not generate any product because of the large size of the wild type gene (size ~21.0 Kb). The PCR product from the *nrps1* mutant (ICMP19477Δ*nrps1*^{2}) was confirmed to be the result of a double crossover event by DNA sequencing (Figure 5.8, Appendix A5.2).

5.4.5.2 Using mutagenic pK18mobsacB plasmids

To generate knockout mutant strains from the chloramphenicol resistant and sucrose sensitive colonies obtained using the mutagenic pK18mobsacB- plasmids, one of the chloramphenicol resistant and sucrose sensitive colonies from each transconjugant was grown overnight on non-selective LB media. Cells were then spread onto minimal sucrose media supplemented with chloramphenicol and sucrose resistant colonies were selected for each gene construct. At least 50 sucrose and chloramphenicol resistant colonies were observed for each gene knockout construct. None of the colonies were able to grow on agar plates containing kanamycin. This indicated that the mutagenic plasmid was likely removed from these colonies. To confirm the removal of plasmid from the colonies, 25 colonies from each gene construct were selected for PCR amplification using primers specific to *sacB* cassette. PCR products corresponding to the size of the *sacB* cassette were amplified from all colonies, confirming that the mutagenic plasmid was still integrated into the chromosome.

To further identify knockout mutant strains, primer pairs (PADSF/PADSR, SIMSF/SIMSR, CFASF/CFASR, NRPSNF/NSPSNR and ABCSF/ABCSR) were used to amplify the external regions upstream and downstream of the deleted genes. A total of 25 chloramphenicol and sucrose resistant colonies were selected for each gene construct and PCR was carried out with external primers specific for each gene

construct. Each time, the amplified PCR product corresponded to the size of wild type gene from ICMP19477.

Although a wild type PCR product was generated, the colonies always remained resistant to chloramphenicol and sucrose and sensitive to kanamycin. In contrast, wild type ICMP19477 did not grow on either of the antibiotics suggesting that the knockout gene mutants had undergone a single crossover event instead of a double-crossover event. The predicted structure of one such single crossover knockout mutant (ICMP19477Δ*sim*[1]) is shown in Appendix A5.8. Sensitivity to kanamycin and resistance to sucrose probably occurred due to point mutations in the *sacB* and kananmycin resistance cassettes. To further confirm the single crossover events, PCR with a combination of primers was carried out. The primer-pair combinations were chlF/gene external reverse and chlR/gene external forward primers. The expected PCR products using chlF/gene external reverse primers (confirmed by sequencing, Appendices 5.3, 5.4, 5.5, 5.6, 5.7; Figures 5.9, 5.10, 5.11, 5.12, 5.13) were observed for all of the knockout gene mutants. However, after repeated attempts, no PCR products were obtained using primer pair chlR/gene external forward primers suggesting the presence of remnants of the pK18mobsacB plasmid, which had not been removed after the counter-selection process using sucrose.

The double crossover mutant (ICMP19477Δ*nrps1*[2]) generated using pKNG101 and the single crossover mutants generated using pK18mobsacB (ICMP19477Δ*pad*[1], ICMP19477Δ*sim*[1], ICMP19477Δ*cfa7*[1], ICMP19477Δ*nrps1*[1] and ICMP19477Δ*abc*[1]) were used in blackleg and soft rot assays to establish whether mutations in these loci had any effect on virulence of ICMP19477. Given the location of many of the targeted genes in distinct clusters, single crossover events were predicted to cause inactivation of any phenotype associated with the target loci in the same manner as double crossover events.

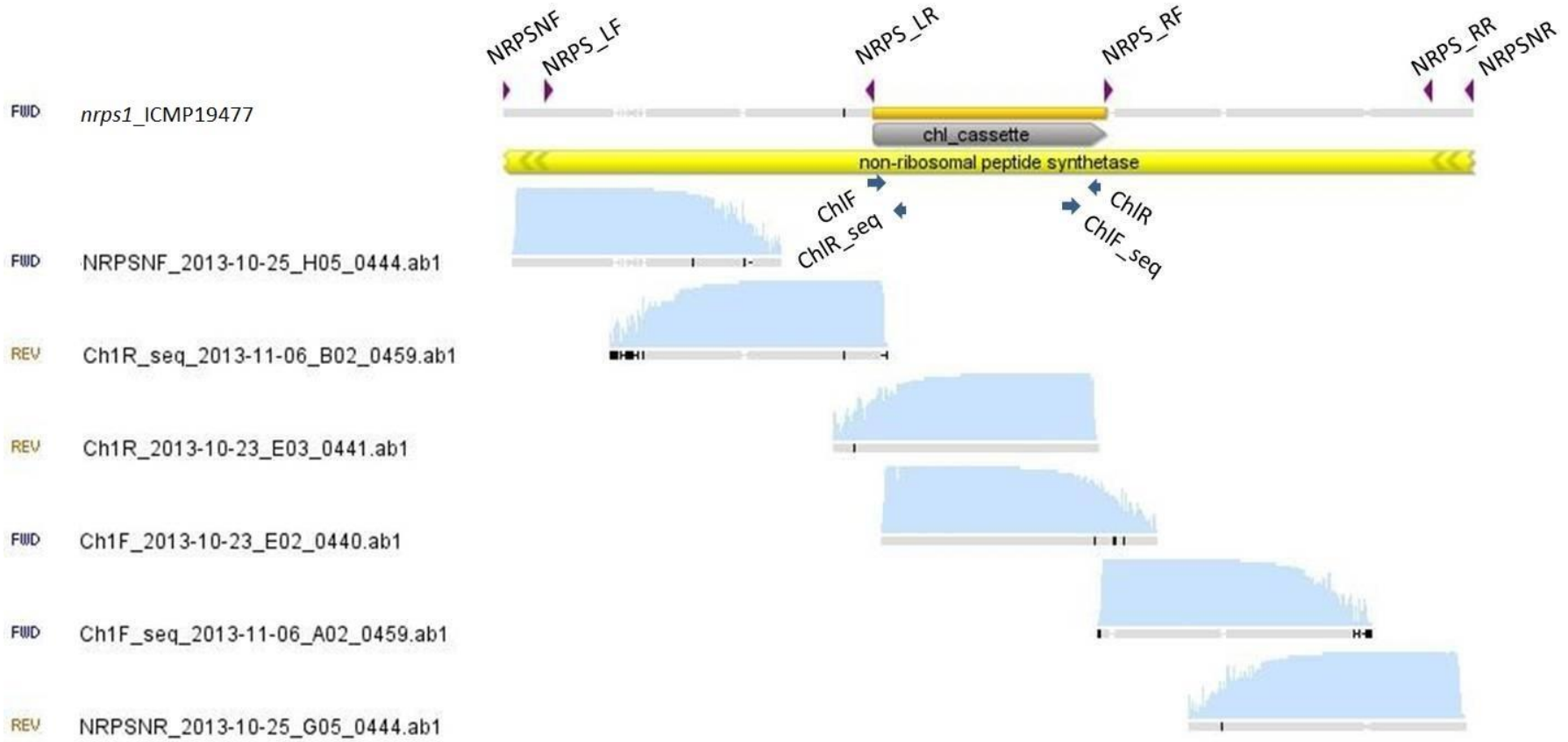


Figure 5.8 A diagrammatic representation of the DNA sequences obtained from Sanger sequencing of the PCR product amplified using the NRPSNF and NRPSNR primers and genomic DNA of ICMP19477Δ*nrps1*[2] as a template. The NRPS_LF/NRPS_LR and NRPS_RF/NRPS_RR primer pairs were used to amplify ~1.4 Kb upstream (left) and downstream (right) regions of the *nrps1* gene respectively, for design of the *nrps1* gene

knockout construct. ChlF and ChlR are the primers flanking the chloramphenicol antibiotic cassette. The primer pair NRPSNF/NRPSNR was designed to bind to the genome outside (externally) the *nrps1* gene knockout construct, to confirm homologous recombination had occurred in ICMP19477Δ*nrps1*[2]. ChlR_seq and ChlF_seq were designed at the ends of the chloramphenicol cassette flanking outwards to obtain the internal sequences of the final gene knockout construct.

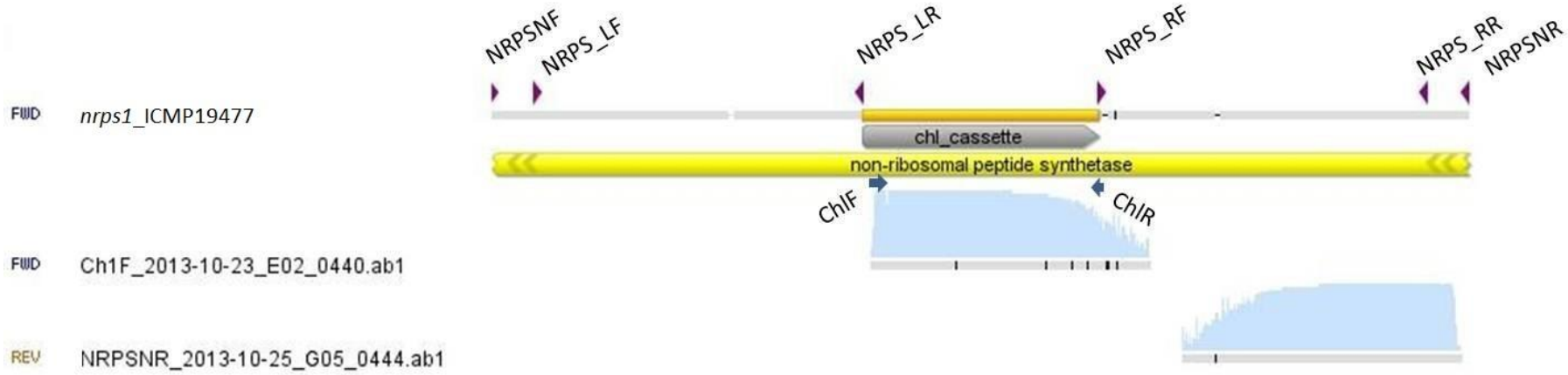


Figure 5.9 A diagrammatic representation of the DNA sequences obtained from Sanger sequencing of the PCR product amplified using the ChlF and NRPSNR primers and genomic DNA of ICMP19477Δ*nrps1*[1] as a template. The NRPS_LF/NRPS_LR and NRPS_RF/NRPS_RR primer pairs were used to amplify ~1.4 Kb upstream (left) and downstream (right) regions of the *nrps1* gene respectively, for design of the *nrps1* gene knockout construct. ChlF and ChlR are the primers flanking the chloramphenicol antibiotic cassette. The NRPSNF and NRPSNR are the primers that were designed outside the regions (upstream and downstream) of the *nrps1* gene knockout construct to confirm homologous recombination had occurred in ICMP19477Δ*nrps1*[1]. No PCR product was amplified using the NRPSNF and ChlR primers.

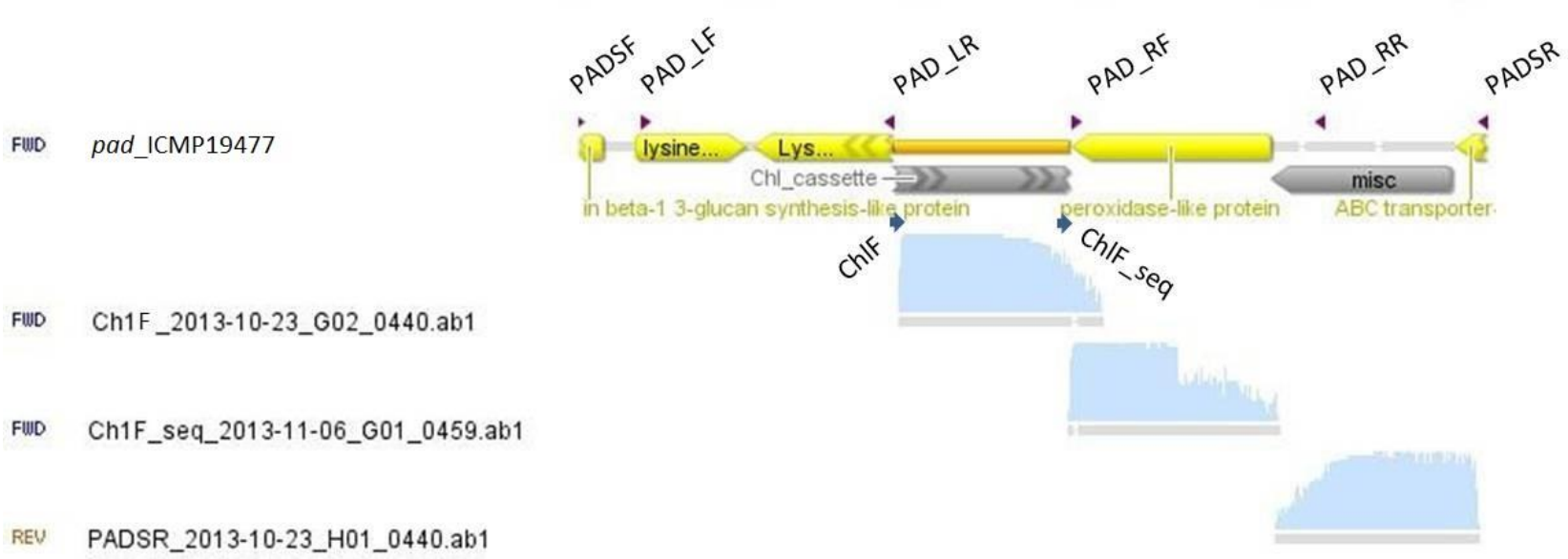


Figure 5.10 A diagrammatic representation of the DNA sequences obtained from Sanger sequencing of the PCR product amplified using the ChlF and PADSR primers and genomic DNA of ICMP19477Δ*pad*[1] as a template. The PAD_LF/PAD_LR and PAD_RF/PAD_RR primer pairs were used to amplify ~1.4 Kb upstream (left) and downstream (right) regions of the *pad* gene respectively, for design of the *pad* gene knockout construct. ChlF and ChlR are the primers flanking the chloramphenicol antibiotic cassette. The primer pair, PADSF/PADSR was designed to bind to the genome outside (externally) *pad* gene knockout construct to confirm homologous recombination had occurred in ICMP19477Δ*pad*[1]. No PCR product was amplified using PADSF and ChlR primers.

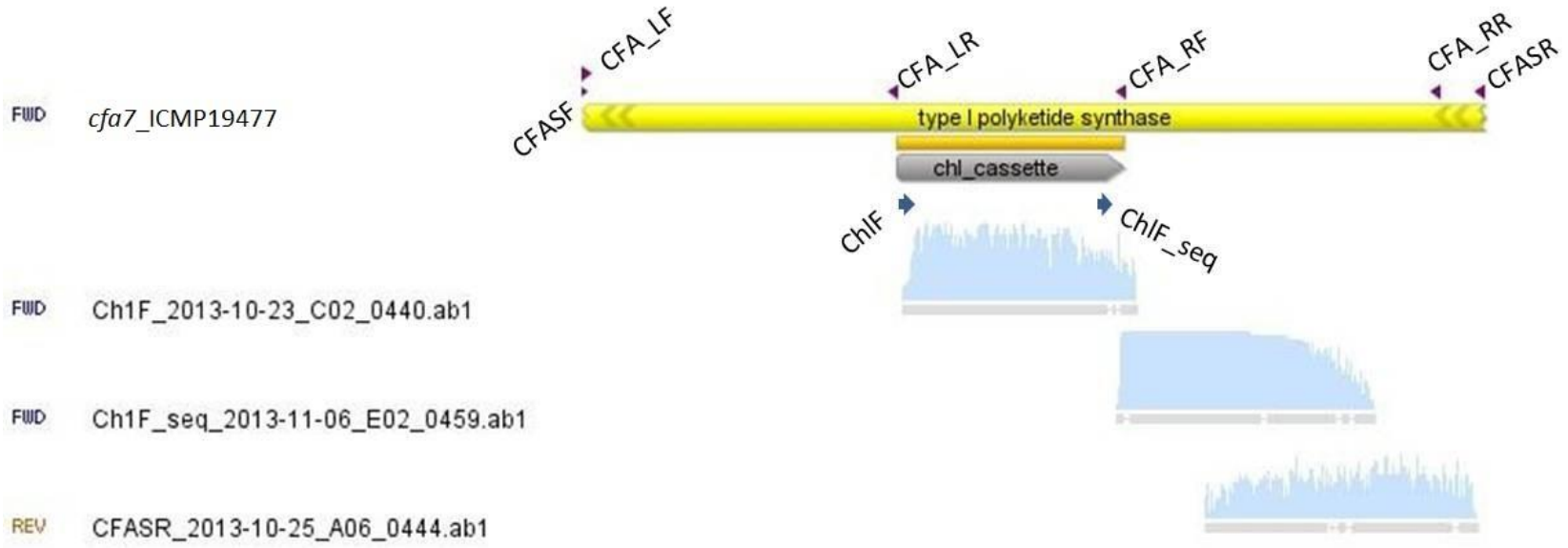


Figure 5.11 A diagrammatic representation of the DNA sequences obtained from Sanger sequencing of the PCR product amplified using the ChlF and CFASR primers and genomic DNA of ICMP19477Δ*cfa7*[1] as a template. The CFA_LF/CFA_LR and CFA_RF/CFA_RR primer pairs were used to amplify ~1.4 Kb upstream (left) and downstream (right) regions of the *cfa7* gene respectively, for design of the *cfa7* gene knockout construct. ChlF and ChlR are the primers flanking the chloramphenicol antibiotic cassette. The primer pair, CFASF/CFASR was designed to bind to the genome outside (externally) *cfa7* gene knockout construct to confirm homologous recombination had occurred in ICMP19477Δ*cfa7*[1]. No PCR product was amplified using the CFASF and ChlR primers.

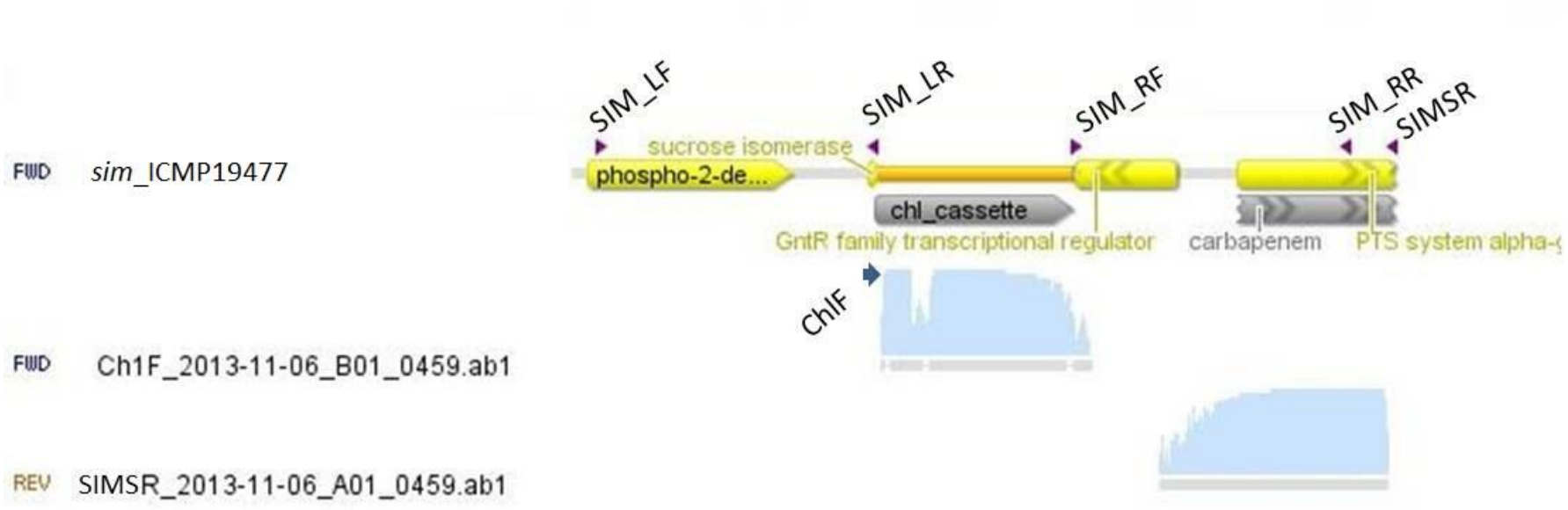


Figure 5.12 A diagrammatic representation of the DNA sequences obtained from Sanger sequencing of the PCR product amplified using the ChlF and SIMSR primers and genomic DNA of ICMP19477Δ*sim*[1] as a template. The SIM_LF/SIM_LR and SIM_RF/SIM_RR primer pairs were used to amplify ~1.4 Kb upstream (left) and downstream (right) regions of the *sim* gene respectively, for design of the *sim* gene knockout construct. ChlF and ChlR are the primers flanking the chloramphenicol antibiotic cassette. The primer pair, SIMSF/SIMSR was designed to bind to the genome outside (externally) the *sim* gene knockout construct, to confirm homologous recombination had occurred in ICMP19477Δ*sim*[1]. No PCR product was amplified using the SIMSF and ChlR primers.

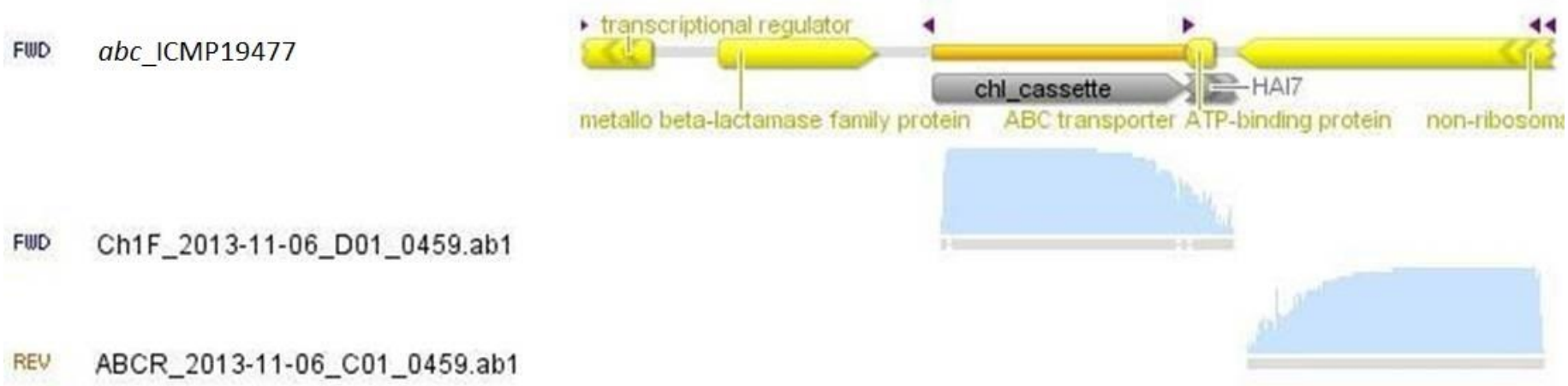


Figure 5.13 A diagrammatic representation of the DNA sequences obtained from Sanger sequencing of the PCR product amplified using the ChlF and ABCSR primers and genomic DNA of ICMP19477Δ*abc*[1] as a template. The ABC_LF/ABC_LR and ABC_RF/ABC_RR primer pairs were used to amplify ~1.4 Kb upstream (left) and downstream (right) regions of the *abc* gene respectively, for design of the *abc* gene knockout construct. ChlF and ChlR are the primers flanking the chloramphenicol antibiotic cassette. The primer pair, ABCSF/ABCR was designed to bind to the genome outside (externally) the *abc* gene knockout construct, to confirm homologous recombination had occurred in the ICMP19477Δ*abc*[1]. No PCR product was amplified using ABCSF and ChlR primers.

5.4.6 Virulence assays

5.4.6.1 Blackleg assays

The virulence of six mutants (ICMP19477Δ*nrps1*[2], ICMP19477Δ*pad*[1], ICMP19477Δ*sim*[1], ICMP19477Δ*cfa7*[1], ICMP19477Δ*nrps1*[1] and ICMP19477Δ*abc*[1]) was compared to that of wild type ICMP19477 in blackleg assays. Two experiments were carried out simultaneously, using bacterial concentrations of 10^2 (1st experiment) or 10^4 per inoculation site.

In both the experiments, two dpi, stems inoculated with ICMP19477Δ*sim*[1] showed a statistically significant reduction in lesion length ($p<0.001$) (Figure 5.14). Indeed, disease scores averaged over the two experiments ranged from 1.18 for the *sim* mutant to 2.68 for the wild type. Stems inoculated with ICMP19477Δ*pad*[1] also showed a reduced lesion length. Plants infected with the remaining mutants, however, showed no significant difference in disease symptoms to those inoculated with the wild type. Plants inoculated with $MgSO_4$ did not show any symptoms.

5.4.6.2 Soft rot assays

The virulence of the six mutants was compared to that of wild type ICMP19477 in soft rot assays. In these assays, the average percentage weight losses for potato tubers inoculated with ICMP19477Δ*sim*[1] and ICMP19477Δ*pad*[1] were significantly lower ($p<0.001$) than those inoculated with the other bacterial strains (Figure 5.15, 5.16). As expected, there was no rotting observed when potato tubers were inoculated with the negative control, $MgSO_4$.

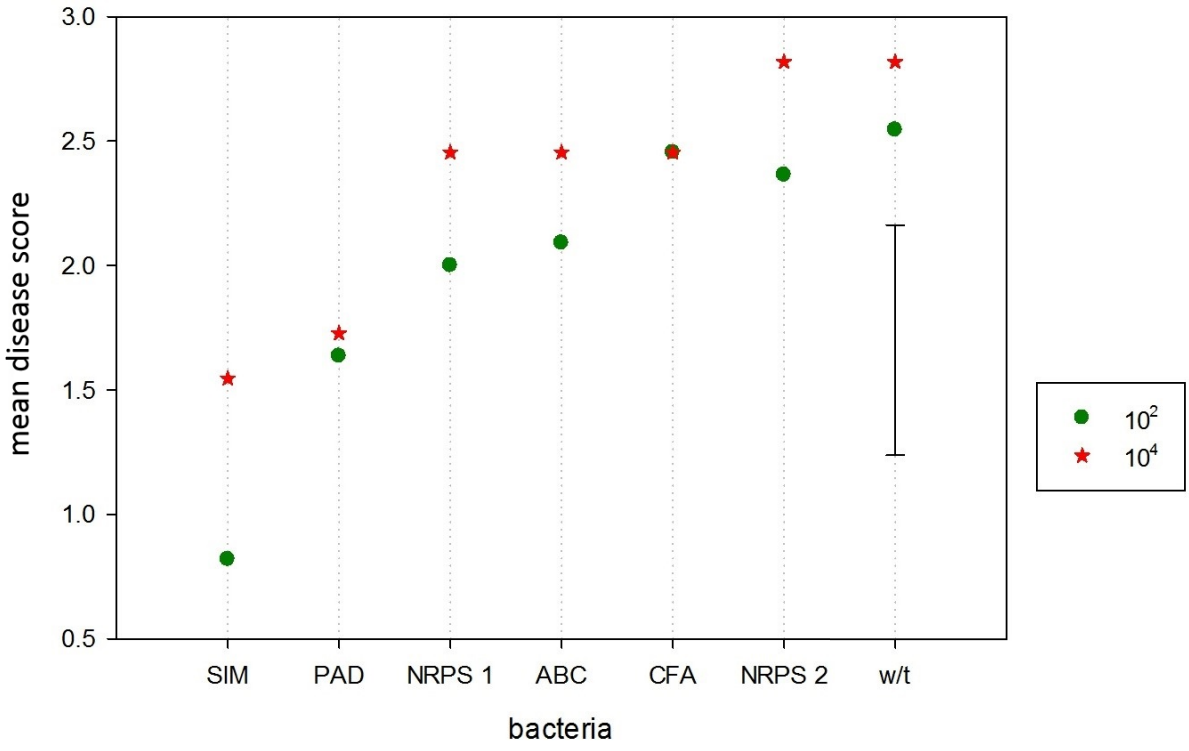


Figure 5.14 Mean scores for blackleg symptoms on potato stems after inoculation with mutants in several candidate virulence loci. Bacteria were inoculated at different concentrations (10^2 and 10^4 colony forming units). Eleven potato plants were used for each bacterial strain and the lesion lengths were recorded two days post inoculation. Error bar is Lsd 5%. SIM, ICMP19477Δ*sim*[1]; PAD, ICMP19477Δ*pad*[1]; NRPS1, ICMP19477Δ*nrps1*[1]; ABC, ICMP19477Δ*abc*[1]; CFA, ICMP19477Δ*cfa7*[1]; NRPS2, ICMP19477Δ*nrps1*[2] and w/t, wild type ICMP19477.

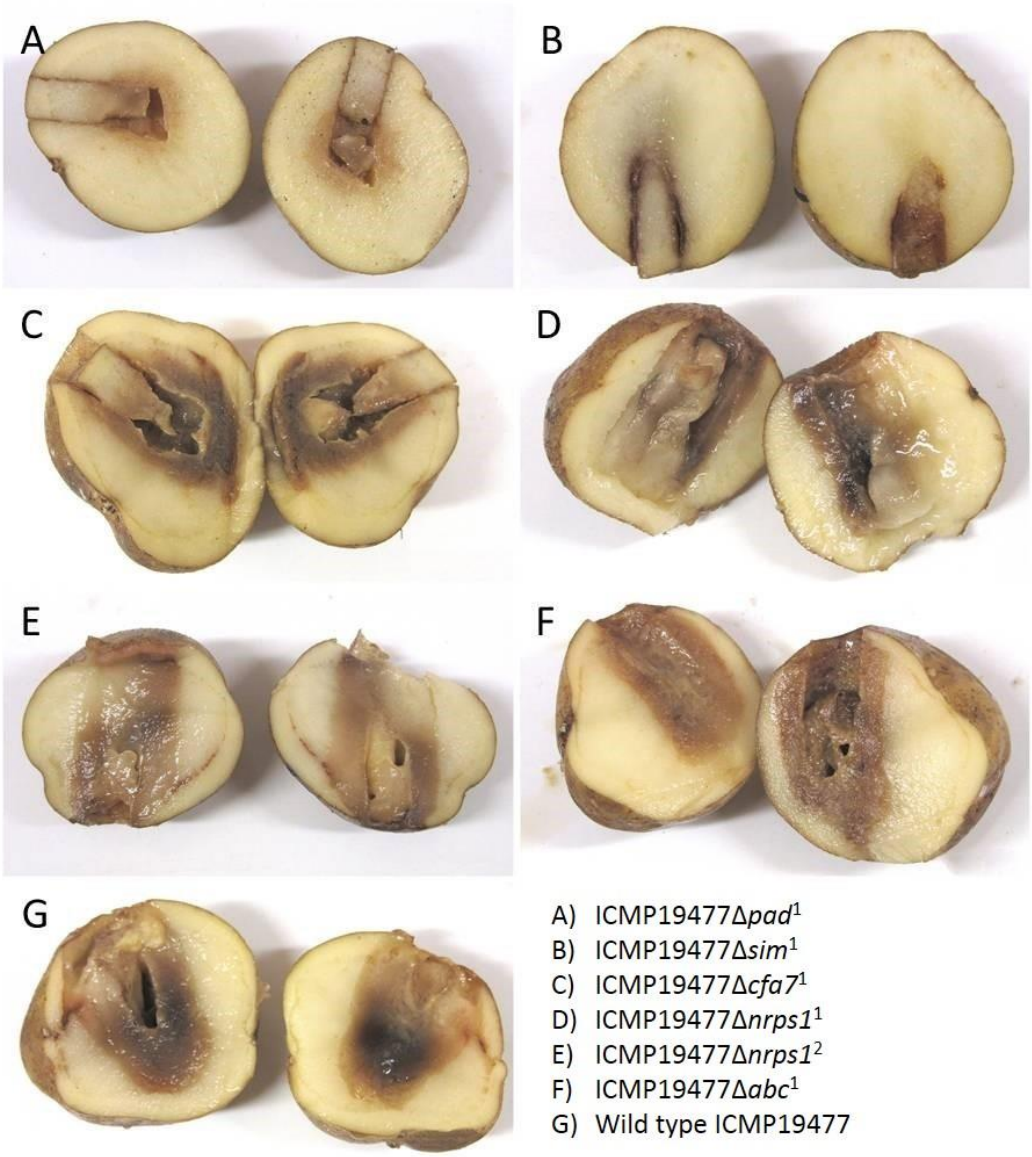


Figure 5.15 Soft rot symptoms on potato tubers two days post inoculation with 10^4 bacterial cells of each of the seven *Pectobacterium* strains (A-G). [1] represents the single crossover mutants of ICMP19477 and [2] represents the double crossover mutant of ICMP19477. Potato tubers inoculated with the *sim* (B) and *pad* (A) mutants showed the lowest amounts of tissue maceration when compared to tubers inoculated with wild type ICMP19477 (G).

