DOCTOR OF PHILOSOPHY

Genomic analysis of bacterial species associated with Acute Oak Decline

Doonan, James

Award date:
2016

Awarding institution:
Bangor University

Genomic analysis of bacterial species associated with Acute Oak Decline

James Doonan

September 2016

PRIFYSGOL
BANGOR
UNIVERSITY

A thesis submitted to Bangor University in candidature
for the degree
Philosophiae Doctor

School of Biological Sciences
Bangor University, Deiniol Road, Bangor, LL57 2UW

Declaration and Consent

Details of the Work

I hereby agree to deposit the following item in the digital repository maintained by Bangor University and/or in any other repository authorized for use by Bangor University.

Author Name: James Doonan

Title: Genomic analysis of bacterial species associated with acute oak decline

Supervisor/Department: Dr James McDonald (School of Biological Sciences)

Funding body (if any): Woodland Heritage

Qualification/Degree obtained: PhD

This item is a product of my own research endeavours and is covered by the agreement below in which the item is referred to as "the Work". It is identical in content to that deposited in the Library, subject to point 4 below.

Non-exclusive Rights

Rights granted to the digital repository through this agreement are entirely non-exclusive. I am free to publish the Work in its present version or future versions elsewhere.

I agree that Bangor University may electronically store, copy or translate the Work to any approved medium or format for the purpose of future preservation and accessibility. Bangor University is not under any obligation to reproduce or display the Work in the same formats or resolutions in which it was originally deposited.

Bangor University Digital Repository

I understand that work deposited in the digital repository will be accessible to a wide variety of people and institutions, including automated agents and search engines via the World Wide Web.

I understand that once the Work is deposited, the item and its metadata may be incorporated into public access catalogues or services, national databases of electronic theses and dissertations such as the British Library's EThOS or any service provided by the National Library of Wales.

I understand that the Work may be made available via the National Library of Wales Online Electronic Theses Service under the declared terms and conditions of use (http://www.llgc.org.uk/index.php?id=4676). I agree that as part of this service the National Library of Wales may electronically store, copy or convert the Work to any approved medium or format for the purpose of future preservation and accessibility. The National Library of Wales is not under any obligation to reproduce or display the Work in the same formats or resolutions in which it was originally deposited.

Statement 1:

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree unless as agreed by the University for approved dual awards.

Signed ..... (candidate)

Date .....

Statement 2:

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction is clearly marked in a footnote(s).

All other sources are acknowledged by footnotes and/or a bibliography.

Signed ..... (candidate)

Date .....

Statement 3:

I hereby give consent for my thesis, if accepted, to be available for photocopying, for inter-library loan and for electronic storage (subject to any constraints as defined in statement 4), and for the title and summary to be made available to outside organisations.

Signed ..... (candidate)

Date .....

Statement 4:

Choose one of the following options

a)	I agree to deposit an electronic copy of my thesis (the Work) in the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorized for use by Bangor University and where necessary have gained the required permissions for the use of third party material.	√

b)	I agree to deposit an electronic copy of my thesis (the Work) in the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorized for use by Bangor University when the approved **bar on access** has been lifted.	
c)	I agree to submit my thesis (the Work) electronically via Bangor University's e-submission system, however I **opt-out** of the electronic deposit to the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorized for use by Bangor University, due to lack of permissions for use of third party material.	

Options B should only be used if a bar on access has been approved by the University.

In addition to the above I also agree to the following:

1. That I am the author or have the authority of the author(s) to make this agreement and do hereby give Bangor University the right to make available the Work in the way described above.
2. That the electronic copy of the Work deposited in the digital repository and covered by this agreement, is identical in content to the paper copy of the Work deposited in the Bangor University Library, subject to point 4 below.
3. That I have exercised reasonable care to ensure that the Work is original and, to the best of my knowledge, does not breach any laws – including those relating to defamation, libel and copyright.
4. That I have, in instances where the intellectual property of other authors or copyright holders is included in the Work, and where appropriate, gained explicit permission for the inclusion of that material in the Work, and in the electronic form of the Work as accessed through the open access digital repository, *or* that I have identified and removed that material for which adequate and appropriate permission has not been obtained and which will be inaccessible via the digital repository.
5. That Bangor University does not hold any obligation to take legal action on behalf of the Depositor, or other rights holders, in the event of a breach of intellectual property rights, or any other right, in the material deposited.
6. That I will indemnify and keep indemnified Bangor University and the National Library of Wales from and against any loss, liability, claim or damage, including without limitation any related legal fees and court costs (on a full indemnity bases), related to any breach by myself of any term of this agreement.

Signature: ..... Date : .....

Contents

List of Figures

List of Tables

Abbreviations

%	Percent
µl	Microliter
µm	Micrometre
^{0}C	Degrees Celsius
A	Adenine
AOD	Acute Oak Decline
Bg	*Brenneria goodwinii* FRB141
BLAST	Basic Local Alignment Search Tool
BLASTn	Nucleotide Basic Local Alignment Search Tool
bp	Base Pair
BR	Broad Range
C	Cytosine
CAZy	Carbohydrate Active Enzyme
CBM	Carbohydrate Binding Module
cDNA	Reverse-Transcribed Ribonucleic Acid
cT4SS	Conjugation Type IV Secretion System
COD	Chronic Oak Decline
cpDNA	Chloroplast DNA
Ct	Threshold Cycle
Da	Daltons
DEG	Differentially Expressed Genes
DNA	Deoxyribonucleic Acid
dNTP	Dinucleotide Triphosphate
dsDNA	Double Stranded Deoxyribonucleic Acid
EC	Enzyme Commission
g	Gram
G	Guanine
GSP	General Secretory Pathway
Gq	*Gibbsiella quercinecans* FRB97
HGT	Horizontal Gene Transfer
HPI	Hours Post Inoculation
HR	Hypersensitivity Response
hs	High Sensitivity

ITS1	Intergenic Transcribed Spacer Region 1
kb	Kilo Base Pairs
l	Litre
LGM	Last Glacial Maximum
Mbp	Mega Base Pairs
ml	Millilitre
mM	Milimole
NB-LRR	Nucleotide-Binding, Leucine Rich Repeat
ng	Nanogram
pT4SS	Protein Secretion Type IV Secretion System
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PCWDE	Plant Cell Wall Degrading Enzyme
qPCR	Quantitative Polymerase Chain Reaction
RD	Recruitment-Destruction
RISA	Ribosomal Intergenic Spacer Analysis
RNA	Ribonucleic Acid
RNA-seq	RNA sequencing
rpm	Revolutions Per Minute
RPKM	Reads Per Kilobase of Gene Per Million Reads Mapped
rRNA	Ribosomal Ribonucleic Acid
SRE	Soft-Rot *Enterobacteriaceae*
SOD	Sudden Oak Death
spp.	Species
subsp.	Sub-species
(T)	Type Strain
T1SS	Type I Secretion System
T2SS	Type II Secretion System
T3SS	Type III Secretion System
T4SS	Type IV Secretion System
T6SS	Type IV Secretion System
WGS	Whole Genome Sequencing
w/v	Weight to Volume Ratio

Acknowledgements

This thesis is dedicated to my parents Jim & Mary Doonan, and my sister Claire. My parents stressed the importance of education and encouraged me throughout my life, which may have appeared to have fallen on deaf ears but ultimately seeped through, this Ph.D thesis is testament to that.

To my wife and best friend Georgina I would not have been able to do this without you.

I owe a huge debt of gratitude to my supervisor Dr James McDonald who is a terrific scientist, a fantastic supervisor and a great guy, a rare combination of characteristics.

I would like to thank members of James McDonald's lab group both past and present, Dr Emma Ransom-Jones for her unique advice on all aspects of life, Dr Tracy Perkins for her Mancunian take on everyday events, and not forgetting, Mallory Diggens, Dr Martin Broberg, and Dr Francis Hassard. Thanks to friends and colleagues from other labs within Bangor University including Dr Tran Hai, Dr Shila Ramayah, Dale Armstrong, Jason Williams and Dr Ade Fewings at HPC Wales.

A huge thanks to Dr Sandra Denman at Forest Research, for her belief in me throughout my Ph.D, which has been a constant source of encouragement, and her fearless commitment to understanding the cause of AOD, no matter the challenges presented. Also, thanks for some great help and really enjoyable conference trips around the UK, and further afield to the Czech Republic, France and Sweden, along with the rest of the Forest Research team, Sarah Plummer, Dr Nathan Brown, Susan Kirk, Dr Maciej Kaczmarek and Dr Gavin Hunter. Furthermore, thanks to the wider AOD research group, Dr Carrie Brady, Prof. Dawn Arnold and Dr Melanie Sapp.

In addition, I would like to thank Dr Fiona Hannah formerly of the gone but never forgotten UMBSM, on the Isle of Cumbrae, and Dr Jim Provan formerly of Queens University Belfast and now at Aberystwyth University.

Finally, I am very grateful for the receipt of a Woodland Heritage Ph.D studentship.

Foreword to the thesis

Upon commencement of my Ph.D project in September 2012, AOD had been identified by Dr Sandra Denman and the Forest Research team at Alice Holt, Surrey. In 2012, the complex nature of AOD may have been imagined but the complexities were not fully realised. To unravel this complex disease, Dr Denman, defined several collaborative projects to tackle the different specificities of the project. For example, the distribution of the beetle *Agrilus biguttatus* in relation to AOD, was primarily investigated by Nathan Brown, then a Ph.D student at Forest Research and later a post-doctoral researcher at Rothamsted Research. In addition, a foundation for the research described here had been built prior to the start of my project, as novel bacteria had been identified from necrotic oak lesions and described by Dr Denman and Dr Carrie Brady. Therefore, to understand the role of these bacteria, in 2012, under the supervision of Dr James McDonald, I began to examine these bacteria, i.e. *Gibbsiella quercinecans* and *Brenneria goodwinii,* at the molecular level. Today, AOD research has significantly expanded, with a substantial knowledge base and a burgeoning team of researchers taking a ground-breaking holistic approach to enable the complete characterisation of AOD, with the ultimate aim of preserving our native oak trees for future generations.

Summary

Acute Oak Decline (AOD) is a complex Decline-disease affecting both native species of oak within the United Kingdom. Two novel species of bacteria, *Gibbsiella quercinecans* and *Brenneria goodwinii,* have been consistently isolated from necrotic lesions of AOD affected trees and are apparently absent from healthy oak trees. The aim of this thesis was to determine the role of *Gibbsiella quercinecans* and *Brenneria goodwinii* in AOD via the application of molecular methods.

The first step in this investigation was to identify an appropriate molecular marker to differentiate between closely related oak lesion isolates belonging to the family *Enterobacteriaceae*. Using ITS1 profiling, a set of *Enterobacteriaceae* strains, primarily *G. quercinecans* and *B. goodwinii* were resolved, and eight strains were selected for sequencing to draft level. This enabled characterisation of strain level variation between sequenced isolates and functional annotation of genome encoded virulence factors, revealing an assortment of phytopathogenic virulence genes As the isolates were sequenced to draft level and no other whole genome sequences of *G. quercinecans* or *B. goodwinii* existed, the type strains of both species, i.e. *G. quercinecans* FRB97 and *B. goodwinii* FRB141 were sequenced to completion using the Pacific Biosciences RSII sequencing platform.

The second component of this investigation characterised the putative phytopathogenic behaviour of *G. quercinecans* and *B. goodwinii*. Data collected from whole genome sequencing was used to align metagenomic and metatranscriptomic data recovered from the necrotic lesions of AOD affected trees, and to inform transcriptomic analysis of virulence gene expression within *in vitro* cultures containing oak tissue. The analysis demonstrated that *G. quercinecans* and *B. goodwinii* alter their behaviour in the presence of oak tissue, and active functionality of *G. quercinecans* and *B. goodwinii* within necrotic lesions AOD-affected trees.

These experiments describe the first whole genome and transcriptome characterisation of *G. quercinecans* and *B. goodwinii*.

CHAPTER 1

Introduction

1.1 British oak

Oak trees are taxonomically located within the family *Fagacea* and the genus *Quercus*, which contains over 500 species worldwide (Zhang *et al.*, 2015). However, the United Kingdom has only two native species, *Quercus robur* (pedunculate oak), and *Quercus petraea* (sessile oak). Aided by the warming waters of the Gulf Stream, these iconic trees have flourished in the temperate British climate. Oaks have been documented in historical literature relating to the British Isles for centuries, from the invasion of England by Julius Cesar in 55 BC, recorded by Pliny the Elder in *The Natural History*, who described druidic sacrifices within sacrosanct oak woodlands, to the construction of ships for use by the British fleet in the Napoleonic wars, providing pedunculate oak with the epochal name of naval oak or true English oak (Pliny, 1991). The historian Simon Schama describes a period of manic pedunculate oak sowing during the Napoleonic wars, as the hardwood tree was, *'the only thing standing between Britain and Bonaparte'* (Schama, 2009). Beyond its cultural importance, oak provides a rich habitat for a diverse range of indigenous flora and fauna, including native bird species such as sparrowhawks, tawny owl, woodpeckers, and smaller birds, such as warblers and flycatchers; small mammals, such as dormice and bats; reptiles such as adders and lizards, and countless plant species (Tyler, 2008). The ecological significance and historical legacy of oak woodlands makes their continued endurance of fundamental importance, both as a keystone species and as a central component of the UK national identity.