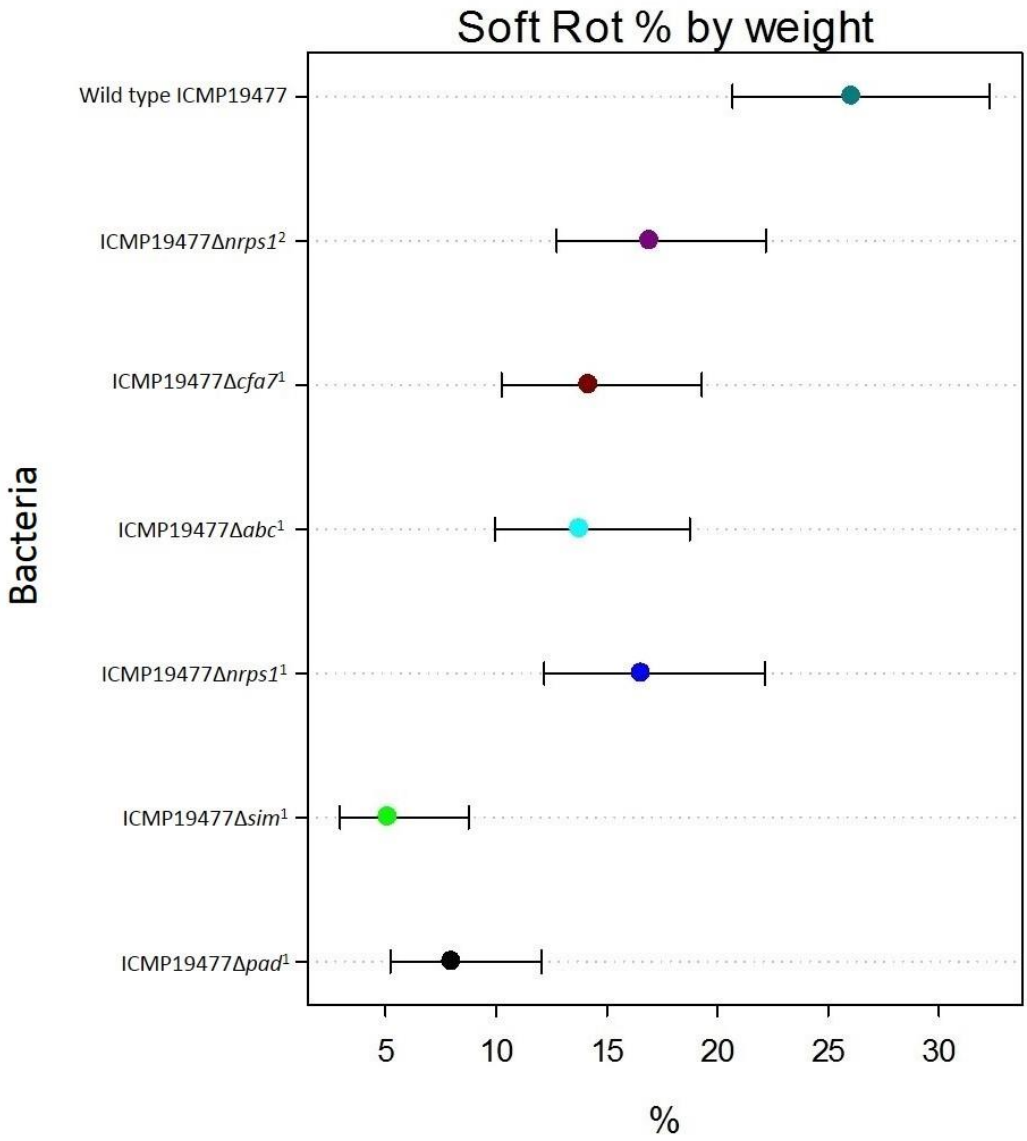


Figure 5.16 Percentage of potato tuber weight macerated when inoculated with each of six *Pectobacterium* mutants or with ICMP19477. For each strain, eleven potato tubers were inoculated with 10^4 bacterial cells. Error bars are 95% confidence limits for the means. Error bar is Lsd 5%. See Figure 5.12 legend for bacterial strain names.

5.5 Discussion

The functions of many unknown genes identified during the sequencing of entire genomes have been investigated using allelic mutation experiments (Aranda *et al*., 2010). In this study, mutants in the candidate virulence genes *nrps1*, *cfa7*, *abc*, *sim* and *pad* were generated in ICMP19477 by allelic exchange mutagenesis using the suicide vectors pK18mobsacB or pKNG101. The mutants were then used in pathogenicity assays, which identified a possible role for at least two of these genes in virulence of *P*. *carotovorum* subsp. *brasiliensis*.

The construction of the allelic exchange cassettes was enabled by the use of overlap PCR and the use of reliable suicide vectors. The advantage of using overlap PCR was that only four PCR reactions with gene specific primers were required to produce each allelic exchange cassette. This proved far less expensive and time consuming than using artificial syntheses of long gene sequences or the sequential cloning of the target flanking sequences and the antibiotic cassettes. Both pK18mobsacB and pKNG101sacB also proved useful vectors for the allelic exchange cassettes; plasmids containing the allelic exchange cassettes for all the targeted loci were successfully cloned into both suicide vectors. The pK18mobsacB plasmid is a derivative of pK18 isolated from *E*. *coli*, with modifications that include the broad host-range transfer machinery of plasmid RP4 (oriT), a multiple cloning site (MCS), and the kanamycin resistance and sucrose sensitive *sacB* genes (Schafer *et al*., 1994). The presence of RP4 (Datta & Hedges, 1971) allows the plasmid to be transferred into a wide range of Gram-negative and Gram-positive bacteria to generate deletion mutants (Okamoto *et al*., 2010; Rosloniec *et al*., 2009; van der Geize *et al*., 2001; Yang *et al*., 2007) as the ColE1 origin of replication of this plasmid is functional only in enteric bacteria and cannot replicate in other bacterial types (Parish & Stoker, 2000). The pKNG101 plasmid also acts as a suicide vector by carrying the *pir*-minus origin of replication (R6K) (Kaniga *et al*., 1991). Suicide vectors carrying the R6K origin of replication can only be maintained in host strains genetically engineered to express the π protein and cannot replicate in strains lacking the essential replicase-encoding *pir* gene. Since, the majority of Gram-negative bacteria do not possess the *pir* gene, R6K pir-dependent suicide vectors are widely used for gene

replacement mutagenesis (Biswas *et al*., 1993; Kaniga *et al*., 1991; Skorupski & Taylor, 1996). The limitation of using R6K based vectors, however, is that cloning of knockout gene constructs can be tedious because of the relatively low number of plasmid copies, which impedes the preparation of large amounts of pure, digestible plasmid DNA (Sarker & Cornelis, 1997).

Attempts were made to create double crossover mutants using both the mutagenic pK18mobsacB and mutagenic pKNG101 plasmids. Despite repeated conjugation and electroporation using the mutagenic plasmids, however, only one double crossover knockout mutant of ICMP19477 was generated. This double crossover mutant, ICMP19477Δ*nrps1*[2], was produced by counter-selection for the pKNG101 plasmid. All other mutants created in this study were generated by a single crossover of the mutagenic pK18mobsacB plasmids. The backbone of pK18mobsacB could not be removed from the single crossover mutants by counterselection on sucrose to generate double crossover mutants. The difficulty in generating mutants of ICMP19477 was unexpected as transformation/conjugation of other SRE has proven successful in previous functional studies (Nykyri *et al*., 2012; Vercoe *et al*., 2013).

Type I restriction-modification (R-M) systems are common among bacteria and are thought to protect bacteria from DNA acquisition, including invading foreign DNA such as bacteriophages and plasmids (Garlant *et al*., 2013; Tock & Dryden, 2005). Variation in plasmid transfer frequency, therefore, could reflect the presence of R-M systems in recalcitrant isolates (Accetto *et al*., 2005; Heurmann & Haas, 1998; Waschkau *et al*., 2008). Given the difficulty in generating mutants of ICMP19477, it was suggested that the poor transfer of plasmid DNA into this strain might be explained by the presence of one or more R-M systems. Annotation of the ICMP19477 genome identified the presence of two clusters of type I R-M systems in this strain. The first cluster encoded a classical methyl-specific restriction system (MSR) consisting of restriction-modification subunits, R (KCO_19417), S (KCO_19422) and M (KCO_19427). Classical R-M systems typically include an endonuclease and a DNA methylase and methylate DNA containing N6-methyladenine or 5-methylcytosine residues (Bickle & Kruger, 1993). A restriction endonuclease CDS (KCO_19437) was present on the same gene loci in the genome of ICMP19477. The

methylase modifies cellular DNA by methylating either adenine or cytosine residues within a specific recognition motif) to provide a defence against the corresponding endonuclease (Zotchev *et al*., 1995). Invading DNA from other bacterial lineages is detected as foreign DNA due to improper methylation and is cleaved by the endonuclease, whilst DNA originating from the same bacterial lineage is detected as self DNA and remains undigested. The first R-M cluster was encoded on a GI, PbN1_GI50, which was absent from other comparator strains. The second R-M cluster was encoded on a GI, PbN1_GI6, which consisted of three CDSs: restriction-modification subunit M (KCO_10640), N-6 DNA methylase (KCO_10645) and, restriction-modification subunit R (KCO_10650). This cluster was unique to ICMP19477 (being absent from the remainder of the strains used in the comparative genomic studies).

Studies by Furuta *et al*. (2010) and Nykyri *et al*. (2012) have shown that the R-M systems are acquired by HGT. It has also been proposed that genomic R-M systems may play a role in stabilizing the host chromosome (i.e., GIs acquired by the bacterium through HGT are not lost (Vasu & Nagaraja, 2013)). To verify whether these R-M systems were having an impact on plasmid transfer to ICMP19477, the mutagenic plasmids were passaged through the *E*. *coli* S17-lambda *pir* strain, to methylate the knockout gene constructs. *In vitro* methylation of the knockout gene constructs was carried out to overcome any degrading effect of the R-M systems in ICMP19477 on the plasmids being used for mutagenesis. Methylation of the plasmid constructs did not, however, have any effects on transformation frequencies. Perhaps ICMP19477 is recalcitrant to transformation for other reasons.

All of the mutations were targeted to genes within clusters on GIs or gene islets, which were likely (or known) to be expressed as operons. Therefore, given that DNA sequencing confirmed the insertion of the allelic exchange cassettes into their target sites, the virulence of the single crossover mutants as well as ICMP19477Δ*nrps1*[2] was compared with that of ICMP19477 in pathogenicity tests. Reductions in lesion lengths on potato stems were observed when potato plants were inoculated with either ICMP19477Δ*sim*[1] or ICMP19477Δ*pad*[1]. No differences were observed, however, when potato plants were inoculated with the *cfa7*, *nrps1* or *abc* mutants

(including the *nrps* double knockout mutant). Though the *nrps1* and *abc* mutants did not confer reduced virulence, their role in pathogenicity of ICMP19477 should not be underestimated as studies have shown their involvement in the production of antibiotics, siderophores or other secondary metabolites (Neilan *et al*., 1999). Since strains of *P*. *carotovorum* subsp. *brasiliensis* are known to inhibit the growth of *P*. *atrosepticum* species *in vitro*, additional experiments including growth inhibition assays, competition assays and toxin production assays would confirm if *nrps1* and/or *abc* are involved in the inhibition of other *Pectobacterium*. In Gram-negative bacteria, ABC transporters can function as importers or exporters of toxins, bacteriocins, proteases and lipases (Dassa & Bouige, 2001; Omori and Idei, 2003; Saurin *et al*., 1999).

Given that a mutation in one of the core *cfa* genes of *P*. *atrosepticum* SCRI1043 exhibited reduced virulence in potato plants (Bell *et al*., 2004), it was anticipated that the *cfa7* mutant of ICMP19477 would show similar results. The virulence data and the absence of the *cfa* cluster in the genome of the previously sequenced *P*. *carotovorum* subsp. *brasiliensis* strain ATCC BAA 417, suggests that the *cfa*-like biosynthetic cluster in ICMP19477 is not essential for plant pathogenicity in this strain. Instead, this cluster may be important for the pathogen on more resistant cultivars (than used in our assays), where suppression of the plant's disease resistance response will be important for successful colonisation.

ICMP19477Δ*sim*[1] and ICMP19477Δ*pad*[1] showed reduced virulence in both blackleg and soft rot assays. These data suggest that the loci encoding the *sim* and *pad* genes may important for virulence of the pathogen on potato. As described in Chapter 4, the *sim* gene is part a gene cluster predicted to encode for the capture and metabolism of specific sugars from the sink organs of the potato plant. This cluster is novel to the *P*. *carotovorum* subsp. *brasiliensis* and *P*. *atrosepticum* strains, which are the most prevalent strains associated with blackleg. Several other studies have shown the importance of a carbohydrate utilization locus (CUT) in the virulence of pathogenic bacterial species. For example, the sucrose CUT of *X. campestris* pv. *campestris* is required for full pathogenicity on Arabidopsis (Blanvallain *et al*., 2007). Likewise, inactivation of the *sim* gene of *E*. *amylovora* reduced dramatically the fire blight symptoms on apple seedlings

caused by this pathogen (Bogs & Geider, 2000). Hence, it is suggested that *P. carotovorum* subsp. *brasiliensis* and *P. atrosepticum* species can cause disease on plant stems because of their ability to scavenge more and different sugars from plant cells.

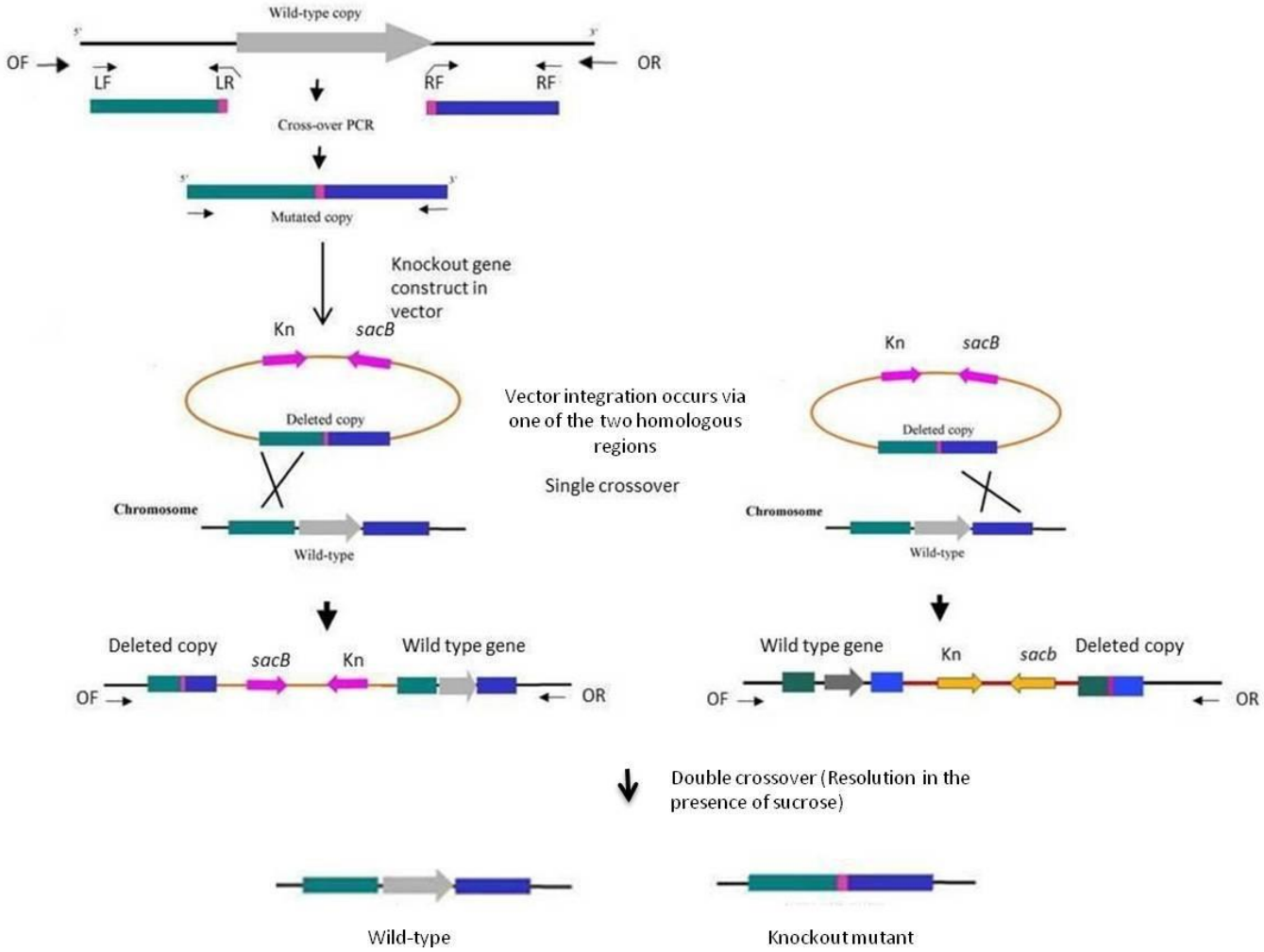
The *pad* gene is located on a gene islet, which encodes genes that are probably important for the degradation of phenolic compounds produced by potato under pathogen attack. This gene cluster was present in blackleg causing strains of *Pectobacterium* (*P. atrosepticum*, *P. carotovorum* subsp. *brasiliensis* and *P. wasabiae*), but absent from other comparator strains used in this study. As described in Chapter 4, phenolic acids are important components of plant cell walls which play an important role in cell defence (Bokern, 1991). In plant associated *A. tumefaciens*, phenolic acids are known to induce virulence gene expression (Lee *et al*., 1995) and strains of *Pseudomonas* and *Acinetobacter* are able to use these acids as the sole source of carbon for growth (Priefert *et al*., 1995; Segura *et al*., 1999). In *Saccharomyces cerevisiae*, the acivity of PAD has been speculated to confer a selective advantage during growth on plants, where the expression of PAD could constitute a stress response induced by phenolic acid (Clausen et al., 1994; Goodey & Tubb, 1982). The presence of *pad* in the blackleg causing strains of *Pectobacterium* (and in *P. wasabiae*, which causes stem diseases), suggests that these bacteria have evolved strategies to co-opt defence-related plant signaling molecules to enhance their resistance to toxic plant chemicals and consequently, to survive in the hostile host environment of the stem. It is hypothesized that PAD manipulates the plant defence system by degrading the phenolic compounds produced by the plants when attacked by pathogens.

Finally, although genome sequencing of ICMP19477 and the subsequent construction of mutants in candidate virulence genes identified several putative virulence factors in this pathogen, complementation assays will be required in the future to confirm that the phenotypes of the mutants were truly due to the inactivation of the genes targeted. Complementation assays involve the reintroduction of native genes into their respective mutants, in which the genes have previously been inactivated. Furthermore, given the difficulties in generating double crossovers in ICMP19477, other more amenable strains could be used for functional studies of virulence in

P. *carotovorum* subsp. *brasiliensis* or even isolates of other species that carry these genes (e.g. *P*. *atrosepticum* SCRI1043). Alternatively, the genes could be cloned into less pathogenic strains to examine whether they increase the virulence of the new host.

5.6 Appendices

Appendix A5.1 Schematic representation of the construction of knockout mutant via allelic mutagenesis



Appendix A5.2 Sequence confirmation of ICMP19477Δ*nrps1*[2] knockout mutant

Sequence of NRPSNF_2013-10-25_H5_0444.abi

CGAGGATACGTACTGATAGCGGCATTGGCGGCATGACGCCACTGCATCGTTGTGCCAGCGCCAGCG
GCGTTTGTTCATGCTCCAGAAGCGAGATGAGTTCATCATAGGTTGAGCGTACGACATCAATTGTGCT
GCGGTCGTTAAGGGAAATGCGCAGCGGCAGTGTGTTAATAAACATCCCCATTACTCTGTCTGCGCCA
TCCGTACCCTGTAAACGCCCAAGCAGGACGGAACCAAACACCACATCATCGCTGTTGCTGACATGGG
CTAACACCTGAGCCCAGGCGATATGGAATAGGACGCTGGGGCTGACGCCCAGTCGGTTTGCCCGGT
GGCGTATTGCTTGTGCTAATGCAGGCTCAAGATACTGCATGGTTTCGTCCACATGCTCGCCAGCCGT
TTTTATATCGATCAGACCAAAAGGCGCGGTTGGTGCCTCTACATCGGCCAACCGGGAGCGGAAATA
CGCTTCATGAGCGGACATGGGCACACTTAAAATCTGAGCGATAAAGTTACGGTAAGGCAGCGGCGT
TGGCAGATCATCATCATGGCCCTGCAAGATTGTGCGTATCTCGTCAATCATTAGGGCGAGCGTGATG
TGGTCACAGACCAGATGATGGAAGCCTAGCGCCAGCAGCCATTCCTGATTGGCCGCATCGTGAGCG
ATATCCGCGGTGAGCAACGGTGCCTGACTCAGATTGATACGGCGCGTATGCGAGCTGGTATGGGCG
CGTAGCTGACCGAGCACATCATCGTCCGAGTCGATCTCCGGCGTGTTTATCGGTAACTGTGCATGAC
GCCAGACCACCTGGACAGGCTGGCTAAGTCCCTGCCAACAGACGGCGGTGTGCAGGATGTCATGAC
GTTCGATAACGCGTTGTAGTGCGCTGAGGAAAGTATCAAGGCGTTCACGGCTATCAAAAGCAATCA
CGGTTTGCGACAGGTAGGTATCTCCCTGTTCCTGAAGCAGATAGTGGAACAAGATCCCTTTCTTGCA
GCGGCGCGAGCGGTAATGTCCTGCACATTGGCCGCGCCGCCGGAACGTCGCGACATCGCGTCATAT
CACGGCTGAGAAAGTGAATCAGCGGCAGCAGTCAGTGTATGGCTGTGCCAGTCCGTCATTCAGGTC
AGAGGAACGAACCCAAAGGT

Sequence of ChlR_seq_2013-11-06_B02_0459.abi

GCCCCTCTCTCTCATACACTCCGGCCCAAACCTGGAGGCGGAAATACGCTCATTGAAGGCGGACCAT
GGCCACCACCTAAAATCTGAAGCCGATAAAAGTTTACCGGTAAGGCAGCGCGTTGGGCAGATCATC
ATCATGGCCCTGCAAAGATTGTGCGTATCTCGTCAATCATTAGGGCGAGCGTGATGTGGTCACAGAC
CAGATGATGGAAGCCTAGCGCCAGCAGCCATTCCTGATTGGCCGCATCGTGAGCGATATCCGCGGT
GAGCAACGGTGCCTGACTCAGATTGATACGGCGCGTATGCGAGCTGGTATGGGCGCGTAGCTGACC
GAGCACATCATCGTCCGAGTCGATCTCCAGCGTGTTTATCGGTAACTGTGCATGACGCCAGACCACC
TGGACAGGCTGGCTAAGTCCCTGCCAACAGACGGCGGTGTGCAGGATGTCATGACGTTCGATAACG
CGTTGTAGTGCGCTGAGGAAAGTATCAAGGCGTTCACGGCTATCAAAAGCAATCACGGTTTGCGAC
AGGTAGGTATCTCCCTGTTCCTGAAGCAGATAGTGGAACAAGATCCCTTCTTGCAGCGGCGCGAGC
GGGTAAATGTCCTGCACATTGGCCGCGCCGCCGGGAACGGTCGCGACAATCGCGTCAATATCACGC
TGAGAAAGTGAAATCAGCGGCAGCAGTTCCGGTGTAATGGCTGTGCAGTCCGGTGAAATCAGGCCA
GAAGGAACGACCAGAGGTTGTTTTGAGCGATTGGTTAGTATGGCCTGTGCCATATCGCAGAGCACA
GGAGCGGAGAACACGCTACGAACGTCCAGTGTCTGGCCGATATTGCGTAACCGCTCAATCAGGCTG
ACTGCCATCAGGGAATGGCCACCAAGCTCGAAGAACTGGTCATGGCGACCAACGCGGTCCAGCCCC
AGGAGATCTTGCCAGATCTGGGCCAATGCGGTTTCGAGTTCTCCTTGCGGTGCGACATAGCCGCGTG
TTGCGACGGCTGATTGGTCCGGCGCGGGTAGCGCTTTGCGGTCAATCTTGCCGTTTGGCGTAAGTG
GGAAGGCATCGAGCGTGACAAAAGCACTGGGGACCATGTACTCAGCCAACTGTTCGGCCAACTGCT
GACGCAGATCGGCGTGTGACGGAAGATCACTTCGCAGAATAAATAAATCCTGGTGTCCTGTGAACC
GAGAG

Sequence of ChlR_2013-10-23_E03_0441.abi

TTTCCTCCCTTGCGTTGCGAACATAGCGCTGTGCGACGACTGATTGTCGCGCGGTAAGCGCTTTGCG
TCAATCTGCCGTTGCGTAAAGTGGGAAAGGCATCGAGCGTGACAAAAGCACTGGGGACCATGTACT
CAGCCAACTGTTCGGCCAACTGCTGACGCAGATCGGCGTGTGACGGAAGATCACTTCGCAGAATAA
ATAAATCCTGGTGTCCCTGTTGATACCGGGAAGCCCTGGGCCAACTTTTGGCGAAAATGAGACGTTG
ATCGGCACGTAAGAGGTTCCAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTTTGA
GTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGATATACCACCGT
TGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATA
ACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTA
TCCGGCCTTTATTCACATTCTTGCCCGCCTGATGGATGCTCATCCGGAATTCCGTATGGCAATGAAAG
ACGGTGAGCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAAC
GTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATG
TGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCA
GCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCC
CGTTTTCACCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTCAGGTT
CATCATGCCGTTTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATG
AGTGGCAGGGCGGGGCGTAATTTTTTTAAGGCAGTTATTGGTGCCCTTAAACGCCTGGTTGCTACGC
CTGAATAAGTGATAATAAGCGGATGAATGGCAGAAATTCGAAAGCAAATCGACCCGGTCGTCGGTC
AGTTTGGA

Sequence of ChlF_2013-10-23_E02_0440.abi

CATCTGGTGTCCTGTTGATACCGGGAAGCCCTGGGCCAACTTTTGGCGAAAATGAGACGTTGATCG
GCACGTAAGAGGTTCCAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTTTGAGTTAT
CGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGATATACCACCGTTGATA
TATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAG
ACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGG
CCTTTATTCACATTCTTGCCCGCCTGATGGATGCTCATCCGGAATTCCGTATGGCAATGAAAGACGGT
GAGCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTC
ATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCG
TGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAA
TCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTT
TCACCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTCAGGTTCATC
ATGCCGTTTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTG
GCAGGGCGGGGCGTAATTTTTTTAAGGCAGTTATTGGTGCCCTTAAACGCCTGGTTGCTACGCCTGA
ATAAGTGATAATAAGCGGATGAATGGCAGAAATTCGAAAGCAAATTCGACCCCGGTCGTCGGTTCA
GGGCAGGGTCGTTAAATAGCCGCTTATGTCTATTGCTGGTGGCAGTCAGTTCGCGGAAGGTGGCCG
TGAACCGGTAATCTCGGTGATGCCGTACATATTAATCAAACCGGGTGTGTTCAACCGGGGTTATCCG
CCTATCCCACGGGGCCAGCATATGTAACCTCCAGCGCCTCCCCCCGCCGAGATGAATGGCAGCGCAG
CGTGTTGTCAGCGTCAATCCCGGTGCGGCATCACCTTGGACGCGGAAAGCGCTC

Sequence of ChlF_seq_2013-11-06_A02_0459.abi

GCGTATGCAGACGTTATAGCCGCTTATGTCTATTGCTGGTGGCAGTCAGTTCGCGGAAGGTGGCGT
GAACGGTAATCTCGGTGATGCCGTACATATTAATCAAACGGGTGTGTTCAACCGGGTTATCCGCTAT
CCACGGGGCCAGCATATGTAACTCCAGCGCCTCCCCGCCGAAGATGATGCAGCGCAGCGTGTGGTC

AGCGTCATCCCGTGCGGCAATCAACTGGCGGAACGCCCCCGGCGTCTGGTTAAGTACCGTTACGTG
CTCCCGACAGAGTAGGGAATAGAACTCCTGAGGTGAACGAGCACACAAGGCTGGAACAACAACCA
ACCTGCCGCCGTAGGTGAGCGCTCCCCATATTTCCCAGACGGAGAAGTCGAACGCAAAGGAGTGGA
ACTGGGTCCAGACGTCGTGGTTGTCAAAGTGGAAATCGTCCTGCGTTGCGTCGAGCAGGCGGGTGA
CACTGGCATGCTCGACCATAACGCCTTTGGGTTTACCGGTGGAGCCGGAGGTGTAGATAACATAGG
CGAGATGACCCGGCGTGACGCCCAGTGCGTTCGCGTCGGGGTTGTGGTCAGGCTGCGCATCAAAGA
CGTCGTGTGCGTCAAGCAGTACGGTTGTCAATGCGGTGTCCGGCAGGACAGCGGTTAGTGCCGACT
GGGTGAGTAGCGCCACGGGAGTGGCGTCCTCCAACATATAGGCAAGGCGCTCGGATGGGTAGGCG
GAGTCTAACGGAAGATAGGCGGCTCCAGCTTTAAGGATAGCCAGAAGGCCAATAATCATCTCAGGG
CTGCGTTCGACGCAAATGGCGACGCGGTCATCGGGTTTAACACCGAGGTCCATGAGGTGATGGGCG
AGCTGGTTGGCGCGACGGTTGAGCGTCTGATAGCTGAGCGAACGTGATTCGAAGACGACGGCAGT
GGCGTCCGGTGAACGTTGCACCTGCGCTTCGAAGCGCGCCTGAATCAGCGCGGGAGCCGGTAGCTC
GGCGTTGGCTGGGTTAAAGTCGACCAGCACTTGCTGGCGCTCGTCTGCCGACAGCATTGGCACGGA
CCACAGAGGTTGTTCCACGTCGGTGGCAATAGCTGTGAGTACATTTACTACGTAGGCAGCACGCGCT
CATGGCTTTCCGATCGAAGATCGACCGCCAATATTTCCAGGTGAGTCAGGTTCGGAAGCA

Sequence of NRPSNR_2013-10-25_G05_0444.abi

GTCCCACAGAACGTTTCCGTGGTTGTCCCAAATGGAATCCGTCTGCGTGCGTCGAGCAAGCGGGTG
ACACTTGCATGCTCCGACCATACGCCCTTGGTTTACCGGTGGGAGCTGGAGGTGTAGATACATAGGC
GAGATGACCCGGCCGTGACGCCCAGTGCGTTCGCGTCGGGGTTGTGGTCAGGCTGCGCATCAAAGA
CGTCGTGTGCGTCAAGCAGTACGGTTGTCAATGCGGTGTCCGGCAGGACAGCGGTTAGTGCCGACT
GGGTGAGTAGCGCCACGGGAGTGGCGTCCTCCAACATATAGGCAAGGCGCTCGGATGGGTAGGCG
GAGTCTAACGGAAGATAGGCGGCTCCAGCTTTAAGGATAGCCAGAAGGCCAATAATCATCTCAGGG
CTGCGTTCGACGCAAATGGCGACGCGGTCATCGGGTTTAACACCGAGGTCCATGAGGTGATGGGCG
AGCTGGTTGGCGCGACGGTTGAGCGTCTGATAGCTGAGCGAACGTGATTCGAAGACGACGGCAGT
GGCGTCCGGTGAACGTTGCACCTGCGCTTCGAAGCGCGCCTGAATCAGCGCGGGAGCCGGTAGCTC
GGCGTTGGCTGGGTTAAAGTCGACCAGCACTTGCTGGCGCTCGTCTGCCGACAGCATTGGCACGGA
CCACAGAGGTTGTTCCACGTCGGTGGCAATAGCTGTGAGTACATTTACTACGTAGCCAGCCACGCGC
TCAATGGTTTCCCGATCGAAAAGATCGACGGCATATTCCAGTGAGCCAGTCAGACCATTATCTGTCT
CGCTAATCGATAACATCAGGTCAAACTGAGCGCTGTGTTTAGGTTGTTCAATAAAAGAGACGTCCAG
ATCGGGCAGCGTTAATCTCTGCGACGACGTATTATTCAGGGCCAGCATCACCTGAAAAATGGGACT
GTAGCTGAGGCTACGTGCGGGTTGCAGCACTTCCACGACCTGATCGAATGGGAGATCCTGATGGGC
ATAAGCGGAAAGTGACTGTTCCTTAACCTGAGCCAGCAGCGCTTCCAGCGTCTTACACTGGTTCAGC
TCAATACGCAGTGCCAGCGTATTAACGAAGAAGCCAATCAGCCCTTCCAGTTCATGGCGGGTACGAT
TGGCGATCGGCGTACCAATGACGATGTCGTCTTGACCACTGAGACGAGACAGCACAGTTCAAT

Appendix A5.3 Sequence confirmation of ICMP19477Δ*nrps1*[1] knockout mutant

Sequence of ChlF_2013-10-23_E02_0440.abi

CATCTGGTGTCCTGTTGATACCGGGAAGCCCTGGGCCAACTTTTGGCGAAAATGAGACGTTGATCG
GCACGTAAGAGGTTCCAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTTTGAGTTAT
CGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGATATACCACCGTTGATA
TATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAG
ACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGG
CCTTTATTCACATTCTTGCCCGCCTGATGGATGCTCATCCGGAATTCCGTATGGCAATGAAAGACGGT
GAGCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTC
ATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCG
TGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAA
TCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTT
TCACCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTCAGGTTCATC
ATGCCGTTTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTG
GCAGGGCGGGGCGTAATTTTTTTAAGGCAGTTATTGGTGCCCTTAAACGCCTGGTTGCTACGCCTGA
ATAAGTGATAATAAGCGGATGAATGGCAGAAATTCGAAAGCAAATTCGACCCCGGTCGTCGGTTCA
GGGCAGGGTCGTTAAATAGCCGCTTATGTCTATTGCTGGTGGCAGTCAGTTCGCGGAAGGTGGCCG
TGAACCGGTAATCTCGGTGATGCCGTACATATTAATCAAACCGGGTGTGTTCAACCGGGGTTATCCG
CCTATCCCACGGGGCCAGCATATGTAACCTCCAGCGCCTCCCCCCGCCGAGATGAATGGCAGCGCAG
CGTGTTGTCAGCGTCAATCCCGGTGCGGCATCACCTTGGACGCGGAAAGCGCTC

Sequence of NRPSNR_2013-10-25_G05_0444.abi

GTCCCACAGAACGTTTCCGTGGTTGTCCCAAATGGAATCCGTCTGCGTGCGTCGAGCAAGCGGGTG
ACACTTGCATGCTCCGACCATACGCCCTTGGTTTACCGGTGGGAGCTGGAGGTGTAGATACATAGGC
GAGATGACCCGGCCGTGACGCCCAGTGCGTTCGCGTCGGGGTTGTGGTCAGGCTGCGCATCAAAGA
CGTCGTGTGCGTCAAGCAGTACGGTTGTCAATGCGGTGTCCGGCAGGACAGCGGTTAGTGCCGACT
GGGTGAGTAGCGCCACGGGAGTGGCGTCCTCCAACATATAGGCAAGGCGCTCGGATGGGTAGGCG
GAGTCTAACGGAAGATAGGCGGCTCCAGCTTTAAGGATAGCCAGAAGGCCAATAATCATCTCAGGG
CTGCGTTCGACGCAAATGGCGACGCGGTCATCGGGTTTAACACCGAGGTCCATGAGGTGATGGGCG
AGCTGGTTGGCGCGACGGTTGAGCGTCTGATAGCTGAGCGAACGTGATTCGAAGACGACGGCAGT
GGCGTCCGGTGAACGTTGCACCTGCGCTTCGAAGCGCGCCTGAATCAGCGCGGGAGCCGGTAGCTC
GGCGTTGGCTGGGTTAAAGTCGACCAGCACTTGCTGGCGCTCGTCTGCCGACAGCATTGGCACGGA
CCACAGAGGTTGTTCCACGTCGGTGGCAATAGCTGTGAGTACATTTACTACGTAGCCAGCCACGCGC
TCAATGGTTTCCCGATCGAAAAGATCGACGGCATATTCCAGTGAGCCAGTCAGACCATTATCTGTCT
CGCTAATCGATAACATCAGGTCAAACTGAGCGCTGTGTTTAGGTTGTTCAATAAAAGAGACGTCCAG
ATCGGGCAGCGTTAATCTCTGCGACGACGTATTATTCAGGGCCAGCATCACCTGAAAAATGGGACT
GTAGCTGAGGCTACGTGCGGGTTGCAGCACTTCCACGACCTGATCGAATGGGAGATCCTGATGGGC
ATAAGCGGAAAGTGACTGTTCCTTAACCTGAGCCAGCAGCGCTTCCAGCGTCTTACACTGGTTCAGC
TCAATACGCAGTGCCAGCGTATTAACGAAGAAGCCAATCAGCCCTTCCAGTTCATGGCGGGTACGAT
TGGCGATCGGCGTACCAATGACGATGTCGTCTTGACCACTGAGACGAGACAGCACAGTTCAAT

Appendix A5.4 Sequence confirmation of ICMP19477Δ*pad*[1] knockout mutant

Sequence of ChlF_2013-10-23_G02_0440.abi

TCATCCTGACCGACGACCGGGTCGATTTGCTTTCGAATTTCTGCCATTCATCCGCTTATTATCACTTAT
TCAGGCGTAGCAACCAGGCGTTTAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGC
CACTCATCGCAGTACTGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCAT
GATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAA
AACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGG
ATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAA
CACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCG
ATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACCAG
CTCACCGTCTTTCATTGCCATACGGAATTCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATA
AAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAAC
GGTCTGGTTATAGGTACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGG
GATATATCAACGGTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTC
GATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGT
GCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAGCAGGCACACCAGG
ATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGCGCTTCATCGCAGCCAGTCTTCGGTCAGCTACC
AGCTTTGCTCTGCTGCAAGAGCGATTGGGCATCGCGCTACTCACGATCCACGGGACGCAAGCGAGC
TCACGCCAGAGGGCGAAAATCTACTCGCGCAGGTGCGGGCCGTATACGGTCCATTCATGGCGCTGG
AAGTCCCGGATCTTA

Sequence of ChlF_seq_2013-11-06_G01_0459.abi

TTAATAGCCGCTTATGTCTATTGCTGGTCGCCTAGCGTGAGCATCAGTCGGTTGGCCCATGAGAAGA
AGGCGGCTGACTGCACCACATCGATAATCTCTAGCGTATCCAATCCCTGTTCTTGCAGCTGTTTCAGG
TCGTTTGCGTTGACCTGCGCGGGCGTGGCGGAAAGGCGTGCCGAGAAATCGATAATCGCCTGCCAG
CGCGGGCTTTGGCCGATGCTCAAATCCCCGCCGGGGACGACGTCCAGCAGCCGCTGCACATCGCTG
TCCTGTTTGGAGAGCTGACTGGCTTTACGCGCGTGTACCGAGGCGCAATAGATGCAGCCGTTGACCT
TGCTGGTGACGGTGGCGATCAGTTCGCGCTCCTTGCGCGGCAGGCCGCCTGCGGTATAAAAAATCC
CTTTATCCGTCAGCGTGCGCTGCTCCAGTACCGGAAGGTTGCGGCCTAGCAAGCGGAAATAGTCGG
AATCGGTGTGGCCGAAGCGCGCCAGAATCGCCTGCTCATCCGCATTGAAGGCGGCCAGCGGTTTGG
GGGCGATCCACGGTTCCCAGCCCAGCTCTGCCTGAGTAAACGCCTGCGGTGCAGACTTGCCGCTGT
GCGTCTGCGATTGGGTATGCCGCTGACGTGCTGCAACAGCCTGCTGATGATGATGGTTGATACCATG
ATCAGCAATCATGCGATAATCGCGCAGCAGTCGGCTGTTATAATTCTTCAATGCAGTGATCAGCTAT
GGCATCACGATGACATCTACCGACCACCCGGCCTGCTGCAGCGCGTTGACGTGGTATGATGATGCTT
GTTGCAGCGTTTTCTTC

Sequence of PADSR_2013-10-23_H01_0440.abi

CTTCCTGTTGACCTGGCGAGCTCCCAGGTCCAATGCCGGTGCGACTGAGTAGGCCATCCAGGCTCCA
AGTGATCGAAGGATCAGCGCATGTGCGGATCCGATGGAATGCACCTGGAATGTTACATCCCGTCGC
CCGTTCTAGTGATGCTGTCCGTCCTTAAGCGACGCGGTGACATCCTGTGCGGTGCCGACATGGCTGT
CGAAGGAGGCGATCAGTGCTTCTATCGAGTCGTGAGGGATCGCGCGGAACGTACCGGAGCGGGCG
GCAGAACGATTAAGCCCTTTCTCCGCCAGACTTAGCGCCAGTTGACGATCGTCAGCCACGAAGACGC
TACGCGAGCTGAGGATACGCGGTGCGACGCCAGCGGGCAGCGCAGCCAGATAGGCGTCGATCATC

GGGTTTTGCAAGTCAGCAAGCGTGGCATCTGGGAAATGTTCTGGGCGTGGCTGGGTACGGGAAAG
CATCAGGCCATCGCCCGCTTTTCCCGCTCGTTCAGCACCCTCAATGGAAAATGTGGCCTGCCAGACG
CGCTTATCTAAATGTGGGGCAGCAGGATATAGGCGATTGCCGTCTTCGCTCAATGCGTCTCCCCGCC
ATGCGGCGCGCAGCTTCTCCAGATAGCGCCCGAGGATCTGCCCGCGCTGTGCGCTGTCGTGACCGA
ACGCGGCAAATGAGGACGGCGTGCCGCCGGAACCCACGCCGACTTCCAGCCTGCCGTTGCTGAGTA
AATCGAGCACGGCGGTGTCTTCCGCCACGCGCAGCGGTTCTTCCATTGGCAGCGTAATTACGCCAGT
GCCGAGTTGAATGCGTTGAGTGCGTGCCGCGACCAGCGCCAGAAACACCAGTGGTGACGGTAACCC
ACCTTCATCAGCGTGAAAGTGGTGCTGTGCGACCCAGGCGCTATCGAATCCTAACTGTTCTGCTTTG
ACGATCTGTTCCGTCGCCAGCCGATAGAGCTGCTGGGCGGAAACGTCATCCAGCAACCGTGTGAAA
AATCCCAGACGTTTCGTTGCCATTTTACAGTCCTTGGTTTAGTGCGAATGTGGATTCAGTATGAATTG
GGGGTTGGTGTGAACGGGCAAAAGCCGGGTGTTGCTGCCCAGGGATAGCCTCGATGAGTTCACGG
GTGTAGCGATGCCCAGGCTGAGCGAAGATGGTCTACTA

Appendix A5.5 Sequence confirmation of ICMP19477Δ*cfa7*[1] knockout mutant

Sequence of ChlF_2013-10-23_C02_0440.abi

TTTCGGGGGTGTCCCTTGTTTGATACCGGGAAGCCCTTGGGCCCAACTTTTTGGCGAAAATGAAGAC
GTTGATCGGCACCTAATAGGTTCCAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTT
TGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGATATACCAC
CGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCT
ATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTT
TTATCCGACCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTCCGTATGGCAATGA
AAGACGGTGAGCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGA
AACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAG
ATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTC
TCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGC
CCCCGTTTTCACCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTCAG
GTTCATCATGCCGTTTTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCG
ATGAGTGGCAGGGCGGGGCGTAATTTTTTTTAAGGCAGTTATTGGTGCCCCTTAAAACGCCTGGTTG
CTACTCCCTGAATAAATGATAATAACCGGATGAATGGGCAGAAAATTCAAAGCAAATTCAACCCGGT
CGTCGGTCAGGGCAGGGTCGTTAAATAACCCGCTTATGTCTATTGCTGGTAATCGGGAACAATCCAG
CAGCCGCGCATCCGGTACTTGTCCTGTTCCGGCG

Sequence of ChlF_seq_2013-11-06_E02_0459.abi

GGGTAATGCAGATCGTTATAGCCGCTTATGTCTATTGCTGGTAATCGGTACGATCCAGCAGCCGCGC
GTCGGTACTGTTCTGTTCGGCGAACATGACGGTTTTAAGGGGTTGCGACAGCAGCGGATCGAGATG
CTGGCAGACGTCGTCGAACGTGTGAGCGAAAACCGGAGATTGACGGTATAGCCCTTGTCCCATTGT
CGGCCACTGTGCACCTTGCCCCGTAAAAAGGAAAGCGACGTTGCCTGACGGAGAACGGACGCCTTG
AATGATGCCGGGAGCGTAGCGATTGGCGGCCAATGCCTCCAGTGCGATAAGCGCATTGTTACTGTC
TTCTGCGGTAATCACGGCACGATAGTCAAAATGGGTGCGGGTGGTGGCTAGCGAATAGGCGATGTC
GGGCATCGAGCTTGCATTCTGAGTCAACCATGCGTGCAACTGCTTGGCCTGAGCGCACAGCCCGTCT
GGTGAGGCCGCAGACAACGGTAGCGCGATGTCAGGCACGGGGAAAGTCCTTGCAGGTGCGGCTTC
TTGCACCGGAGGCGAAATTGACGGAGCCTCTTCGAGAATAAGGTGAGCGTTGGTGCCGGAAAAGC
CAAATGATGATACCGCCGCGCGTTTGGGACGCCCGGTATCAGGCCAGTTACGTGCCGTGTTGAGTA
ACTTAATGTTGCCGCGTGACCAGTCTATTTTGCCGGACGGCGTCTCGGCATGTAGCGTTTTGGGCAA
GCGCCCGTGCCTGATAGCCATCACCATCTTGATGACGCTGGCGACGCCAGCGGCAAGCTGGGTATG
GCCGATATTGGATTTTAGTGAGCCAAGCCATAGCGGTTGATCGTCAGACCGATGTTTGCCATAGGTG
TTGAGCAGTGCATGTGCTTCAATGGGGTCGCCAAGCGTGGTGCCGGTGCCGTGCGCTTCTACGACA
TCAATATCGGAAAAGGTCAGCTCCGCATTATTCAGCGCCCTGACGGATGACCTGTTTCCTGCGCAGG
CCGTTTGGCGCAGTCAGCCCGTTGCTGGCACCATCCTGATTAAATTGCGCTCCTCTATCACGGCGAGT
ACGCTGTGACAGCTCGTTATTGCGTCGGAGAGTTTCTCAGGCCAAATAAGCCCCACGCTCAGAAACC
TCGGTTCGTTCGCAATGGATCTCAGGGAAG

Sequence of CFASR_2013-10-25_A06_0444.abi

CTGATAGTCACTGTCATTTCGAGTTCAAGCATTGCTATGCCAACTGCTGACTGAGGCCACAGGCAGT
CTGTGAGCGCAGCAAACGGTAGCGCGATGTCAGTCACGGGAAGTCTGCAAGTGCGCTTCTTGCACC
GAGGTGCAAATGACGGAGCGTCATCGAGAATATGCTGAGCGTTGGTGCCAGAAAAGCCAAATGAT
GATTTCCTGCAGCGGCGTATGAGACAGCTGCTATCAGACCACTAGAGTGCCGTGTTGAGTAAGTTG

AGTGTTGCATTCGTGACCAGTCTATTTCTCCGGCCGGAGTCTCGGCACGCAGCGTTTTGGACAAGCG
ACCGTGCGTGAGAGCCATCACCATATCGATGACGCTGGCGATGCCAGCGACAAGTTGGGTATGCCC
GATATTGGATTTTAGTGAGCCAAGCCATAGCGGTTGATAGTCAGACTGATGTCTGCCATAGGTGTTG
AGCAGTGCATGTGCTTCAATGGGGTCGCCAAGCGTGGTGCCGGTGCCGTGCGCTTGTACGACATCA
ATATCGGAAAAGGTCAGCTCCGCATTATTCAGCGCCTGACGGATGACCTGTTCCTGCGCAGGGCCGT
TTGGCGCAGTCAGCCCGTTGCTGGCACCATCCTGATTAATTGCGCTCCCTCTAATCACGGCGAGTAC
GCTGTGACCAGCTCGTAATGCGTCGGAGAGTTTCTCCAGCAAAATAAGCCCAACGCCTTCAGAAAAC
CCCGTTCCGTCGGCATGGTCAGAAAAGGCTTTACATCTTCCGTCTGCGGCAAGACCGCTTTGACGGG
TGAACTCGGCAAAGACCTCTGGAGAGGCCATAATCGTTGCTCCGCCTGCGATAGCAAGCGAGCACT
CGCCGCTCTTCAGTGAGCGAAGGGCTTGATGGATCGCAGTGAGAGACGCAGAGCAGGCTGTATCTA
CCGTGATAGCCGGACCGTTAAGTCCCAATGCGTAAGCGATGCGCCCGGAACTGATGCTGCTCAGGC
TGCCTTGTAAGCGATAACCGTCGGCGGCGGGGCAGCGCCGCTCGATATCGCTGAGATAGCCGTTGA
GCGTGGATCCAATAAAGACCCCGCCGTGTGTGTTTCATCGCGTGGCGGCGCGATTCCCCCACTGTCC
AGCGCGCCCCAAGTCCCATCCACCGTGCAGCCTTTCCACGCATAAATCGCCTCTCCTCAACGTAATGA
CAATCTATGAAGAATTT

Appendix A5.6 Sequence confirmation of ICMP19477Δ*sim*[1] knockout mutant

Sequence of ChlF_2013-11-06_B01_0459.abi

CTTCTGGTGTCCTGTTGATACCGGGAGCCCTGGGCCAACTTTTGGCGAAAATGAGACGTTGATCGGC
ACGTAAGAGGTTCCAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTTTGAGTTATCG
AGATTTTCAGGAGCTGATGAAAACAAAATGCACAGAAAAATCTCAGAAAATACCTCCGATGATATAT
CCCTATGGTGGCAGCATCAACCTTTTCAGGTTGTTCCGAATATTGCTCTGTGACCCTATAACCAGACC
GTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCT
TTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTCCGTATGGCAATGAAAGACGGTGA
GCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCAT
CGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTG
TTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATC
CCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTC
ACCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTCAGGTTCATCAT
GCCGTTTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGC
AGGGCGGGGCGTAATTTTTTTAAGGCAGTTATTGGTGCCCTTAAACGCCTGGTTGCTACGCCTGAAT
AAGTGATAATAAGCGGATGAATGGCAGAAATTCGAAAGCAAATTCGACCCGGTCGTCGGTTCAGGC
AGGGTCGTTAAATAGCCGCTTATGTCTATTGCTGGTCCAGAATTACGATCCAACGTAACGGAATACG
GTCGGCAAAGTAGATCTTTTCATCTGTTTCGCGAGAAGCCTGCTCTGCA

Sequence of SIMSR_2013-11-06_A01_0459.abi

TCTTTCGGCCAGGAATGTTTTCTCGCTGGCACACTCATACGCGGAAATCTGCGATATGCCGTGAGCG
AATCTTGTTCGGCAATAAATTTATAGATCACTTTTAGACCTCCATTTCCTCTTCCGATGCCATAAGATT
CACTGACATCATTAATTTTATTTCCTAATCATAGTATTACGTTTCTCTCACCGCCAACGCTCTCCTGCAC
AGAATAAACAGGTCTGCAAGACGGCCAAAAGCTATACCTCTTCCCGCGCGCCAACTTGTTGATTGAG
TTCACAAATCATCAATGCATGCTCTGTTCGAGTGAACCAACCTGCTTATTCTCACTACAGAACGCGTT
ATTTAGACCAAAAAAAATAAATTTCCCTACAGACATAAAAATGAAGGACACCATTATGCTCAGTCAA
ATTCAGCGGTTTGGTGGCGCAATGTTCACACCCGTGCTGCTATTTCCTTTTGCCGGTATCGTCGTCGG
GTTGGCGATCATGCTGTCGAATCCCATACTTGTCGGTGAGTTAGCCAACCCGAACAGTCTATTTTTCC
AGATAGTTTATGTGATTGAAGAAGGCGGATGGACCATCTTTCGCAATATGCCACTGATTTTTGCTTTC
TGTCTGCCGATTGGATTGGCAAATCACTCACCTGCCAGAGCGTGTCTTGCAGCACTGGTTTCCTACCT
GACCTTTAATTATCTGATTCAAGCTATGGCGACCCAATGGGGCGGCTATTTCAATATTGATTTCAGCG
CCGAGATCGGCGGCATCAGTGGATTAACCACGATCGCAGGCATCAAAACGCTGGATACCAGCATTA
TTGGCGCTATCGCCATCTCCGCCTTTATCACCGCGCTTCACAACCGTTATTACAATAAAACGCTGCCA
GTTTATCTCGGCATATTTCAAGGCACTGCTTTTGTGGTCATCATCAGCTTCCTGCTGATGCTACCAGCC
GCCTGGTTGACCTTACTGATCTGGCCTAAAATTCAGATGGGAATCGCATCCCTTCAGGGATTACTGCT
GGCATCGGGCGCTGCCGGAGTCTGGGTATACACGTTTCTGGAACGCATTTTAATCCCCACTGGACTG
CATCATTTCGTTTACGGACCGTTTATCTATGGTCCAGTCGCAGTAGAAGGCGGTATTCAGGTTAACTG
GATCGCCAAATCAGAG

Appendix A5.7 Sequence confirmation of ICMP19477Δ*abc*[1] knockout mutant

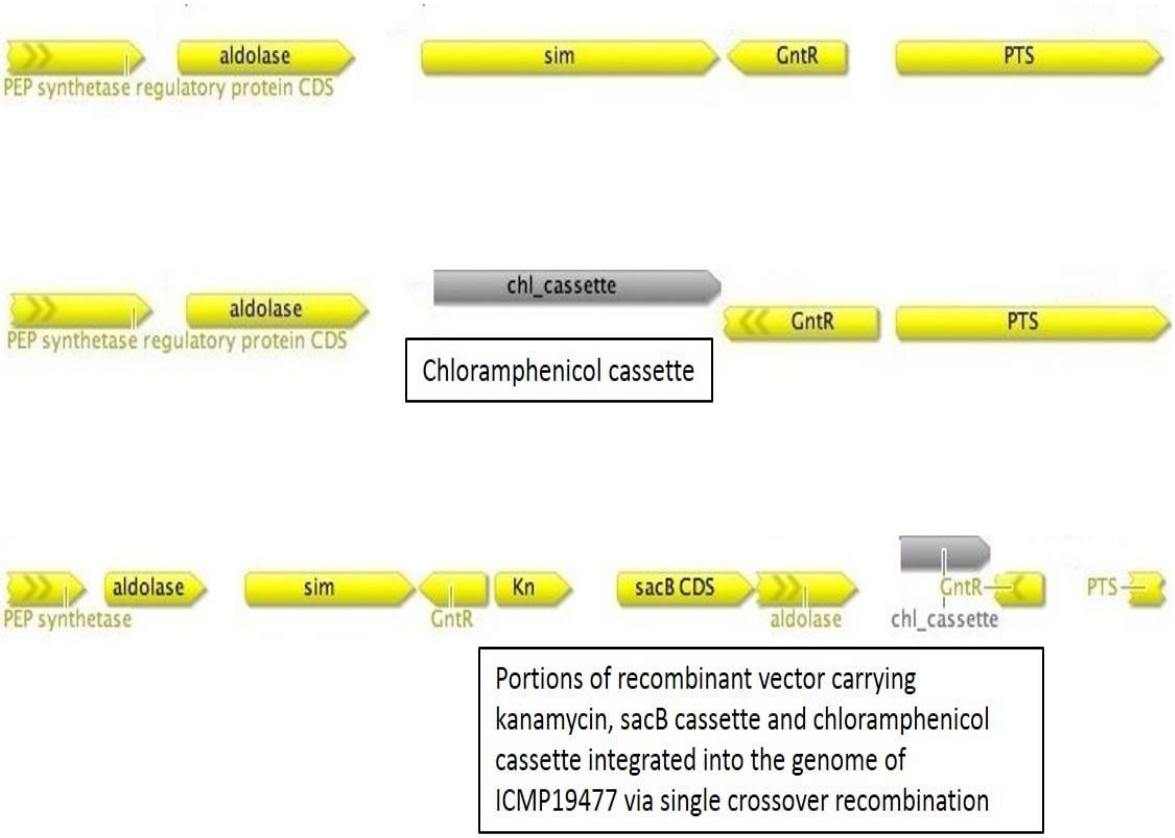
Sequence of ChlF_2013-11-06_D01_0459.abi

CTGTCTGGTGTCCTGTTGATACCGGGAGCCCTGGGCCAACTTTTGGCGAAAATGAGACGTTGATCGGCA
CGTAAGAGGTTCCAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTTTGAGTTATCGAGA
TTTTCAGGAGCTAAGGAGGCTAAAATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCCCAA
TGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGC
TGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATT
CTTGCCCGCCTGATGAATGCTCATCCGGAATTCCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGG
GATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAATA
CCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCC
TATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTTGAGTTTCACCAGTT
TTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACCATGGGCAAATATTATACGCAA
GGCGACAAGGTGCTGATGCCGCTGGCGATTCAGGTTCATCATGCCGTTTGTGATGGCTTCCATGTCGGC
AGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAATTTTTTTAAGGCAGT
TATTGGTGCCCTTAAACGCCTGGTTGCTACGCCTGAATAAGTGATAATAAGCGGATGAATGGCAGAAAT
TCGAAAGCAAATTCGACCCGGTCGTCGGTTCATGGCAGGTCGTTAAATAGCCGCTTATGTCTATTGCTGG
TGAATGCCATTGGTGACGCTGGCTATTGTGGCAATCGCCAGTATCCATCGTGAATCGAACAGGAGTGAG
AGTAAAAGCTGTATCGCTGGGATTTTTTTGGGTCGTCGACGTTTTCATACTATATGTCCACACGGTTAATT
GGACGTACACGATGACGAGATCGCGAACCAGCCCCTGAAATACG

Sequence of ABCR_2013-11-06_C01_0459.abi

TATCACACTTTTGAAGATGCAATCGGTTCATGTGGATGACCGGCATCCATACCCGGTCTAAACCTGGGTA
ACATGCTCTTCATTTCGCTGTCGTCGTCCGTCTTCATTCGGTATCCCCCTCTTTGGCACTGACCAGATAGA
TGTCACCTGCATAGTGAGTGCGTGGTGTATACACCGTGTTCAGATTGGGCTTGCATAACCTGAACAATTT
CCCTGCAGCAGCGATGTGGTCGTGTGCGTAGAGAAAATACCGGCTTTTACCAGAGCGCGGTGCAGCTGT
TTGATCTGTTCCTCTTGTGCCATATCCACAAACTGAGATCGCGTCAGCGCTAAAGGCTGCGTCAGAATCA
TGTTGTAGATGTCGATCAATTTCATGATCGTCTCAACGCGATCGATTGATTTTGGCGGGCACCCCTGAAT
GTCGGGAGCATCGGTATCCACCAGTATCAGGCTGGCAACCGATTCTCCCTGTGCCTGTAGCTGTAATGC
GAGTTCAAAGGCGACCCATCCGCCAAAGGAATGCCCGAGCAGGTGATAAGGCCCATGAGGTTGTTTTT
GCCGTATTCCTTGAAGATAGGCTCGTGCAGCCCCCTCTACGCTGATGTAAGGCGGTTGGTTGATTTCTGT
CAGGCCGCGTGCCTGTAAGGCGCATACGGGTTGTTGCGTCGGCAGCGCCAGCGCCAGTTCAATAAAGC
TGGAAGCATTGGCACCCGCCCCAGGGATACAAAAGAGCGGAGGCGCGCCCGGCGTACCGTTCTGGATA
ACGATGTTTGGCACATAGTCATTCGGTAACAGGAATTCGCTAAACGAGTCGGCAACCTGATTCAGCAAG
GGTGGCTGCATTATCGTGTAGTGAGTTCCACCGATAGGATGAAGCACGGAATCCTTACCGACGATGCCC
TGCCATTCTCCCCAGATATCGTCATCGGTGGTGTGTTCTGAACGGTAGAGGTGAACCGGCAGGATTGAA
GCCGGAGCCCGGTAATGTTCAGACAACAGCAAGATCGCTCTTTGGCTGTTGATGCGTAGAAGAATATCC
TCTTTTGTAATACCGGACGGAAGCCATTGACGTTCATGACAGCGCGCAATGATTAGCTCGATGTCTCCCA
GCGTGCGCAGATCCTCCAACTCGTGTTCATCGGCCGTCGTCATGTGATTACGTAAGTAGAAGAGGATGA
ACC

Appendix A5.8 Structure of single crossover knockout mutant, ICMP19477Δ*sim*[1]



wild type *sim* gene cluster

proposed structure of double crossover *sim* knockout mutant (unsuccessful)

structure of single crossover knockout mutant, ICMP19477Δ*sim*[1]

Chapter 6

Conclusion

6.1 Summary

The SRE are an economically important group of bacterial pathogens that cause blackleg of growing potato plants and soft rotting of potato tubers. Until recently, they have largely been characterized as 'brute force' broad host range pathogens (Toth & Birch, 2005), although genome sequencing of several SRE has identified a wide range a virulence factors that contribute to virulence (Bell *et al.*, 2004; Corbett *et al.*, 2005; del Pilar Marquez-Villavicencio *et al.*, 2011; Evans *et al.*, 2010; Kim *et al.*, 2009; Lee *et al.*, 2013; Liu *et al.*, 2008; Mattinen *et al.*, 2004; Mole *et al.*, 2010; Sjöblom *et al.*, 2008; Nykyri *et al*., 2012; Toth *et al.*, 2003; Urbany & Neuhaus, 2008) and phylogenetic analysis of the SRE has shown host specificity among strains and genera (Ma *et al*. 2007). These and other findings have revealed a more complex life style for the SRE than previously thought. This lifestyle includes: i) survival as an epiphyte or on alternative hosts, ii) invasion of the primary host and onset of a latent phase, iii) suppression of plant defence responses and colonization of different plant organs and, iv) acquisition of nutrients via the necrotrophic phase. Little is known about how SRE successfully adapt to these different life stages, nor how these are coordinated. Yet, this is likely to have a major impact on the success of the pathogen in colonizing potato and variability in their ability to cause blackleg and soft rotting of tubers. In particular, no evidence exists to indicate why some SRE variants have emerged to cause blackleg whilst others have not.

Pectobacterium atrosepticum and *P. carotovorum* subsp*. brasiliensis* are generally considered to be the main SRE to invade the xylem of potato, which ultimately results in blackleg. In this study, *P*. *carotovorum* subsp. *brasiliensis* was identified for the first time as a major cause of blackleg disease and soft rotting of potato in New Zealand (Chapter 2). Using a combination of molecular and phylogenetic analyses, including subspecies-specific PCR, 16S rRNA gene sequencing analysis

and MLSA of concatenated sequences of eight housekeeping genes, the taxonomic status of a highly aggressive strain ICMP19477 from potato was confirmed as *P. carotovorum* subsp. *brasiliensis*. The analyses also strongly differentiated ICMP19477 from *P. carotovorum* subsp. *carotovorum*. This in itself was a major breakthrough towards understanding the etiology of these diseases in New Zealand, as ICMP19477 and other closely related isolates have previously been classified as *P. carotovorum* subsp. *carotovorum* by Pitman *et al*. (2008). *Pectobacterium carotovorum* subsp. *carotovorum* has rarely been shown to cause blackleg. The more accurate identification of ICMP19477 now allows a much more insightful assessment of its potential hosts, its geographic and environmental ranges and mechanisms of infection.

The particularly aggressive nature of ICMP19477 makes it an attractive target for academic scrutiny of *P. carotovorum* subsp. *brasiliensis*. In particular, it is a good candidate to help understand the evolution of *P. carotovorum* subsp. *brasiliensis* and other SRE, the traits that are critical for causing severe disease, the mechanisms that underlie those traits, and ultimately the potential targets for engineering disease resistance or other methods of control. With this in mind, the aspiration for this study was to identify potential virulence factors encoded by ICMP19477, such that their importance and role could be examined. Of equal importance was the identification of potential MGEs involved in the acquisition of the virulence factors, in particular GIs. GIs have been shown to be central to the evolution of plant and animal pathogens alike by mediating transmission of virulence genes between strains of the same bacterium or between species (Davies & Davies, 2010; Dobrindt *et al*., 2004; Jackson et al., 2011). A comparative genome sequencing approach was taken to identify genes that might encode these virulence factors and the GIs that might be responsible for their acquisition. As little relevant genomic data was publically available for comparative genomics of *P. carotovorum subsp. brasiliensis* at the onset of this project, de novo sequencing, assembly and annotation of the IMP19477 genome was undertaken together with that of the type strain of *P. atrosepticum* (ICMP1526) (Chapter 3). Sequencing of the type strain was considered important given type strains form the basis for the designation of relevant species or subspecies. Few type strains had been sequenced for the SRE.

At the time the project was initiated, the sequencing of multiple genomes was still considered a major undertaking, both financially and in terms of the computational resources required to assemble and annotate a genome. 454 sequencing was the primary method of choice for effectively obtaining these genomes. During the thesis, other sequencing platforms became prevalent due to their decreased costs and higher associated coverage, and improvements in the length of their sequencing reads and the informatics tools available for assembly and annotation. As a result, the genomes were sequenced for ICMP5702 (type strain for *P. carotovorum* subsp. *carotovorum*), ICMP1526 (type strain for *P*. *atrosepticum*) and, UGC32 using Illumina sequencing. These sequences provided additional genetic information. Indeed, the core genome of the SRE defined using these SRE genomes and those retrieved from the public databases (14 in total) was very similar to the theoretical core genome predicted for the SRE using the core development plot analysis (Chapter 4).

Subsequent comparison of the core genome from the sequenced strains and those in the public databases, predicted to be inherited via vertical transmission, reiterated the identity of ICMP19477 as *P. carotovorum subsp. brasiliensis*. It also re-classified UGC32 as belonging to the *P. atrosepticum* lineage rather than that of *P. carotovorum*. The taxonomic re-classification of UGC32 was important given the initial identification of this strain as *P. carotovorum*, providing a probable explanation as to why this strain can cause blackleg. Taxonomic revision of UGC32 and several other strains based on their genomes (Nabhan *et al*., 2012; Nykyri *et al*., 2012) demonstrates the limitations of previous tools such as biochemical assays, RFLP and even MLSA, and the advantages of using a genome-based process for genetically differentiating bacteria.