Oak woodlands are the most common in Britain. The south and east of England have mineral-rich, clay soil, with a slightly alkali pH, conditions favoured by the dominant pedunculate oak. Further north and to the west, sessile oaks are better able to thrive in the acidic, dry soils of Scotland and Wales (Brookes and Wigston, 1979). Sessile and pedunculate oak can be differentiated taxonomically through leaf morphology, with sessile oak having 5 to 8 lobes per

side and a long leaf stalk or petiole, whereas pedunculate oak has 3 to 5 lobes per side and a short leaf stalk (Mosedale and Feuillat, 1998). However, taxonomic identification is obfuscated by hybridisation between the species, which occurs with increasing frequency from south to north. Hybrids are known as *Quercus* x *rosacea* and are able to back-cross with parent strains, further confusing their classification and often requiring taxonomic resolution with molecular markers for clarification when classical methods have failed (Ferris *et al.*, 1997). Although habitual variation between species can be defined, today many randomly introduced foreign and native oaks can be found around Britain. Prior to these anthropogenic introductions, native trees descended from glacial refugia at lower latitudes after the last glacial maximum (LGM) around 23,000-18,000 years ago, and colonised the landscape (Provan and Bennett, 2008). Using chloroplast DNA (cpDNA) as molecular markers, the provenance of distinct haplotypes can be attributed to historical geographical locations (Cottrell *et al.*, 2002). Despite introductions, cpDNA markers have revealed that around 98% of oak currently in Britain, migrated from their Iberian Peninsula refugia after the LGM (Cottrell *et al.*, 2002).

Oak tissue has low variability between species, and a transverse section can be broadly categorised into five layers (Fig. 1.1) (Doughari, 2015): 1) outer-bark or periderm, consisting of cork, cork cambium and secondary phloem. 2) inner-bark, consisting of a living layer of phloem transport tissue 3) vascular cambium, located at the phloem-xylem interface, and produces new xylem and phloem transport tissue 4) xylem or sapwood, the second transport tissue after phloem, which carries water and some nutrients from the roots, throughout the tree 5) heartwood, the durable core of the oak tree.

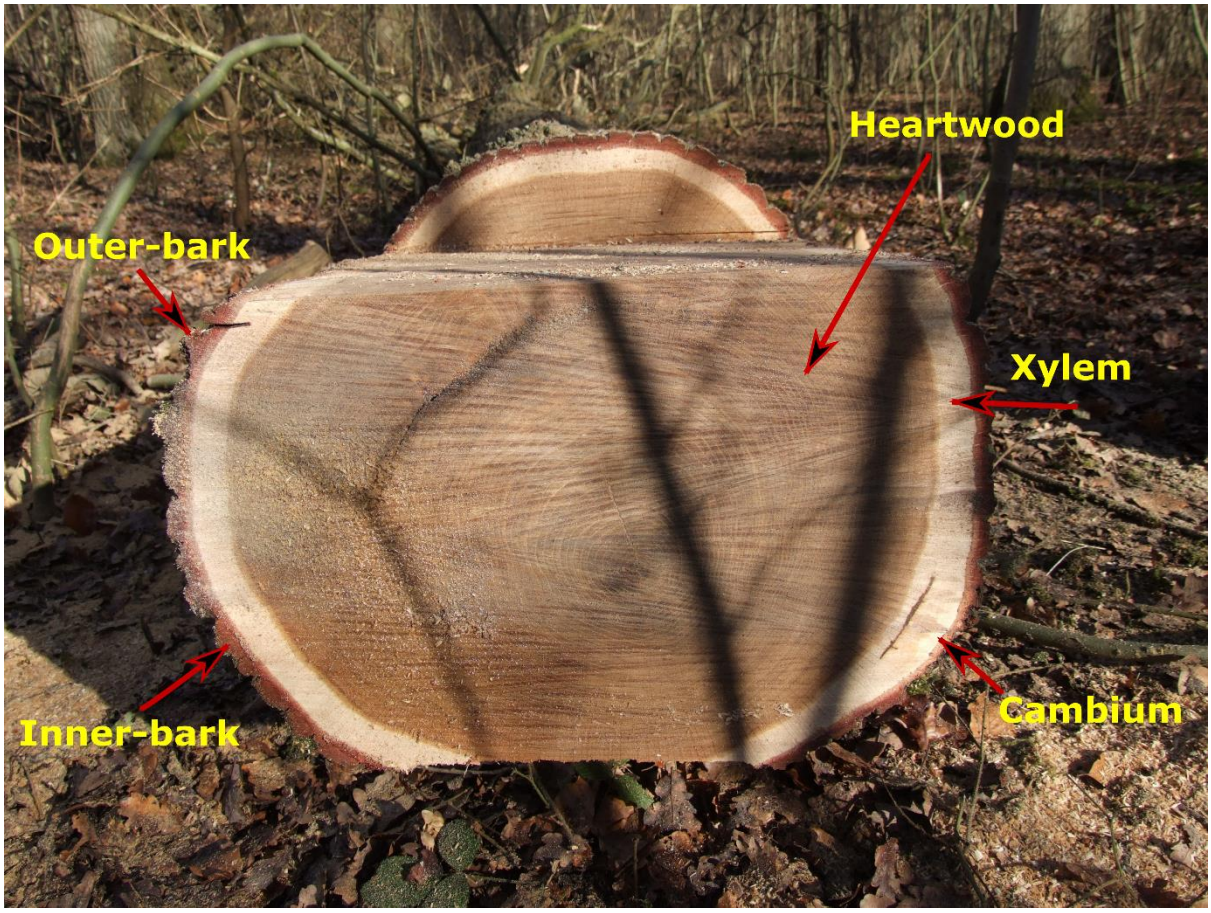


Figure 1.1. Transverse section of a pedunculate oak (*Quercus robur*). Photograph courtesy of Dr Sandra Denman.

1.3 The composition of oak tissue

Plants are characterised by a dynamic, complex plant cell wall. The oak cell wall is a lignocellulosic matrix of interwoven polymers; the dry weight of which is formed of cellulose (~40%), hemicellulose (~25%), lignin (~20%) and proteins (~10%) (Zhang *et al.*, 2015). These tissues form rigid bonds; cellulose forms hydrogen bonds with hemicellulose, which is covalently bonded through ester and ether linkages to lignin. Additional plant polymers including tannin and pectin, are embedded at varying quantities and constitute significant levels of cellular mass throughout the tree tissue. Plant cell walls can be functionally and compositionally differentiated into primary and secondary cell walls; primary cell walls surround growing and dividing cells, whereas secondary walls are mature, recalcitrant structures with a greater ability to resist hydrolytic invasion. Dicotyledonous plants or dicots, which include angiosperms, have group I cell walls; this group have a higher proportion of xyloglucan as a component of hemicellulose

and a higher composition of pectins (Carpita and Gibeaut, 1993). Below is a description of the main features of the oak cell wall and examples of the hydrolytic enzymes used by phytopathogens to convert them into residual nutrients.

1.3 Cellulose

Cellulose is the most abundant organic polymer on earth, and has a simple linear structure of β-(1,4)-linked glucan chains (Klemm *et al.*, 2005). Cellulose is the central component of the lignocellulosic plant cell wall and is organised into tightly bound highly crystalline microfibrils, providing the plant with flexible strength (Gibson *et al.*, 2011). The compact structure of cellulose instils plants with an obdurate core, providing resistance to biodegradation with tensile strength comparable to steel (Eckardt, 2003).

1.3.1 Cellulase

The recalcitrant nature of cellulose limits the number of microorganisms encoding cellulolytic ability. There are at least five distinct mechanisms of cellulolytic activity, within three classes of cellulase, these classes are 1) endo-(1,4)-β-D-glucanase, which randomly hydrolyses accessible O-glycosidic sites across the cellulose chain, reducing the density of cellulose, 2) exo-(1,4)-β-D-glucanase, which cleaves from the chain end, progressively removing β-cellobiose and has minimal effect on density, and 3) β-glucosidases, which target β-cellobiose, reducing the disaccharide to glucose (Kuhad *et al.*, 2011). Cellulolytic activity can be tested on various forms of cellulose, including carboxymethylcellulose (CMC), a soluble cellulose derivative which does not require a carbohydrate binding module (CBM) for hydrolysis, and crystalline cellulose (e.g. Avicel or bacterial cellulose), this is a representative variant of the insoluble plant cell wall form, requiring a CBM and cumulative action of all three classes of cellulase for effective hydrolysis (Barras *et al.*, 1994; Wilson, 2011).

Cellulases and other enzymes which degrade plant cell walls fall within three classes of carbohydrate active enzyme (CAZymes), these are glycoside hydrolases, polysaccharide lyases and carboxyl esterases, and an associated group of enzymes - carbohydrate binding modules, which enhance enzymatic activity (Lombard *et al.*, 2014). CAZymes which can degrade plant cell walls are classified as plant cell wall degrading enzymes (PCWDEs) (Gibson *et al.*, 2011). Cellulolytic PCWDEs fall within several groups of glycoside hydrolases including GH5. Within

the GH5 group several reactions can be catalysed, therefore the CAZyme grouping provides a broad definition of the enzymes catalytic potential, this can be refined through the enzyme commission (EC) grouping. EC classification is a more specific mechanism of classification than the general CAZy scheme, for example, endoglucanase belongs to EC class 3.2.1.4, and the hydrolysis mechanism of this reaction is clear from the EC definition. (Badieyan *et al.*, 2012).

1.4 Hemicellulose

Hemicellulose is a plant cell wall polysaccharide which entwines other lignocellulosic components to provide stability, shape and tensile strength (Scheller and Ulvskov, 2010). Hemicelluloses vary between plants and are chemically divided into four groups: xylan, xyloglucans, mannans, and glucomannans. The dominant hemicellulose in dicots is xyloglucan in primary cell walls, and xylan in secondary cell walls. Xyloglucan is constituted of repetitive units of β-(1,4)-linked glucan backbone, and various oligosaccharide side chains substituents. In higher plants xyloglucan tethers cellulose microfibrils allowing cells to grow, differentiate and form into their final maturation shape (Hayashi and Kaida, 2011). As a component of the primary cell wall, xyloglucan has an important role in biosynthesis and decreases in mature secondary cell walls to be replaced by xylan (Collins *et al.*, 2005). Xylan is a complex structural polysaccharide and is formed of a homopolymeric backbone of β-(1,4)-linked xylose residues. In dicots these residues can be substituted with α-(1,2)-linked glucuronosyl and 4-O-methyl glucuronosyl (these xylans are often known as glucuronoxylans) (Scheller and Ulvskov, 2010; Rennie and Scheller, 2014). Xylan strengthens, transports water, and provides fibre cohesion at the interface between lignin and cellulose within the secondary cell wall (Collins *et al.*, 2005). Hemicelluloses are key to the dynamic structure of plant cell walls, providing stability through hydrogen bonds to other plant polymers and tensile strength allowing structural adaptations within a changing environment.

1.4.1 Xyloglucanase

Degradation of xyloglucan tethered to cellulose microfibrils loosens the rigid bonds of the plant cell wall (Hayashi and Kaida, 2011). A plant hormone - auxin, controls the size of cells through the measured degradation of xyloglucan into oligosaccharides (Benko *et al.*, 2008). Bacterial phytopathogens can replicate auxin induced degradation through digesting of the cross-linked xyloglucan-cellulose bonds, using PCWDEs, such as endoglucanases, xyloglucan

endotransglycosylases and exoglycosidases, including α-fucosidases and β-galactosidases (Pitzschke, 2013).

1.4.2 Xylanase

The heterogeneity of xylan structures requires putative pathogens to encode a repertoire of enzymes to facilitate hydrolysis. Glycoside hydrolase families 5, 7, 8, 10, 11 and 43 have xylanase catalytic domains (Collins *et al.*, 2005). Xylanases belong to EC 3.2.1, however to completely degrade xylan the hydrolases need support from esterases. For example, β-D-xylanases (EC 3.2.1.8) first randomly split the xylan backbone, then β-D-xylosidases reduce xylose monomers, and finally, cleavage of the side groups requires more diverse enzymes including exo-α-L-arabinofuranosidase (EC 3.2.1.55) (Fig. 1.2) and feruloyl esterases (EC 3.1.1.73) (Fig. 1.3).

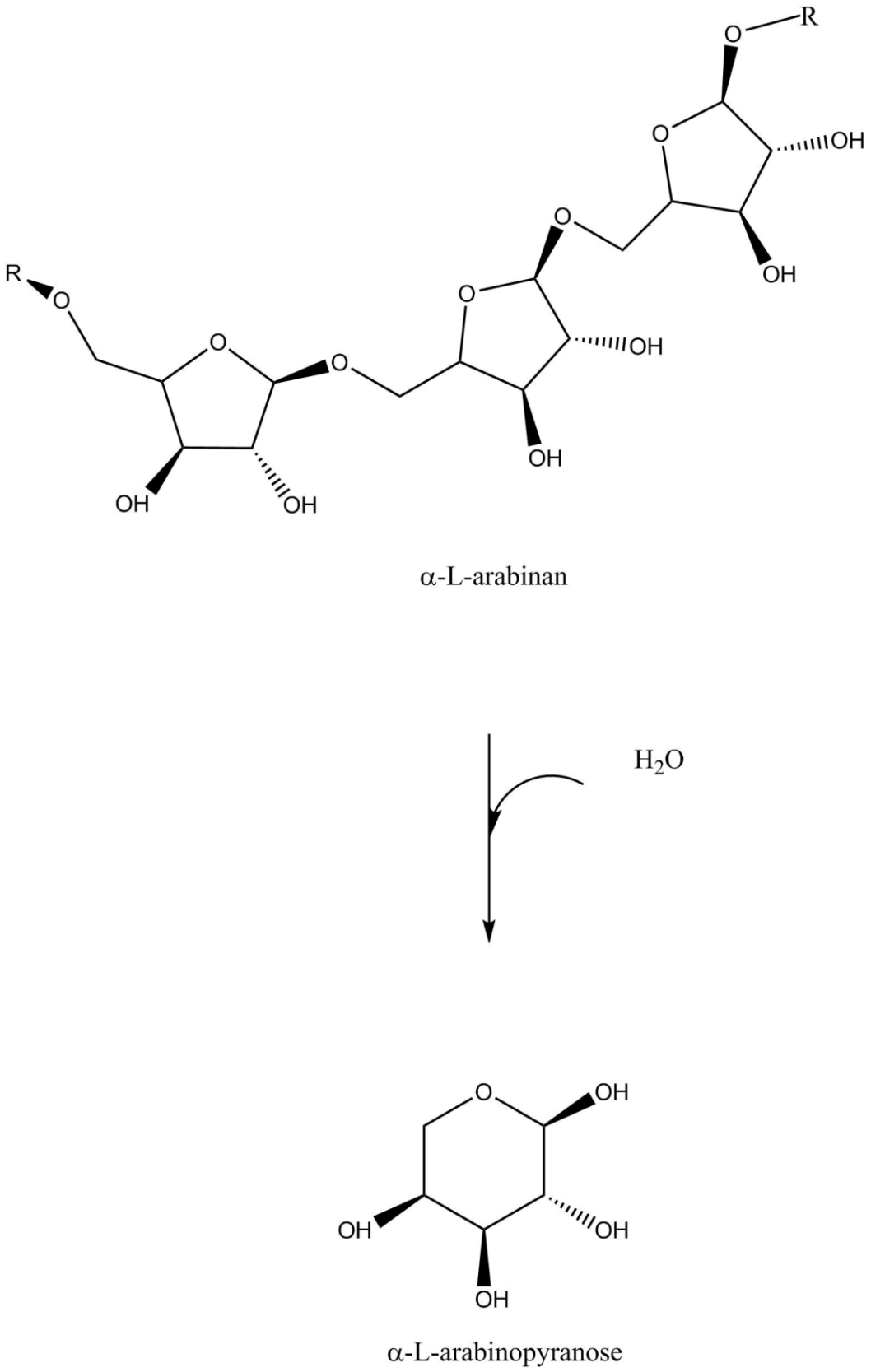
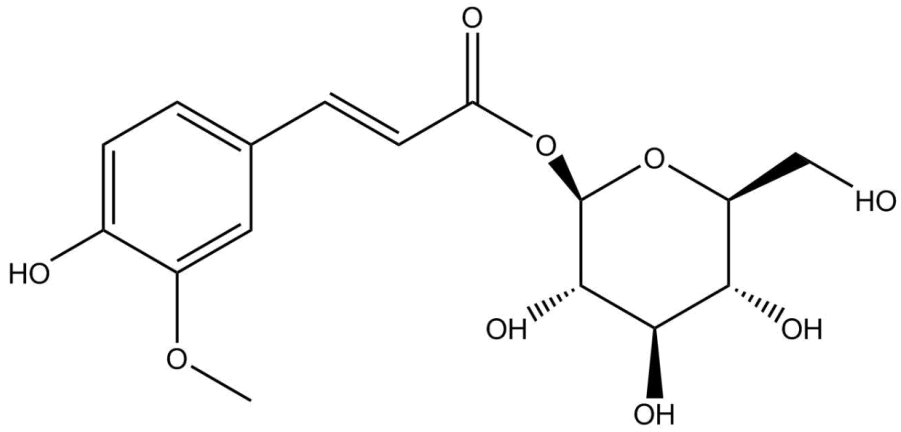
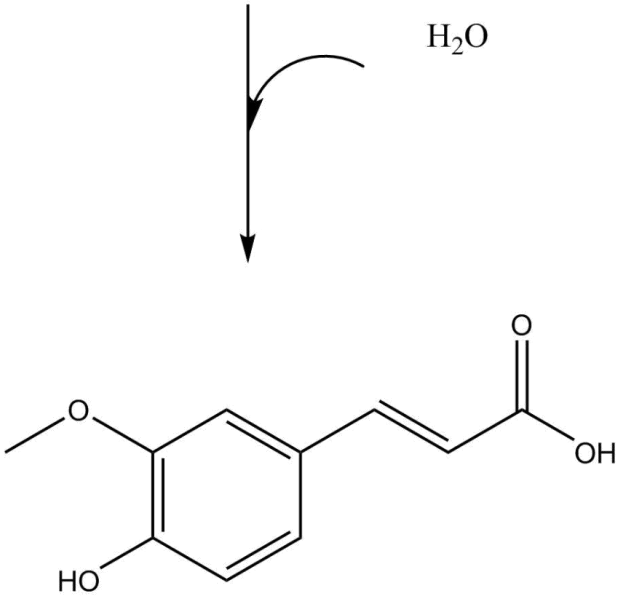


Figure 1.2. Simplified illustration of a arabinosidase reaction from an enzyme of the bacterium *Bremneria goodwinii* FRB141. α-L-arabinofuranosidase (EC 3.2.1.55), hydrolyses α-L-arabinans to α-L-arabinopyranose. R: α-L-arabinopyranose.

Feruloyl esterases (FAEs) are hemicellulolytic PCWDEs which hydrolyse the L-arabinofuranose containing polysaccharide ester bonds of hemicelluloses and pectins, releasing phenolic acids and their dimers (Topakas *et al.*, 2007). FAEs are commonly known as hemicellulase accessory enzymes as they assist xylanases and pectinases to degrade hemicellulose.



1-O-Feruloyl-β-D-glucose



Ferulic acid

Figure 1.3. Illustrated feruloyl esterase reaction from an enzyme of the bacterium *Shigella flexneri*. Feruloyl esterase (EC 3.1.1.73), hydrolyses a feruloyl group from an esterified sugar.

1.5 Lignin

Lignin is an aromatic biopolymer, formed of oxidised cross-linked monolignols, namely, p-coumaryl, coniferyl, and sinapyl alcohols, which may constitute up 50% of plant phenolic material (Witzell and Martín, 2008). One of the key benefits conferred to plants by lignin is durability. Lignin has high recalcitrance to chemical and biological hydrolysis, it does not rot easily, and decomposes slowly. Within the lignocellulosic matrix, lignin is cross-linked with hemicellulose, and forms a shield between the more readily hydrolysable cellulose and hemicellulose preventing ease of access to phytopathogenic microbes. Unlike other plant polymers lignin is industrially inert, with most ligninolytic enzymatic investigations focusing on separating lignin from valuable biopolymers (Eudes *et al.*, 2014).

1.5.1 Ligninolytic microorganisms

Microbial degradation of lignin requires several enzymatic reactions, where peroxidase enzymes synergistically combine to oxidise benzene rings (Ruiz-Dueñas and Martínez, 2009). The peroxidases generate hydrogen peroxide to break-down lignin, a reaction which has been described as enzymatic combustion (Kirk and Farrell, 1987). The only characterised microbial group which can extensively mineralise lignin are aerobic white-rot and brown-rot fungi from the phylum *Basidiomycota* (Bugg *et al.*, 2011). *Basidiomycota* secrete interacting peroxidases, which are necessary for ligninolytic activity. Other microbial phyla have indirect ligninolytic activity, enzymes such as laccases can contribute to lignin degradation. Laccases are a group of copper-dependent, microbial oxidoreductases which are produced by many *Basidiomycota* but are also found in the bacterial phyla, actinomycetes, γ-proteobacteria, and α-proteobacteria and are induced on lignocellulosic matter (Bugg *et al.*, 2011). These enzymes have no direct influence on lignin in isolation, but when combined with redox mediators they can degrade the rigid structures of lignin.

1.6 Tannin

Tannins are defined as water-soluble polyphenolic secondary metabolites of higher plants, with molecular weights ranging from 500 to 30,000 Da (Serrano *et al.*, 2009; Jiménez *et al.*, 2014). *In planta*, they form large hydrogen bonded macromolecular structures with proteins and to a lesser extent with cellulose and pectin (Bhat *et al.*, 1998). Alongside their structural role, tannins are involved in defence against pathogens (particularly fungal pathogens) (Scalbert, 1991). Tannin

mediated defence interferes with microbial metabolism which inhibits growth; this defence mechanism astringently complexes with target substrates of the invading microorganism, including iron. Tannin-mediated defence is amplified upon infection, with tannins accumulating at the infection site as part of the compartmentalising process, which isolates affected areas from healthy tissue (Sallé *et al.*, 2014). There are many structural variances between tannins, but all are formed of either galloyl esters or their derivatives. They have the industrially important ability to precipitate proteins from solution, a function which has many uses in leather production, turning animal skin into leather and brewing, where tannins precipitate unstable proteins from wine and beer (Chowdhury *et al.*, 2004). Tannins can be broadly classified into two groups, 1) condensed and 2) hydrolysable, see below for more details.

1.6.1 Hydrolysable tannins

Hydrolysable tannins are formed by esters of gallic or ellagic acid (gallotannins or ellagitannins) and a sugar core (usually D-glucose, but can be xylose, fructose, hamamelose and saccharose) (Serrano *et al.*, 2009; Barbehenn and Constabel, 2011). A high proportion of hydrolysable tannins are found in the outer-bark, making up around 20% of the tissue. Deeper into the tree tissue, hydrolysable tannins progressively reduce in concentration (Hernes and Hedges, 1992). Hydrolysable tannins are less resistant to hydrolysis than condensed tannins and are easily degraded enzymatically to glucose and its constituent acid (Bhat *et al.*, 1998).

1.6.2 Condensed tannins

Condensed tannins are polyflavonols joined by C6-C3-C6 catechin or epicatechin monomers (Puech *et al.*, 1999). Highly polymerised, condensed tannins are less soluble, more resistant to basic and acidic degradation than the hydrolysable form. Oak leaves contain high concentrations of condensed tannins which increase upon maturity and are toxic to animals if eaten in large quantities. However, some herbivores such as goats and sheep can tolerate tannins as they are degraded by the rumen microbiota constituents *Streptococcus caprinus* and *Selenomona ruminatium* subsp. *ruminantium* (Makkar *et al.*, 1991; Perez-Maldonado and Norton, 1996).

The plastic structure of tannins complicates generalised predictions on their enzymatic degradation, inhibition of microbes and toxicity to herbivores (Ayres *et al.*, 1997). This idiosyncratic nature makes tannins problematic to classify and requires tannin polymers to be

experimentally tested for susceptibility to environmental stresses before classification can be assigned (Barbehenn and Constabel, 2011).

1.6.3 Tannase

Tannin acyl hydrolase (EC 3.1.1.20), commonly known as tannase, catalyses the breakdown of hydrolysable tannin ester bonds into oligomeric tannins or derivatives, such as gallic acid (Fig. 1.4). Tannase activity is induced by methyl gallate and tannic acid but not by other phenols such as gallic acid. Tannases have various degrees of degradative ability; yeast tannases are weak degraders of tannin and are mostly saprophytic, whereas bacterial tannases can hydrolyse condensed tannins found in hardwood trees, such as oak and chestnut (Deschamps and Lebeault, 1984). Bacterial tannins have been reported in a number of *Enterobacteriaceae* species including, *Streptococcus bovis*, *Klebsiella* spp., *Serratia* spp. and *Pantoea agglomerans* (Pepi *et al.*, 2010).