The division of *Pectobacterium* into two distinct lineages as part of the analysis of the core genome revealed that strains within each lineage have the capacity to cause blackleg while others do not. Thus, the capacity to cause blackleg probably evolved by the acquisition of genes by HGT as opposed to vertical, generational evolution. Comparative genomics using the four genomes sequenced in this study, and another 10 genomes that were available at the time, identified the accessory genome of the SRE. Many of the genes in the accessory genome were ‘cluttered’ at

distinct locations in the genome, indicative of HGT of GIs or bacteriophages. The evolutionary advantages of acquiring GIs are obvious. Regarded as genetic 'building blocks', these GIs when horizontally transferred from one organism to the next can confer a complete phenotypic trait to the acceptor enabling, for instance, adaptation to a novel ecological niche. Comparative genomic data here identified over 60 GIs and 5 smaller gene islets in ICMP19477 alone, potentially involved in the virulence of this bacterium. These islands and islets constituted about 37.4% of the genome (Chapter 4), a percentage that is considered high when compared to other species. For example, HGT of foreign DNA accounts for about 7% of *R*. *solanacearum* genome, and at least 10% of *P*. *aeruginosa* genomes (Fall *et al*., 2007; Kung *et al*., 2010; Shen *et al*., 2006). The average proportion of horizontally transferred genes per genome (in a comparative study using 116 prokaryotic genomes) was ~ 12% of all ORFs, ranging from 0.5% to 25% depending on prokaryotic lineage (Nakamura *et al*., 2004). The high proportion of the genome defined as 'accessory' suggests that the genomes of *Pectobacterium* isolates are highly divergent, and probably have easy access to a common pool of virulence factors that enables them to evolve rapidly to their diverse environment. It also suggests GIs are likely to play an important role in many adaptive traits in *P. carotovorum subsp. brasiliensis*.

Although many GIs and islets were identified that encoded known virulence factors, a definitive set of islands (and virulence genes) responsible for blackleg disease or contributing to the aggressiveness of *P. carotovorum subsp. brasiliensis* was not identified in this study. This was because many GIs and their associated virulence genes were present in ICMP19477, but were either not in other blackleg causing strains or were present in all blackleg causing strains but were also detected in the genomes of one or more non-blackleg strains. For example, a *cfa* biosynthetic cluster was found on a GI PbN1_GI15, in the genome of ICMP19477. The *cfa* cluster is also encoded by HAI2 in *P*. *atrosepticum* SCRI1043. This suggested CFA might be important for full expression of blackleg, yet the absence of the *cfa* cluster from the genome of *P*. *carotovorum* subsp. *brasiliensis* strain ATCC BAA 417 indicated that it was unlikely to be essential for pathogenicity or responsible for eliciting blackleg.

A second cluster of genes (on PbN1_GI24 in ICMP19477), encoding for an NRP with similarity to the phytotoxin syringomycin was identified exclusively in the genomes of *P. atrosepticum* and *P. carotovorum* subsp*. brasiliensis* by comparative genomics. On closer inspection, however, the two NRPS genes and the ABC transporter gene in each cluster showed significant nucleotide divergence from the other cluster (< 85% similarity). Furthermore, the GIs (PbN1_GI24, HAI6) encoding these clusters were also divergent in the two species/subspecies; the genetic structure, organization and nucleotide sequences of the coding domains showing significant variation. This led to the suggestion that, although these NRPS clusters might be important for *P. atrosepticum* and *P. carotovorum* subsp*. brasiliensis*, they were unlikely to produce identical products. This placed doubt on their possible role in blackleg development and suggested they may have independent functions as antimicrobials instead.

Thirdly, *pad*, a gene encoding a phenolic acid decarboxylase, was identified on a gene islet in all *P*. *carotovorum* subsp. *brasiliensis*, *P. atrosepticum* strains including UGC32, *P*. *wasabiae* and *Dickeya* strains used in this study but was absent in *P*. *carotovorum* subsp. *carotovorum* strains ICMP5702, WPP14 and PC1. This suggested that *pad* is not essential for the survival of *P*. *carotovorum* subsp. *carotovorum* species or SRE. Indeed, it is possible that the *pad* gene has been lost during the evolution of *P*. *carotovorum* subsp. *carotovorum* species rather than gained by the other species and subspecies of SRE. Since, *P*. *carotovorum* subsp. *carotovorum* is considered to be broad host-range pathogen, the cause of gene loss might be related to the change of living niches, where the existence of such genes in the host genome confers a fitness cost. The function of PAD in Gram-negative bacteria is unknown. The presence of *pad* in only a subset of SRE suggests, however, that these bacteria have evolved strategies to co-opt defence-related plant signaling molecules to enhance their resistance to toxic plant chemicals.

The only candidate virulence genes encoded exclusively on a GI or gene islet by all strains of *P. atrosepticum* and *P. carotovorum* subsp*. brasiliensis*, were those in the *sim* cluster. The *sim* cluster encodes four CDSs: 6-phospho-alpha-glucosidase, PTS system alpha-glucoside-specific transporter subunit IIBC, a transcriptional regulator and SIM. SIM catalyzes the isomerization of

sucrose into isomaltulose, trehalulose and other glucose and fructose byproducts (Salvucci, 2003). These sucrose isomers cannot be metabolized by plant cells and are therefore, likely to be advantageous to the infecting pathogen (Bornke *et al*., 2001). Studies have also shown that the sugar:PTS system is a key factor in the pathogenicity of *Spiroplasma citri*, a plant-pathogenic mollicute (Andre *et al*., 2003). Furthermore, sucrose isomerase mutants of *E*. *amylovora* did not produce significant fire blight symptoms on apple seedlings (Bogs & Geider, 2000). The presence of islets carrying the sucrose isomerase gene in *P*. *carotovorum* subsp. *brasiliensis* and *P*. *atrosepticum* suggests that these species are likely to metabolize more and different sugars from plant cells than other SRE and that these strains may be better adapted to capture the limited sugars available in plant stems.

Given that many of the virulence genes were either present in both blackleg causing strains and non-blackleg causing strains or were only carried by some blackleg-causing strains, it was hypothesized that a discrete set of virulence determinants does not enable SRE to elicit this disease. Instead, it seems that multiple virulence genes have been acquired from the pool of genetic information available in the environment inhabited by SRE. The cumulative acquisition of these virulence factors, likely by HGT, enables virulence to constantly evolve until the pathogen is able to overcome the plant to cause disease. Particular genes or gene clusters may still provide significant benefits for a pathogen of potato, however. For example, the presence of the *cfa* genes on a different island in *P. atrosepticum* and *P. carotovorum* subsp*. brasiliensis* suggests the independent acquisition of these virulence genes. Another bacterial pathogen of potato, *S. scabies* has also acquired the *cfa* genes suggesting this cluster must have some role to play in infection of this host.

The idea that a plant pathogen interacts solely with its host plant is slowly being overshadowed by awareness that the success of a pathogen may also be dictated by its ability to out-compete other microbes that colonise the plant environment. For example, the recent emergence of SRE has been predicted to have arisen through the production of antimicrobials that enable the pathogen to out-compete related species or strains (Garlant *et al*., 2013). Many genes encoding

for putative antimicrobials were identified in the genome of ICMP19477. For example, the novel bacteriocin colicin V was identified in the genomes of both *P*. *carotovorum* subsp. *brasiliensis* ICMP19477 and ATCC BAA 417. Given that colicin V was absent from the remainder of the comparator strains, it was hypothesized that colicin V might contribute to the antagonism of closely related *Pectobacterium* by *P*. *carotovorum* subsp. *brasiliensis*. Recent studies have suggested that *P*. *carotovorum* subsp. *brasiliensis* produces an antimicrobial with activity against other *Pectobacterium* species (del Pilar Marquez-Villavicencio *et al*., 2011). Thus, it appears that *P*. *carotovorum* subsp. *brasiliensis* may have to compete with a variety of competitors to ensure it can successfully survive as a pathogen on potato.

The role of GIs (or gene islets) in ICMP19477 was initially only *predicted*, based on the presence of similar islands in other *Pectobacterium* genomes or on the functional annotation of the genes they carry. However, the *actual* role of the GIs or the genes encoded on these islands can only be confirmed by functional studies. In this study, to understand the role of GIs and/or gene islets in virulence of *P*. *carotovorum* subsp. *brasiliensis*, pathogenicity tests were carried out using knockout mutants in genes predicted to be important in virulence (Chapter 5). The development of knockout mutants proved difficult, however, possibly due to the presence of a restriction-modification system in ICMP19477 that rendered it recalcitrant to transformation with the mutagenic plasmids constructed as part of the study. Nevertheless, this study confirmed that the NRPS cluster has no involvement in virulence of *P*. *carotovorum* subsp. *brasiliensis* on potato (cv. Ilam Hardy). Furthermore, single cross-over mutants in the *sim* and *pad* genes indicated that these gene clusters might be involved in virulence. Double cross-over mutants and complementation assays need to be carried out to confirm these results, however.

An alternative strategy to investigate the role of the GIs or the candidate virulence genes they carry, may have been to remove the entire GI of interest or to target individual genes using CRISPR (clustered regularly interspaced short palindromic repeats)-mediated genome editing. Using CRISPR, an entire GI HAI2, was removed to study its acquisition in the genome of *P*. *atrosepticum* SCRI1043 (Dy *et al*., 2013). Alternatively, the GI of interest could be transferred to

a non-pathogenic strain to study if there is any increase in the virulence of that strain. GIs have been successfully transferred between strains of plant pathogenic bacterial species including *Pseudomonas* and *Ralstonia in planta*, as plants provide an ecological niche where specific organisms can exchange genetic information (Bertolla *et al*., 1999; Coupat-Goutaland *et al*., 2010; Lovell *et al*., 2010). This depends on the conditions *in planta* and the competence of the bacterial species in these conditions, however.

Finally, one of the major limitations of comparative genomics in the identification of GIs is that it can tell you events that have happened, but does not necessarily reveal whether the islands continue to play a role in the evolution of the pathogen. GIs have played an important role in accessorizing the genome of bacterial pathogens, either by introducing new virulence factors into the genome or provisioning bacteria with new mechanisms for genome restructuring or gene capture (Jackson *et al*., 2011). Whether these islands are readily available to the genomic pool (i.e. whether they are mobile) or permanently anchored (stably integrated) to the bacterial chromosome is unknown and cannot be predicted by bioinformatics analysis. An illustrative example is SPI-1, which lost all features of a mobile element and virtually became integral with the *Salmonella* genome (Ilyina & Romanova, 2001). Thus, functional assays to examine the mobilization of the GIs out of their chromosomal locations and their transmission between bacteria are important in determining the ongoing role of a GI in the evolution of the SRE.

6.2 Future perspectives

Whole genome sequencing studies are powerful hypothesis generators, and therefore a full understanding of the mechanisms involved in virulence of *P*. *carotovorum* subsp. *brasiliensis*, may be very possible through future functional studies of candidate virulence factors and the GIs that harbor them. Additional *Pectobacterium* genomes, however, need to be sequenced before the pan genome is closed and to confirm the core genome of SRE. These genomes would provide an even greater understanding of sub-specific variation and provide possible new clues as to the portfolio of GIs (and their genes) required for blackleg. Indeed, building a comprehensive

annotated GI database, which will host all discovered GIs to this date, would be helpful in the future. The construction of such databases might also improve computational prediction tools by re-evaluating the GI-associated features based on the more recently discovered islands.

Improving protocols and/or experimental design for the creation of successful knockout mutants (for functional analysis of *P. carotovorum* subsp. *brasiliensis*) is of utmost importance for functional studies, as even after repeated attempts, only one double cross-over knockout mutant of ICMP19477 was successfully obtained. The CRISPR-based method can be used as an alternative strategy to create large-scale genomic alterations, including remodeling or deletion of pathogenicity islands and, other non-mobile chromosomal regions as shown in studies by Dy *et al*. (2013) and Wang *et al*. (2013). Islandless strains created in this way could also be used to identify cryptic virulence factors, not identified using bioinformatics tools due to their novelty.

To examine the transfer of GIs (carrying virulence genes or other beneficial traits) between bacterial species, mobility assays *in vitro* and *in planta* can be conducted between virulent strains (donor) and non-virulent strains (without GIs, recipient). This can be achieved by tagging a marker in a GI of a pathogenic bacterial strain (donor) and transferring this tagged island to the recipient strain (islandless) to study if there is any increase or decrease in the virulence of the recipient. This will help to further our understanding of the processes or mechanisms by which bacterial plant pathogens evolve to cause diseases in their host.

A whole genome microarray of the bacteria pathogen can be constructed to analyze the expression patterns of different genes in free-living cells (*in vitro*) versus that in the pathogen grown in diseased/infected potato stems (*in planta*). This will assist in the identification of target genes induced during infection of the host plant. Previous studies by Mattinen *et al*. (2008) and Liu *et al*. (2008) have successfully used microarray-based approaches to identify putative new virulence genes of *P. atrosepticum* during plant infection or in the presence of host extract. Studying whether these genes are clustered on GIs would also demonstrate whether expression of island-encoded genes is co-ordinately regulated, as they are in other pathogenic bacteria such

as *Salmonella* (Heithoff *et al*., 1999; Main-Hester *et al*., 2008). In particular, transcriptome studies on bacteria grown under different conditions (such as differences in media composition, environmental stimuli, etc.) and during its different stages of infection could be carried out to reveal if the expression of these horizontally transferred genes within GIs are coordinately regulated. Conditional transcription of genes within GIs will reveal whether these genes are regulated in an independent or concerted fashion in response to different conditions. Hacker & Carniel, (2001) noted that not all laterally transferred genes encoded on GIs are functionally active, as newly acquired elements will only be expressed in the host cell if recognized by the transcriptional machinery and their regulation is efficiently coordinated with the rest of the genome.

Newer sequencing technologies such as dual RNA deep-sequencing (RNA-seq) might also be used to analyse changes in gene expression in both the pathogen and the host in parallel. This will aid in a comprehensive understanding of host-pathogen interactions. Using dual RNA-seq, disease-related genes of the sudden oak death pathogen *Phytophthora ramorum* and its most susceptible forest host, tanoak, were identified in a recent study by Hayden *et al*. (2012). The design and data analysis of a dual RNA-seq experiment is a time consuming and complex task, however and relies on aligning reads to a reference genome and is thus unsuitable for samples from organisms with genomes that are partially or fully unsequenced. For bacteria, a more cost-effective and practical alternative is to combine analysis of RNA-Seq data with a draft genome sequence derived from NGS. Since, the half-life of bacterial RNAs is very short (not stable), studies of bacteria associated with their hosts have relied on bacterial enrichment to help with subsequent steps of enriching for bacterial RNA (Bomar *et al*., 2011; Mandlik *et al*., 2011).

General Appendix

Media, buffers and antibiotics used in this study

Component	Amount
Luria-Bertani (LB) media (1 L)	
Trypton	10.0 g
Yeast extract	5.0 g
Sodium chloride	5.0 g
LB and Agar (1 L)	
Trypton	10.0 g
Yeast extract	5.0 g
Sodium chloride	5.0 g
Agar	16.0 g
50X Phosphate buffer (500 mL)	
Di potassium phosphate	35.0 g
Potassium phosphate	100.0 g
Minimal Glucose Media (1 L)	
50X Phosphate buffer	20.0 mL
10% (w/v) Ammonium sulphate	10.0 mL
1 M Magnesium sulphate	400 µL
20% Glucose	10.0 mL
Agar	16.0 g
Minimal Sucrose Media (1 L)	
50X Phosphate buffer	20.0 mL
10% (w/v) Ammonium sulphate	10.0 mL
1 M Magnesium sulphate	400 µL
50% Sucrose	200.0 mL
Agar	16.0 g

Component	Amount
10X Tris/Borate/EDTA buffer (TBE) (1 L)	
Tris Bse	108.0 g
Boric acid	55.0 g
Component	**Amount**
EDTA	9.3 g
Streptomycin stock solution (100 mg/mL)*	
Streptomycin sulphate	1.0 g
Double distilled water	10.0 mL
Kanamycin stock solution (50 mg/mL)*	
Kanamycin sulphate	0.5 g
Double distilled water	10.0 mL
Chloramphenicol stock solution (34 mg/mL)*	
Chloramphenicol	0.34 g
100% Ethanol	10.0 mL

*All antibiotic stock solutions are filter sterilised using a 50 mL capacity syringe fitted with a 0.22 µm Millipore filter before storing at -20^{0}C for further use.

References

Abo-Elyousr KaM, Sallam MA, Hassan MH, Allam AD, 2010. Effect of certain cultural practices on susceptibility of potato tubers to soft rot disease caused by *Erwinia carotovora* pv. *carotovora*. *Archives of Phytopathology and Plant Protection* **43**, 1625-635.

Accetto T, Peterka M, Avguštin G, 2005. Type II restriction modification systems of *Prevotella bryantii* TC1–1 and *Prevotella ruminicola* 23 strains and their effect on the efficiency of DNA introduction via electroporation. *FEMS microbiology letters* **247**, 177-83.

Achaz G, Boyer F, Rocha EPC, Viari A, Coissac E, 2007. Repseek, a tool to retrieve approximate repeats from large DNA sequences. *Bioinformatics* **23**, 119-21.

Adhya SL, Shapiro JA, 1969. The galactose operon of *E*. *coli* K-12. I. Structural and pleiotropic mutations of the operon. *Genetics* **62**, 231.

Adler J, 1974. Chemotaxis in bacteria. In. *Biochemistry of Sensory Functions.* Springer, 107-31.

Aggarwal G, Ramaswamy R, 2002. Ab initio gene identification: prokaryote genome annotation with GeneScan and GLIMMER. *Journal of biosciences* **27**, 7-14.

Agrios GN, 1997. *Plant Pathology*. Fourth Edition, Academic Press, London, 65.

Ahmadian A, Gharizadeh B, Gustafsson AC, Sterky F, Nyrén P, Uhlén M, Lundeberg J, 2000. Single-nucleotide polymorphism analysis by pyrosequencing. *Analytical biochemistry* **280**, 103-10.

Alamuri P, Löwer M, Hiss JA, Himpsl SD, Schneider G, Mobley HLT, 2010. Adhesion, invasion, and agglutination mediated by two trimeric autotransporters in the human uropathogen *Proteus mirabilis*. *Infection and immunity* **78**, 4882-894.

Alderborn A, Kristofferson A, Hammerling U, 2000. Determination of single-nucleotide polymorphisms by real-time pyrophosphate DNA sequencing. *Genome research* **10**, 1249-258.

Alfano JR, Collmer A, 2004. Type III secretion system effector proteins: double agents in bacterial disease and plant defense. *Annual Review of Phytopathol.* **42**, 385-414.

Almeida NF, Yan S, Lindeberg M, Studholme DJ, Schneider DJ, Condon B, Liu H, Viana CJ, Warren A, Evans C, 2009. A draft genome sequence of *Pseudomonas syringae* pv. tomato T1 reveals a type III effector repertoire significantly divergent from that of *Pseudomonas syringae* pv. tomato DC3000. *Molecular plant-microbe interactions* **22**, 52-62.

Altenhoff AM, Dessimoz C, 2009. Phylogenetic and functional assessment of orthologs inference projects and methods. *PLoS computational biology* **5**, e1000262.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, 1990. Basic local alignment search tool. *Journal of molecular biology* **215**, 403-10.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ, 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* **25**, 3389-402.

Alvarez-Martinez CE, Christie PJ, 2009. Biological diversity of prokaryotic type IV secretion systems. *Microbiology and molecular biology reviews* **73**, 775-808.

Ammar GA, Tryono R, Döll K, Karlovsky P, Deising HB, Wirsel SGR, 2013. Identification of ABC Transporter Genes of *Fusarium graminearum* with Roles in Azole Tolerance and/or Virulence. *PloS one* **8**, e79042.

An CL, Lim WJ, Hong SY, Kim EJ, Shin EC, Kim MK, Lee JR, Park SR, Woo JG, Lim YP, 2004. Analysis of bgl operon structure and characterization of β-glucosidase from *Pectobacterium carotovorum* subsp. *carotovorum* LY34. *Bioscience, biotechnology, and biochemistry* **68**, 2270-278.

Andersson RA, Eriksson ARB, Heikinheimo R, Mäe A, Pirhonen M, Kõiv V, Hyytiäinen H, Tuikkala A, Palva ET, 2000. Quorum sensing in the plant pathogen *Erwinia carotovora* subsp. *carotovora*: the role of expREcc. *Molecular plant-microbe interactions* **13**, 384-93.

Andre A, Maccheroni W, Doignon F, Garnier M, Renaudin J, 2003. Glucose and trehalose PTS permeases of *Spiroplasma citri* probably share a single IIA domain, enabling the spiroplasma to adapt quickly to carbohydrate changes in its environment. Microbiology **149**, 2687–696.

André A, Maucourt M, Moing A, Rolin D, Renaudin J, 2005. Sugar import and phytopathogenicity of *Spiroplasma citri*: glucose and fructose play distinct roles. *Molecular plant-microbe interactions* **18**, 33-42.

Andro T, Chambost J-P, Kotoujansky A, Cattaneo J, Bertheau Y, Barras F, Van Gijsegem F, Coleno A, 1984. Mutants of *Erwinia chrysanthemi* defective in secretion of pectinase and cellulase. *Journal of bacteriology* **160**, 1199-203.

Antunez-Lamas M, Cabrera E, Lopez-Solanilla E, Solano R, González-Melendi P, Chico JM, Toth I, Birch P, Pritchard L, Liu H, 2009. Bacterial chemoattraction towards jasmonate plays a role in the entry of *Dickeya dadantii* through wounded tissues. *Molecular microbiology* **74**, 662-71.

Aoki SK, Diner EJ, De Roodenbeke CTK, Burgess BR, Poole SJ, Braaten BA, Jones AM, Webb JS, Hayes CS, Cotter PA, 2010. A widespread family of polymorphic contact-dependent toxin delivery systems in bacteria. *Nature* **468**, 439-42.

Apweiler R, Biswas M, Fleischmann W, Kanapin A, Karavidopoulou Y, Kersey P, Kriventseva EV, Mittard V, Mulder N, Phan I, 2001. Proteome Analysis Database: online application of InterPro

and CluSTr for the functional classification of proteins in whole genomes. *Nucleic acids research* **29**, 44-48.

Aranda J, Poza M, Pardo BG, Rumbo S, Rumbo C, Parreira JR, Rodríguez-Velo P, Bou G, 2010. A rapid and simple method for constructing stable mutants of *Acinetobacter baumannii*. *BMC microbiology* **10**, 279.

Arellano-Reynoso B, Lapaque N, Salcedo S, Briones G, Ciocchini AE, Ugalde R, Moreno E, Moriyón I, Gorvel J-P, 2005. Cyclic β-1, 2-glucan is a *Brucella* virulence factor required for intracellular survival. *Nature immunology* **6**, 618-25.

Arnold DL, Jackson RW, Fillingham AJ, Goss SC, Taylor JD, Mansfield JW, Vivian A, 2001. Highly conserved sequences flank avirulence genes: isolation of novel avirulence genes from *Pseudomonas syringae* pv. *pisi*. *Microbiology* **147**, 1171-182.

Asante A, Tahara S, Hashidoko Y, 2008. Screening of rhizobacteria possessing phenolic acid-decarboxylation abilities from several plant families. *Journal of the Research Faculty of Agriculture, Hokkaido University=* 北海道大学大学院農学研究院紀要 **72**, 1-19.

Avrova AO, Hyman LJ, Toth RL, Toth IK, 2002. Application of amplified fragment length polymorphism fingerprinting for taxonomy and identification of the soft rot bacteria *Erwinia carotovora* and *Erwinia chrysanthemi*. *Applied and Environmental Microbiology* **68**, 1499-508.

Aziz R, Bartels D, Best A, Dejongh M, Disz T, Edwards R, Formsma K, Gerdes S, Glass E, Kubal M, 2008. The RAST Server: rapid annotations using subsystems technology. *BMC genomics* **9**, 75.

Babić A, Lindner AB, Vulić M, Stewart EJ, Radman M, 2008. Direct visualization of horizontal gene transfer. *Science* **319**, 1533-536.

Baghaee-Ravari S, Rahimian H, Shams-Bakhsh M, Lopez-Solanilla E, Antúnez-Lamas M, Rodríguez-Palenzuela P, 2011. Characterization of *Pectobacterium* species from Iran using biochemical and molecular methods. *European Journal of Plant Pathology* **129**, 413-25.

Bain RA, Pérombelon MCM, Tsror L, Nachmias A, 1990. Blackleg development and tuber yield in relation to numbers of *Erwinia carotovora* subsp. *atroseptica* on seed potatoes. *Plant pathology* **39**, 125-33.

Bainton NJ, Stead P, Chhabra SR, Bycroft BW, Salmond GP, Stewart GS, Williams P, 1992. N-(3-oxohexanoyl)-L-homoserine lactone regulates carbapenem antibiotic production in *Erwinia carotovora*. *Biochemical Journal* **288**, 997-1004.

Baltrus DA, Nishimura MT, Romanchuk A, Chang JH, Mukhtar MS, Cherkis K, Roach J, Grant SR, Jones CD, Dangl JL, 2011. Dynamic evolution of pathogenicity revealed by sequencing and comparative genomics of 19 *Pseudomonas syringae* isolates. *PLoS Pathogens* **7**, e1002132.

Barabote RD, Johnson OL, Zetina E, San Francisco SK, Fralick JA, San Francisco MJD, 2003. Erwinia chrysanthemi tolC Is Involved in Resistance to Antimicrobial Plant Chemicals and Is Essential for Phytopathogenesis†. *Journal of bacteriology* **185**, 57727-7728.

Barnard AML, Salmond GPC, 2007. Quorum sensing in *Erwinia* species. *Analytical and bioanalytical chemistry* **387**, 415-23.

Barnett MJ, Fisher RF, Jones T, Komp C, Abola AP, Barloy-Hubler F, Bowser L, Capela D, Galibert F, Gouzy J, 2001. Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti* pSymA megaplasmid. *Proceedings of the National Academy of Sciences* **98**, 9883-888.

Barras F, Van Gijsegem F, Chatterjee AK, 1994. Extracellular enzymes and pathogenesis of soft-rot Erwinia. *Annual Review of Phytopathology* **32**, 201-34.

Barthelmebs L, Divies C, Cavin J-F, 2000. Knockout of the p-coumarate decarboxylase gene from *Lactobacillus plantarum* reveals the existence of two other inducible enzymatic activities involved in phenolic acid metabolism. *Applied and environmental microbiology* **66**, 3368-375.

Bartram AK, Lynch MDJ, Stearns JC, Moreno-Hagelsieb G, Neufeld JD, 2011. Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end Illumina reads. *Applied and environmental microbiology* **77**, 3846-852.

Bastide M, Mccombie WR, 2007. Assembling genomic DNA sequences with PHRAP. *Current Protocols in Bioinformatics*, 11.4. 1-.4. 5.

Bejerano-Sagie M, Xavier KB, 2007. The role of small RNAs in quorum sensing. *Current opinion in microbiology* **10**, 189-98.

Bell KS, Sebaihia M, Pritchard L, Holden MTG, Hyman LJ, Holeva MC, Thomson NR, Bentley SDL, Churcher JC, Mungall K, 2004. C., Birch, PRJ, Parkhill, J., and Toth, IK, Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora* subsp. *atroseptica* and characterization of virulence factors. *Proceedings of the National Academy of Sciences. USA* **101**, 11105-1110.

Bender CL, Alarcón-Chaidez F, Gross DC, 1999. *Pseudomonas syringae* phytotoxins: mode of action, regulation, and biosynthesis by peptide and polyketide synthetases. *Microbiology and molecular biology reviews* **63**, 266-92.

Bennett S, 2004. Solexa ltd. *Pharmacogenomics* **5**, 433-38.

Bentley SD, Parkhill J, 2004. Comparative genomic structure of prokaryotes. *Annual Review of Genetics* **38**, 771-91.

Bertolla F, Frostegård A, Brito B, Nesme X, Simonet P, 1999. During infection of its hosts, the plant pathogen *Ralstonia solanacearum* naturally develops a state of competence and exchanges genetic material. *Molecular Plant Microbe Interactions* **12**, 467–472.

Besemer J, Borodovsky M, 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic acids research* **33**, W451-54.

Besemer J, Lomsadze A, Borodovsky M, 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic acids research* **29**, 2607-618.

Bhagwat AA, Gross KC, Tully RE, Keister DL, 1996. Beta-glucan synthesis in *Bradyrhizobium japonicum*: characterization of a new locus (ndvC) influencing beta-(1--> 6) linkages. *Journal of bacteriology* **178**, 4635-642.

Bhagwat AA, Mithöfer A, Pfeffer PE, Kraus C, Spickers N, Hotchkiss A, Ebel J, Keister DL, 1999. Further Studies of the Role of Cyclic β-Glucans in Symbiosis. An ndvC Mutant of *Bradyrhizobium japonicum* Synthesizes Cyclodecakis-(1→ 3)-β-Glucosyl. *Plant physiology* **119**, 1057-564.

Bickle TA, Krüger DH, 1993. Biology of DNA restriction. *Microbiological reviews* **57**, 434-50.

Bignell DRD, Seipke RF, Huguet-Tapia JC, Chambers AH, Parry RJ, Loria R, 2010. *Streptomyces scabies* 87-22 contains a coronafacic acid-like biosynthetic cluster that contributes to plant-microbe interactions. *Molecular plant-microbe interactions* **23**, 161-75.

Binet R, Létoffé S, Ghigo JM, Delepelaire P, Wandersman C, 1997. Protein secretion by Gram-negative bacterial ABC exporters–a review. *Gene* **192**, 7-11.

Birch PRJ, Bryan G, Fenton B, Gilroy EM, Hein I, Jones JT, Ankush P, Taylor MA, Torrance L, Toth IK, 2012. Crops that feed the world 8: potato: are the trends of increased global production sustainable? *Food Security* **4**, 477-508.

Biswas I, Gruss A, Ehrlich SD, Maguin E, 1993. High-efficiency gene inactivation and replacement system for gram-positive bacteria. *Journal of bacteriology* **175**, 3628-635.

Blair DF, 1995. How bacteria sense and swim. *Annual Reviews in Microbiology* **49**, 489-520.

Blanvillain S, Meyer D, Boulanger A, Lautier M, Guynet C, Denancé N, Vasse Lauber E, Arlat M, 2007. Plant carbohydrate scavenging through TonB-dependent receptors: a feature shared by phytopathogenic and aquatic bacteria. *PLoS one*, **2(2)**, e224.

Blom J, Albaum SP, Doppmeier D, Pühler A, Vorhölter F-J, Zakrzewski M, Goesmann A, 2009. EDGAR: a software framework for the comparative analysis of prokaryotic genomes. *BMC bioinformatics* **10**, 154.

Boetzer M, Pirovano W, 2012. Toward almost closed genomes with GapFiller. *Genome biology* **13**, R56.

Bogdanove AJ, Beer SV, Bonas U, Boucher CA, Collmer A, Coplin DL, Cornelis GR, Huang HC, Hutcheson SW, Panopoulos NJ, 1996. Unified nomenclature for broadly conserved hrp genes of phytopathogenic bacteria. *Molecular microbiology* **20**, 681-83.

Bogs J, Geider K, 2000. Molecular Analysis of Sucrose Metabolism of *Erwinia amylovora* and Influence on Bacterial Virulence. *Journal of bacteriology* **182**, 5351-358.

Boissy R, Ahmed A, Janto B, Earl J, Hall BG, Hogg JS, Pusch GD, Hiller LN, Powell E, Hayes J, 2011. Comparative supragenomic analyses among the pathogens *Staphylococcus aureus*, *Streptococcus pneumoniae*, and *Haemophilus influenzae* using a modification of the finite supragenome model. *BMC genomics* **12**, 187.

Bokern M, Wray V, Strack D, 1991. Accumulation of phenolic acid conjugates and betacyanins, and changes in the activities of enzymes involved in feruloylglucose metabolism in cell-suspension cultures of *Chenopodium rubrum* L. *Planta* **184**, 261-70.

Bomar L, Maltz M, Colston S, Graf J, 2011. Directed Culturing of Microorganisms Using Metatranscriptomics. *mBio* **2**, e00012–11.

Bönemann G, Pietrosiuk A, Diemand A, Zentgraf H, Mogk A, 2009. Remodelling of VipA/VipB tubules by ClpV-mediated threading is crucial for type VI protein secretion. *The EMBO journal* **28**, 315-25.

Bornke F, Hajirezaei M, Sonnewald U, 2001. Cloning and characterization of the gene cluster for Palatinose metabolism from the phytopathogenic bacterium *Erwinia* rhapontici. *Journal of bacteriology* **183**, 2425-430.

Borodovsky M, Mcininch J, 1993. GENMARK - PARALLEL GENE RECOGNITION FOR BOTH DNA STRANDS. *Computers & Chemistry* **17**, 123-33.

Boyd D, Peters GA, Cloeckaert A, Boumedine KS, Chaslus-Dancla E, Imberechts H, Mulvey MR, 2001. Complete nucleotide sequence of a 43-kilobase genomic island associated with the multidrug resistance region of *Salmonella enterica* serovar Typhimurium DT104 and its identification in phage type DT120 and serovar Agona. *Journal of bacteriology* **183**, 5725-732.

Boyd DA, Peters GA, Ng LK, Mulvey MR, 2000. Partial characterization of a genomic island associated with the multidrug resistance region of *Salmonella enterica* Typhymurium DT104. *FEMS microbiology letters* **189**, 285-91.

Boyd EF, Brüssow H, 2002. Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. *Trends in microbiology* **10**, 521-29.