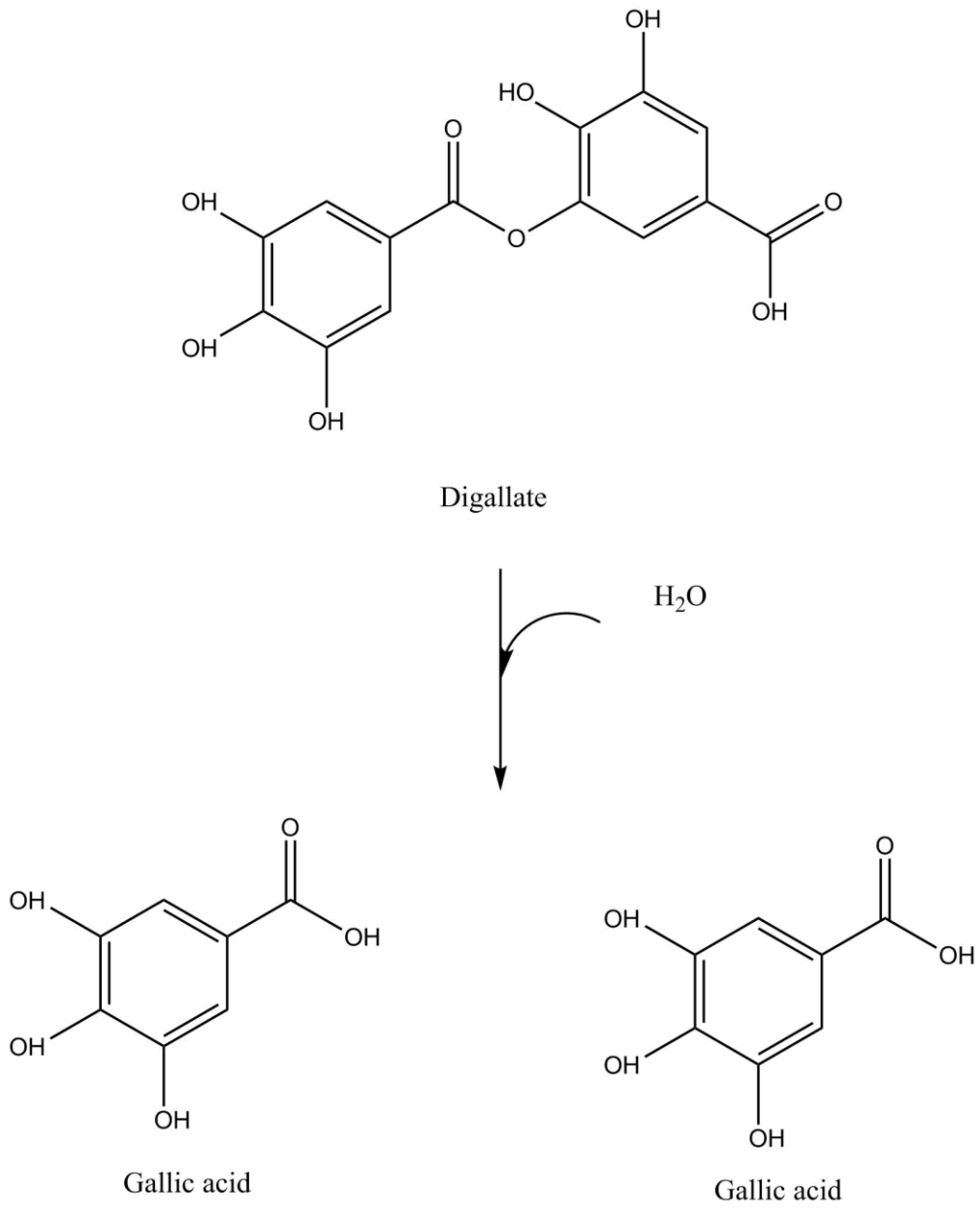


Figure 1.4. Simplified illustration of a tannase reaction from an enzyme of the fungus *Aspergillus niger*. Tannase (EC 3.1.1.20) hydrolyses the central ester linkage between the two aromatic rings of digallate, producing two gallic acid molecules (Aguilar *et al.*, 2007).

1.7 Pectin

Pectin is a gelatinous heterogeneous plant cell wall polysaccharide, which consists of homogalacturonan, rhamnogalacturonan I, and rhamnogalacturonan II. Pectin polymers provide structural stability to the cell wall through linking cellulose and hemicellulose fibres. Pectin constitutes around one third of the dry wet weight of higher plants and is most abundant in the middle lamella, acting as 'molecular glue' for the cellulosic network of the primary cell

wall. Industrially, it is used as a thickening and gelling agent, and is most frequently extracted from citrus peel or apple pomace which contain 20-30% and 10-15% dry matter of pectin, respectively (Thakur *et al.*, 2009). Commercial pectin is typically 70% esterified, other forms can be purchased but require further processing such as addition of acid to remove ester groups (May, 1990).

1.7.1 Pectic enzymes

Pectic enzymes catalyse reactions which degrade the pectin molecule, these are; hydrolysis reactions (substitution reaction, adding water to cleave chemical bonds) and lyase reactions (elimination reaction, causing single or double break to bonds, in this case carbon-oxygen bonds), these reactions are capable of degrading oligogalacturonate units on any of the three pectin polymers (Herron *et al.*, 2000). They are a central component of the virulence arsenal of many bacterial phytopathogens and are found in the microbiome of herbivores (Linhardt *et al.*, 1987; Herron *et al.*, 2000) (Fig. 1.5). The soft rot *Enterobacteriaceae* (SRE) are a family of phytopathogenic bacteria which target plant cell wall components, but particularly pectin and pectate (distinguished by the methyl groups of their backbone, which alter the surface charge of the polymer and are recognised by binding pockets of correctly conformed enzymes). The SRE release PCWDEs belonging to multiple polysaccharide lyase, glycoside hydrolase and carboxyl esterase (family 1 only) CAZy families (May, 1990; Toth *et al.*, 2006). Expression of PCWDEs can be induced by pectin fragments or quorum sensing signalling, and target various linkages across the pectin molecule, depolymerising the polysaccharide into ingestible energy rich monomers (Charkowski *et al.*, 2012).

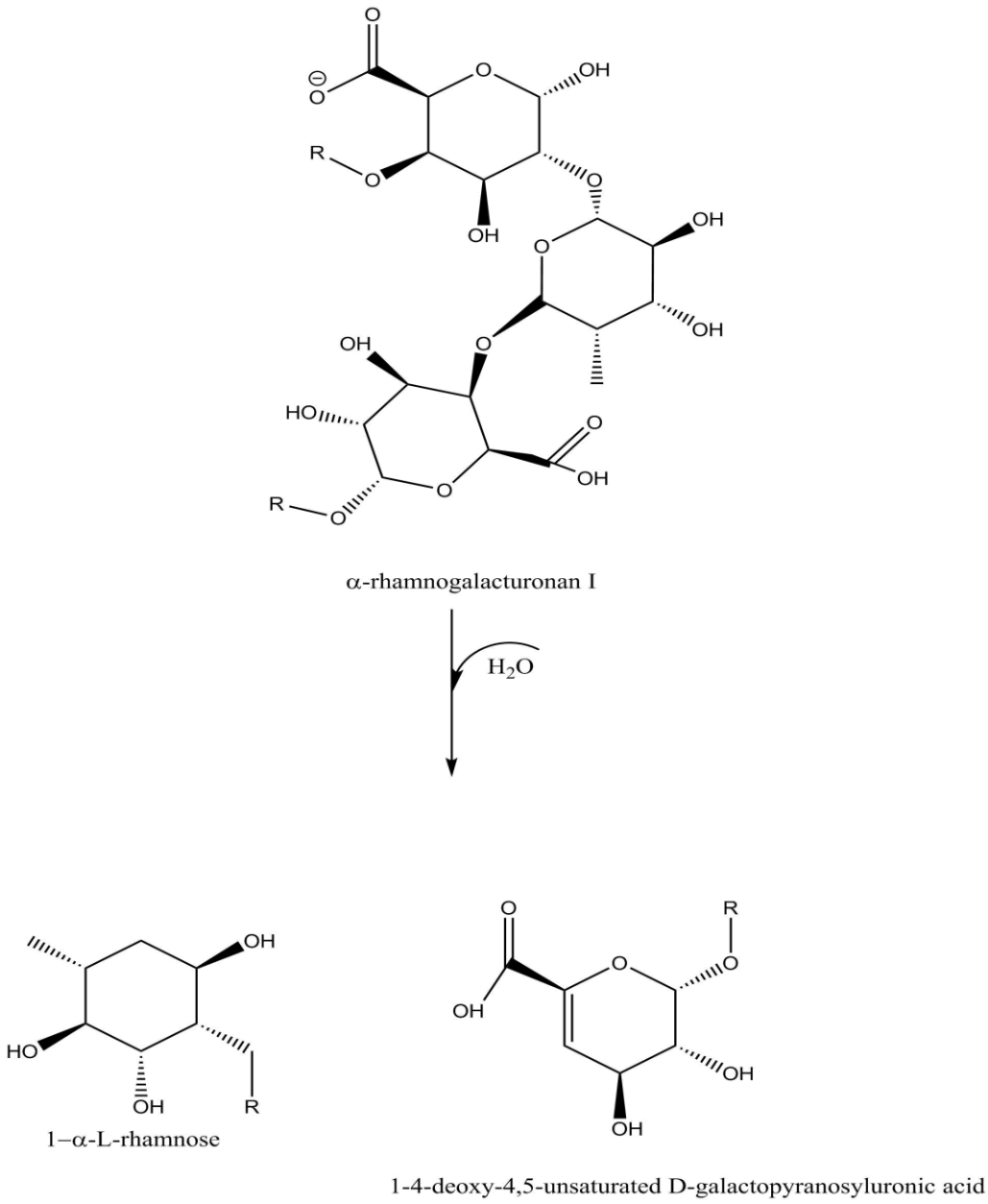


Figure 1.5. Simplified illustration of a rhamnogalacturonan endolyase reaction from an enzyme of the bacterium *Gibbsiella quercinecans* FRB97. Rhamnogalacturonan endolyase (EC 4.2.2.23) cleaves rhamnogalacturonan I of the pectin polysaccharide, producing α-rhamnogalacturonan and 1-α-L-rhamnose at the reducing end and α-rhamnogalacturonan and 1-4-deoxy-4,5-unsaturated D-galactopyranosyluronic acid at the non-reducing end (Azadi *et al.*, 1995). R: α-L- rhamnogalacturonan.

1.8 Tree Declines

Tree Declines have been reported since the inception of forest science over two centuries ago (Ciesla and Donaubauer, 1994). Classic Decline symptoms result in loss of tree vigour, branch dieback, and in some cases, death. The causal agents are biotic and abiotic and have been described as a '*chain of negative influences'*, as there is no clear emphasis on any one factor (Buck *et al.*, 2004). Stress factors may predispose trees to Decline, allowing one or more pathogens to thrive in a host which was previously unavailable. To facilitate recognition of Declines, three distinct stages have been defined (Fig. 1.6); 1) predisposing factors, long term stress factors, including climate, soil, air quality, genetic diversity and age; 2) inciting factors, short term factors, such as insect defoliation, frost, drought, mechanical injury, and air pollutants; 3) contributing factors, long-term endemic factors such as, bark beetles, canker fungi, and root decay fungi (Manion, 1981). However, alternative models of Decline, have been proposed, including that Declines are a natural process and part of forest dynamics and succession (Mueller-Dombois, 1992), and a model which proposes that all Declines are a result of climatic anomalies, causing physical damage to xylem, resulting in moisture stress and the characteristic Decline symptoms (Auclair *et al.*, 1992). However, the Manion model and the alternative models of Decline share the belief that climate change is a predisposing factor and that the potential for anthropogenic climate change to graduate the number and intensity of Decline events is a significant threat.

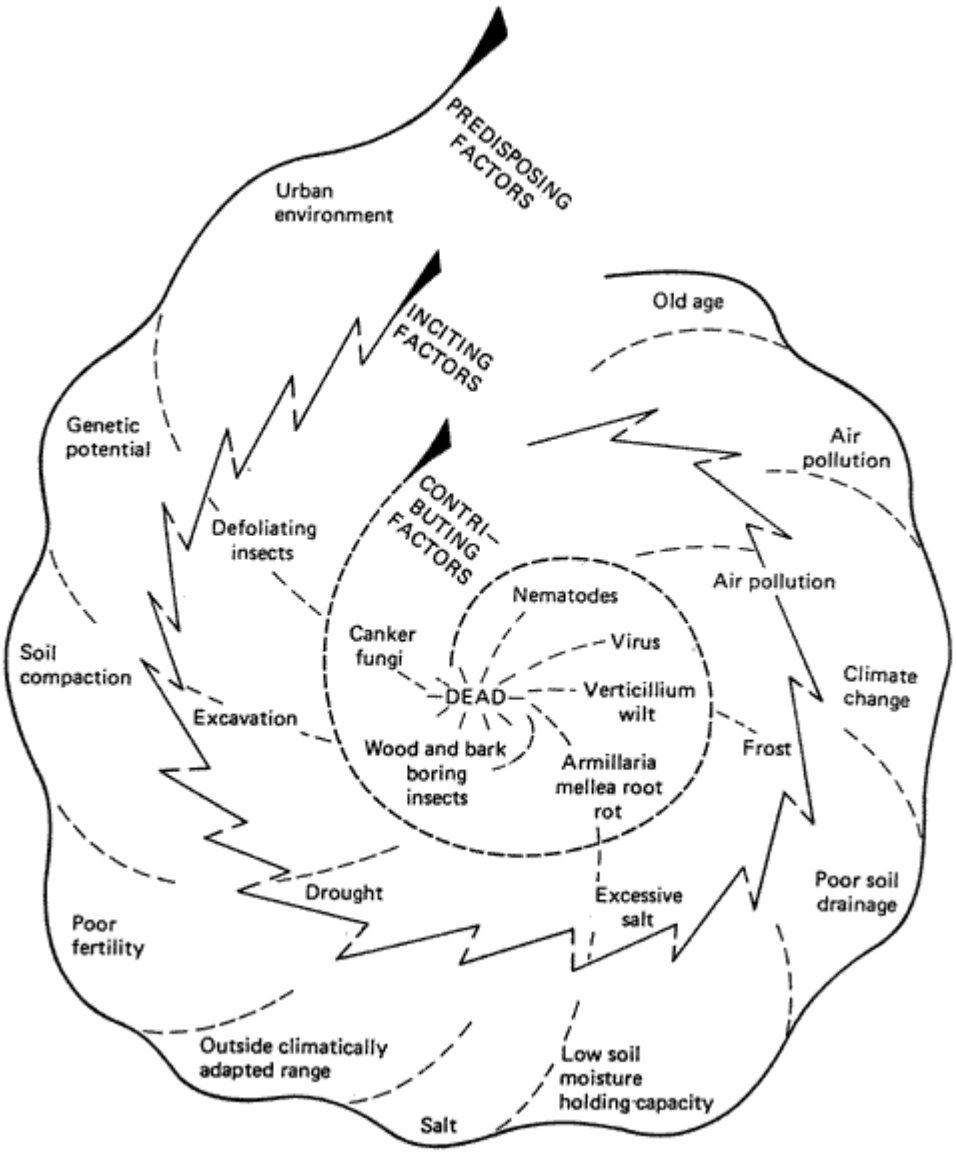


Figure 1.6. Manion's Decline-disease spiral. Interacting causal effects of Decline sequentially progress an affected tree towards death. Each causal effect is categorised into the three tiered ring of predisposing, inciting and contributing Decline-disease factors. Distance from the centre represent levels of involvement, with those further from the centre having less direct involvement (Manion, 1981).