Breedveld MW, Miller KJ, 1994. Cyclic beta-glucans of members of the family Rhizobiaceae. *Microbiological reviews* **58**, 145-61.

Brenner DJ, Mcwhorter AC, Kai A, Steigerwalt AG, Farmer JJ, 1986. *ENTEROBACTER-ASBURIAE* SP-NOV, A NEW SPECIES FOUND IN CLINICAL SPECIMENS, AND REASSIGNMENT OF *ERWINIA-DISSOLVENS* AND *ERWINIA-NIMIPRESSURALIS* TO THE GENUS *ENTEROBACTER* AS *ENTEROBACTER-DISSOLVENS* COMB-NOV AND *ENTEROBACTER-NIMIPRESSURALIS* COMB-NOV. *Journal of Clinical Microbiology* **23**, 1114-120.

Brooks DM, Bender CL, Kunkel BN, 2005. The *Pseudomonas syringae* phytotoxin coronatine promotes virulence by overcoming salicylic acid-dependent defences in *Arabidopsis thaliana*. *Molecular Plant Pathology* **6**, 629-39.

Brooks DM, Hernández-Guzmán G, Kloek AP, Alarcón-Chaidez F, Sreedharan A, Rangaswamy V, Peñaloza-Vázquez A, Bender CL, Kunkel BN, 2004. Identification and characterization of a well-defined series of coronatine biosynthetic mutants of *Pseudomonas syringae* pv. tomato DC3000. *Molecular plant-microbe interactions* **17**, 162-74.

Brosch R, Pym AS, Gordon SV, Cole ST, 2001. The evolution of mycobacterial pathogenicity: clues from comparative genomics. *Trends in microbiology* **9**, 452-58.

Brown JS, Gilliland SM, Holden DW, 2001. A *Streptococcus pneumoniae* pathogenicity island encoding an ABC transporter involved in iron uptake and virulence. *Molecular microbiology* **40**, 572-85.

Brown SP, Le Chat L, De Paepe M, Taddei F, 2006. Ecology of microbial invasions: amplification allows virus carriers to invade more rapidly when rare. *Current Biology* **16**, 2048-052.

Buchrieser C, Brosch R, Bach S, Guiyoule A, Carniel E, 1998. The high-pathogenicity island of *Yersinia pseudotuberculosis* can be inserted into any of the three chromosomal asn tRNA genes. *Molecular microbiology* **30**, 965-78.

Buell CR, Joardar V, Lindeberg M, Selengut J, Paulsen IT, Gwinn ML, Dodson RJ, Deboy RT, Durkin AS, Kolonay JF, 2003. The complete genome sequence of the *Arabidopsis* and tomato pathogen *Pseudomonas syringae* pv. tomato DC3000. *Proceedings of the National Academy of Sciences* **100**, 10181-0186.

Bulach DM, Zuerner RL, Wilson P, Seemann T, Mcgrath A, Cullen PA, Davis J, Johnson M, Kuczek E, Alt DP, 2006. Genome reduction in *Leptospira borgpetersenii* reflects limited transmission potential. *Proceedings of the National Academy of Sciences* **103**, 14560-5.

Burkholder WH, 1957. Genus VI. Erwinia Winslow et al. 1917. *Bergey's Manual of Determinative Bacteriology*, 349-59.

Burkholder WR, Mcfadden LA, Dimock EW, 1953. A bacterial blight ol Chrysanthemums. *Phytopathology* **43**, 522-26.

Burr T, Barnard AML, Corbett MJ, Pemberton CL, Simpson NJL, Salmond GPC, 2006. Identification of the central quorum sensing regulator of virulence in the enteric phytopathogen, *Erwinia carotovora*: the VirR repressor. *Molecular microbiology* **59**, 113-25.

Burrus V, Pavlovic G, Decaris B, Guédon G, 2002. Conjugative transposons: the tip of the iceberg. *Molecular microbiology* **46**, 601-10.

Burrus V, Waldor MK, 2003. Control of SXT integration and excision. *Journal of bacteriology* **185**, 5045-054.

Büttner D, He SY, 2009. Type III protein secretion in plant pathogenic bacteria. *Plant physiology* **150**, 1656-664.

Cahill MJ, Köser CU, Ross NE, Archer JaC, 2010. Read length and repeat resolution: exploring prokaryote genomes using next-generation sequencing technologies. *PloS one* **5**, e11518.

Calcutt MJ, Schmidt FJ, 1994. Gene organization in the bleomycin-resistance region of the producer organism *Streptomyces verticillus*. *Gene* **151**, 17-21.

Cambray G, Guerout A-M, Mazel D, 2010. Integrons. *Annual review of genetics* **44**, 141-66.

Carter MQ, Chen J, Lory S, 2010. The *Pseudomonas aeruginosa* pathogenicity island PAPI-1 is transferred via a novel type IV pilus. *Journal of bacteriology* **192**, 3249-258.

Cascales E, Buchanan SK, Duché D, Kleanthous C, Lloubes R, Postle K, Riley M, Slatin S, Cavard D, 2007. Colicin biology. *Microbiology and molecular biology reviews* **71**, 158-229.

Cascales E, Christie PJ, 2003. The versatile bacterial type IV secretion systems. *Nature Reviews Microbiology* **1**, 137-49.

Casjens S, 2000. *Borrelia* genomes in the year 2000. *Journal of molecular microbiology and biotechnology* **2**, 401-10.

Casjens S, 2003. Prophages and bacterial genomics: what have we learned so far? *Molecular microbiology* **49**, 277-300.

Casjens S, Palmer N, Van Vugt R, Mun Huang W, Stevenson B, Rosa P, Lathigra R, Sutton G, Peterson J, Dodson RJ, 2000. A bacterial genome in flux: the twelve linear and nine circular

extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete *Borrelia burgdorferi*. *Molecular microbiology* **35**, 490-516.

Castro OA, Zorreguieta A, Ielmini V, Vega G, Ielpi L, 1996. Cyclic beta-(1, 2)-glucan synthesis in *Rhizobiaceae*: roles of the 319-kilodalton protein intermediate. *Journal of bacteriology* **178**, 6043-048.

Cavin J-F, Barthelmebs L, Divies C, 1997. Molecular characterization of an inducible p-coumaric acid decarboxylase from Lactobacillus plantarum: gene cloning, transcriptional analysis, overexpression in *Escherichia coli*, purification, and characterization. *Applied and environmental microbiology* **63**, 1939-944.

Chaisson M, Pevzner P, Tang H, 2004. Fragment assembly with short reads. *Bioinformatics* **20**, 2067-074.

Chan Y-C, Wu J-L, Wu H-P, Tzeng K-C, Chuang D-Y, 2011. Cloning, purification, and functional characterization of Carocin S2, a ribonuclease bacteriocin produced by *Pectobacterium carotovorum*. *BMC microbiology* **11**, 99.

Chapon V, Czjzek M, El Hassouni M, Py B, Juy M, Barras F, 2001. Type II protein secretion in gram-negative pathogenic bacteria: the study of the structure/secretion relationships of the cellulase cel5 (formerly EGZ) from *Erwinia chrysanthemi*. *Journal of molecular biology* **310**, 1055-066.

Charkowski A, Blanco C, Condemine G, Expert D, Franza T, Hayes C, Hugouvieux-Cotte-Pattat N, Solanilla EL, Low D, Moleleki L, 2012. The role of secretion systems and small molecules in soft-rot enterobacteriaceae pathogenicity. *Annual Review of Phytopathology* **50**, 425-49.

Charkowski AO, Alfano JR, Preston G, Yuan J, He SY, Collmer A, 1998. The Pseudomonas syringae pv. tomato HrpW protein has domains similar to harpins and pectate lyases and can elicit the plant hypersensitive response and bind to pectate. *Journal of bacteriology* **180**, 5211-217.

Chatterjee A, Cui Y, Hasegawa H, Leigh N, Dixit V, Chatterjee AK, 2005. Comparative analysis of two classes of quorum-sensing signaling systems that control production of extracellular proteins and secondary metabolites in *Erwinia carotovora* subspecies. *Journal of bacteriology* **187**, 8026-038.

Chatterjee AK, Thurn KK, Tyrell DJ, 1985. Isolation and characterization of Tn5 insertion mutants of *Erwinia chrysanthemi* that are deficient in polygalacturonate catabolic enzymes oligogalacturonate lyase and 3-deoxy-D-glycero-2, 5-hexodiulosonate dehydrogenase. *Journal of bacteriology* **162**, 708-14.

Chen I, Dubnau D, 2004. DNA uptake during bacterial transformation. *Nature Reviews Microbiology* **2**, 241-49.

Cheng Y, Liu X, An S, Chang C, Zou Y, Huang L, Zhong J, Liu Q, Jiang Z, Zhou J, 2013. A Nonribosomal Peptide Synthase Containing a Stand-Alone Condensation Domain Is Essential for Phytotoxin Zeamine Biosynthesis. *Molecular plant-microbe interactions* **26**, 1294-301.

Christie PJ, 1997. *Agrobacterium tumefaciens* T-complex transport apparatus: a paradigm for a new family of multifunctional transporters in eubacteria. *Journal of bacteriology* **179**, 3085-094.

Christie PJ, 2001. Type IV secretion: intercellular transfer of macromolecules by systems ancestrally related to conjugation machines. *Molecular microbiology* **40**, 294-305.

Christie PJ, 2004. Type IV secretion: the *Agrobacterium* VirB/D4 and related conjugation systems. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **1694**, 219-34.

Christie PJ, Cascales E, 2005. Structural and dynamic properties of bacterial type IV secretion systems (review). *Molecular membrane biology* **22**, 51-61.

Chuang D-Y, Chien Y-C, Wu H-P, 2007. Cloning and expression of the *Erwinia carotovora* subsp. *carotovora* gene encoding the low-molecular-weight bacteriocin carocin S1. *Journal of bacteriology* **189**, 620-26.

Cianciotto NP, 2005. Type II secretion: a protein secretion system for all seasons. *Trends in microbiology* **13**, 581-88.

Clarke J, Wu H-C, Jayasinghe L, Patel A, Reid S, Bayley H, 2009. Continuous base identification for single-molecule nanopore DNA sequencing. *Nature nanotechnology* **4**, 265-70.

Clausen M, Lamb CJ, Megnet R, Doerner PW, 1994. *PAD1* encodes phenylacrylic acid decarboxylase which confers resistance to cinnamic acid in *Saccharomyces cerevisiase*. *Gene* **142**, 107-12.

Coburn B, Sekirov I, Finlay BB, 2007. Type III secretion systems and disease. *Clinical microbiology reviews* **20**, 535-49.

Collmer A, Bateman DF, 1981. Impaired induction and self-catabolite repression of extracellular pectate lyase in *Erwinia chrysanthemi* mutants deficient in oligogalacturonide lyase. *Proceedings of the National Academy of Sciences* **78**, 3920-924.

Collmer A, Keen NT, 1986. The role of pectic enzymes in plant pathogenesis. *Annual Review of Phytopathology* **24**, 383-409.

Collmer A, Schneider DJ, Lindeberg M, 2009. Lifestyles of the effector rich: genome-enabled characterization of bacterial plant pathogens. *Plant physiology* **150**, 1623-630.

Condemine G, Castillo A, Passeri F, Enard C, 1999. The PecT repressor coregulates synthesis of exopolysaccharides and virulence factors in *Erwinia chrysanthemi*. *Molecular plant-microbe interactions* **12**, 45-52.

Condemine G, Dorel C, Hugouvieux-Cotte-Pattat N, Robert-Baudouy J, 1992. Some of the out genes involved in the secretion of pectate lyases in *Erwinia chrysanthemi* are regulated by kdgR. *Molecular microbiology* **6**, 3199-211.

Condemine G, Hugouvieux-Cotte-Pattat N, Robert-Baudouy J, 1986. Isolation of *Erwinia chrysanthemi* kduD mutants altered in pectin degradation. *Journal of bacteriology* **165**, 937-41.

Condemine G, Robert-Baudouy J, 1987. *Tn5* insertion in *kdgR*, a regulatory gene of the polygalacturonate pathway in *Erwinia chrysanthemi*. *FEMS microbiology letters* **42**, 39-46.

Corbett M, Virtue S, Bell K, Birch P, Burr T, Hyman L, Lilley K, Poock S, Toth I, Salmond G, 2005. Identification of a new quorum-sensing-controlled virulence factor in *Erwinia carotovora* subsp. *atroseptica* secreted via the type II targeting pathway. *Molecular plant-microbe interactions* **18**, 334-42.

Costa AB, Eloy M, Cruz L, Janse JD, Oliveira H, 2006. Studies on pectolytic *Erwinia* spp. in Portugal reveal unusual strains of *E*. *carotovora* subsp. *atroseptica*. *Journal of Plant Pathology*, 161-69.

Cote CK, Cvitkovitch D, Bleiweis AS, Honeyman AL, 2000. A novel β-glucoside-specific PTS locus from *Streptococcus mutans* that is not inhibited by glucose. *Microbiology* **146**, 1555-563.

Coulthurst SJ, Barnard AML, Salmond GPC, 2005. Regulation and biosynthesis of carbapenem antibiotics in bacteria. *Nature Reviews Microbiology* **3**, 295-306.

Coupat B, Chaumeille-Dole F, Fall S, Prior P, Simonet P, Nesme X, Bertolla F, 2008. Natural transformation in the *Ralstonia solanacearum* species complex: number and size of DNA that can be transferred. *FEMS microbiology ecology* **66**, 14-24.

Coupat-Goutaland B, Bernillon D, Guidot A, Prior P, Nesme X, Bertolla F, 2011. *Ralstonia solanacearum* virulence increased following large interstrain gene transfers by natural transformation. *Molecular Plant Microbe Interactions* **24**, 497–505.

Crowhurst RN, Wright PJ, 1988. Subspecies of *Erwinia carotovora* causing blackleg of potato in Pukekohe and Pukekawa and their survival in soil. *New Zealand Potato Bulletin*, 20-22.

Cubitt MF, Hedley PE, Williamson NR, Morris JA, Campbell E, Toth IK, Salmond GPC, 2013. A metabolic regulator modulates virulence and quorum sensing signal production in *Pectobacterium atrosepticum*. *Molecular plant-microbe interactions* **26**, 356-66.

Cui Y, Chatterjee A, Chatterjee AK, 1991. Effects of the two compomemt system cpmprising *GacA* and *GacS* of *Erwinia carotovora* subspecies *carotovora* on the production of global regulatory rmsB RNA, extracellular enzymes and hairpin (*Ecc*). *Molecular Plant-Microbe Interactions* **14**, 516-26.

Cui Y, Chatterjee A, Hasegawa H, Dixit V, Leigh N, Chatterjee AK, 2005. ExpR, a LuxR homolog of *Erwinia carotovora* subsp. *carotovora*, activates transcription of rsmA, which specifies a global regulatory RNA-binding protein. *Journal of bacteriology* **187**, 4792-803.

Cui Y, Chatterjee A, Liu Y, Dumenyo CK, Chatterjee AK, 1995. Identification of a global repressor gene, rsmA, of *Erwinia carotovora* subsp. *carotovora* that controls extracellular enzymes, N-(3-oxohexanoyl)-L-homoserine lactone, and pathogenicity in soft-rotting *Erwinia* spp. *Journal of bacteriology* **177**, 5108-115.

Curie C, Cassin G, Couch D, Divol F, Higuchi K, Le Jean M, Misson J, Schikora A, Czernic P, Mari S, 2009. Metal movement within the plant: contribution of nicotianamine and yellow stripe 1-like transporters. *Annals of botany* **103**, 1-11.

Czajkowski R, Perombelon MCM, Van Veen JA, Van Der Wolf JM, 2011. Control of blackleg and tuber soft rot of potato caused by *Pectobacterium* and *Dickeya* species: a review. *Plant pathology* **60**, 999-1013.

Czajkowski R, Van Veen JA, Van Der Wolf JM, 2009. New biovar 3 *Dickeya* spp. strain (syn. *Erwinia chrysanthemi*) as a causative agent of blackleg in seed potato in Europe. *Phytopathology* **99**, S27.

Da Silva ACR, Ferro JA, Reinach FC, Farah CS, Furlan LR, Quaggio RB, Monteiro-Vitorello CB, Van Sluys MA, Almeida NF, Alves LMC, 2002. Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* **417**, 459-63.

Dahler GS, Barras F, Keen NT, 1990. Cloning of genes encoding extracellular metalloproteases from *Erwinia chrysanthemi* EC16. *Journal of bacteriology* **172**, 5803-815.

Darling ACE, Mau B, Blattner FR, Perna NT, 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome research* **14**, 1394-403.

Darling AE, Mau B, Perna NT, 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PloS one* **5**, e11147.

Darrasse A, Priou S, Kotoujansky A, Bertheau Y, 1994. PCR and restriction fragment length polymorphism of a pel gene as a tool to identify *Erwinia carotovora* in relation to potato diseases. *Applied and environmental microbiology* **60**, 1437-443.

Dassa E, Bouige P, 2001. The ABC of ABCs: A phylogenetic and functional classification of ABC systems in living organisms. *Research in Microbiology* **152**, 211-29.

Datta N, Hedges RW, 1972. Host ranges of R factors. *Journal of General Microbiology* **70**, 453-60.

Davidsson PR, Kariola T, Niemi O, Palva ET, 2013. Pathogenicity of and plant immunity to soft rot pectobacteria. *Frontiers in plant science* **4, 191**.

Davies J, Davies D, 2010. Origins and evolution of antibiotic resistance. *Microbiology and Molecular Biology Reviews* **74**, 417-33.

Daw MA, Falkiner FR, 1996. Bacteriocins: nature, function and structure. *Micron* **27**, 467-79.

Dazzo FB, Truchet GL, Hollingsworth RI, Hrabak EM, Pankratz HS, Philip-Hollingsworth S, Salzwedel JL, Chapman K, Appenzeller L, Squartini A, 1991. *Rhizobium* lipopolysaccharide modulates infection thread development in white clover root hairs. *Journal of bacteriology* **173**, 5371-384.

De Boer SH, 2002. Relative incidence of *Erwinia carotovora* subsp. *atroseptica* in stolon end and peridermal tissue of potato tubers in Canada. *Plant Disease* **86**, 960-64.

De Boer SH, 2003. Characterization of pectolytic erwinias as highly sophisticated pathogens of plants. *European Journal of Plant Pathology* **109**, 893-99.

De Boer SH, 2004. Blackleg of potato. *The Plant Health Instructor*.

De Boer SH, Copeman RJ, Vruggink H, 1979. Serogroups of *Erwinia carotovora* potato strains determined with diffusible somatic antigens. *Phytopathology* **69**, 316-19.

De Boer SH, Li X, Ward LJ, 2012. *Pectobacterium* spp. associated with bacterial stem rot syndrome of potato in Canada. *Phytopathology* **102**, 937-47.

De Boer SH, Mcnaughton ME, 1987. Monoclonal antibodies to the lipopolysaccharide of *Erwinia carotovora* subsp. *atroseptica* serogroup I. *Phytopathology* **77**, 828-32.

De Boer SH, Verdonck L, Vruggink H, Harju P, Bång HO, Ley JD, 1987. Serological and biochemical variation among potato strains of *Erwinia carotovora* subsp. *atroseptica* and their taxonomic relationship to other *E*. *carotovora* strains. *Journal of Applied Bacteriology* **63**, 487-95.

De Boer SH, Ward LJ, 1995. PCR detection of *Erwinia carotovora* subsp. *atroseptica* associated with potato tissue. *Phytopathology* **85**, 854-58.

De Haan EG, Dekker-Nooren TCEM, Van Den Bovenkamp GW, Speksnijder AGCL, Van Der Zouwen PS, Van Der Wolf JM, 2008. *Pectobacterium carotovorum* subsp. *carotovorum* can cause potato blackleg in temperate climates. *European Journal of Plant Pathology* **122**, 561-69.

De La Cruz F, Frost LS, Meyer RJ, Zechner EL, 2010. Conjugative DNA metabolism in Gram-negative bacteria. *FEMS microbiology reviews* **34**, 18-40.

Dean D, 1981. A plasmid cloning vector for the direct selection of strains carrying recombinant plasmids. *Gene* **15**, 99-102.

Del Pilar Marquez-Villavicencio M, Weber B, Witherell RA, Willis DK, Charkowski AO, 2011. The 3-hydroxy-2-butanone pathway is required for *Pectobacterium carotovorum* pathogenesis. *PloS one* **6**, e22974.

Delcher AL, Bratke KA, Powers EC, Salzberg SL, 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23**, 673-79.

Delcher AL, Harmon D, Kasif S, White O, Salzberg SL, 1999a. Improved microbial gene identification with GLIMMER. *Nucleic Acids Research* **27**, 4636-641.

Delcher AL, Kasif S, Fleischmann RD, Peterson J, White O, Salzberg SL, 1999b. Alignment of whole genomes. *Nucleic acids research* **27**, 2369-376.

Delcher AL, Phillippy A, Carlton J, Salzberg SL, 2002. Fast algorithms for large-scale genome alignment and comparison. *Nucleic acids research* **30**, 2478-483.

Delepelaire P, 2004. Type I secretion in gram-negative bacteria. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **1694**, 149-61.

Delepelaire P, Wandersman C, 1989. Protease secretion by *Erwinia chrysanthemi*. Proteases B and C are synthesized and secreted as zymogens without a signal peptide. *Journal of Biological Chemistry* **264**, 9083-089.

Delepelaire P, Wandersman C, 1991. Characterization, localization and transmembrane organization of the three proteins PrtD, PrtE and PrtF necessary for protease secretion by the Gram-negative bacterium *Erwinia chrysanthemi*. *Molecular microbiology* **5**, 2427-434.

Dellagi A, Brisset M-N, Paulin J-P, Expert D, 1998. Dual role of desferrioxamine in *Erwinia amylovora* pathogenicity. *Molecular plant-microbe interactions* **11**, 734-42.

Dellagi A, Reis D, Vian B, Expert D, 1999. Expression of the ferrioxamine receptor gene of *Erwinia amylovora* CFBP 1430 during pathogenesis. *Molecular plant-microbe interactions* **12**, 463-66.

Deng W, Burland V, Plunkett Iii G, Boutin A, Mayhew GF, Liss P, Perna NT, Rose DJ, Mau B, Zhou S, 2002. Genome sequence of Yersinia pestis KIM. *Journal of bacteriology* **184**, 4601-611.

Derzelle S, Duchaud E, Kunst F, Danchin A, Bertin P, 2002. Identification, characterization, and regulation of a cluster of genes involved in carbapenem biosynthesis in *Photorhabdus luminescens*. *Applied and environmental microbiology* **68**, 3780-789.

Devaux PF, Lopez-Montero I, Bryde S, 2006. Proteins involved in lipid translocation in eukaryotic cells. *Chemistry and Physics of Lipids* **141**, 119-32.

Dillard JP, Seifert HS, 2001. A variable genetic island specific for *Neisseria gonorrhoeae* is involved in providing DNA for natural transformation and is found more often in disseminated infection isolates. *Molecular microbiology* **41**, 263-77.

Dobrindt U, Hochhut B, Hentschel U, Hacker J, 2004. Genomic islands in pathogenic and environmental microorganisms. *Nature Reviews Microbiology* **2**, 414-24.

Dominguez H, Lindley ND, 1996. Complete Sucrose Metabolism Requires Fructose Phosphotransferase Activity in *Corynebacterium glutamicum* To Ensure Phosphorylation of Liberated Fructose. *Applied and environmental microbiology* **62**, 3878-880.

Dominguez H, Rollin C, Guyonvarch A, Guerquin-Kern JL, Cocaign-Bousquet M, Lindley ND, 1998. Carbon-flux distribution in the central metabolic pathways of *Corynebacterium glutamicum* during growth on fructose. *European Journal of Biochemistry* **254**, 96-102.

Douet V, Expert D, Barras F, Py B, 2009. *Erwinia chrysanthemi* iron metabolism: the unexpected implication of the inner membrane platform within the type II secretion system. *Journal of bacteriology* **191**, 795-804.

Drigues P, Demery-Lafforgue D, Trigalet A, Dupin P, Samain D, Asselineau J, 1985. Comparative studies of lipopolysaccharide and exopolysaccharide from a virulent strain of *Pseudomonas solanacearum* and from three avirulent mutants. *Journal of bacteriology* **162**, 504-09.

Droege M, Hill B, 2008. The Genome Sequencer FLX™ System—Longer reads, more applications, straight forward bioinformatics and more complete data sets. *Journal of biotechnology* **136**, 3-10.

Duarte V, De Boer SH, Ward LJ, De Oliveira AMR, 2004. Characterization of atypical *Erwinia carotovora* strains causing blackleg of potato in Brazil. *Journal of Applied Microbiology* **96**, 535-45.

Dubnau D, 1999. DNA uptake in bacteria. *Annual Reviews in Microbiology* **53**, 217-44.

Dudley EG, Thomson NR, Parkhill J, Morin NP, Nataro JP, 2006. Proteomic and microarray characterization of the AggR regulon identifies a pheU pathogenicity island in enteroaggregative *Escherichia coli*. *Molecular microbiology* **61**, 1267-282.

Dunlap J, Minami E, Bhagwat AA, Keister DL, Stacey G, 1996. Nodule development induced by mutants of *Bradyrhizobium japonicum* defective in cyclic B-glucan synthesis. *MPMI-Molecular Plant Microbe Interactions* **9**, 546-55.

Dutta PR, Cappello R, Navarro-García F, Nataro JP, 2002. Functional comparison of serine protease autotransporters of *Enterobacteriaceae*. *Infection and immunity* **70**, 7105-113.

Dy RL, Pitman AR, Fineran PC, 2013. Chromosomal targeting by CRISPR-Cas systems can contribute to genome plasticity in bacteria, *Mobile Genetic Elements*, **3**:5, e26831.

Dykes GA, Hastings JW, 1997. Selection and fitness in bacteriocin–producing bacteria. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **264**, 683-87.

Dyrløv Bendtsen J, Nielsen H, Von Heijne G, Brunak S, 2004. Improved prediction of signal peptides: SignalP 3.0. *Journal of molecular biology* **340**, 783-95.

Eddy SR, 2002. A memory-efficient dynamic programming algorithm for optimal alignment of a sequence to an RNA secondary structure. *BMC bioinformatics* **3**, 18.

Edgar RC, 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792-797.

Elphinstone JG, 1987. Soft Rot and Blackleg of Potato; *Erwinia* spp. *Technical Information Bulletin* **21**. International Potato Center, Lima, Peru.

Elphinstone J, Toth I, 2007. *Erwinia chrysanthemi* (*Dickeya* spp.) The Facts. British Potato Council Publications, Oxford. www.potato.org.uk.

English AC, Richards S, Han Y, Wang M, Vee V, Qu J, Qin X, Muzny DM, Reid JG, Worley KC, 2012. Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PloS one* **7**, e47768.

Evans TJ, Ind A, Komitopoulou E, Salmond GPC, 2010. Phage-selected lipopolysaccharide mutants of *Pectobacterium atrosepticum* exhibit different impacts on virulence. *Journal of Applied Microbiology* **109**, 505-14.

Ewing B, Green P, 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome research* **8**, 186-94.

Expert D, 1999. Withholding and exchanging iron: interactions between *Erwinia* spp. and their plant hosts. *Annual Review of Phytopathology* **37**, 307-34.

Expert D, O'brian MR, 2012. Molecular Aspects of Iron Metabolism in Pathogenic and Symbiotic Plant-microbe Associations. *SpringerBriefs in Molecular Science*, 2191-5407.

Fakruddin M, Chowdhury A, 2012. Pyrosequencing: An alternative to traditional sanger sequencing. *American Journal of Biochemistry and Biotechnology* **8**, 14-20.

Fall S, Mercier A, Bertolla F, Calteau A, Gueguen L, Perriere G, Vogel TM, Simonet P, 2007. Horizontal gene transfer regulation in bacteria as a "spandrel" of DNA repair mechanisms. *PloS one* **2**, e1055.

Farizo KM, Cafarella TG, Burns DL, 1996. Evidence for a ninth gene, ptlI, in the locus encoding the pertussis toxin secretion system of *Bordetella pertussis* and formation of a PtlI-PtlF complex. *Journal of Biological Chemistry* **271**, 31643-1649.

Federici L, Walas F, Luisi B, 2004. The structure and mechanism of the TolC outer membrane transport protein. *Current science-bangalore-* **87**, 190-96.

Ferrante P, Scortichini M, 2010. Molecular and phenotypic features of *Pseudomonas syringae* pv. *actinidiae* isolated during recent epidemics of bacterial canker on yellow kiwifruit (*Actinidia chinensis*) in central Italy. *Plant pathology* **59**, 954-62.

Fessehaie A, De Boer SH, Lévesque CA, 2002. Molecular characterization of DNA encoding 16S 23S rRNA intergenic spacer regions and 16S rRNA of pectolytic *Erwinia* species. *Canadian journal of microbiology* **48**, 387-98.

Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, 2009. The Pfam protein families database. *Nucleic acids research*, gkp985.

Fitzpatrick DA, 2009. Lines of evidence for horizontal gene transfer of a phenazine producing operon into multiple bacterial species. *Journal of molecular evolution* **68**, 171-85.

Flannery EL, 2011. *Contribution of Horizontal Gene Transfer to Virulence and Antibiotic Resistance in Pathogens Transmitted in Healthcare Facilities*: University of Michigan.

Flego D, Pirhonen M, Saarilahti H, Palva TK, Palva ET, 1997. Control of virulence gene expression by plant calcium in the phytopathogen *Erwinia carotovora*. *Molecular microbiology* **25**, 831-38.

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb J-F, Dougherty BA, Merrick JM, 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496-512.

Foerstner KU, Von Mering C, Hooper SD, Bork P, 2005. Environments shape the nucleotide composition of genomes. *EMBO reports* **6**, 1208-213.

Forst CV, Schulten K, 2001. Phylogenetic analysis of metabolic pathways. *Journal of molecular evolution* **52**, 471-89.

Forward JA, Behrendt MC, Wyborn NR, Cross R, Kelly DJ, 1997. TRAP transporters: a new family of periplasmic solute transport systems encoded by the dctPQM genes of *Rhodobacter capsulatus* and by homologs in diverse gram-negative bacteria. *Journal of bacteriology* **179**, 5482-493.

Francia M, Varsaki A, Garcillán-Barcia MP, Latorre A, Drainas C, Cruz F, 2004. A classification scheme for mobilization regions of bacterial plasmids. *FEMS microbiology reviews* **28**, 79-100.

Franza T, Mahé B, Expert D, 2005. *Erwinia chrysanthemi* requires a second iron transport route dependent of the siderophore achromobactin for extracellular growth and plant infection. *Molecular microbiology* **55**, 261-75.

Franza T, Michaud-Soret I, Piquerel P, Expert D, 2002. Coupling of iron assimilation and pectinolysis in *Erwinia chrysanthemi* 3937. *Molecular plant-microbe interactions* **15**, 1181-191.

Franza T, Sauvage C, Expert D, 1999. Iron regulation and pathogenicity in *Erwinia chrysanthemi* 3937: role of the Fur repressor protein. *Molecular plant-microbe interactions* **12**, 119-28.

Fraser CM, Eisen JA, Nelson KE, Paulsen IT, Salzberg SL, 2002. The value of complete microbial genome sequencing (you get what you pay for). *Journal of bacteriology* **184**, 6403-405.

Fraser CM, Fleischmann RD, 1997. Strategies for whole microbial genome sequencing and analysis. *Electrophoresis* **18**, 1207-216.

Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM, 1995. The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**, 397-404.

Frechon D, Exbrayat P, Helias V, Hyman LJ, Jouan B, Llop P, Lopez MM, Payet N, Perombelon MCM, Toth IK, 1998. Evaluation of a PCR kit for the detection of *Erwinia carotovora* subsp. *atroseptica* on potato tubers. *Potato research* **41**, 163-73.

Frederick RD, Ahmad M, Majerczak DR, Arroyo-Rodríguez AS, Manulis S, Coplin DL, 2001. Genetic organization of the *Pantoea stewartii* subsp. *stewartii* hrp gene cluster and sequence analysis of the *hrpA*, *hrpC*, *hrpN*, and *wtsE* operons. *Molecular plant-microbe interactions* **14**, 1213-222.

Frost LS, Ippen-Ihler K, Skurray RA, 1994. Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiological reviews* **58**, 162-210.

Frost LS, Leplae R, Summers AO, Toussaint A, 2005. Mobile genetic elements: the agents of open source evolution. *Nature Reviews Microbiology* **3**, 722-32.

Fujimori S, Hirai N, Ohashi H, Masuoka K, Nishikimi A, Fukui Y, Washio T, Oshikubo T, Yamashita T, Miyamoto-Sato E, 2012. Next-generation sequencing coupled with a cell-free display technology for high-throughput production of reliable interactome data. *Scientific reports* **2, 691**.

Fujiwara A, Kawasaki T, Usami S, Fujie M, Yamada T, 2008. Genomic characterization of *Ralstonia solanacearum* phage ϕRSA1 and its related prophage (ϕRSX) in strain GMI1000. *Journal of bacteriology* **190**, 143-56.

Fukiya S, Mizoguchi H, Tobe T, Mori H, 2004. Extensive genomic diversity in pathogenic *Escherichia coli* and Shigella strains revealed by comparative genomic hybridization microarray. *Journal of bacteriology* **186**, 3911-921.

Fuqua WC, Winans SC, Greenberg EP, 1994. Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. *Journal of bacteriology* **176**, 269.