1.8.1 Oak Declines

Oak Declines, like other Declines, are syndromes caused by an array of interactive factors which vary between sites and impart differing influences within the Decline; crucially however, Declines have no definitive causative agent. Oak Declines can be categorised into two groups, chronic oak Decline (COD) and acute oak Decline (AOD) (Denman and Webber, 2009).

1.8.2 Chronic Oak Decline

Chronic Oak Decline (COD) is caused by biotic and abiotic factors, and affects both species of British oak (Denman and Webber, 2009). The associated biotic pathogens include herbivorous insects and endemic interchangeable fungal species, such as the honey fungus *Armillaria spp.*, the root rot pathogens *Phytophthora* spp. and *Collybia fusipes*, and the powdery mildew *Erisyphe alphitoides* (Thomas *et al.*, 2002; Thomas, 2008). Among the abiotic inducing factors are extended drought, frost, increased ozone levels, and climate change. COD rarely results in mortality, and is characterised by a relatively slow rate of development, contrasting to disease caused by devastating pathogens, such as *Ophiostoma* spp., the causative agent of Dutch elm disease, but rather produces a general deterioration in health, including: progressive canopy thinning, branch dieback, reduced growth, reduced photosynthetic activity, and depleted carbohydrate reserves (Brasier, 1991; Marçais and Desprez-Loustau, 2014). COD can harm the natural regeneration of oak, endanger the general health of indigenous oak forests, and requires a concerted effort over a number of years to understand its incitement, prevalence and development of effective fungicides or preventative measure.

1.8.2 Acute Oak Decline

Acute Oak Decline (AOD) is a recently described Decline-disease affecting both native species of British oak, *Quercus robur* (pedunculate oak) and *Quercus petraea* (sessile oak) (Denman *et al.*, 2014). Similar to other Decline-diseases, AOD is marked by a decrease in tree health and vitality, caused by interacting biotic and abiotic factors (Brown *et al.*, 2014). In contrast to COD which affects the oak root, AOD is an above ground pathogen, characterised by rapid Decline in health over 3-5 years (hence, the epidemiological title, acute) and its near endemic co-occurrence with the European bark boring buprestid beetle, *Agrilus biguttatus*. The main symptoms of AOD are stem bleeding and necrosis of inner-bark tissue, which can ultimately lead to death. As with other Declines, there is no single causative agent and the interaction of several

putative pathogens has been proposed. AOD is a syndrome which mainly affects mature trees and it has become evident that it has existed in Britain (Denman and Webber, 2009) and the same or a similar Decline-disease in continental Europe (Gibbs and Greig, 1997; Holdenrieder *et al.*, 2004), for some time. The first recorded incidence of AOD in Britain, occurred in the early 1920's, caused by defoliation of newly formed leaves in the canopy (as opposed to the current episode which affects the oak stem) by caterpillar of the oak roller moth (*Tortrix viridana*), which was highly abundant after the end of the first world war until the mid-1920's. During this period, it was reported that affected oaks had high mortality and were dying at an alarming rate. Mortality was caused by a succession of secondary pathogens, with initial defoliation followed by the oak mildew *Microsphera quercina* and finally by the honey fungus, *Armillaria mellea*. Unlike the first recorded incidence of AOD, the recent outbreak affects the stem as opposed to the canopy.

Recently, there has been an effort to describe the current outbreak of AOD and present an understanding of the causes, effects and guidelines for preventative measures (Brown *et al.*, 2014; Denman *et al.*, 2014; Doonan *et al.*, 2015). Characterising the causes of AOD is challenging, as unravelling the interactive effects of multiple putative pathogens requires comprehensive levels of time and resources. AOD has four indicative symptoms (Denman *et al.* 2014); 1) stem bleeds on the outer bark; 2) necrotic lesions on the underlying inner bark; 3) weeping patches on oak tree trunks; 4) larval galleries of a putative pathogen, the buprestid beetle, *Agrilus biguttatus*, which are frequently found (greater than 90% co-occurrence in affected trees) within the vascular tissue. It is thought that weakened oaks are vulnerable to impregnation with *Agrilus* larvae, however healthy trees are able to offer resistance in the form of dark exudate (Gibbs and Greig, 1997).

Bacterial isolates, including novel genera and species have been consistently recovered from necrotic lesions of AOD affected trees, but significantly, not from healthy trees. It has been hypothesised that two bacteria, *Gibbsiella quercinecans* (Brady *et al.*, 2010) and *Brenneria goodwinii* (Denman *et al.*, 2012), cause necrosis of oak tree tissue. Furthermore, there has been accumulating evidence within the United Kingdom and worldwide, that bacteria play a role in Decline-disease of oak trees (Scortichini *et al.*, 1993; Biosca *et al.*, 2003; Denman *et al.*, 2014).

1.9. Bacterial phytopathogens

1.9.1 Plant immunity

The study of plant-pathogen co-evolutionary interaction reveals the mechanisms of infectious disease processes (Burdon and Thrall, 2009). However, plant-pathogen interactions are dynamic, as plant resistance genes must circumvent pathogen induced disease, and the pathogen must avoid or suppress this response to prevent limitations on their growth (Dangl and Jones, 2001), this is commonly referred to as the evolutionary arms race and underpins the Red Queen hypothesis (van Valen, 1973; Anderson *et al.*, 2010). Unlike mammals, plants have no adaptive immunity and rely on their innate immune system, which is activated as pathogens enter the plant through gas and water pores, or wounds (Jones and Dangl, 2006). There are two components of the plant innate immune system, the first of which has a two-step response to challenge invading pathogens; 1) plant cell surface receptors which bind conserved bacterial epitopes such as flagellin, these are known as pathogen associated molecular patterns (PAMPs) (Liu *et al.*, 2010; Arnold and Jackson, 2011); 2) activated PAMP receptors act as a molecular switch to mobilise immune signalling pathways, triggering systemic, endogenous salicylic acid, ethylene and jasmonate phytohormones, stimulating specific resistance genes, such as the bactericidal secondary metabolite, putrescine (Kim *et al.*, 2013), reactive oxygen species and the hypersensitive cell death response (Glazebrook, 2005). The second component of the immune response is characterised by nucleotide-binding, leucine rich repeat (NB-LRR) domains within most plant resistance genes, which recognise and bind pathogen effectors. The NB-LRR domain recognisntion system is is most effective against biotrophic or hemibiotrophic pathogens, but offers no resistance to necrotrophic pathogens (Jones and Dangl, 2006). To overcome host immunity, natural selection forces invading bacterial pathogens to alter their nucleotide sequences within conserved domains, and to select strains with beneficial horizontally acquired regions, which are subsequently able to avoid detection by the innate immune system.

1.9.2 Classes of bacterial phytopathogen

Bacterial phytopathogens have several commonalities: they require a suitable host, an appropriate environment, and a suitable mode of infection (Doughari, 2015). Bacterial phytopathogens can be broadly divided into three classes according to their mode of infection: 1) biotrophs, 2), necrotrophs and 3) hemibiotrophs. Biotrophs are obligate parasites and require

continuous interaction with a living host to ensure a continuous supply of nutrients; necrotrophs are pathogens which acquire nutrients from dead or dying tissue, and hemibiotrophs are pathogens which exist as biotrophs and necrotrophs at different stages of their life (Glazebrook, 2005).

1.9.3 Pathogenic lifestyles

Bacterial phytopathogens do not continuously exist in a pathogenic state, and can live as commensals on the tree surface (epiphytic lifestyle), within a plant (endophytic lifestyle), on decomposing plant matter (saprophytic lifestyle), but can enter a latent infection stage as environmental conditions change (Toth and Birch, 2005). Transition between lifestyles is a response to nutrient availability, as opportunistic bacteria can be part of the plant microbiome and persist as harmless commensals, until nutrients deplete or the microbiome is altered, at which point a transition to a pathogenic lifestyle may occur (Mendes *et al.*, 2013). Opportunistic phytopathogens within the plant microbiota typically graduate away from the harsh external environment, towards the nutrient rich inner plant tissue, which contains sugars and minerals. Once within plant tissue, pathogenic bacteria reside alongside the plant cell wall and within the extracellular space (apoplast), or within phloem and xylem cells (Fatima and Senthil-Kumar, 2015). Ingression of bacteria to the inner tissues typically elicits activation of the salicylic acid pathway, for biotrophic pathogens, whereas detection of a necrotrophic pathogen, typically results in expression of the jasmonate/ethylene pathway.

1.9.4 Biotrophic pathogens

Agrobacterium tumefaciens is a biotrophic plant pathogen; it manipulates the plant host's morphology and physiology, via type IV secretion system mediated translocation of oncogenic T-DNA (transferred DNA) and *vir* regulon encoded virulence proteins into the host nucleus (Pitzschke, 2013). This biotroph blocks the innate immune response through intercepting the MPK3 signal kinase, thereby obstructing the positive regulator from activating the subsequent cascade, and further utilises this pathway by hijacking the downstream host protein VIP1, which functions to activate defence protein transcription factors but instead shuttles foreign DNA and proteins into the nucleus. Resultant agrobacteria crown galls are tumours induced by suppression of the immune signalling cascade and promotion of auxin expression, which antagonistically

effects salicylic acid production, which in a healthy plant would resist invasion through the inactivation of the *vir* regulon and prevent pathogen colonisation (Yuan *et al.*, 2007).

1.9.5 Necrotrophic pathogens

Necrotrophic plant pathogens include the Soft-Rot *Enterobacteriacae* (SRE), e.g. *Pectobacterium* spp. and *Dickeya* spp. (Charkowski *et al.*, 2012), *Xanthomonas* spp. (da Silva *et al.*, 2002) and *Xylella fastidiosa* (Janse and Obradovic, 2010). These are brute force pathogens which break down the plant cell wall through an assortment of cellulolytic, pectinolytic and hemicellulolytic PCWDEs (Toth *et al.*, 2006). Extensive release of PCWDEs degrades host cell integrity and rots living and dead tissue. Necrotrophic pathogens have strain and plant specific pathologies, they have different repertoires of PCWDEs and consequently variable catalytic activity (Barras *et al.*, 1994). Despite these specificities, many necrotrophs such as the SRE have a broad host range, and affect around half of angiosperm plants worldwide, causing significant damage to crop and ornamental plants (Ma *et al.*, 2007). Although the SRE are generally considered to be necrotrophs, they encode a Type III secretion system (T3SS), which is widely thought to assist in evasion or manipulation of host defences, shifting the classification of the SRE from necrotrophs to hemibiotrophs (Toth and Birch, 2005). This may be true of many other phytopathogens classified as necrotrophs, as they may have evolved mechanisms to avoid immune detection, and are not exclusively brute force pathogens.

1.9.6 Hemibiotrophic pathogens

Pseudomonas syringae is the archetypal bacterial hemibiotrophic pathogen, which uses T3SS effectors to suppress and manipulate the host immune system, allowing widespread growth of the bacterium, before latently transforming to a highly destructive necrotrophic phase (Toth and Birch, 2005; Lee and Rose, 2010). The T3SS machinery is critical to the virulence of *P. syringae*; it is a highly conserved nanomachine with an array of hypervariable secreted effectors, which vary in quantity and between pathovars (Collmer *et al.*, 2002). The diversity of effectors between pathovars reflects their host specificity, where the complete genomes may be very similar but diverge in their effector repertoire, adapting to host biology (Arnold and Jackson, 2011). *P. syringae* pathovars have a narrow host range, which is governed by their host specific virulence genes, variation in the nucleotide sequences of their encoded effectors or acquisition of novel virulence mechanisms through horizontal gene transfer, allows the pathogen to enter and

colonise a novel host which is unable to detect its presence due to a lack of shared evolutionary history, thereby a new pathogenic strain (pathovar) is formed (Vinatzer *et al.*, 2014).

1.9.7 Bacteria isolated from necrotic lesions of AOD affected trees

Several genera and multiple species of bacteria have been consistently isolated and others infrequently isolated from necrotic lesions of AOD affected trees (Denman *et al.*, 2014). Frequently isolated species include *Gibbsiella quercinecans* (Brady *et al.*, 2010), *Brenneria goodwinii* (Denman *et al.*, 2012), infrequently isolated species include *Lonsdalea quercina* (Brady *et al.*, 2014a), *Rahnella victoriana* and *Rahnella variigena* (Brady *et al.*, 2014b). Several authorities have suggested that bacterial species including *G. quercinecans* and *B. goodwinii* are secondary necrogenic pathogens, and have proposed that bacteria consistently isolated from necrotic lesions of affected oaks are secondary pathogens (Vansteenkiste *et al.*, 2004; Denman *et al.*, 2014).

1.10 Koch's postulates: proving disease causation by an infectious agent

Koch's postulates, sometimes known as the Henle-Koch postulates, are the foundation of pathology research and the standard criteria for proving disease causation by an infectious agent (Evans, 1976). The intellecutual framework for Koch's postulates was developed in 1840 by Jakob Henle, in his seminal book, *On Miasmata and Contaigae* (Henle, 1840) which associated micro-organisms with disease, however the ideas were developed and introduced by his colleague Robert Koch, who presented to the Physiological Society in Berlin on the 24th of March 1882, entitled '*Die Aetiologie der Tuberculose*', translated as 'On the origin of Tuberculosis'. To prove the aetiology of a disease there are 3 postulates to satisfy (from Evans, 1976, translated from the original German): 1) The parasite occurs in every case of the disease in question and under circumstances which can account for the pathological changes and clinical course of the disease; 2) it occurs in no other disease as a fortuitous and non-pathogenic parasite; 3) after being fully isolated from the body and repeatedly grown in pure culture, it can induce the disease anew. A fourth postulate was widely advocated shortly after and has been extensively applied over the years; this fourth postulate specifies that after the pathogen has been inoculated into an unaffected host, it can be re-isolated (Fredricks and Relman, 1996). However, shortly after publication, Koch realised his postulates had limitations, as unlike his study organism *Mycobacterium tuberculosis*, the causative agent of tuberculosis, many organisms such as *M. leprae* cannot be

cultured in the laboratory. Furthermore, some pathogens could exist asymptomatically, such as the bacterium *Vibrio cholera* (the causative agent of cholera), which can be isolated and cultured from an individual who continues to shed the pathogen whilst remaining asymptomatic. In the cases of cholera, postulates 1 and 2 were not fulfilled, as the pathogen is not found in every case of the disease and there is no definitive association between the pathogen and host. Koch himself described this in a subsequent conference address in 1891, revealing that the postulates could not be rigidly applied to all pathogens where the causative agent was unequivocal, but satisfaction of all three postulates was not achieved (Rivers, 1937). However, the aim of Koch's research was not to assert implacable scientific laws, but to prove that microorganisms were the causative agents of disease and present a scientifically rigorous framework which would bring together cumulative evidence from microbiological, pathological, and clinical sources to ultimately deduce the agents of disease (Fredricks and Relman, 1996).

Since their publication over a hundred years ago, Koch's postulates have provided a clear framework for linking disease to conventional pathogens (Aguzzi *et al.*, 2008). However, adaptations to the postulates have been proposed several times over the past century (Hill, 1965; Fredricks and Relman, 1996; Falkow, 2004; Walker *et al.*, 2006). It is now known that infectious disease can be caused by organisms which comply with few if any of the postulates but are unequivocally the causative agents of disease, these include inanimate viruses and prions, and parasites, such as *Plasmodium falciparum*, which causes malaria (Aguzzi *et al.*, 2008). The increasing limitations of Koch's postulates became more apparent with the advent of molecular technologies, and led to the development of molecular Koch's postulates or Falkow's postulates, which were based on the original postulates, but proposed that a gene was capable of causing disease (Falkow, 1988). A recent update of the postulates extends their definition to encompass the broad spectrum of pathogens which are outside the definitions of the original postulates, and those of Falkow, to encompass disease caused by multiple organisms (polymicrobial infection) (Byrd and Segre, 2016). These updated postulates recommend using contemporary sequencing in combination with traditional culture methods to identify individual pathogens and infection by microbial consortia. Byrd and Serge proposed four postulates: 1) sequencing to classify all members of the microbial community; 2) computational models to assess microbes, both necessary and sufficient for disease induction; 3) targeted culturing to isolate microbes of interest from the disease host; 4) testing primary isolates and consortia in relevant disease models. Stringent application of the original postulates in the light of modern knowledge and technologies would clearly be wrong; it may therefore be apt to periodically revisit the postulates as novel

technologies reveal greater nuances to infectious disease. The postulates should be continuously refined but crucially, they should be conducted with scientific rigour, this would satisfy Koch and Falkow, as this was how they intended their postulates to be investigated.

There is clear evidence for extension of the postulates, to move from the scientific orthodoxy of pathogens as monospecies, to the inclusion of microbial communities as putative pathogens (the pathobiome). This is evident in phytopathogenic polymicrobial infections such as olive knot disease, caused by interaction between an interchangeable consortia of putative pathogenic bacteria or pathobiota, including the genera: *Pantoea, Erwinia*, and *Enterobacter*, and *Pseudomonas savastanoi* pv. *savastanoi* (*Pss*), (Buonaurio *et al.*, 2015; Lamichhane and Venturi, 2015). When the aetiology of olive knot disease was first described, the investigators named *P. agglomerans* as the pathogen, as it fulfilled Koch's postulates, however it has since been revealed that *P. agglomerans* is incapable of pathogenicity in olive knot alone, indicating that the original researches must have experimentally co-inoculated *P. agglomerans* alongside the true pathogen *Pss* (Marchi *et al.*, 2006). This example highlights the difficultly of characterising polymicrobial infections, however, using Byrd and Serge' postulates, the erroneous identification of the original olive knot disease would likely be avoided. A modern discovery of olive knot disease would apply metagenomics to extract a taxonomic inventory revealing potential pathobiota and their pathogenicity potential, metatranscriptomics would reveal community activity and could potentially highlight the necessity of the identified community members and reject the single pathogen hypothesis. However, the understanding of the pathobiome is in development, and full understanding is restrained by insufficient tools and methods for characterisation of polymicrobial pathogens, particularly in the interpretation of biochemical interactions within a disease, which are affected by substantial and complex interactions, therefore the study of polymicrobial infections represent a significant research challenge (Vayssier-Taussat *et al.*, 2014). The biochemical contribution to pathogenicity of disease inducing pathobiota, such as in olive knot disease, are complex and may be extensive; for example, *Pss* is unable to degrade salicylic acid, whereas *E. toletana* and *P. agglomerans* convert salicylic acid to catechol, debilitating the plants immune response (da Silva *et al.*, 2014). The study of pathobiome related disease is at an early stage, however, the ability of interacting pathobiota, to cause disease is established (Marchi *et al.*, 2006). Synergy of microbial community members to advance a diseased state is unsurprising, as bacteria rarely exist in isolation in the natural environment and typically interact in multi-species communities, perhaps it is surprising that with the extensive recognition of

interacting microorganisms, scientific research has extensively focused on single species infections.

1.11 Identification and characterisation of bacterial pathogens, from the environment to the laboratory

1.11.1 Taxonomic identification

Taxonomic relationships between bacterial species are phylogenetically determined by comparing homologous nucleotide or amino acid sequences, known as molecular markers, to reconstruct their evolutionary history (Patwardhan *et al.*, 2014). Determining evolutionary relationships using molecular markers was first proposed in 1977 and identified ribosomal RNA as an operon containing universally comparable genes, reflecting ancestral history, and the 16S rRNA gene as a suitable marker for the delineation of bacterial species (Woese and Fox, 1977). This method revolutionised the inference of phylogenetic relationships and our understanding of the tree of life, by matching gene or protein similarities and graphically representing these relationships within a phylogenetic tree. However, it is recognised that the 16S rRNA gene has discriminatory limitations between many closely related bacteria, for example *Lactobacillus* spp. have between 98.7% and 99.9% homology in the 16S gene across species, negating the resolving power of the marker. In contrast, the conserved 40-kDa heat-shock chaperone protein *dnaJ* has 87.1% similarity across species, and is therefore a more appropriate marker for this genus (Huang *et al.*, 2015). Alternative molecular markers such as *dnaJ* have been sought to reflect evolutionary diversity between clades which cannot be resolved using the 16S rRNA gene, some commonly used alternatives include DNA gyrase B (*gyrB*) a topoisomerase and the 70-kDa heat-shock chaperone protein *dnaK* (Huang and Lee, 2011).

To increase phylogenetic resolving power, molecular markers can be combined, and two common methods have been widely adopted; multi-locus sequence typing (MLST), uses a combination of conserved maker loci or housekeeping genes, which are evolutionary conserved and hence under purifying selection, this method has resolved the phylogeny of the xylem-limited plant pathogen *Xylella fastidiosa*, using seven conserved housekeeping genes (Scally *et al.*, 2005). An alternative approach is multi-locus sequence analysis (MLSA), which targets conserved and

divergent loci, a method which has resolved the phylogeny of the bacterial wilt pathogen *Ralstonia solanacearum*, using four housekeeping genes and three plasmid variable loci (Castillo and Greenberg, 2007). Molecular marker resolution of evolutionary relationships was a landmark scientific breakthrough, and is still widely used, however it does not reflect genome-wide genomic diversity and may fail to resolve newly discovered bacteria which have identical conserved marker genes but are divergent at other critical loci (Lang *et al.*, 2013). An example of these constraints is found in resolving the phylogeny of the endophyte *Erwinia billingae* which has a similar repertoire of conserved genes to the pathogen *E. tasmaniensis* and would be difficult to distinguish without prior knowledge of the recently acquired virulence genes (Kube *et al.*, 2010). The concept of taxonomic resolution using marker genes was proposed prior to the genomics revolution and the realisation of the extent of horizontal gene transfer between bacterial species, therefore with whole genome sequence data a more informed, high resolving power phylogeny can be inferred.

1.11.2 DNA sequencing

Whole genome sequencing (WGS) and expeditious annotation methods have enabled the rapid characterisation of putative and established bacterial pathogens. Previously, phenotypic characterisation of bacterial pathogens would take a team of scientists several years. The introduction of next generation sequencing (NGS) methods triggered a 'gold rush' in the uptake of this technology, exponentially increasing the number of sequenced bacterial genomes combined with a reduction in associated consumable and labour costs, as large sequencing centres produced data on an industrial scale (Hall, 2013). Research on bacterial phytopathogens has benefited significantly from the advancement of sequencing technologies, allowing the molecular determinants of pathogenicity to be revealed (Lindeberg, 2012). For novel pathogens, WGS uncovers fundamental aspects of their biology which may otherwise remain hidden. For example, the economically important crop pathogen *Pectobacterium atrosepticum* WGS project was published in 2004 (Bell *et al.*, 2004), revealing previously unknown pathogenicity determinants, which ultimately led to production of a resistant transgenic potato plant (Toth *et al.*, 2015). WGS provides insights on the underlying biology of bacterial pathogens, including virulence potential and allows the generation of novel hypotheses to facilitate research on the mechanisms of infection and ultimately inform prevention strategies.

1.11.3 RNA sequencing

RNA sequencing (RNA-seq) is an NGS enabled method for the high-resolution analysis of genome wide expression data (transcriptomics), at a specific time (Wang *et al.*, 2009). Since its inception, RNA-seq has been widely applied to reveal the mechanisms of bacteria-host interactions, however, studies of bacterial phytopathogens have largely remained unexplored (Chapelle *et al.*, 2015). Adaptation of bacteria to a suitable host or within a novel environment can be investigated through analysis of changes to their gene expression. For example, a recent study revealed the central role of small RNAs (sRNAs) in the adaptive response of *P. atrosepticum* to starvation, and their involvement in the nutrient deficiency signalling cascade (Kwenda *et al.*, 2016). Previously, mutagenesis and targeted gene expression studies were the methods of choice for identification of pathogenicity related genes, however, virulence genes are often controlled by multiple interactive regulators, therefore a global analysis is required to gain a comprehensive understanding of this process (Creecy and Conway, 2015). RNA-seq technology is currently the foremost method for measuring complex changes in global expression within a polymicrobial infection.

1.11.4 Environmental DNA and RNA sequencing

Random sequencing of environmental DNA and RNA reveals community membership, functionality and activity. These methods known as metagenomics (community DNA sequencing) and metatranscriptomics (community RNA sequencing), have exponentially increased the ability to gather community data on taxonomy, function, and activity. Metagenomics provides data on environmental community membership and functional potential. The method allows characterisation of unculturable bacterial species, annotation of previously unknown functional genes and the *in silico* availability of enzymes of potential industrial importance (Knight *et al.*, 2012). Furthermore, a predicative understanding of an ecosystem can be inferred from genome-wide coding domains revealed by metagenomics, allowing predictions on, for example, pathogenicity, nitrogen fixation or photosynthetic ability. Metatranscriptomics is a developing technology which provides a snapshot of community gene expression at a given time point. The widespread uptake of this information rich technology has been hindered by the short half-life of RNA, the lack of transferable techniques from eukaryotic RNA sequencing due to the absence of polyA tails on bacterial mRNAs and insufficient

standardised post-sequencing processing methods (Filiatrault, 2011; Westreich *et al.*, 2016). However, unlike metagenomics, metatranscriptomics provides data on actual community activity, and the fluctuation of this activity in response to environmental stimuli (Moran, 2009). Metatranscriptomics offers *in situ* gene expression data of complex communities, unravelling reservoirs of information on bacterial gene expression responses to changing environments (Jiang *et al.*, 2016).