Furuta Y, Abe K, Kobayashi I, 2010. Genome comparison and context analysis reveals putative mobile forms of restriction–modification systems and related rearrangements. *Nucleic acids research* **38**, 2428-443.

Galán JE, Collmer A, 1999. Type III secretion machines: bacterial devices for protein delivery into host cells. *Science* **284**, 1322-328.

Gallois A, Samson R, Ageron E, Grimont PAD, 1992. *Erwinia carotovora* subsp. *odorifera* subsp. nov., associated with odorous soft rot of chicory (*Cichorium intybus* L.). *International Journal of Systematic Bacteriology* **42**, 582-88.

Garcia CA, Ahmadian A, Gharizadeh B, Lundeberg J, Ronaghi M, Nyrén P, 2000. Mutation detection by pyrosequencing: sequencing of exons 5–8 of the p53 tumor suppressor gene. *Gene* **253**, 249-57.

Gardan L, Gouy C, Christen R, Samson R, 2003. Elevation of three subspecies of *Pectobacterium carotovorum* to species level: *Pectobacterium atrosepticum* sp nov., *Pectobacterium betavasculorum* sp nov and *Pectobacterium wasabiae* sp nov. *International Journal of Systematic and Evolutionary Microbiology* **53**, 381-91.

Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, Wilkinson AC, Finn RD, Griffiths-Jones S, Eddy SR, 2009. Rfam: updates to the RNA families database. *Nucleic acids research* **37**, D136-40.

Gardy JL, Laird MR, Chen F, Rey S, Walsh CJ, Ester M, Brinkman FSL, 2005. PSORTb v. 2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics* **21**, 617-23.

Garlant L, Koskinen P, Rouhiainen L, Laine P, Paulin L, Auvinen P, Holm L, Pirhonen M, 2013. Genome Sequence of *Dickeya solani*, a New soft Rot Pathogen of Potato, Suggests its Emergence May Be Related to a Novel Combination of Non-Ribosomal Peptide/Polyketide Synthetase Clusters. *Diversity* **5**, 824-42.

Gavini F, Mergaert J, Beji A, Mielcarek C, Izard D, Kersters K, Deley J, 1989. TRANSFER OF *ENTEROBACTER-AGGLOMERANS* (BEIJERINCK 1888) EWING AND FIFE 1972 TO *PANTOEA* GEN-NOV AS *PANTOEA-AGGLOMERANS* COMB NOV AND DESCRIPTION OF *PANTOEA-DISPERSA* SP-NOV. *International Journal of Systematic Bacteriology* **39**, 337-45.

Gay P, Le Coq D, Steinmetz M, Ferrari E, Hoch JA, 1983. Cloning structural gene *sacB*, which codes for exoenzyme levansucrase of *Bacillus subtilis*: expression of the gene in *Escherichia coli*. *Journal of bacteriology* **153**, 1424-431.

Gemski P, Lazere JR, Casey T, 1980. Plasmid associated with pathogenicity and calcium dependency of *Yersinia enterocolitica*. *Infection and immunity* **27**, 682-85.

Gharizadeh B, Ghaderi M, Nyren P, 2007. Pyrosequencing technology for short DNA sequencing and whole genome sequencing. *Technology* **47**, 129-32.

Gharizadeh B, Käller M, Nyrén P, Andersson A, Uhlén M, Lundeberg J, Ahmadian A, 2003. Viral and microbial genotyping by a combination of multiplex competitive hybridization and specific extension followed by hybridization to generic tag arrays. *Nucleic acids research* **31**, e146.

Giddens SR, Houliston GJ, Mahanty HK, 2003. The influence of antibiotic production and pre-emptive colonization on the population dynamics of *Pantoea agglomerans* (*Erwinia herbicola*) Eh1087 and *Erwinia amylovora in planta*. *Environmental microbiology* **5**, 1016-121.

Gilks WR, Audit B, De Angelis D, Tsoka S, Ouzounis CA, 2005. Percolation of annotation errors through hierarchically structured protein sequence databases. *Mathematical biosciences* **193**, 223-34.

Glasner JD, Marquez-Villavicencio M, Kim HS, Jahn CE, Ma B, Biehl BS, Rissman AI, Mole B, Yi X, Yang CH, 2008. Niche-specificity and the variable fraction of the *Pectobacterium* pan genome. *Molecular plant-microbe interactions* **21**, 1549-560.

Glazebrook J, 2005. Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annual Reviews of Phytopathology* **43**, 205-27.

Göhre V, Robatzek S, 2008. Breaking the barriers: microbial effector molecules subvert plant immunity. *Annu.al Reviews of. Phytopathology* **46**, 189-215.

Goodner B, Hinkle G, Gattung S, Miller N, Blanchard M, Qurollo B, Goldman BS, Cao Y, Askenazi M, Halling C, 2001. Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. *Science* **294**, 2323-328.

Goodey AR, Tubb RS, 1982. Genetioc and biochemical analysis of the ability of *Saccharomyces cerevisiaeb* to decarboxylate cinnamic acids. *Journal of General Mcirobiology* **128**, 2615-620.

Gorb TE, Tovkach FI, 1997. Typing *Erwinia carotovora* phytopathogenic strains on the basis of pectinolytic activity and sensitivity to bacteriocins (carotovoricins). *Microbiology* **66**, 690-95.

Görke B, Stülke J, 2008. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nature Reviews Microbiology* **6**, 613-24.

Goto M, Matsumoto K, 1987. *Erwinia carotovora* subsp. *wasabiae* subsp. nov. isolated from diseased rhizomes and fibrous roots of Japanese horseradish (Eutrema wasabi Maxim.). *International Journal of Systematic Bacteriology* **37**, 130-35.

Gotoh O, 1982. An improved algorithm for matching biological sequences. *Journal of molecular biology* **162**, 705-08.

Gowri VS, Pandit SB, Karthik PS, Srinivasan N, Balaji S, 2003. Integration of related sequences with protein three-dimensional structural families in an updated version of PALI database. *Nucleic acids research* **31**, 486-88.

Graham DC. Identification of soft rot coliform bacteria. *Proceedings of the Proceedings of the 3rd International Conference on Plant Pathogenic Bacteria*, *1972*, 273-79.

Grant SR, Fisher EJ, Chang JH, Mole BM, Dangl JL, 2006. Subterfuge and manipulation: type III effector proteins of phytopathogenic bacteria. *Annual Reviews of Phytopathology* **60**, 425-49.

Greenberg JT, Vinatzer BA, 2003. Identifying type III effectors of plant pathogens and analyzing their interaction with plant cells. *Current opinion in microbiology* **6**, 20-28.

Greenleaf WJ, Block SM, 2006. Single-molecule, motion-based DNA sequencing using RNA polymerase. *Science* **313**, 801.

Griffith F, 1928. The significance of *pneumococcal* types. *Journal of Hygiene* **27**, 113-59.

Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR, 2003. Rfam: an RNA family database. *Nucleic acids research* **31**, 439-41.

Grinter NJ, 1983. A broad-host-range cloning vector transposable to various replicons. *Gene* **21**, 133-43.

Grinter R, Milner J, Walker D, 2012. Ferredoxin containing bacteriocins suggest a novel mechanism of iron uptake in *Pectobacterium* spp. *PloS one* **7**, e33033.

Groisman EA, Ochman H, 1996. Pathogenicity islands: bacterial evolution in quantum leaps. *Cell* **87**, 791-94.

Gross DC, Powelson ML, Regner KM, Radamaker GK, 1991. A bacteriophage-typing system for surveying the diversity and distribution of strains of *Erwinia carotovora* in potato fields. *Phytopathology* **81**, 220-26.

Groth AC, Calos MP, 2004. Phage integrases: biology and applications. *Journal of molecular biology* **335**, 667-78.

Gulick AM, 2009. Ironing out a new siderophore synthesis strategy. *Nature chemical biology* **5**, 143-44.

Gury J, Barthelmebs L, Tran NP, Divies C, Cavin J-F, 2004. Cloning, deletion, and characterization of PadR, the transcriptional repressor of the phenolic acid decarboxylase-encoding *padA* gene of *Lactobacillus plantarum*. *Applied and environmental microbiology* **70**, 2146-153.

Hacker J, Blum-Oehler G, Mühldorfer I, Tschäpe H, 1997. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Molecular microbiology* **23**, 1089-097.

Hacker J, Carniel E, 2001. Ecological fitness, genomic islands and bacterial pathogenicity. *EMBO reports* **2**, 376-81.

Hacker J, Kaper JB, 2000. Pathogenicity islands and the evolution of microbes. *Annual Reviews in Microbiology* **54**, 641-79.

Harris TD, Buzby PR, Babcock H, Beer E, Bowers J, Braslavsky I, Causey M, Colonell J, Dimeo J, Efcavitch JW, 2008. Single-molecule DNA sequencing of a viral genome. *Science* **320**, 106-09.

Harrison EM, Carter MEK, Luck S, Ou H-Y, He X, Deng Z, O'callaghan C, Kadioglu A, Rajakumar K, 2010. Pathogenicity islands PAPI-1 and PAPI-2 contribute individually and synergistically to the virulence of *Pseudomonas aeruginosa* strain PA14. *Infection and immunity* **78**, 1437-446.

Hashimoto Y, Ezaki T, Li N, Yamamoto H, 1991. Molecular cloning of the *ViaB* region of *Salmonella typhi*. *FEMS microbiology letters* **90**, 53-56.

Hauben L, Moore ERB, Vauterin L, Steenackers M, Mergaert J, Verdonck L, Swings J, 1998. Phylogenetic position of phytopathogens within the *Enterobacteriaceae*. *Systematic and Applied Microbiology* **21**, 384-97.

Hayden KJ, Garbelotto M, Knaus BJ, Cronn RC, Rai H, Wright JW, 2014. Dual RNA-seq of the plant pathogen *Phytophthora ramorum* and its tanoak host. *Tree Genetics and Genomes* **10**, 489-502.

Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Umayam L, 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* **406**, 477-83.

Heilbronn J, Johnston DJ, Dunbar B, Lyon GD, 1995. Purification of a metalloprotease produced by *Erwinia carotovora* ssp. *carotovora* and the degradation of potato lectin *in vitro*. *Physiological and molecular plant pathology* **47**, 285-92.

Heithoff DM, Conner CP, Hentschel U, Govantes F, Hanna PC, Mahan MJ, 1999. Coordinate Intracellular Expression of *Salmonella* Genes Induced during Infection. *Journal of Bacteriology* **181**, 799–807.

Hélias V, Andrivon D, Jouan B, 2000. Internal colonization pathways of potato plants by *Erwinia carotovora* ssp. *atroseptica*. *Plant pathology* **49**, 33-42.

Hélias V, Hamon P, Huchet E, Wolf JVD, Andrivon D, 2012. Two new effective semiselective crystal violet pectate media for isolation of *Pectobacterium* and *Dickeya*. *Plant pathology* **61**, 339-45.

Hélias V, Le Roux A-C, Bertheau Y, Andrivon D, Gauthier J-P, Jouan B, 1998. Characterisation of *Erwinia carotovora* subspecies and detection of *Erwinia carotovora* subsp. *atroseptica* in potato plants, soil and water extracts with PCR-based methods. *European Journal of Plant Pathology* **104**, 685-99.

Henderson IR, Navarro-Garcia F, Desvaux M, Fernandez RC, Ala'aldeen D, 2004. Type V protein secretion pathway: the autotransporter story. *Microbiology and molecular biology reviews* **68**, 692-744.

Herrero M, De Lorenzo V, Timmis KN, 1990. Transposon vectors containing non-antibiotic resistance selection markers for cloning and stable chromosomal insertion of foreign genes in gram-negative bacteria. *Journal of bacteriology* **172**, 6557-567.

Heuermann D, Haas R, 1998. A stable shuttle vector system for efficient genetic complementation of *Helicobacter pylori* strains by transformation and conjugation. *Molecular and General Genetics MGG* **257**, 519-28.

Himpsl SD, Pearson MM, Arewång CJ, Nusca TD, Sherman DH, Mobley HLT, 2010. Proteobactin and a yersiniabactin-related siderophore mediate iron acquisition in *Proteus mirabilis*. *Molecular microbiology* **78**, 138-57.

Hofreuter D, Odenbreit S, Haas R, 2001. Natural transformation competence in *Helicobacter pylori* is mediated by the basic components of a type IV secretion system. *Molecular microbiology* **41**, 379-91.

Hogan CS, Mole BM, Grant SR, Willis DK, Charkowski AO, 2013. The type III secreted effector DspE is required early in *Solanum tuberosum* Leaf infection by *Pectobacterium carotovorum* to cause cell death, and requires Wx (3–6) D/E motifs. *PloS one* **8**, e65534.

Hogg JS, Hu FZ, Janto B, Boissy R, Hayes J, Keefe R, Post JC, Ehrlich GD, 2007. Characterization and modeling of the *Haemophilus influenzae* core and supragenomes based on the complete genomic sequences of Rd and 12 clinical nontypeable strains. *Genome Biology* **8**, R103.

Holeva MC, Bell KS, Hyman LJ, Avrova AO, Whisson SC, Birch PRJ, Toth IK, 2004. Use of a pooled transposon mutation grid to demonstrate roles in disease development for *Erwinia carotovora* subsp. *atroseptica* putative type III secreted effector (DspE/A) and helper (HrpN) proteins. *Molecular plant-microbe interactions* **17**, 943-50.

Hommais F, Oger-Desfeux C, Van Gijsegem F, Castang S, Ligori S, Expert D, Nasser W, Reverchon S, 2008. PecS is a global regulator of the symptomatic phase in the phytopathogenic bacterium *Erwinia chrysanthemi* 3937. *Journal of bacteriology* **190**, 7508-522.

Horner DS, Pavesi G, Castrignanò T, De Meo PDO, Liuni S, Sammeth M, Picardi E, Pesole G, 2009. Bioinformatics approaches for genomics and post genomics applications of next-generation sequencing. *Briefings in bioinformatics*, bbp046.

Hossain MM, Shibata S, Aizawa S-I, Tsuyumu S, 2005. Motility is an important determinant for pathogenesis of *Erwinia carotovora* subsp. *carotovora*. *Physiological and molecular plant pathology* **66**, 134-43.

Hoyos ME, Stanley CM, He SY, Pike S, Pu XA, Novacky A, 1996. The interaction of harpinPss, with plant cell walls. *MPMI-Molecular Plant Microbe Interactions* **9**, 608-16.

Hsiao WWL, Ung K, Aeschliman D, Bryan J, Finlay BB, Brinkman FSL, 2005. Evidence of a large novel gene pool associated with prokaryotic genomic islands. *PLoS Genetics* **1**, e62.

Huang X, Madan A, 1999. CAP3: A DNA sequence assembly program. *Genome research* **9**, 868-77.

Hueck CJ, 1998. Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiology and molecular biology reviews* **62**, 379-433.

Huelsenbeck JP, Ronquist F, 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-55.

Hugouvieux-Cotte-Pattat N, Condemine G, Nasser W, Reverchon S, 1996. Regulation of pectinolysis in *Erwinia chrysanthemi*. *Annual Reviews in Microbiology* **50**, 213-57.

Hugouvieux-Cotte-Pattat N, Robert-Baudouy J, 1987. Hexuronate catabolism in *Erwinia chrysanthemi*. *Journal of bacteriology* **169**, 1223-31.

Hugouvieux-Cotte-Pattat N, Robert-Baudouy J, 1989. Isolation of *Erwinia chrysanthemi* mutants altered in pectinolytic enzyme production. *Molecular microbiology* **3**, 1587-597.

Huguet-Tapia JC, Bignell DRD, Loria R, 2014. Characterization of the Integration and Modular Excision of the Integrative Conjugative Element PAISt in *Streptomyces turgidiscabies* Car8. *PloS one* **9**, e99345.

Hunt M, Newbold C, Berriman M, Otto TD, 2014. A comprehensive evaluation of assembly scaffolding tools. *Genome biology* **15**, R42.

Hyman ED, 1988. A new method of sequencing DNA. *Analytical biochemistry* **174**, 423-36.

Ibrahim A, Goebel BM, Liesack W, Griffiths M, Stackebrandt E, 1993. The phylogeny of the genus *Yersinia* based on 16S rDNA sequences. *FEMS microbiology letters* **114**, 173-77.

Ilyina TS, Romanova YM, 2002. Bacterial genomic islands: organization, function, and evolutionary role. *Journal of Molecular Biology* **36**, 171–79.

Inglis TJJ, Robertson T, Woods DE, Dutton N, Chang BJ, 2003. Flagellum-mediated adhesion by Burkholderia pseudomallei precedes invasion of *Acanthamoeba astronyxis*. *Infection and immunity* **71**, 2280-282.

Ishimaru CA, Loper JE, 1992. High-affinity iron uptake systems present in *Erwinia carotovora* subsp. *carotovora* include the hydroxamate siderophore aerobactin. *Journal of bacteriology* **174**, 2993-3003.

Ingram-Smith C, Miller KJ, 1998. Effects of ionic and osmotic strength on the glucosyltransferase of *Rhizobium meliloti* responsible for cyclic β-(1, 2)-glucan biosynthesis. *Applied and environmental microbiology* **64**, 1290-297.

Itoh Y, Izaki K, Takahashi H, 1978. Purification and characterization of a bacteriocin from *Erwinia carotovora*. *Journal of general and applied microbiology* **24**, 27-39.

Jack RW, Tagg JR, Ray B, 1995. Bacteriocins of gram-positive bacteria. *Microbiological reviews* **59**, 171-200.

Jackson RW, Johnson LJ, Clarke SR, Arnold DL, 2011. Bacterial pathogen evolution: breaking news. *Trends in genetics* **27**, 32-40.

Jacob-Dubuisson F, Fernandez R, Coutte L, 2004. Protein secretion through autotransporter and two-partner pathways. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **1694**, 235-57.

Janeway CA. Approaching the asymptote? Evolution and revolution in immunology. *Proceedings of the Cold Spring Harbor symposia on quantitative biology*, *1989*: Cold Spring Harbor Laboratory Press, 1-13.

Janssen P, Goldovsky L, Kunin V, Darzentas N, Ouzounis CA, 2005. Genome coverage, literally speaking. *EMBO reports* **6**, 397-99.

Jensen MA, Webster JA, Straus N, 1993. Rapid identification of bacteria on the basis of polymerase chain reaction-amplified ribosomal DNA spacer polymorphisms. *Applied and environmental microbiology* **59**, 945-52.

Jin Q, He S-Y, 2001. Role of the Hrp pilus in type III protein secretion in *Pseudomonas syringae*. *Science* **294**, 2556-558.

Johnson TL, Abendroth J, Hol WGJ, Sandkvist M, 2006. Type II secretion: from structure to function. *FEMS microbiology letters* **255**, 175-86.

Jones LR, 1901. A soft rot of carrot and other vegetables caused by *Bacillus carotovora*. *Jones Vt. Agric. Exp. Stat. Ann. Rept* **13**, 299-332.

Jones DaC, Mcleod A, Hyman LJ, Perombelon MCM, 1993. Specificity of an antiserum against *Erwinia carotovora* subsp. *atroseptica* in indirect ELISA. *Journal of Applied Bacteriology* **74**, 620-24.

Josenhans C, Suerbaum S, 2002. The role of motility as a virulence factor in bacteria. *International Journal of Medical Microbiology* **291**, 605-14.

Juhas M, Crook DW, Hood DW, 2008. Type IV secretion systems: tools of bacterial horizontal gene transfer and virulence. *Cellular microbiology* **10**, 2377-386.

Juhas M, Der Meer V, Roelof J, Gaillard M, Harding RM, Hood DW, Crook DW, 2009. Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS microbiology reviews* **33**, 376-93.

Kang H-W, Park D-S, Go S-J, Eun M-Y, 2002. Fingerprinting of diverse genomes using PCR with universal rice primers generated from repetitive sequence of Korean weedy rice. *Molecules and cells* **13**, 281-87.

Kang HW, Kwon SW, Go SJ, 2003. PCR-based specific and sensitive detection of *Pectobacterium carotovorum* ssp. *carotovorum* by primers generated from a URP-PCR fingerprinting-derived polymorphic band. *Plant pathology* **52**, 127-33.

Kaniga K, Delor I, Cornelis GR, 1991. A wide-host-range suicide vector for improving reverse genetics in Gram-negative bacteria: inactivation of the *blaA* gene of *Yersinia enterocolitica*. *Gene* **109**, 137-41.

Karlin S, 2001. Detecting anomalous gene clusters and pathogenicity islands in diverse bacterial genomes. *Trends in microbiology* **9**, 335-43.

Karlin S, Mrázek J, Campbell AM, 1998. Codon usages in different gene classes of the *Escherichia coli* genome. *Molecular microbiology* **29**, 1341-355.

Kay S, Bonas U, 2009. How *Xanthomonas* type III effectors manipulate the host plant. *Current opinion in microbiology* **12**, 37-43.

Keen EC, 2012. Paradigms of pathogenesis: targeting the mobile genetic elements of disease. *Frontiers in cellular and infection microbiology* **2, 161**.

Kelman A, Dickey RS, 1988. Soft rot of 'catotovora' group. In: Laboratoty guide for identification of Plant Pathogenic Bacteria. Second edition, St. Paul. *American Phytopathological Society*, 44-59.

Khoo SK, Loll B, Chan WT, Shoeman RL, Ngoo L, Yeo CC, Meinhart A, 2007. Molecular and structural characterization of the PezAT chromosomal toxin-antitoxin system of the human pathogen *Streptococcus pneumoniae*. *Journal of Biological Chemistry* **282**, 19606-9618.

Kim B-S, Yi H, Chun J, Cha C-J, 2014. Genome sequence of type strain of *Staphylococcus aureus* subsp. *aureus*. *Gut pathogens* **6**, 6.

Kim H-S, Ma B, Perna NT, Charkowski AO, 2009. Phylogeny and virulence of naturally occurring type III secretion system-deficient *Pectobacterium* strains. *Applied and environmental microbiology* **75**, 4539-49.

Kim H-S, Thammarat P, Lommel SA, Hogan CS, Charkowski AO, 2011. Pectobacterium carotovorum elicits plant cell death with DspE/F but the *P. carotovorum* DspE does not suppress callose or induce expression of plant genes early in plant-microbe interactions. *Molecular plant-microbe interactions* **24**, 773-86.

Kircher M, Kelso J, 2010. High-throughput DNA sequencing–concepts and limitations. *Bioessays* **32**, 524-36.

Klappenbach JA, Dunbar JM, Schmidt TM, 2000. rRNA operon copy number reflects ecological strategies of bacteria. *Applied and environmental microbiology* **66**, 1328-333.

Knirel YA, Zdorovenko GM, 1997. Structures of O-Polysaccharide Chains of Lipopolysaccharid es as the Basis for Classification of *Pseudomonas syringae* and Related Strains. In. *Pseudomonas syringae pathovars and related pathogens.* Springer, 475-80.

Kõiv V, Andresen L, Broberg M, Frolova J, Somervuo P, Auvinen P, Pirhonen M, Tenson T, Mäe A, 2013. Lack of RsmA-mediated control results in constant hypervirulence, cell elongation, and hyperflagellation in *Pectobacterium wasabiae*. *PloS one* **8**, e54248.

Koonin EV, Makarova KS, Aravind L, 2001. Horizontal gene transfer in prokaryotes: quantification and classification 1. *Annual Reviews in Microbiology* **55**, 709-42.

Korlach J, Bjornson KP, Chaudhuri BP, Cicero RL, Flusberg BA, Gray JJ, Holden D, Saxena R, Wegener J, Turner SW, 2010. Real-time DNA sequencing from single polymerase molecules. *Methods in enzymology* **472**, 431-55.

Kokjohn TA, 1989. Transduction: mechanism and potential for gene transfer in the environment. In: Levy SB, Miller RV (eds.), Gene Transfer in the Environment. McGraw-Hill, New York, 73-98.

Koskinen JP, Laine P, Niemi O, Nykyri J, Harjunpää H, Auvinen P, Paulin L, Pirhonen M, Palva T, Holm L, 2012. Genome sequence of *Pectobacterium* sp. strain SCC3193. *Journal of bacteriology* **194**, 21.

Koskiniemi S, Lamoureux JG, Nikolakakis KC, De Roodenbeke CTK, Kaplan MD, Low DA, Hayes CS, 2013. Rhs proteins from diverse bacteria mediate intercellular competition. *Proceedings of the National Academy of Sciences* **110**, 7032-037.

Kotoujansky A, 1987. Molecular genetics of pathogenesis by soft-rot Erwinias. *Annual Review of Phytopathology* **25**, 405-30.

Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL, 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology* **305**, 567-80.

Kundig W, Ghosh S, Roseman S, 1964. Phosphate bound to histidine in a protein as an intermediate in a novel phospho-transferase system. *Proc Natl Acad Sci U S A* **52**, 1067.

Kung VL, Ozer EA, Hauser AR, 2010. The accessory genome of *Pseudomonas aeruginosa*. *Microbiology and Molecular Biology Reviews* **74**, 621-41.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL, 2004. Versatile and open software for comparing large genomes. *Genome biology* **5**, R12.

Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW, 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic acids research* **35**, 3100-108.

Lahaye T, Bonas U, 2001. Molecular secrets of bacterial type III effector proteins. *Trends in plant science* **6**, 479-85.

Langille MGI, Hsiao WWL, Brinkman FSL, 2008. Evaluation of genomic island predictors using a comparative genomics approach. *BMC bioinformatics* **9**, 329.

Lanquar V, Lelièvre F, Bolte S, Hamès C, Alcon C, Neumann D, Vansuyt G, Curie C, Schröder A, Krämer U, 2005. Mobilization of vacuolar iron by AtNRAMP3 and AtNRAMP4 is essential for seed germination on low iron. *The EMBO journal* **24**, 4041-051.

Laurila J, Ahola V, Lehtinen A, Joutsjoki T, Hannukkala A, Rahkonen A, Pirhonen M, 2008. Characterization of *Dickeya* strains isolated from potato and river water samples in Finland. *European Journal of Plant Pathology* **122**, 213-25.

Lawrence JG, Ochman H, 1997. Amelioration of bacterial genomes: rates of change and exchange. *Journal of molecular evolution* **44**, 383-97.

Lederberg J, Tatum EL, 1946. Gene recombination in *Escherichia coli*. *Nature* **158**, 558.

Lee DH, Lim J-A, Lee J, Roh E, Jung K, Choi M, Oh C, Ryu S, Yun J, Heu S, 2013. Characterization of genes required for the pathogenicity of *Pectobacterium carotovorum* subsp. *carotovorum* Pcc21 in Chinese cabbage. *Microbiology* **159**, 1487-496.

Lee YW, Jin S, Sim WS, Nester EW, 1995. Genetic evidence for direct sensing of phenolic compounds by the virA protein of *Agrobacterium tumefacians*. *Proceedings of Natural Acadamies of Science,* **USA 92,** 12245-2249.

Lee WH, Vega VB, 2004. Heterogeneity detector: finding heterogeneous positions in Phred/Phrap assemblies. *Bioinformatics* **20**, 2863-864.

Lee Y-A, Yu C-P, 2006. A differential medium for the isolation and rapid identification of a plant soft rot pathogen, *Erwinia chrysanthemi*. *Journal of microbiological methods* **64**, 200-06.

Lehtimäki S, Rantakari A, Routtu J, Tuikkala A, Li J, Virtaharju O, Palva ET, Romantschuk M, Saarilahti HT, 2003. Characterization of the hrp pathogenicity cluster of *Erwinia carotovora* subsp. *carotovora*: high basal level expression in a mutant is associated with reduced virulence. *Molecular Genetics and Genomics* **270**, 263-72.

Leite LN, De Haan EG, Krijger M, Kastelein P, Van Der Zouwen PS, Van Den Bovenkamp GW, Tebaldi ND, Van Der Wolf JM, 2014. First report of potato blackleg caused by *Pectobacterium carotovorum* subsp. *brasiliensis* in the Netherlands. *New Disease Reports* **29**, 24.

Lelliott RA, Dickey RS, 1984. Genus VII. Erwinia. In: *Bergey's manual of systematic bacteriology*, Williams and Wilkins, Baltimore, USA, 469-476.

Létoffé S, Ghigo JM, Wandersman C, 1994. Secretion of the *Serratia marcescens* HasA protein by an ABC transporter. *Journal of bacteriology* **176**, 5372-737.

Lewis JA, Hatfull GF, 2001. Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins. *Nucleic acids research* **29**, 2205-216.

Li L, Stoeckert CJ, Roos DS, 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome research* **13**, 2178-189.

Li P-L, Everhart DM, Farrand SK, 1998. Genetic and sequence analysis of the pTiC58 trb locus, encoding a mating-pair formation system related to members of the type IV secretion family. *Journal of bacteriology* **180**, 6164-172.

Li Y, Peng Q, Selimi D, Wang Q, Charkowski AO, Chen X, Yang C-H, 2009. The plant phenolic compound p-coumaric acid represses gene expression in the *Dickeya dadantii* type III secretion system. *Applied and environmental microbiology* **75**, 1223-228.

Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, 2010. De novo assembly of human genomes with massively parallel short read sequencing. *Genome research* **20**, 265-72.

Lindeberg M, 2012. Genome-enabled perspectives on the composition, evolution, and expression of virulence determinants in bacterial plant pathogens. *Annual Review of Phytopathology* **50**, 111-32.

Lindow SE, Brandl MT, 2003. Microbiology of the phyllosphere. *Applied and environmental microbiology* **69**, 1875-883.

Liu H, Coulthurst SJ, Pritchard L, Hedley PE, Ravensdale M, Humphris S, Burr T, Takle G, Brurberg M-B, Birch PRJ, 2008. Quorum sensing coordinates brute force and stealth modes of infection in the plant pathogen *Pectobacterium atrosepticum*. *PLoS Pathogens* **4**, e1000093.

Liu Y, Cui Y, Mukherjee A, Chatterjee AK, 1998. Characterization of a novel RNA regulator of *Erwinia carotovora* ssp. *carotovora* that controls production of extracellular enzymes and secondary metabolites. *Molecular microbiology* **29**, 219-34.

Livermore DM, 1995. beta-Lactamases in laboratory and clinical resistance. *Clinical microbiology reviews* **8**, 557-84.

Llosa M, Gomis-Rüth FX, Coll M, Cruz FDL, 2002. Bacterial conjugation: a two-step mechanism for DNA transport. *Molecular microbiology* **45**, 1-8.

Llosa M, Zunzunegui S, De La Cruz F, 2003. Conjugative coupling proteins interact with cognate and heterologous VirB10-like proteins while exhibiting specificity for cognate relaxosomes. *Proceedings of the National Academy of Sciences* **100**, 10465-0470.

Llosa M, O'callaghan D, 2004. Euroconference on the Biology of Type IV Secretion Processes: bacterial gates into the outer world. *Molecular microbiology* **53**, 1-8.

Lorenz MG, Wackernagel W, 1994. Bacterial gene transfer by natural genetic transformation in the environment. *Microbiological reviews* **58**, 563.

Lovell HC, Jackson RW, Mansfield JW, Godfrey SaC, Hancock JT, Desikan R, Arnold DL, 2011. *In planta* conditions induce genomic changes in *Pseudomonas syringae* pv. *phaseolicola*. *Molecular Plant Pathology* **12**, 167-76.

Lovell HC, Mansfield JW, Godfrey SaC, Jackson RW, Hancock JT, Arnold DL, 2009. Bacterial evolution by genomic island transfer occurs via DNA transformation *in planta*. *Current Biology* **19**, 1586-590.

Lowe TM, Eddy SR, 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* **25**, 955-64.

Lumb VM, Perombelon MCM, Zutra D, 1986. Studies of a wilt disease of the potato plant in Israel caused by *Erwinia chrysanthemi*. *Plant pathology* **35**, 196-202.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 18.

Lwoff A, 1953. Lysogeny. *Bacteriological Reviews* **17**, 269.

Lynd LR, Weimer PJ, Van Zyl WH, Pretorius IS, 2002. Microbial cellulose utilization: fundamentals and biotechnology. *Microbiology and molecular biology reviews* **66**, 506-77.

Ma B, Hibbing ME, Kim H-S, Reedy RM, Yedidia I, Breuer J, Breuer J, Glasner JD, Perna NT, Kelman A, 2007a. Host range and molecular phylogenies of the soft rot enterobacterial genera *Pectobacterium* and *Dickeya*. *Phytopathology* **97**, 1150-163.

Maclean D, Jones JDG, Studholme DJ, 2009. Application of'next-generation'sequencing technologies to microbial genetics. *Nature Reviews Microbiology* **7**, 287-96.

Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K, 2012. Cluster: cluster analysis basics and extensions. *R package version* **1**.

Mahillon J, Chandler M, 1998. Insertion sequences. *Microbiology and molecular biology reviews* **62**, 725-74.

Main-Hester KL, Colpitts KM, Thomas GA, Fang FC, Libby SJ, 2008. Coordinate Regulation of Salmonella Pathogenicity Island 1 (SPI1) and SPI4 in *Salmonella enterica* Serovar Typhimurium. *Infection and Immunity* **76**, 1024–035.

Malachowa N, Deleo FR, 2010. Mobile genetic elements of *Staphylococcus aureus*. *Cellular and molecular life sciences* **67**, 3057-071.