1.12 Aims and objectives

The aim of this project is to characterise the role of *Gibbsiella quercinecans* and *Brenneria goodwinii* in AOD, using molecular methods to understand their virulence potential. Within that aim are three objectives;

1) to sequence and annotate the whole genomes of G. quercinecans and B. goodwinii.

Prior to whole genome sequencing, the bacterial isolates were recovered from necrotic lesion of AOD affected trees and distinguished taxonomically, initially using the *gyrB* molecular marker and a novel method using the hypervariable non-coding intergenic spacer region 1 (ITS1) of the rRNA operon, reported in Chapter 2. The whole genomes from strains of *G. quercinecans*, *B. goodwinii*, and *Ewingella americana*, and two positive controls were then sequenced to draft status and their genomic potential for virulence was assessed, reported in Chapter 3. *G. quercinecans* FRB97 and *B. goodwinii* FRB141 were sequenced to finished, reference genome status, reported in Chapter 4.

2) To understand the relationship between oak tree tissue and *G. quercinecans and B. goodwinii* virulence.

Transcripts from *in vitro* cultures of *B. goodwinii* and *G. quercinecans* were sequenced in the presence and absence of oak tissue and aligned against the reference genomes, this allowed the measurement of differential virulence expression induced by oak tree tissue, reported in Chapter 5.

3) To characterise the importance of *G. quercinecans and B. goodwinii* in active necrotic lesions.

Environmental DNA and RNA from active necrotic lesions were aligned against the reference genomes to understand the importance of the *G. quercinecans* and *B. goodwinii* within active AOD necrotic lesions, reported in Chapter 6.

1.13 References

Aguilar, C.N., Rodríguez, R., Gutiérrez-Sánchez, G., Augur, C., Favela-Torres, E., Prado-Barragan, L.A., *et al.* (2007) Microbial tannases: Advances and perspectives. *Appl. Microbiol. Biotechnol.* **76**: 47–59.

Aguzzi, A., Baumann, F., and Bremer, J. (2008) The prion's elusive reason for being. *Annu. Rev. Neurosci.* **31**: 439–77.

Anderson, J.P., Gleason, C.A., Foley, R.C., Thrall, P.H., Burdon, J.B., and Singh, K.B. (2010) Plants versus pathogens: An evolutionary arms race. *Funct. Plant Biol.* **37**: 499–512.

Arnold, D.L. and Jackson, R.W. (2011) Bacterial genomes: evolution of pathogenicity. *Curr. Opin. Plant Biol.* **14**: 385–91.

Auclair, A.N., Worrest, R., Lachance, D., and Martin, H. (1992) Climatic perturbation as a general mechanism of forest dieback. *Forest Decline Concepts*. The American Phytopathological Society, pp. 38–58.

Ayres, M.P., Clausen, T.P., Maclean, S.F., Redman, A.M., and Reichardt, P.B. (1997) Diversity of structure and antiherbivore activity in condensed tannins. *Ecology* **78**: 1696–1712.

Azadi, P., O' Neill, M.A., Bergmann, C., Darvill, A.G., and Albersheim, P. (1995) The backbone of the pectic polysaccharide rhamnogalacturonan I is cleaved by an endohydrolase and an endolyase. *Glycobiology* **5**: 783–789.

Badieyan, S., Bevan, D.R., and Zhang, C. (2012) Study and design of stability in GH5 cellulases. *Biotechnol. Bioeng.* **109**: 31–44.

Barbehenn, R. V. and Constabel, C.P. (2011) Tannins in plant-herbivore interactions. *Phytochemistry* **72**: 1551–1565.

Barras, F., van Gijsegem, F., and Chatterjee, A. (1994) Extracellular enzymes and pathogenesis of soft-rot *Erwinia*. *Annu. Rev. Phytopathol.* **32**: 201–34.

Bell, K.S., Sebaihia, M., Pritchard, L., Holden, M.T.G., Hyman, L.J., Holeva, M.C., *et al.* (2004) Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora* subsp. *atroseptica* and characterization of virulence factors. *Proc. Natl. Acad. Sci. U. S. A.* **101**: 11105–10.

Benko, Z., Siika-aho, M., Viikari, L., and Reczey, K. (2008) Evaluation of the role of xyloglucanase in the enzymatic hydrolysis of lignocellulosic substrates. *Enzyme Microb. Technol.* **43**: 109–114.

Bhat, T.K., Singh, B., and Sharma, O.P. (1998) Microbial degradation of tannins - A current perspective. *Biodegradation* **9**: 343–357.

Biosca, E.G., González, R., López-López, M.J., Soria, S., Montón, C., Pérez-Laorga, E., and López, M.M. (2003) Isolation and Characterization of *Brenneria quercina*, Causal Agent for bark canker and drippy Nut of *Quercus* spp. in Spain. *Phytopathology* **93**: 485–92.

Brady, C., Denman, S., Kirk, S., Venter, S., Rodríguez-Palenzuela, P., and Coutinho, T. (2010) Description of *Gibbsiella quercinecans* gen. nov., sp. nov., associated with Acute Oak Decline. *Syst. Appl. Microbiol.* **33**: 444–50.

Brady, C., Hunter, G., Kirk, S., Arnold, D., and Denman, S. (2014a) Description of *Brenneria roseae* sp. nov. and two subspecies, *Brenneria roseae* subspecies *roseae* ssp. nov and *Brenneria roseae* subspecies *americana* ssp. nov. isolated from symptomatic oak. *Syst. Appl. Microbiol.*

Brady, C., Hunter, G., Kirk, S., Arnold, D., and Denman, S. (2014b) *Rahnella victoriana* sp. nov., *Rahnella bruchi* sp. nov., *Rahnella woolbedingensis* sp. nov., classification of *Rahnella* genomospecies 2 and 3 as *Rahnella variigena* sp. nov. and *Rahnella inusitata* sp. nov., respectively and emended description of the genus. *Syst. Appl. Microbiol.* **37**: 545–552.

Brasier, C.M. (1991) *Ophiostoma novo-ulmi* sp. nov., causative agent of current Dutch elm disease pandemics. *Mycopathologia* **115**: 151–161.

Brookes, P.C. and Wigston, D.L. (1979) Variation of morphological and chemical characteristics of acorns from population of *Quercus petraea* (Matt.) Liebl., *Quercus robur* L. and their hybrids. *Watsonia* **12**: 315–324.

Brown, N., Inward, D.J.G., Jeger, M., and Denman, S. (2014) A review of *Agrilus biguttatus* in UK forests and its relationship with acute oak decline. *Forestry* 1–11.

Buck, A., Burger, B., and Wolfrum, G. (2004) Forest research – challenges and concepts in a changing world. *IUFRO Occassional Pap.* **16**: 1–79.

Bugg, T., Ahmad, M., Hardiman, E., and Singh, R. (2011) The emerging role for bacteria in lignin degradation and bio-product formation. *Curr. Opin. Biotechnol.* **22**: 394–400.

Bugg, T.D.H., Ahmad, M., Hardiman, E.M., and Rahmanpour, R. (2011) Pathways for degradation of lignin in bacteria and fungi. *Nat. Prod. Rep.* **28**: 1883–96.

Buonaurio, R., Moretti, C., da Silva, D.P., Cortese, C., Ramos, C., and Venturi, V. (2015) The olive knot disease as a model to study the role of interspecies bacterial communities in plant disease. *Front. Plant Sci.* **6**: 1–12.

Burdon, J.J. and Thrall, P.H. (2009) Coevolution of plants and their pathogens in natural habitats. *Science* **324**: 755–756.

Byrd, B.A.L. and Segre, J.A. (2016) Adapting Koch's postulates. *Sci. Mag.* **351**: 224–226.

Carpita, N.C. and Gibeaut, D.M. (1993) Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *Plant J.* **3**: 1–30.

Castillo, J.A. and Greenberg, J.T. (2007) Evolutionary dynamics of *Ralstonia solanacearum*. *Appl. Environ. Microbiol.* **73**: 1225–1238.

Chapelle, E., Alunni, B., Malfatti, P., Solier, L., Pédron, J., Kraepiel, Y., and Van Gijsegem, F. (2015) A straightforward and reliable method for bacterial in planta transcriptomics: Application to the *Dickeya dadantii/Arabidopsis thaliana* pathosystem. *Plant J.* **82**: 352–362.

Charkowski, A., Blanco, C., Condemine, G., Expert, D., Franza, T., Hayes, C., *et al.* (2012) The role of secretion systems and small molecules in soft-rot *Enterobacteriaceae* pathogenicity. *Annu. Rev. Phytopathol.* **50**: 425–49.

Chowdhury, S.P., Khanna, S., Verma, S.C., and Tripathi, A.K. (2004) Molecular diversity of tannic acid degrading bacteria isolated from tannery soil. *J. Appl. Microbiol.* **97**: 1210–1219.

Ciesla, W.M. and Donaubauer, E. (1994) Decline and dieback of trees and forests. A global overview. FAO Forestry Paper.

Collins, T., Gerday, C., and Feller, G. (2005) Xylanases, xylanase families and extremophilic xylanases. *FEMS Microbiol. Rev.* **29**: 3–23.

Collmer, A., Lindeberg, M., Petnicki-Ocwieja, T., Schneider, D.J., and Alfano, J.R. (2002) Genomic mining type III secretion system effectors in *Pseudomonas syringae* yields new picks for all TTSS

prospectors. *Trends Microbiol.* **10**: 462–469.

Cottrell, J.E., Munro, R.C., Tabbener, H.E., Gillies, A.C.M., Forrest, G.I., Deans, J.D., and Lowe, A.J. (2002) Distribution of chloroplast DNA variation in British oaks (*Quercus robur* and *Q. petraea*): The influence of postglacial colonisation and human management. *For. Ecol. Manage.* **156**: 181–195.

Creecy, J.P. and Conway, T. (2015) Quantitative bacterial transcriptomics with RNA-seq. *Curr. Opin. Microbiol.* **23**: 133–140.

Dangl, J.L. and Jones, J.D.G. (2001) Defence responses to infection. *Nature* **411**: 826–833.

Denman, S., Brady, C., Kirk, S., Cleenwerck, I., Venter, S., Coutinho, T., and De Vos, P. (2012) *Brenneria goodwinii* sp. nov., associated with acute oak decline in the UK. *Int. J. Syst. Evol. Microbiol.* **62**: 2451–2456.

Denman, S., Brown, N., Kirk, S., Jeger, M., and Webber, J. (2014) A description of the symptoms of Acute Oak Decline in Britain and a comparative review on causes of similar disorders on oak in Europe. *Forestry* **87**: 535–551.

Denman, S. and Webber, J. (2009) Oak declines: new definitions and new episodes in Britain. *Q. J. For.* **103**: 285–290.

Deschamps, A.. and Lebeault, J. (1984) Production of gallic acid from tara tannin by bacterial strains. *Biotechnol. Lett.* **6**: 237–242.

Doonan, J., Denman, S., Gertler, C., Pachebat, J. A, Golyshin, P.N., and McDonald, J.E. (2015) The intergenic transcribed spacer region 1 as a molecular marker for identification and discrimination of *Enterobacteriaceae* associated with acute oak decline. *J. Appl. Microbiol.* **118**: 193–201.

Doughari, J. (2015) An overview of plant immunity. *J. Plant Pathol. Microbiol.* **6**: 1000322.

Eckardt, N.A. (2003) Cellulose synthesis takes the CesA train. *Plant Cell* **15**: 1685–1687.

Eudes, A., Liang, Y., Mitra, P., and Loqué, D. (2014) Lignin bioengineering. *Curr. Opin. Biotechnol.* **26**: 189–198.

Evans, A.S. (1976) Causation and disease: the Henle-Koch postulates revisited. *Yale J. Biol. Med.* **49**: 175–195.

Falkow, S. (2004) Molecular Koch's postulates applied to bacterial pathogenicity--a personal recollection 15 years later. *Nat. Rev. Microbiol.* **2**: 67–72.

Falkow, S. (1988) Molecular Koch's postulates applied to microbial pathogenicity. *Rev. Infect. Dis.* **10**: 7–10.

Fatima, U. and Senthil-Kumar, M. (2015) Plant and pathogen nutrient acquisition strategies. *Front. Plant Sci.* **6**: 750.

Ferris, C., Davy, A J., and Hewitt, G.M. (1997) A strategy for identifying introduced provenances and translocations. *Forestry* **70**: 211–222.

Filiatrault, M.J. (2011) Progress in prokaryotic transcriptomics. *Curr. Opin. Microbiol.* **14**: 579–586.