Malnoy M, Faize M, Venisse J-S, Geider K, Chevreau E, 2005. Expression of viral EPS-depolymerase reduces fire blight susceptibility in transgenic pear. *Plant cell reports* **23**, 632-83.

Maloy SR, Nunn WD, 1981. Selection for loss of tetracycline resistance by *Escherichia coli*. *Journal of bacteriology* **145**, 1110-111.

Maltsev N, Glass E, Sulakhe D, Rodriguez A, Syed MH, Bompada T, Zhang Y, D'souza M, 2006. PUMA2—grid-based high-throughput analysis of genomes and metabolic pathways. *Nucleic acids research* **34**, D369-72.

Mandlik A, Livny J, Robins WP, Ritchie JM, Mekalanos JJ, Waldor MK, 2011. RNA-Seq-based monitoring of infection-linked changes in *Vibrio cholerae* gene expression. *Cell Host Microbe* **10**, 165–174.

Mann RA, Smits THM, Bühlmann A, Blom J, Goesmann A, Frey JE, Plummer KM, Beer SV, Luck J, Duffy B, 2013. Comparative genomics of 12 strains of *Erwinia amylovora* identifies a pan genome with a large conserved core. *PloS one* **8**, e55644.

Marcus SL, Brumell JH, Pfeifer CG, Finlay BB, 2000. *Salmonella* pathogenicity islands: big virulence in small packages. *Microbes and Infection* **2**, 145-56.

Mardis E, Mcpherson J, Martienssen R, Wilson RK, Mccombie WR, 2002. What is finished, and why does it matter. *Genome research* **12**, 669-71.

Mardis ER, 2008. Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics* **9**, 387-402.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z, 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-80.

Marits R, Köiv V, Laasik E, Mäe A, 1999. Isolation of an extracellular protease gene of *Erwinia carotovora* subsp. *carotovora* strain SCC3193 by transposon mutagenesis and the role of protease in phytopathogenicity. *Microbiology* **145**, 1959-966.

Marits R, Tshuikina M, Pirhonen M, Laasik E, Mäe A, 2002. Regulation of the expression of prtW:: gusA fusions in *Erwinia carotovora* subsp. *carotovora*. *Microbiology* **148**, 835-42.

Mattinen L, Nissinen R, Riipi T, Kalkkinen N, Pirhonen M, 2007. Host-extract induced changes in the secretome of the plant pathogenic bacterium *Pectobacterium atrosepticum*. *Proteomics* **7**, 3527-537.

Mattinen L, Tshuikina M, Mäe A, Pirhonen M, 2004. Identification and characterization of Nip, necrosis-inducing virulence protein of *Erwinia carotovora* subsp. *carotovora*. *Molecular plant-microbe interactions* **17**, 1366-375.

Mavrodi DV, Peever TL, Mavrodi OV, Parejko JA, Raaijmakers JM, Lemanceau P, Mazurier S, Heide L, Blankenfeldt W, Weller DM, 2010. Diversity and evolution of the phenazine biosynthesis pathway. *Applied and environmental microbiology* **76**, 866-79.

Mazel D, Dychinco B, Webb VA, Davies J, 1998. A distinctive class of integron in the *Vibrio cholerae* genome. *Science* **280**, 605-08.

Mazodier P, Davies J, 1991. Gene transfer between distantly related bacteria. *Annual review of genetics* **25**, 147-71.

Mazzaglia A, Studholme DJ, Taratufolo MC, Cai R, Almeida NF, Goodman T, Guttman DS, Vinatzer BA, Balestra GM, 2012. *Pseudomonas syringae* pv. *actinidiae* (PSA) isolates from recent bacterial canker of kiwifruit outbreaks belong to the same genetic lineage. *PloS one* **7**, e36518.

Mccann HC, Rikkerink EHA, Bertels F, Fiers M, Lu A, Rees-George J, Andersen MT, Gleave AP, Haubold B, Wohlers MW, 2013. Genomic analysis of the kiwifruit pathogen *Pseudomonas syringae* pv. *actinidiae* provides insight into the origins of an emergent plant disease. *PLoS Pathogens* **9**, e1003503.

Mccarthy AJ, Lindsay JA, 2012. The distribution of plasmids that carry virulence and resistance genes in *Staphylococcus aureus* is lineage associated. *BMC microbiology* **12**, 104.

Mccullagh P, Nelder JA, 1989. Generalized linear models. Second edition, London: Chapman and Hall.

Mcdonald JH, Kreitman M, 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* **351**, 652-54.

Mcevoy CR, Seshadri R, Firgaira FA, 1998. Large DNA fragment sizing using native acrylamide gels on an automated DNA sequencer and GENESCAN software. *BioTechniques* **25**, 464-70.

Mcgowan S, Sebaihia M, Jones S, Yu B, Bainton N, Chan PF, Bycroft B, Stewart G, Williams P, Salmond GPC, 1995. Carbapenem antibiotic production in *Erwinia carotovora* is regulated by CarR, a homologue of the LuxR transcriptional activator. *Microbiology* **141**, 541-50.

Mckenna F, El-Tarabily KA, Petrie S, Chen C, Dell B, 2002. Application of actinomycetes to soil to ameliorate water repellency. *Letters in applied microbiology* **35**, 107-12.

Médigue C, Moszer I, 2007. Annotation, comparison and databases for hundreds of bacterial genomes. *Research in microbiology* **158**, 724-36.

Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R, 2005. The microbial pan genome. *Current opinion in genetics & development* **15**, 589-94.

Meinhart A, Alonso JC, Sträter N, Saenger W, 2003. Crystal structure of the plasmid maintenance system ε/ζ: Functional mechanism of toxin ζ and inactivation by ε2ζ2 complex formation. *Proceedings of the National Academy of Sciences* **100**, 1661-666.

Melamede RJ, 1989. Automatable process for sequencing nucleotide. US Patent (4,863,849).

Mellgren EM, Kloek AP, Kunkel BN, 2009. Mqo, a tricarboxylic acid cycle enzyme, is required for virulence of *Pseudomonas syringae* pv. tomato strain DC3000 on *Arabidopsis thaliana*. *Journal of bacteriology* **191**, 3132-141.

Meng Y, Zhang Q, Ding W, Shan W, 2014. Phytophthora parasitica: a model oomycete plant pathogen. *Mycology*, 1-9.

Merriman B, Torrent I, Rothberg JM, Team D, 2012. Progress in Ion Torrent semiconductor chip based sequencing. *Electrophoresis* **33**, 3397-417.

Metzker ML, 2009. Sequencing technologies—the next generation. *Nature Reviews Genetics* **11**, 31-46.

Michel-Briand Y, Baysse C, 2002. The pyocins of *Pseudomonas aeruginosa*. *Biochimie* **84**, 499-510.

Miller JR, Koren S, Sutton G, 2010. Assembly algorithms for next-generation sequencing data. *Genomics* **95**, 315-27.

Milne I, Bayer M, Cardle L, Shaw P, Stephen G, Wright F, Marshall D, 2010. Tablet—next generation sequence assembly visualization. *Bioinformatics* **26**, 401-02.

Mithöfer A, Fliegmann J, Daxberger A, Ebel C, Neuhaus-Url G, Bhagwat AA, Keister DL, Ebel J, 2001. Induction of H_2O_2 synthesis by β-glucan elicitors in soybean is independent of cytosolic calcium transients. *FEBS letters* **508**, 191-95.

Mole B, Habibi S, Dangl JL, Grant SR, 2010. Gluconate metabolism is required for virulence of the soft-rot pathogen *Pectobacterium carotovorum*. *Molecular plant-microbe interactions* **23**, 1335-344.

Moleleki LN, Onkendi EM, Mongae A, Kubheka GC, 2013. Characterisation of *Pectobacterium wasabiae* causing blackleg and soft rot diseases in South Africa. *European Journal of Plant Pathology* **135**, 279-88.

Molina JJ, Harrison MD, 1977. The role of *Erwinia carotovora* in the epidemiology of potato blackleg I. Relationship of *E*. *carotorora* var. *carotovora* and *E*. *carotovora* var. *atroseptica* to potato blackleg in Colorado. *American Potato Journal* **54**, 587-91.

Molina JJ, Harrison MD, 1980. The role of *Erwinia carotovora* in the epidemiology of potato blackleg. II. The effect of soil temperature on disease severity. *American Potato Journal* **57**, 351-63.

Mora L, Klepsch M, Buckingham RH, Heurgué-Hamard V, Kervestin S, De Zamaroczy M, 2008. Dual roles of the central domain of colicin D tRNase in TonB-mediated import and in immunity. *Journal of Biological Chemistry* **283**, 4993-5003.

Morales G, Wiehlmann L, Gudowius P, Van Delden C, Tümmler B, Martínez JL, Rojo F, 2004. Structure of *Pseudomonas aeruginosa* populations analyzed by single nucleotide polymorphism and pulsed-field gel electrophoresis genotyping. *Journal of bacteriology* **186**, 4228-237.

Moreira LM, Almeida NF, Potnis N, Digiampietri LA, Adi SS, Bortolossi JC, Da Silva AC, Da Silva AM, De Moraes FE, De Oliveira JC, 2010. Novel insights into the genomic basis of citrus canker based on the genome sequences of two strains of *Xanthomonas fuscans* subsp. *aurantifolii*. *BMC genomics* **11**, 238.

Morello JE, Collmer A, 2009. Pseudomonas syringae HrpP is a type III secretion substrate specificity switch domain protein that is translocated into plant cells but functions atypically for a substrate-switching protein. *Journal of bacteriology* **191**, 3120-131.

Morrissey J, Guerinot ML, 2009. Iron uptake and transport in plants: the good, the bad, and the ionome. *Chemical reviews* **109**, 4553-567.

Mougous JD, Cuff ME, Raunser S, Shen A, Zhou M, Gifford CA, Goodman AL, Joachimiak G, Ordoñez CL, Lory S, 2006. A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. *Science* **312**, 1526-130.

Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Barrell D, Bateman A, Binns D, Biswas M, Bradley P, Bork P, 2003. The InterPro Database, 2003 brings increased coverage and new features. *Nucleic acids research* **31**, 315-18.

Mulholland V, Hinton JCD, Sidebotham J, Toth IK, Hyman LJ, Perombelon M, Reeves PJ, Salmond GPC, 1993. A pleiotropic reduced virulence (*Rvi*−) mutant of *Erwinia carotovora* subspecies *atroseptica* is defective in flagella assembly proteins that are conserved in plant and animal bacterial pathogens. *Molecular microbiology* **9**, 343-56.

Muñoz-López M, García-Pérez JL, 2010. DNA transposons: nature and applications in genomics. *Current genomics* **11**, 115.

Musto H, Naya H, Zavala A, Romero H, Alvarez-Valín F, Bernardi G, 2006. Genomic GC level, optimal growth temperature, and genome size in prokaryotes. *Biochemical and biophysical research communications* **347**, 1-3.

Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, Flanigan MJ, Kravitz SA, Mobarry CM, Reinert KHJ, Remington KA, 2000. A whole-genome assembly of *Drosophila*. *Science* **287**, 2196-204.

Nabhan S, Boer SH, Maiss E, Wydra K, 2012a. Taxonomic relatedness between *Pectobacterium carotovorum* subsp. *carotovorum*, *Pectobacterium carotovorum* subsp. *odoriferum* and *Pectobacterium carotovorum* subsp. *brasiliense* subsp. nov. *Journal of Applied Microbiology* **113**, 904-13.

Nabhan S, Wydra K, Linde M, Debener T, 2012b. The use of two complementary DNA assays, AFLP and MLSA, for epidemic and phylogenetic studies of pectolytic enterobacterial strains with focus on the heterogeneous species *Pectobacterium carotovorum*. *Plant pathology* **61**, 498-508.

Nakai K, Horton P, 1999. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends in biochemical sciences* **24**, 34-35.

Nakamura Y, Itoh T, Matsuda H, Gojobori T, 2004. Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nature Genetics* **36**, 760-66.

Nassar A, Darrasse A, Lemattre M, Kotoujansky A, Dervin C, Vedel R, Bertheau Y, 1996. Characterization of *Erwinia chrysanthemi* by pectinolytic isozyme polymorphism and restriction fragment length polymorphism analysis of PCR-amplified fragments of *pel* genes. *Applied and environmental microbiology* **62**, 2228-235.

Nasser W, Reverchon S, Condemine G, Robert-Baudouy J, 1994. Specific Interactions of *Erwinia chrysanthemi* KdgR Repressor with Different Operators of Genes Involved in Pectinolysis. *Journal of molecular biology* **236**, 427-40.

Naum M, Brown EW, Mason-Gamer RJ, 2008. Is 16S rDNA a reliable phylogenetic marker to characterize relationships below the family level in the enterobacteriaceae? *Journal of molecular evolution* **66**, 630-42.

Needleman SB, Wunsch CD, 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of molecular biology* **48**, 443-53.

Neilan BA, Dittmann, Rouhiainen L, Bass RA, Schaub V, Sivonen K, Borner T, 1999. Nonribosomal peptide synthesis and toxigenicity of cyanobacteria. *Journal of Bacteriology* **181**, 4089-097.

Newton JA, Fray RG, 2004. Integration of environmental and host-derived signals with quorum sensing during plant–microbe interactions. *Cellular microbiology* **6**, 213-24.

Ngadze E, Icishahayo D, Coutinho TA, Van Der Waals JE, 2012. Role of polyphenol oxidase, peroxidase, phenylalanine ammonia lyase, chlorogenic acid, and total soluble phenols in resistance of potatoes to soft rot. *Plant Disease* **96**, 186-92.

Nguyen AH, Tomita T, Hirota M, Sato T, Kamio Y, 1999. A simple purification method and morphology and component analyses for carotovoricin Er, a phage-tail-like bacteriocin from the plant pathogen *Erwinia carotovora* Er. *Bioscience, biotechnology, and biochemistry* **63**, 1360-369.

Nguyen HA, Kaneko J, Kamio Y, 2002. Temperature-dependent production of carotovoricin Er and pectin lyase in phytopathogenic *Erwinia carotovora* subsp. *carotovora* Er. *Bioscience, biotechnology, and biochemistry* **66**, 444-47.

Nguyen HA, Tomita T, Hirota M, Kaneko J, Hayashi T, Kamio Y, 2001. DNA inversion in the tail fiber gene alters the host range specificity of carotovoricin Er, a phage-tail-like bacteriocin of phytopathogenic *Erwinia carotovora* subsp. *carotovora* Er. *Journal of bacteriology* **183**, 6274-281.

Nguyen TKC, Tran NP, Cavin J-F, 2011. Genetic and biochemical analysis of PadR-padC promoter interactions during the phenolic acid stress response in *Bacillus subtilis* 168. *Journal of bacteriology* **193**, 4180-191.

Nielubowicz GR, Smith SN, Mobley HLT, 2008. Outer membrane antigens of the uropathogen *Proteus mirabilis* recognized by the humoral response during experimental murine urinary tract infection. *Infection and immunity* **76**, 4222-231.

Nikaido H, Vaara M, 1985. Molecular basis of bacterial outer membrane permeability. *Microbiological reviews* **49**, 1.

Nordström T, Ronaghi M, Forsberg L, De Faire U, Morgenstern R, Nyrén P, 2000. Direct analysis of single-nucleotide polymorphism on double-stranded DNA by pyrosequencing. *Biotechnology and applied biochemistry* **31**, 107-12.

Novick RP, 2003. Mobile genetic elements and bacterial toxinoses: the superantigen-encoding pathogenicity islands of *Staphylococcus aureus*. *Plasmid* **49**, 93-105.

Nykyri J, Niemi O, Koskinen P, Nokso-Koivisto J, Pasanen M, Broberg M, Plyusnin I, Törönen P, Holm L, Pirhonen M, 2012. Revised phylogeny and novel horizontally acquired virulence determinants of the model soft rot phytopathogen *Pectobacterium wasabiae* SCC3193. *PLoS Pathogens* **8**, e1003013.

Ochman H, Lawrence JG, Groisman EA, 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299-304.

Oerke EC, 2006. Crop losses to pests. *The Journal of Agricultural Science* **144**, 31-43.

Oh CS, Kim JF, Beer SV, 2005. The Hrp pathogenicity island of *Erwinia amylovora* and identification of three novel genes required for systemic infection‡. *Molecular Plant Pathology* **6**, 125-38.

Ohnishi M, Kurokawa K, Hayashi T, 2001. Diversification of *Escherichia coli* genomes: are bacteriophages the major contributors? *Trends in microbiology* **9**, 481-85.

Okamoto S, Van Petegem F, Patrauchan MA, Eltis LD, 2010. AnhE, a metallochaperone involved in the maturation of a cobalt-dependent nitrile hydratase. *Journal of Biological Chemistry* **285**, 25126-5133.

Okinaka Y, Yang C-H, Perna NT, Keen NT, 2002. Microarray profiling of *Erwinia chrysanthemi* 3937 genes that are regulated during plant infection. *Molecular plant-microbe interactions* **15**, 619-29.

Oliver DC, Huang G, Fernandez RC, 2003. Identification of secretion determinants of the *Bordetella pertussis* BrkA autotransporter. *Journal of bacteriology* **185**, 489-95.

Omori K, Idei A, 2003. Gram-negative bacterial ATP-binding cassette protein exporter family and diverse secretory proteins. *Journal of Bioscience and Bioengineering* **95**, 1-12.

Onkendi EM, Maluleke LN, Moleleki LN, 2014. First Report of *Pectobacterium carotovorum* subsp. *brasiliense* Causing Soft Rot and Blackleg of Potatoes in Kenya. *Plant Disease* **98**, 684.

Palacios JL, Zaror I, MartíNez P, Uribe F, Opazo P, SocíAs T, Gidekel M, Venegas A, 2001. Subset of hybrid eukaryotic proteins is exported by the type I secretion system of *Erwinia chrysanthemi*. *Journal of bacteriology* **183**, 1346-358.

Palmer DA, Bender CL, 1995. Ultrastructure of tomato leaf tissue treated with the pseudomonad phytotoxin coronatine and comparison with methyl jasmonate. *MPMI-Molecular Plant Microbe Interactions* **8**, 683-92.

Palva TK, Holmstrôm K-O, Heino P, Palva ET, 1993. Induction of plant defense response by exoenzymes of *Erwinia carotovora* subsp. *carotovora*. *Molecular Plant Microbe Interactions* **6**, 190-96.

Parish T, Stoker NG, 2000. Use of a flexible cassette method to generate a double unmarked *Mycobacterium tuberculosis* tlyA plcABC mutant by gene replacement. *Microbiology* **146**, 1969-975.

Parker WL, Rathnum ML, Wells Jr JS, Trejo WH, Principe PA, Sykes RB, 1982. SQ 27,860, a simple carbapenem produced by species of *Serratia* and *Erwinia*. *The Journal of antibiotics* **35**, 653-60.

Parkhill J, Dougan G, James KD, Thomson NR, Pickard D, Wain J, Churcher C, Mungall KL, Bentley SD, Holden MTG, 2001. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* **413**, 848-52.

Parkinson N, DeVos P, Pirhonen M, Elphinstone J, 2014. *Dickeya aquatica* sp. nov., isolated from waterways. *International Journal of Systematic and Evolutionary Microbiology* **64**, 2264-266.

Pasanen M, Laurila J, Brader G, Palva ET, Ahola V, Van Der Wolf J, Hannukkala A, Pirhonen M, 2013. Characterisation of *Pectobacterium wasabiae* and *Pectobacterium carotovorum* subsp. *carotovorum* isolates from diseased potato plants in Finland. *Annals of Applied Biology* **163**, 403-19.

Payne JH, Schoedel C, Keen NT, Collmer A, 1987. Multiplication and virulence in plant tissues of *Escherichia coli* clones producing pectate lyase isozymes PLb and PLe at high levels and of an *Erwinia chrysanthemi* mutant deficient in PLe. *Applied and environmental microbiology* **53**, 2315-320.

Pearson LA, Hisbergues M, Börner T, Dittmann E, Neilan BA, 2004. Inactivation of an ABC transporter gene, mcyH, results in loss of microcystin production in the cyanobacterium *Microcystis aeruginosa* PCC 7806. *Applied and environmental microbiology* **70**, 6370-378.

Pédron J, Mondy S, Des Essarts YR, Van Gijsegem F, Faure D, 2014. Genomic and metabolic comparison with *Dickeya dadantii* 3937 reveals the emerging *Dickeya solani* potato pathogen to display distinctive metabolic activities and T5SS/T6SS-related toxin repertoire. *BMC genomics* **15**, 283.

Pei AY, Oberdorf WE, Nossa CW, Agarwal A, Chokshi P, Gerz EA, Jin Z, Lee P, Yang L, Poles M, 2010. Diversity of 16S rRNA genes within individual prokaryotic genomes. *Applied and environmental microbiology* **76**, 3886-897.

Pelicic V, 2008. Type IV pili: e pluribus unum? *Molecular microbiology* **68**, 827-37.

Peltzer S, Sivasithamparam K, 1985. Soft-rot erwinias and stem rots in potatoes. *Animal Production Science* **25**, 693-96.

Pemberton CL, Whitehead NA, Sebaihia M, Bell KS, Hyman LJ, Harris SJ, Matlin AJ, Robson ND, Birch PRJ, Carr JP, 2005. Novel quorum-sensing-controlled genes in *Erwinia carotovora* subsp. *carotovora*: identification of a fungal elicitor homologue in a soft-rotting bacterium. *Molecular plant-microbe interactions* **18**, 343-53.

Penfold RJ, Pemberton JM, 1992. An improved suicide vector for construction of chromosomal insertion mutations in bacteria. *Gene* **118**, 145-46.

Pérez-Mendoza D, Coulthurst SJ, Humphris S, Campbell E, Welch M, Toth IK, Salmond GPC, 2011. A multi-repeat adhesin of the phytopathogen, *Pectobacterium atrosepticum*, is secreted by a Type I pathway and is subject to complex regulation involving a non-canonical diguanylate cyclase. *Molecular microbiology* **82**, 719-33.

Pérombelon MCM, 1992. Potato blackleg: epidemiology, host-pathogen interaction and control. *Netherlands Journal of Plant Pathology* **98**, 135-46.

Pérombelon MCM, 2002. Potato diseases caused by soft rot erwiniae: an overview of pathogenesis. *Plant pathology* **51**, 1-12.

Perombelon MCM, Kelman A, 1980. Ecology of the soft rot erwiniae. *Annual Review of Phytopathology* **18**, 361-87.

Pérombelon MCM, Kelman A, 1987. Blackleg and other potato diseases caused by soft rot erwiniae: proposal for revision of terminology. *Plant Disease* **71**, 283-85.

Pérombelon MCM, Lopez MM, Carbonell J, Hyman LJ, 1988. Effects of contamination by *Erwinia carotovora* subsp. *carotovora* and *E*. *carotovora* subsp. *atroseptica* of potato seed tubers and of cultivar resistance on blanking or nonemergence and blackleg development in Valencia, Spain. *Potato research* **31**, 591-99.

Pérombelon MCM, Salmond GPC, 1995. Bacterial soft rots. *in Pathogenesis and Host Specificity in Plant Diseases, US Singh, RP Singh, and K. Kohmoto, Eds* **1**, 1-20.

Pérombelon MCM, Van Der Wolf JM, 2002. Methods for the detection and quantification of *Erwinia carotovora* subsp. *atroseptica* (*Pectobacterium carotovorum* subsp. *atrosepticum*) on potatoes: a laboratory manual. *Dundee, Scotland: Scottish Crop Research Institute Occasional Publication*.

Petnicki-Ocwieja T, Schneider DJ, Tam VC, Chancey ST, Shan L, Jamir Y, Schechter LM, Janes MD, Buell CR, Tang X, 2002. Genomewide identification of proteins secreted by the Hrp type III protein

secretion system of *Pseudomonas syringae* pv. tomato DC3000. *Proceedings of the National Academy of Sciences* **99**, 7652-657.

Petnicki-Ocwieja T, Van Dijk K, Alfano JR, 2005. The hrpK operon of *Pseudomonas syringae* pv. tomato DC3000 encodes two proteins secreted by the type III (Hrp) protein secretion system: HopB1 and HrpK, a putative type III translocator. *Journal of bacteriology* **187**, 649-63.

Phoebe Lostroh C, Lee CA, 2001. The *Salmonella* pathogenicity island-1 type III secretion system. *Microbes and Infection* **3**, 1281-291.

Pickard D, Wain J, Baker S, Line A, Chohan S, Fookes M, Barron A, Gaora PÓ, Chabalgoity JA, Thanky N, 2003. Composition, acquisition, and distribution of the Vi exopolysaccharide-encoding *Salmonella enterica* pathogenicity island SPI-7. *Journal of bacteriology* **185**, 5055-065.

Pirhonen M, Flego D, Heikinheimo R, Palva ET, 1993. A small diffusible signal molecule is responsible for the global control of virulence and exoenzyme production in the plant pathogen *Erwinia carotovora*. *The EMBO journal* **12**, 2467.

Pirhonen M, Saarilahti H, Karlsson M-B, Palva ET, 1991. Identification of pathogenicity determinants of *Erwinia carotovora* subsp. *carotovora* by transposon mutagenesis. *Molecular Plant Microbe Interactions* **4**, 276-83.

Pitman AR, Harrow SA, Visnovsky SB, 2010. Genetic characterisation of *Pectobacterium wasabiae* causing soft rot disease of potato in New Zealand. *European Journal of Plant Pathology* **126**, 423-35.

Pitman AR, Jackson RW, Mansfield JW, Kaitell V, Thwaites R, Arnold DL, 2005. Exposure to host resistance mechanisms drives evolution of bacterial virulence in plants. *Current Biology* **15**, 2230-235.

Pitman AR, Wright PJ, Galbraith MD, Harrow SA, 2008. Biochemical and genetic diversity of pectolytic enterobacteria causing soft rot disease of potatoes in New Zealand. *Australasian Plant Pathology* **37**, 559-68.

Põllumaa L, Alamäe T, Mäe A, 2012. Quorum sensing and expression of virulence in pectobacteria. *Sensors* **12**, 3327-349.

Poole SJ, Diner EJ, Aoki SK, Braaten BA, De Roodenbeke CTK, Low DA, Hayes CS, 2011. Identification of functional toxin/immunity genes linked to contact-dependent growth inhibition (CDI) and rearrangement hotspot (Rhs) systems. *PLoS Genetics* **7**, e1002217.

Pop M, 2009. Genome assembly reborn: recent computational challenges. *Briefings in bioinformatics* **10**, 354-66.

Pop M, Salzberg SL, 2008. Bioinformatics challenges of new sequencing technology. *Trends in genetics* **24**, 142-49.

Porreca GJ, Shendure J, Church GM, 2006. Polony DNA sequencing. *Current Protocols in Molecular Biology* **76**, 7.8:7.8.1–7.8.22

Postma PW, Lengeler JW, Jacobson GR, 1993. Phosphoenolpyruvate: carbohydrate phosphotransferase systems of bacteria. *Microbiological reviews* **57**, 543.

Pray L, Zhaurova K, 2008. Barbara McClintock and the discovery of jumping genes (transposons). *Nature Education* **1**, 169.

Prentki P, Krisch HM, 1984. In vitro insertional mutagenesis with a selectable DNA fragment. *Gene* **29**, 303-13.

Priefert H, Rabenhorst J, Steinbuchel A, 1997. Molecular characterization of genes of *Pesudomonas* sp. strain HR199 involved in bioconversion of vanillin to protocatechuate. Journal of *Bacteriology* **179**, 2595-607.

Pukatzki S, Ma AT, Revel AT, Sturtevant D, Mekalanos JJ, 2007. Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. *Proceedings of the National Academy of Sciences* **104**, 15508-5513.

Pukatzki S, Ma AT, Sturtevant D, Krastins B, Sarracino D, Nelson WC, Heidelberg JF, Mekalanos JJ, 2006. Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the *Dictyostelium* host model system. *Proceedings of Natural Academy of Sciences U S A* **103**, 1528-533.

Py B, Barras F, Harris S, Robson N, Salmond GPC, 1998. 5.5 Extracellular Enzymes and Their Role in *Erwinia* Virulence. *Methods in Microbiology* **27**, 157-68.

Qiu X, Gurkar AU, Lory S, 2006. Interstrain transfer of the large pathogenicity island (PAPI-1) of *Pseudomonas aeruginosa*. *Proceedings of the National Academy of Sciences* **103**, 19830-9835.

Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R, 2005. InterProScan: protein domains identifier. *Nucleic acids research* **33**, W116-20.

Ragan MA, 2001. Detection of lateral gene transfer among microbial genomes. *Current opinion in genetics & development* **11**, 620-26.

Ramsay JP, Sullivan JT, Stuart GS, Lamont IL, Ronson CW, 2006. Excision and transfer of the *Mesorhizobium loti* R7A symbiosis island requires an integrase IntS, a novel recombination directionality factor RdfS, and a putative relaxase RlxS. *Molecular microbiology* **62**, 723-34.

Rankin DJ, Rocha EPC, Brown SP, 2010. What traits are carried on mobile genetic elements, and why&quest. *Heredity (Edinb)* **106**, 1-10.

Rantakari A, Virtaharju O, Vähämiko S, Taira S, Palva ET, Saarilahti HT, Romantschuk M, 2001. Type III secretion contributes to the pathogenesis of the soft-rot pathogen *Erwinia carotovora*: partial characterization of the hrp gene cluster. *Molecular plant-microbe interactions* **14**, 962-68.

Rashid A, Fahad MaB, Khan MA, Mateen A, Farooq M, Ashraf M, Ahmad F, Latif M, Ahmad M, 2012. Incidence of potato blackleg caused by *Pectobacterium atrosepticum* in district Chiniot and its management through bio-products. *African Journal of Agricultural Research* **7**, 6035-048.

Rasko DA, Rosovitz MJ, Myers GSA, Mongodin EF, Fricke WF, Gajer P, Crabtree J, Sebaihia M, Thomson NR, Chaudhuri R, 2008. The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E*. *coli* commensal and pathogenic isolates. *Journal of bacteriology* **190**, 6881-893.

Ratledge C, Dover LG, 2000. Iron metabolism in pathogenic bacteria. *Annual Reviews in Microbiology* **54**, 881-941.

Ravatn R, Studer S, Zehnder AJB, Van Der Meer JR, 1998. Int-B13, an Unusual Site-Specific Recombinase of the Bacteriophage P4 Integrase Family, Is Responsible for Chromosomal Insertion of the 105-Kilobase clc Element of *Pseudomonas* sp. Strain B13. *Journal of bacteriology* **180**, 5505-514.

Ravirala RS, Barabote RD, Wheeler DM, Reverchon S, Tatum O, Malouf J, Liu H, Pritchard L, Hedley PE, Birch PRJ, 2007. Efflux pump gene expression in *Erwinia chrysanthemi* is induced by exposure to phenolic acids. *Molecular plant-microbe interactions* **20**, 313-20.

Reed JL, Patel TR, Chen KH, Joyce AR, Applebee MK, Herring CD, Bui OT, Knight EM, Fong SS, Palsson BO, 2006. Systems approach to refining genome annotation. *Proceedings of the National Academy of Sciences* **103**, 17480-7484.

Reeves PJ, Whitcombe D, Wharam S, Gibson M, Allison G, Bunce N, Barallon R, Douglas P, Mulholland V, Stevens S, 1993. Molecular cloning and characterization of 13 out genes from *Erwinia carotovora* subspecies *carotovora*: genes encoding members of a general secretion pathway (GSP) widespread in Gram-negative bacteria. *Molecular microbiology* **8**, 443-56.

Remenant B, Coupat-Goutaland B, Guidot A, Cellier G, Wicker E, Allen C, Fegan M, Pruvost O, Elbaz M, Calteau A, 2010. Genomes of three tomato pathogens within the *Ralstonia solanacearum* species complex reveal significant evolutionary divergence. *BMC genomics* **11**, 379.

Reyrat J-M, Pelicic V, Gicquel B, Rappuoli R, 1998. Counterselectable markers: untapped tools for bacterial genetics and pathogenesis. *Infection and immunity* **66**, 4011-017.

Richardson EJ, Watson M, 2013. The automatic annotation of bacterial genomes. *Briefings in bioinformatics* **14**, 1-12.

Richter M, Rosselló-Móra R, 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences* **106**, 19126-9131.

Rietschel ET, Brade H, 1992. Bacterial endotoxins. *Scientific American* **267**, 54-61.

Rigano LA, Payette C, Brouillard G, Marano MR, Abramowicz L, Torres PS, Yun M, Castagnaro AP, El Oirdi M, Dufour V, 2007. Bacterial cyclic β-(1, 2)-glucan acts in systemic suppression of plant immune responses. *The Plant Cell Online* **19**, 2077-089.

Riley MA, 1998. Molecular mechanisms of bacteriocin evolution. *Annual review of genetics* **32**, 255-78.

Riley MA, Wertz JE, 2002. Bacteriocins: evolution, ecology, and application. *Annual Reviews in Microbiology* **56**, 117-37.

Roach JC, Boysen C, Wang K, Hood L, 1995. Pairwise end sequencing: a unified approach to genomic mapping and sequencing. *Genomics* **26**, 345-53.

Roberts P, Weingartner P, Kucharek, 2007. 2007 Florida plant disease management guide: potato, Irish, Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, University of Florida.

Rodríguez-Palenzuela P, Matas IM, Murillo J, López-Solanilla E, Bardaji L, Pérez-Martínez I, Rodríguez-Moskera ME, Penyalver R, López MM, Quesada JM, 2010. Annotation and overview of the *Pseudomonas savastanoi* pv. *savastanoi* NCPPB 3335 draft genome reveals the virulence gene complement of a tumour-inducing pathogen of woody hosts. *Environmental microbiology* **12**, 1604-620.

Roh E, Park T-H, Kim M-I, Lee S, Ryu S, Oh C-S, Rhee S, Kim D-H, Park B-S, Heu S, 2010. Characterization of a new bacteriocin, Carocin D, from *Pectobacterium carotovorum* subsp. *carotovorum* Pcc21. *Applied and environmental microbiology* **76**, 7541-549.

Roine E, Wei W, Yuan J, Nurmiaho-Lassila E-L, Kalkkinen N, Romantschuk M, He SY, 1997. Hrp pilus: an hrp-dependent bacterial surface appendage produced by *Pseudomonas syringae* pv. tomato DC3000. *Proceedings of the National Academy of Sciences* **94**, 3459-464.

Rojas CM, Ham JH, Deng W-L, Doyle JJ, Collmer A, 2002. HecA, a member of a class of adhesins produced by diverse pathogenic bacteria, contributes to the attachment, aggregation, epidermal cell killing, and virulence phenotypes of *Erwinia chrysanthemi* EC16 on *Nicotiana clevelandii* seedlings. *Proceedings of the National Academy of Sciences* **99**, 13142-3147.

Ronaghi M, 2001. Pyrosequencing sheds light on DNA sequencing. *Genome research* **11**, 3-11.

Ronaghi M, Nygren M, Lundeberg J, Nyrén P, 1999. Analyses of secondary structures in DNA by pyrosequencing. *Analytical biochemistry* **267**, 65-71.

Rose RE, 1988. The nucleotide sequence of pACYC184. *Nucleic acids research* **16**, 355.

Rosłoniec KZ, Wilbrink MH, Capyk JK, Mohn WW, Ostendorf M, Van Der Geize R, Dijkhuizen L, Eltis LD, 2009. Cytochrome P450 125 (CYP125) catalyses C26-hydroxylation to initiate sterol side-chain degradation in *Rhodococcus jostii* RHA1. *Molecular microbiology* **74**, 1031-043.

Ross H, 1986. Potato Breeding – Problems and Perspectives. Verlag Paul Parey, Berlin and Hamburg, Germany, 132.

Rowe RC, Miller A, Riedel RM, 1995. Blackleg, aerial stem rot and tuber soft rot of potato. The Ohio State University Extension Fact Sheet. Ohio State University, HYG-3106-95. http://ohioline.osu.edu/hyg-fact/3000/3106.html

Rudolph K, 2001. Thirty-five Years of Phytobacteriology Research with Special Emphasis on Pathogenicity of *Pseudomonas syringae*. In. *Plant pathogenic bacteria.* Springer, 109-17.

Rutberg B, 1997. Antitermination of transcription of catabolic operons. *Molecular microbiology* **23**, 413-21.

Salanoubat M, Genin S, Artiguenave F, Gouzy J, Mangenot S, Arlat M, Billault A, Brottier P, Camus JC, Cattolico L, 2002. Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature* **415**, 497-502.

Salmond GPC, Bycroft BW, Stewart G, Williams P, 1995. The bacterial 'enigma': cracking the code of cell–cell communication. *Molecular microbiology* **16**, 615-24.

Salvucci ME, 2003. Distinct sucrose isomerases catalyze trehalulose synthesis in whiteflies, *Bemisia argentifolii*, and *Erwinia rhapontici*. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **135**, 385-95.

Sam MD, Cascio D, Johnson RC, Clubb RT, 2004. Crystal structure of the excisionase-DNA complex from bacteriophage lambda. *Journal of Molecular Biology* **338(2),** 229–240.

Samson R, Legendre JB, Christen R, Fischer-Le Saux M, Achouak W, Gardan L, 2005. Transfer of *Pectobacterium chrysanthemi* (Burkholder et al. 1953) Brenner et al. 1973 and *Brenneria paradisiaca* to the genus *Dickeya* gen. nov as *Dickeya chrysanthemi* comb. nov and *Dickeya paradisiaca* comb. nov and delineation of four novel species, *Dickeya dadantii* sp nov., *Dickeya dianthicola* sp nov., *Dickeya dieffenbachiae* sp nov and *Dickeya zeae* sp nov. *International Journal of Systematic and Evolutionary Microbiology* **55**, 1415-427.

Sanger F, Coulson AR, 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of molecular biology* **94**, 441-48.

Sarker MR, Cornelis GR, 1997. An improved version of suicide vector pKNG101 for gene replacement in Gram-negative bacteria. *Molecular microbiology* **23**, 410-11.

Sasakawa C, Kamata K, Sakai T, Makino S, Yamada M, Okada N, Yoshikawa M, 1988. Virulence-associated genetic regions comprising 31 kilobases of the 230-kilobase plasmid in *Shigella flexneri* 2a. *Journal of bacteriology* **170**, 2480-484.

Saurin W, Hofnung M, Dassa E, 1999. Getting in or out: Early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters. *Journal of Molecular Evolution* **48**, 22-41.

Sawyer SA, Hartl DL, 1992. Population genetics of polymorphism and divergence. *Genetics* **132**, 1161-176.

Schadt EE, Turner S, Kasarskis A, 2010. A window into third-generation sequencing. *Human molecular genetics* **19**, R227-40.

Schatz MC, Delcher AL, Salzberg SL, 2010. Assembly of large genomes using second-generation sequencing. *Genome research* **20**, 1165-173.

Schechter LM, Vencato M, Jordan KL, Schneider SE, Schneider DJ, Collmer A, 2006. Multiple approaches to a complete inventory of *Pseudomonas syringae* pv. tomato DC3000 type III secretion system effector proteins. *Molecular plant-microbe interactions* **19**, 1180-192.
Scheibye-Alsing K, Hoffmann S, Frankel A, Jensen P, Stadler PF, Mang Y, Tommerup N, Gilchrist MJ, Nygård AB, Cirera S, 2009. Sequence assembly. *Computational Biology and Chemistry* **33**, 121-36.

Schell MA, Ulrich RL, Ribot WJ, Brueggemann EE, Hines HB, Chen D, Lipscomb L, Kim HS, Mrázek J, Nierman WC, 2007. Type VI secretion is a major virulence determinant in *Burkholderia mallei*. *Molecular microbiology* **64**, 1466-485.

Schmidt H, Hensel M, 2004. Pathogenicity islands in bacterial pathogenesis. *Clinical microbiology reviews* **17**, 14-56.

Schmidt H, Zhang WL, Hemmrich U, Jelacic S, Brunder W, Tarr PI, Dobrindt U, Hacker J, Karch H, 2001. Identification and characterization of a novel genomic island integrated at selC in locus of enterocyte effacement-negative, Shiga toxin-producing *Escherichia coli*. *Infection and immunity* **69**, 6863-873.

Schneider J, Vorhölter F-J, Trost E, Blom J, Musa YR, Neuweger H, Niehaus K, Schatschneider S, Tauch A, Goesmann A, 2010. CARMEN-Comparative analysis and in silico reconstruction of organism-specific metabolic networks. *Genet Mol Res* **9**, 1660-672.

Schwecke T, Aparicio JF, Molnar I, König A, Khaw LE, Haydock SF, Oliynyk M, Caffrey P, Cortes J, Lester JB, 1995. The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proceedings of the National Academy of Sciences* **92**, 7839-843.

Serfontein S, Logan C, Swanepoel AE, Boelema BH, Theron DJ, 1991. A potato wilt disease in South Africa caused by *Erwinia carotovora* subspecies *carotovora* and *E*. *chrysanthemi*. *Plant pathology* **40**, 382-86.

Segura A, Bunz PV, Argenio DA, Omston LN, 1999. Genetic analysis of a chromosomal region containing *vanA* and *vanB*, genes required for conversion of either ferulate or vanillate to protocatechuate in *Acinetobacter*. *Journal of Bacteriology* **181**, 3494-504.

Seubert A, Hiestand R, De La Cruz F, Dehio C, 2003. A bacterial conjugation machinery recruited for pathogenesis. *Molecular microbiology* **49**, 1253-266.

Shapiro JA, 1969. Mutations caused by the insertion of genetic material into the galactose operon of *Escherichia coli*. *Journal of molecular biology* **40**, 93-105.

Shapiro JA, Adhya SL, 1969. The galactose operon of *E*. *coli* K-12. II. A deletion analysis of operon structure and polarity. *Genetics* **62**, 249.

Shaw JG, Hamblin MJ, Kelly DJ, 1991. Purification, characterization and nucleotide sequence of the periplasmic C4-dicarboxylate-binding protein (DctP) from *Rhodobacter capsulatus*. *Molecular microbiology* **5**, 3055-062.

Shen K, Sayeed S, Antalis P, Gladitz J, Ahmed A, Dice B, Janto B, Dopico R, Keefe R, Hayes J, Johnson S, Yu S, Ehrlich N, Jocz J, Kropp L, Wong R, Wadowsky R, Slifkin M, Preston R, Erdos G, Post JC, Ehrlich GD, Hu FZ, 2006. Extensive genomic plasticity in *Pseudomonas aeruginosa* revealed by identification and distribution studies of novel genes among clinical isolates. *Infection and Immunity* **74**, 5272-283.

Shendure J, Ji H, 2008. Next-generation DNA sequencing. *Nature biotechnology* **26**, 1135-145.

Shirasu K, Morel P, Kado CI, 1990. Characterization of the *virB* operon of an *Agrobacterium tumefaciens* Ti plasmid: nucleotide sequence and protein analysis. *Molecular microbiology* **4**, 1153-63.

Shrivastava S, Reddy CVSK, Mande SS, 2010. INDeGenIUS, a new method for high-throughput identification of specialized functional islands in completely sequenced organisms. *Journal of biosciences* **35**, 351-64.

Silipo A, Molinaro A, Sturiale L, Dow JM, Erbs G, Lanzetta R, Newman M-A, Parrilli M, 2005. The elicitation of plant innate immunity by lipooligosaccharide of *Xanthomonas campestris*. *Journal of Biological Chemistry* **280**, 33660-3668.

Silverman JM, Brunet YR, Cascales E, Mougous JD, 2012. Structure and regulation of the type VI secretion system. *Annual review of microbiology* **66**, 453.

Simon R, Priefer U, Pühler A, 1983. A broad host range mobilization system for *in vivo* genetic engineering: transposon mutagenesis in gram negative bacteria. *Nature biotechnology* **1**, 784-91.

Simpson AJG, Reinach FC, Arruda P, Abreu FA, Acencio M, Alvarenga R, Alves LMC, Araya JE, Baia GS, Baptista CS, 2000. The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature* **406**, 151-57.

Singh U, Trevors CM, De Boer SH, Janse JD, 2000. Fimbrial-specific monoclonal antibody-based ELISA for European potato strains of *Erwinia chrysanthemi* and comparison to PCR. *Plant Disease* **84**, 443-48.

Sjöblom S, Harjunpää H, Brader G, Palva ET, 2008. A novel plant ferredoxin-like protein and the regulator Hor are quorum-sensing targets in the plant pathogen *Erwinia carotovora*. *Molecular plant-microbe interactions* **21**, 967-78.

Skerman VBD, Mcgowan V, Sneath PHA, 1980. Approved lists of bacterial names. *International Journal of Systematic Bacteriology* **30**, 225-30.

Skorupski K, Taylor RK, 1996. Positive selection vectors for allelic exchange. *Gene* **169**, 47-52.
Skovgaard M, Jensen LJ, Brunak S, Ussery D, Krogh A, 2001. On the total number of genes and their length distribution in complete microbial genomes. *Trends in genetics* **17**, 425-28.

Slawiak M, Lojkowska E, 2009. Genes responsible for coronatine synthesis in *Pseudomonas syringae* present in the genome of soft rot bacteria. *European Journal of Plant Pathology* **124**, 353-61.

Sledz W, Jafra S, Waleron M, Lojkowska E, 2000. Genetic diversity of *Erwinia carotovora* strains isolated from infected plants grown in Poland. *OEPP/EPPO Bulletin* **30**, 403–407.

Smith C, Bartz JA, 1990. Variation in the pathogenicity and aggressiveness of strains of *Erwinia carotovora* subsp. *carotovora* isolated from different hosts. *Plant Disease* **74**, 505-09.

Smith TF, Waterman MS, 1981. Comparison of biosequences. *Advances in applied mathematics* **2**, 482-89.

Smoot JC, Barbian KD, Van Gompel JJ, Smoot LM, Sylva GL, Sturdevant DE, Ricklefs SM, Porcella SF, Parkins LD, Beres SB, 2002. Genome sequence and comparative microarray analysis of serotype M18 group A *Streptococcus* strains associated with acute rheumatic fever outbreaks. *Proceedings of the National Academy of Sciences* **99**, 4668-673.

Sonnhammer ELL, Eddy SR, Durbin R, 1997. Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins-Structure Function and Genetics* **28**, 405-20.

Spiridonov NA, Wilson DB, 2001. Cloning and biochemical characterization of BglC, a β-glucosidase from the cellulolytic actinomycete *Thermobifida fusca*. *Current microbiology* **42**, 295-301.

Spröer C, Mendrock U, Swiderski J, Lang E, Stackebrandt E, 1999. The phylogenetic position of *Serratia*, *Buttiauxella* and some other genera of the family *Enterobacteriaceae*. *International Journal of Systematic and Evolutionary Microbiology* **49**, 1433-438.

Staley JT, 2006. The bacterial species dilemma and the genomic–phylogenetic species concept. *Philosophical Transactions of the Royal Society B: Biological Sciences* **361**, 1899-909.

Stein MA, Leung KY, Zwick M, Portillo FGD, Finlay BB, 1996. Identification of a *Salmonella* virulence gene required for formation of filamentous structures containing lysosomal membrane glycoproteins within epithelial cells. *Molecular microbiology* **20**, 151-64.

Stokes HWT, Hall RM, 1989. A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. *Molecular microbiology* **3**, 1669-683.

Strom MS, Lory S, 1993. Structure-function and biogenesis of the type IV pili. *Annual Reviews in Microbiology* **47**, 565-96.
Sutcliffe IC, Russell RR, 1995. Lipoproteins of gram-positive bacteria. *Journal of bacteriology* **177**, 1123.

Talaga P, Fournet B, Bohin J-P, 1994. Periplasmic glucans of *Pseudomonas syringae* pv. *syringae*. *Journal of bacteriology* **176**, 6538-544.

Talaga P, Stahl B, Wieruszeski J-M, Hillenkamp F, Tsuyumu S, Lippens G, Bohin J-P, 1996. Cell-associated glucans of *Burkholderia solanacearum* and *Xanthomonas campestris* pv. *citri*: a new family of periplasmic glucans. *Journal of bacteriology* **178**, 2263-271.

Talavera G, Castresana J, 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic biology* **56**, 564-77.

Tam R, Saier MH, 1993. Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiological reviews* **57**, 320-46.

Tamas I, Klasson L, Canbäck B, Näslund AK, Eriksson A-S, Wernegreen JJ, Sandström JP, Moran NA, Andersson SGE, 2002. 50 million years of genomic stasis in endosymbiotic bacteria. *Science* **296**, 2376-379.

Tanaka H, Kawai T, 2009. Partial sequencing of a single DNA molecule with a scanning tunnelling microscope. *Nature nanotechnology* **4**, 518-22.

Tang X, Xiao Y, Zhou J-M, 2006. Regulation of the type III secretion system in phytopathogenic bacteria. *Molecular plant-microbe interactions* **19**, 1159-166.

Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan genome". *Proceedings of Natural Academy of Sciences U S A* **102**, 13950-3955.

Tettelin H, Nelson KE, Paulsen IT, Eisen JA, Read TD, Peterson S, Heidelberg J, Deboy RT, Haft DH, Dodson RJ, 2001. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* **293**, 498-506.

Tettelin H, Riley D, Cattuto C, Medini D, 2008. Comparative genomics: the bacterial pan genome. *Current opinion in microbiology* **11**, 472-77.

Thomas JD, Reeves PJ, Salmond GPC, 1997. The general secretion pathway of *Erwinia* i subsp. *carotovora*: analysis of the membrane topology of OutC and OutF. *Microbiology* **143**, 713-20.

Thomson NR, Thomas JD, Salmond GPC, 1999. 12 Virulence Determinants in the Bacterial Phytopathogen *Erwinia*. *Methods in Microbiology* **29**, 347-426.

Thomson SV, Hildebrand DC, Schroth MN, 1981. Identification and nutritional differentiation of the *Erwinia* sugar beet pathogen from members of *Erwinia carotovora* and *Erwinia chrysanthemi*. *Phytopathology* **71**, 1037-042.

Tock MR, Dryden DTF, 2005. The biology of restriction and anti-restriction. *Current opinion in microbiology* **8**, 466-72.

Torrence ME, Isaacson RE, 2008. *Microbial food safety in animal agriculture: current topics*. John Wiley & Sons.

Toth I, Avrova A, Hyman L, 2001. Rapid identification and differentiation of the soft rot erwiniae by 16S-23S intergenic transcribed spacer-PCR and restriction fragment length polymorphism analyses. *Applied and Environmental Microbiology* **67**, 4070-076.

Toth IK, Bell KS, Holeva MC, Birch PRJ, 2003. Soft rot erwiniae: from genes to genomes. *Molecular Plant Pathology* **4**, 17-30.

Toth IK, Bertheau Y, Hyman LJ, Laplaze L, López MM, Mcnicol J, Niepold F, Persson P, Salmond GPC, Sletten A, 1999. Evaluation of phenotypic and molecular typing techniques for determining diversity in *Erwinia carotovora* subsp. *atroseptica*. *Journal of Applied Microbiology* **87**, 770-81.

Toth IK, Birch PRJ, 2005. Rotting softly and stealthily. *Current opinion in plant biology* **8**, 424-29.

Toth IK, Pritchard L, Birch PRJ, 2006. Comparative genomics reveals what makes an enterobacterial plant pathogen. *Annual Review of Phytopathology* **44**, 305-36.

Toth IK, Van Der Wolf JM, Saddler G, Lojkowska E, Hélias V, Pirhonen M, Elphinstone JG, 2011. *Dickeya* species: an emerging problem for potato production in Europe. *Plant pathology* **60**, 385-99.

Tovkach FI, Mukvich NS, 2003. The Study of *Erwinia carotovora* Bacteriocins with the Aid of Nalidixic Acid–Resistant Bacterial Indicator Cells. *Microbiology* **72**, 167-72.

Tran NP, Gury J, Dartois V, Nguyen TKC, Seraut H, Barthelmebs L, Gervais P, Cavin J-F, 2008. Phenolic acid-mediated regulation of the *padC* gene, encoding the phenolic acid decarboxylase of *Bacillus subtilis*. *Journal of bacteriology* **190**, 3213-224.

Truernit E, Schmid J, Epple P, Illig J, Sauer N, 1996. The sink-specific and stress-regulated *Arabidopsis STP4* gene: enhanced expression of a gene encoding a monosaccharide transporter by wounding, elicitors, and pathogen challenge. *The Plant Cell Online* **8**, 2169-182.

Tseng T-T, Tyler BM, Setubal JC, 2009. Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC microbiology* **9**, S2.

Tsror L, Erlich O, Lebiush S, Hazanovsky M, Zig U, Slawiak M, Grabe G, Van Der Wolf JM, Van De Haar JJ, 2009. Assessment of recent outbreaks of *Dickeya* sp.(syn. *Erwinia chrysanthemi*) slow wilt in potato crops in Israel. *European Journal of Plant Pathology* **123**, 311-20.

Urbany C, Neuhaus HE, 2008. Citrate uptake into *Pectobacterium atrosepticum* is critical for bacterial virulence. *Molecular plant-microbe interactions* **21**, 547-54.

Ussery DW, Jensen MS, Poulsen TR, Hallin PF, 2004. Genome Update: alignment of bacterial chromosomes. *Microbiology* **150**, 2491-493.

Valentini M, Storelli N, Lapouge K, 2011. Identification of C4-Dicarboxylate Transport Systems in *Pseudomonas aeruginosa* PAO1. *Journal of bacteriology* **193**, 4307-316.

Van Der Geize R, Hessels GI, Van Gerwen R, Van Der Meijden P, Dijkhuizen L, 2001. Unmarked gene deletion mutagenesis of *kstD*, encoding encoding 3-ketosteroid Δ 1-dehydrogenase, *Rhodococcus erythropolis* SQ1 using *sacB* as counter-selectable marker. *FEMS microbiology letters* **205**, 197-202.

Van Der Merwe JJ, Coutinho TA, Korsten L, Van Der Waals JE, 2010. *Pectobacterium carotovorum* subsp. *brasiliensis* causing blackleg on potatoes in South Africa. *European Journal of Plant Pathology* **126**, 175-85.

Van Der Ploeg JR, 2005. Regulation of bacteriocin production in *Streptococcus mutans* by the quorum-sensing system required for development of genetic competence. *Journal of bacteriology* **187**, 3980-989.

Van Der Wolf JM, De Boer SH, 2007. Bacterial pathogens of potato. *Potato Biology and Biotechnology, Advances and Perspectives. Oxford, UK: Elsevier*, 595-619.

Van Gijsegem F, Vasse J, Camus JC, Marenda M, Boucher C, 2000. *Ralstonia solanacearum* produces Hrp-dependent pili that are required for PopA secretion but not for attachment of bacteria to plant cells. *Molecular microbiology* **36**, 249-60.

Van Loon LC, 1997. Induced resistance in plants and the role of pathogenesis-related proteins. *European Journal of Plant Pathology* **103**, 753-65.

Van Sluys MA, Monteiro-Vitorello CB, Camargo LEA, Menck CFM, Da Silva ACR, Ferro JA, Oliveira MC, Setubal JC, Kitajima JP, Simpson AJ, 2002. Comparative genomic analysis of plant-associated bacteria. *Annual Review of Phytopathology* **40**, 169-89.

Vanga BR, Butler RC, Toth IK, Ronson CW, Pitman AR, 2012. Inactivation of PbTopo IIIβ causes hyper-excision of the Pathogenicity Island HAI2 resulting in reduced virulence of *Pectobacterium atrosepticum*. *Molecular microbiology* **84**, 648-63.

Vanneste JL, Poliakoff F, Audusseau C, Cornish DA, Paillard S, Rivoal C, Yu J, 2012. First report of *Pseudomonas syringae* pv. *actinidiae*, the causal agent of bacterial canker of kiwifruit in France. *Molecular Plant Pathology* **13**, 631-40.

Varani AM, Monteiro-Vitorello CB, Nakaya HI, Van Sluys M-A, 2013. The role of prophage in plant-pathogenic bacteria. *Annual Review of Phytopathology* **51**, 429-51.

Vasu K, Nagaraja V, 2013. Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiology and molecular biology reviews* **77**, 53-72.

Vercoe RB, Chang JT, Dy RL, Taylor C, Gristwood T, Clulow JS, Richter C, Przybilski R, Pitman AR, Fineran PC, 2013. Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. *PLoS Genet* **9**, e1003454.

Vernikos GS, Parkhill J, 2006. Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. *Bioinformatics* **22**, 2196-203.

Vidaver AK, Lambrecht PA, 2004. Bacteria as plant pathogens. *The Plant Health Instructor*.

Vodovar N, Vallenet D, Cruveiller S, Rouy Z, Barbe V, Acosta C, Cattolico L, Jubin C, Lajus A, Segurens B, 2006. Complete genome sequence of the entomopathogenic and metabolically versatile soil bacterium *Pseudomonas entomophila*. *Nature biotechnology* **24**, 673-79.

Vojnov AA, Slater H, Newman M-A, Daniels MJ, Dow MJ, 2001. Regulation of the synthesis of cyclic glucan in *Xanthomonas campestris* by a diffusible signal molecule. *Archives of microbiology* **176**, 415-20.

Von Bodman SB, Bauer WD, Coplin DL, 2003. Quorum sensing in plant-pathogenic bacteria. *Annual Review of Phytopathology* **41**, 455-82.

Waack S, Keller O, Asper R, Brodag T, Damm C, Fricke WF, Surovcik K, Meinicke P, Merkl R, 2006. Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. *BMC bioinformatics* **7**, 142.

Wallis FM, Truter SJ, 1978. Histopathology of tomato plants infected with *Pseudomonas solanacearum*, with emphasis on ultrastructure.*Physiological Plant Pathology* **13(3**), 307-317.
Wagner PL, Waldor MK, 2002. Bacteriophage control of bacterial virulence. *Infection and immunity* **70**, 3985-93.

Waldor MK, Friedman DI, 2005. Phage regulatory circuits and virulence gene expression. *Current opinion in microbiology* **8**, 459-65.

Waleron M, Waleron K, Lojkowska E, 2013. Occurrence of *Pectobacterium wasabiae* in potato field samples. *European Journal of Plant Pathology* **137**, 149-58.
Waleron M, Waleron K, Lojkowska E, 2014. Characterization of *Pectobacterium carotovorum* subsp. *odoriferum* causing soft rot of stored vegetables. *European Journal of Plant Pathology* **139**, 457-69.

Wall D, Kaiser D, 1999. Type IV pili and cell motility. *Molecular microbiology* **32**, 01-10.

Wallace A, Pérombelon MCM, 1992. Haemagglutinins and fimbriae of soft rot erwiniae. *Journal of Applied Bacteriology* **73**, 114-19.

Wandersman C, Delepelaire P, Letoffe S, 1990. Secretion processing and activation of *Erwinia chrysanthemi* proteases. *Biochimie* **72**, 143-46.

Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, Jaenisch R, 2013. One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. *Cell* **153**, 910–918.

Ward LJ, De Boer SH, 1994. Specific detection of *Erwinia carotovora* subsp. *atroseptica* with a digoxigenin-labeled DNA probe. *Phytopathology* **84**, 180-86.

Weber E, Ojanen-Reuhs T, Huguet E, Hause G, Romantschuk M, Korhonen TK, Bonas U, Koebnik R, 2005. The type III-dependent Hrp pilus is required for productive interaction of *Xanthomonas campestris* pv. *vesicatoria* with pepper host plants. *Journal of bacteriology* **187**, 2458-468.

Wells JM, Moline HE, 1991. Differentiation of the Soft-Rotting Erwinias (the Carotovora Group) by Fatty Acid Composition. *Journal of Phytopathology* **131**, 22-32.

Whitehead NA, Byers JT, Commander P, Corbett MJ, Coulthurst SJ, Everson L, Harris AKP, Pemberton CL, Simpson NJL, Slater H, 2002. The regulation of virulence in phytopathogenic *Erwinia* species: quorum sensing, antibiotics and ecological considerations. *Antonie Van Leeuwenhoek* **81**, 223-31.

Williamson NR, Commander P, Salmond GPC, 2010. Quorum sensing-controlled Evr regulates a conserved cryptic pigment biosynthetic cluster and a novel phenomycin-like locus in the plant pathogen, *Pectobacterium carotovorum*. *Environmental microbiology* **12**, 1811-827.

Winans SC, Burns DL, Christie PJ, 1996. Adaptation of a conjugal transfer system for the export of pathogenic macromolecules. *Trends in microbiology* **4**, 64-68.

Winstanley C, Langille MGI, Fothergill JL, Kukavica-Ibrulj I, Paradis-Bleau C, Sanschagrin F, Thomson NR, Winsor GL, Quail MA, Lennard N, 2009. Newly introduced genomic prophage islands are critical determinants of *in vivo* competitiveness in the Liverpool Epidemic Strain of *Pseudomonas aeruginosa*. *Genome research* **19**, 12-23.

Wolfe KH, Li W-H, 2003. Molecular evolution meets the genomics revolution. *nature genetics* **33**, 255-65.

Wood DE, Lin H, Levy-Moonshine A, Swaminathan R, Chang Y-C, Anton BP, Osmani L, Steffen M, Kasif S, Salzberg SL, 2012. Thousands of missed genes found in bacterial genomes and their analysis with COMBREX. *Biol Direct* **7**, 37.

Wood DW, Setubal JC, Kaul R, Monks DE, Kitajima JP, Okura VK, Zhou Y, Chen L, Wood GE, Almeida NF, 2001. The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. *Science* **294**, 2317-323.

Wozniak RaF, Fouts DE, Spagnoletti M, Colombo MM, Ceccarelli D, Garriss G, Déry C, Burrus V, Waldor MK, 2009. Comparative ICE genomics: insights into the evolution of the SXT/R391 family of ICEs. *PLoS Genetics* **5**, e1000786.

Wu J, Zhang H-B, Xu J-L, Cox RJ, Simpson TJ, Zhang L-H, 2010. 13C labeling reveals multiple amination reactions in the biosynthesis of a novel polyketide polyamine antibiotic zeamine from *Dickeya zeae*. *Chemical Communications* **46**, 333-35.

Xu G-W, Gross DC, 1988. Evaluation of the role of syringomycin in plant pathogenesis by using Tn5 mutants of *Pseudomonas syringae* pv. *syringae* defective in syringomycin production. *Applied and environmental microbiology* **54**, 1345-353.

Xu X, Melcher J, Basak S, Reifenberger R, Raman A, 2009. Compositional contrast of biological materials in liquids using the momentary excitation of higher eigenmodes in dynamic atomic force microscopy. *Physical Review Letters* **102**, 060801.

Yahiaoui-Zaidi R, Jouan B, Andrivon D, 2003. Biochemical and molecular diversity among *Erwinia* isolates from potato in Algeria. *Plant pathology* **52**, 28-40.

Yamaguchi T, Hayashi T, Takami H, Ohnishi M, Murata T, Nakayama K, Asakawa K, Ohara M, Komatusuzawa H, Sugai M, 2001. Complete nucleotide sequence of a *Staphylococcus aureus* exfoliative toxin B plasmid and identification of a novel ADP-ribosyltransferase, EDIN-C. *Infection and immunity* **69**, 7760-771.

Yang S, Peng Q, Zhang Q, Zou L, Li Y, Robert C, Pritchard L, Liu H, Hovey R, Wang Q, 2010. Genome-wide identification of HrpL-regulated genes in the necrotrophic phytopathogen *Dickeya dadantii* 3937. *PloS one* **5**, e13472.

Yang X, Liu X, Song L, Xie F, Zhang G, Qian S, 2007. Characterization and functional analysis of a novel gene cluster involved in biphenyl degradation in *Rhodococcus* sp. strain R04. *Journal of Applied Microbiology* **103**, 2214-224.

Yap M-N, Barak JD, Charkowski AO, 2004. Genomic diversity of *Erwinia carotovora* subsp. *carotovora* and its correlation with virulence. *Applied and environmental microbiology* **70**, 3013-23.

Young JM, Park DC, 2007. Relationships of plant pathogenic enterobacteria based on partial *atpD*, *carA*, and *recA* as individual and concatenated nucleotide and peptide sequences. *Systematic and Applied Microbiology* **30**, 343-54.

Young JM, Saddler GS, Takikawa Y, De Boer SH, Vauterin L, Gardan L, Gvozdyak RI, Stead DE, 1996. Names of plant pathogenic bacteria 1864-1995. *Review of Plant Pathology* **75**, 721-63.

Yu J, Peñaloza-Vázquez A, Chakrabarty AM, Bender CL, 1999. Involvement of the exopolysaccharide alginate in the virulence and epiphytic fitness of *Pseudomonas syringae* pv. *syringae*. *Molecular microbiology* **33**, 712-20.

Zerbino DR, Birney E, 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research* **18**, 821-29.

Zhang H-B, Wang L-H, Zhang L-H, 2002. Genetic control of quorum-sensing signal turnover in *Agrobacterium tumefaciens*. *Proceedings of the National Academy of Sciences* **99**, 4638-643.

Zhang J, 2010. Evolutionary genetics: Progress and challenges. *Evolution Since Darwin: The First 150 Years*, 87-118.

Zhang J-H, Quigley NB, Gross DC, 1995. Analysis of the *syrB* and *syrC* genes of *Pseudomonas syringae* pv. *syringae* indicates that syringomycin is synthesized by a thiotemplate mechanism. *Journal of bacteriology* **177**, 4009-020.

Zheng J, Ho B, Mekalanos JJ, 2011. Genetic analysis of anti-amoebae and anti-bacterial activities of the type VI secretion system in *Vibrio cholerae*. *PloS one* **6**, e23876.

Zhou H, Pastore L, Beaudet A, 2002. Helper-dependent adenoviral vectors. *Methods in Enzymology* **346**, 177-198.

Zhou J, Zhang H, Wu J, Liu Q, Xi P, Lee J, Liao J, Jiang Z, Zhang L-H, 2011. A novel multidomain polyketide synthase is essential for zeamine production and the virulence of *Dickeya zeae*. *Molecular plant-microbe interactions* **24**, 1156-164.

Zhu J, Winans SC, 2001. The quorum-sensing transcriptional regulator TraR requires its cognate signaling ligand for protein folding, protease resistance, and dimerization. *Proceedings of the National Academy of Sciences* **98**, 1507-512.

Zink RT, Engwall JK, Mcevoy JL, Chatterjee AK, 1985. recA is required in the induction of pectin lyase and carotovoricin in *Erwinia carotovora* subsp. *carotovora*. *Journal of bacteriology* **164**, 390-96.

Zotchev SB, Schrempf H, Hutchinson CR, 1995. Identification of a methyl-specific restriction system mediated by a conjugative element from *Streptomyces bambergiensis*. *Journal of bacteriology* **177**, 4809-812.