Fredricks, D.N. and Relman, D.A. (1996) Sequence-based identification of microbial pathogens : a reconsideration of Koch's postulates. *Clin Microbiol Rev* **9**: 18–33.

Gibbs, J.N. and Greig, B.J.W. (1997) Biotic and abiotic factors affecting the dying back of pedunculate oak *Quercus robur* L. *Forestry* **70**: 399–406.

Gibson, D.M., King, B.C., Hayes, M.L., and Bergstrom, G.C. (2011) Plant pathogens as a source of diverse enzymes for lignocellulose digestion. *Curr. Opin. Microbiol.* **14**: 264–270.

Glazebrook, J. (2005) Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu. Rev. Phytopathol.* **43**: 205–227.

Hall, N. (2013) After the gold rush. *Genome Biol.* **14**: 115.

Hayashi, T. and Kaida, R. (2011) Functions of xyloglucan in plant cells. *Mol. Plant* **4**: 17–24.

Henle, J. (1840) On miasmata and contagie. Johns Hopkins Press, Baltimore.

Hernes, P. and Hedges, J. (1992) Geochemistry of Tannin: Methods and Applications. *Plant Polyphenols 2.*, pp. 853–865.

Herron, S.R., Benen, J. A, Scavetta, R.D., Visser, J., and Jurnak, F. (2000) Structure and function of pectic enzymes: virulence factors of plant pathogens. *Proc. Natl. Acad. Sci. U. S. A.* **97**: 8762–8769.

Hill, A.B. (1965) the Environment and Disease: Association or Causation? *Proc. R. Soc. Med.* **58**: 295–300.

Holdenrieder, O., Pautasso, M., Weisberg, P.J., and Lonsdale, D. (2004) Tree diseases and landscape processes: The challenge of landscape pathology. *Trends Ecol. Evol.* **19**: 446–452.

Huang, C.-H., Chang, M.-T., Huang, L., and Chu, W.-S. (2015) The *dnaJ* gene as a molecular discriminator to differentiate among species and strain within the *Lactobacillus casei* group. *Mol. Cell. Probes* **29**: 479–484.

Huang, C.H. and Lee, F.L. (2011) The *dnaK* gene as a molecular marker for the classification and discrimination of the *Lactobacillus casei* group. *Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol.* **99**: 319–327.

Janse, J.D. and Obradovic, A. (2010) Xylella fastidiosa: Its biology, diagnosis, control and risks. *J. Plant Pathol.* **92**:.

Jiang, Y., Xiong, X., Danska, J., and Parkinson, J. (2016) Metatranscriptomic analysis of diverse microbial communities reveals core metabolic pathways and microbiome-specific functionality. *Microbiome* **4**: 2.

Jiménez, N., Esteban-Torres, M., Mancheño, J.M., de Las Rivas, B., and Muñoz, R. (2014) Tannin degradation by a novel tannase enzyme present in some *Lactobacillus plantarum* strains. *Appl. Environ. Microbiol.* **80**: 2991–7.

Jones, J.D.G. and Dangl, J.L. (2006) The plant immune system. *Nat. Rev.* **444**: 323–329.

Kim, S.H., Kim, S.H., Yoo, S.J., Min, K.H., Nam, S.H., Cho, B.H., and Yang, K.Y. (2013) Putrescine regulating by stress-responsive MAPK cascade contributes to bacterial pathogen defense in Arabidopsis. *Biochem. Biophys. Res. Commun.* **437**: 502–508.

Kirk, T.K. and Farrell, R.L. (1987) Enzymatic "Combustion": the micrbial degradation of lignin. *Ann Rev Microbiol* **41**: 465–505.

Klemm, D., Heublein, B., Fink, H.P., and Bohn, A. (2005) Cellulose: Fascinating biopolymer and sustainable raw material. *Angew. Chemie - Int. Ed.* **44**: 3358–3393.

Knight, R., Jansson, J., Field, D., Fierer, N., Desai, N., Fuhrman, J. A., *et al.* (2012) Unlocking the potential of metagenomics through replicated experimental design. *Nat. Biotechnol.* **30**: 513–520.

Kube, M., Migdoll, A.M., Gehring, I., Heitmann, K., Mayer, Y., Kuhl, H., *et al.* (2010) Genome comparison of the epiphytic bacteria *Erwinia billingiae* and *E. tasmaniensis* with the pear pathogen *E. pyrifoliae*. *BMC Genomics* **11**: 393.

Kuhad, R.C., Gupta, R., and Singh, A. (2011) Microbial cellulases and their industrial applications. *Enzyme Res.* **2011**: 280696.

Kwenda, S., Gorshkov, V., Ramesh, A.M., Naidoo, S., Rubagotti, E., Birch, P.R.J., and Moleleki, L.N. (2016) Discovery and profiling of small RNAs responsive to stress conditions in the plant pathogen *Pectobacterium atrosepticum*. *BMC Genomics* **47**:.

Lamichhane, J.R. and Venturi, V. (2015) Synergisms between microbial pathogens in plant disease complexes: a growing trend. *Front Plant Sci* **6**: 1–12.

Lang, J.M., Darling, A.E., and Eisen, J.A. (2013) Phylogeny of Bacterial and Archaeal Genomes Using Conserved Genes: Supertrees and Supermatrices. *PLoS One* **8**: e62510.

Lee, S.-J. and Rose, J.K.C. (2010) Mediation of the transition from biotrophy to necrotrophy in hemibiotrophic plant pathogens by secreted effector proteins. *Plant Signal. Behav.* **5**: 769–772.

Lindeberg, M. (2012) Genome-Enabled Perspectives on the Composition, Evolution, and Expression of Virulence Determinants in Bacterial Plant Pathogens. *Annu. Rev. Phytopathol.* **50**: 111–132.

Linhardt, R.J., Galliher, P.M., and Cooney, C.L. (1987) Polysaccharide lyases. *Appl. Biochem. Biotechnol.* **12**: 135–176.

Liu, P.-P., Yang, Y., Pichersky, E., and Klessig, D.F. (2010) Altering expression of benzoic acid/salicylic acid carboxyl methyltransferase 1 compromises systemic acquired resistance and PAMP-triggered immunity in arabidopsis. *Mol. Plant. Microbe. Interact.* **23**: 82–90.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**: D490-5.

Ma, B., Hibbing, M.E., Kim, H., Reedy, R.M., Yedidia, I., Breuer, J., *et al.* (2007) Host Range and Molecular Phylogenies of the Soft Rot Enterobacterial Genera *Pectobacterium* and *Dickeya*. *Phytopathology* **97**: 1150–1163.

Makkar, H.P.S., Dawra, R.K., and Singh, B. (1991) Tannins levels in leaves of some oak species at different stage of maturity. *J. Sci. Food Agric.* **54**: 513–519.

Manion, P. (1981) Tree Disease Concepts Prentice Hall, Englewood Cliffs, NJ.

Marçais, B. and Desprez-Loustau, M.L. (2014) European oak powdery mildew: Impact on trees, effects of environmental factors, and potential effects of climate change. *Ann. For. Sci.* **71**: 633–642.

Marchi, G., Sisto, A., Cimmino, A., Andolfi, A., Cipriani, M.G., Evidente, A., and Surico, G. (2006) Interaction between Pseudomonas savastanoi pv. savastanoi and Pantoea agglomerans in olive knots. *Plant Pathol.* **55**: 614–624.

May, C.D. (1990) Industrial pectins: Sources, production and applications. *Carbohydr. Polym.* **12**: 79–99.

Mendes, R., Garbeva, P., and Raaijmakers, J.M. (2013) The rhizosphere microbiome: Significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. *FEMS Microbiol. Rev.* **37**: 634–663.

Moran, M.A. (2009) Metatranscriptomics: Eavesdropping on Complex Microbial Communities. *Microbe* **4**: 329–335.

Mosedale, J. and Feuillat, F. (1998) Variability of wood extractives among *Quercus robur* and *Quercus petraea* trees from mixed stands and their relation to wood anatomy and leaf morphology. *Can. J. For. Res.* **28**: 994–1006.

Mueller-Dombois, D. (1992) A natural dieback theory, cohort senescence as an alternative to the decline disease theory. *Forest Decline Concepts*. The American Phytopathological Society, MN, USA, pp. 26–37.

Patwardhan, A., Ray, S., and Roy, A. (2014) Molecular markers in phylogenetic studies-A review. *J. Phylogenetics Evol. Biol.* **2**: 1–9.

Pepi, M., Lampariello, L.R., Altieri, R., Esposito, A., Perra, G., Renzi, M., *et al.* (2010) Tannic acid degradation by bacterial strains *Serratia* spp. and *Pantoea* spp. isolated from olive mill waste mixtures. *Int. Biodeterior. Biodegrad.* **64**: 73–80.

Perez-Maldonado, R. A and Norton, B.W. (1996) Digestion of ^{14}C-labelled condensed tannins from *Desmodium intortum* in sheep and goats. *Br. J. Nutr.* **76**: 501–13.

Pitzschke, A. (2013) *Agrobacterium* infection and plant defense-transformation success hangs by a thread. *Front. Plant Sci.* **4**: 519.

Pliny, the E. (1991) Natural History: A selection. Penguin Classics, London, UK.

Provan, J. and Bennett, K.D. (2008) Phylogeographic insights into cryptic glacial refugia. *Trends Ecol. Evol.* **23**: 564–571.

Puech, J.L., Feuillat, F., and Mosedale, J.R. (1999) The tannins of oak heartwood: Structure, properties, and their influence on wine flavor. *Am. J. Enol. Vitic.* **50**: 469–478.

Rennie, E.A. and Scheller, H.V. (2014) Xylan biosynthesis. *Curr. Opin. Biotechnol.* **26**: 100–107.

Rivers, T.M. (1937) Viruses and Koch's Postulates. *J. Bacteriol.* **33**: 1–12.

Ruiz-Dueñas, F.J. and Martínez, Á.T. (2009) Microbial degradation of lignin: How a bulky recalcitrant polymer is efficiently recycled in nature and how we can take advantage of this. *Microb. Biotechnol.* **2**: 164–177.

Sallé, A., Nageleisen, L.M., and Lieutier, F. (2014) Bark and wood boring insects involved in oak declines in Europe: Current knowledge and future prospects in a context of climate change. *For. Ecol. Manage.* **328**: 79–93.

Scalbert, A. (1991) Antimicrobial properties of tannins. *Phytochemistry* **30**: 3875–3883.

Scally, M., Schuenzel, E.L., Stouthamer, R., and Nunney, L. (2005) Multilocus sequence type system for the plant pathogen *Xylella fastidiosa* and relative contributions of recombination and point mutation to clonal diversity. *Appl. Environ. Microbiol.* **71**: 8491–8499.

Schama, S. (2009) A History of Britain: The Fate of Empire 1776-2000. Bodley Head, London, UK.

Scheller, H.V. and Ulvskov, P. (2010) Hemicelluloses. *Annu. Rev. Plant Biol.* **61**: 263–289.

Scortichini, M., Stead, D.E., and Pia Rossi, M. (1993) Oak decline: Aerobic bacteria associated with declining Quercus cerris in Central Italy. *Eur. J. For. Pathol.* **23**: 120–127.

Serrano, J., Puupponen-Pimiä, R., Dauer, A., Aura, A.M., and Saura-Calixto, F. (2009) Tannins: Current knowledge of food sources, intake, bioavailability and biological effects. *Mol. Nutr. Food Res.* **53**: 310–329.

da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R., Quaggio, R.B., *et al.* (2002)

Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* **417**: 459–63.

da Silva, D.P., Castaneda-Ojeda, M.P., Moretti, C., Buonaurio, R., Ramos, C., and Venturi, V. (2014) Bacterial multispecies studies and microbiome analysis of a plant disease. *Microbiol. (United Kingdom)* **160**: 556–566.

Thakur, B.R., Singh, R.K., Handa, A.K., and Rao, M. A (2009) Chemistry and uses of pectin — A review. *Crit. Rev. Food Sci. Nutr.* **37(1)**: 47–73.

Thomas, F. (2008) Recent advances in cause-effect research on oak decline in Europe. *CAB Rev. Perspect. Agric. Vet. Sci. Nutr. Nat. Resour.* **3**: 1–12.

Thomas, F.M., Blank, R., and Hartmann, G. (2002) Abiotic and biotic factors and their interactions as causes of oak decline in Central Europe. *For. Pathol.* **32**: 277–307.

Topakas, E., Vafiadi, C., and Christakopoulos, P. (2007) Microbial production, characterization and applications of feruloyl esterases. *Process Biochem.* **42**: 497–509.

Toth, I., Humphris, S., Campbell, E., and Pritchard, L. (2015) Why genomics research on *Pectobacterium* and *Dickeya* makes a difference. Am. *J. Potato Res.* **92**:218

Toth, I.K. and Birch, P.R.J. (2005) Rotting softly and stealthily. *Curr. Opin. Plant Biol.* **8**: 424–429.

Toth, I.K., Pritchard, L., and Birch, P.R.J. (2006) Comparative genomics reveals what makes an enterobacterial plant pathogen. *Annu. Rev. Phytopathol.* **44**: 305–36.

Tyler, M. (2008) British Oaks 1st ed. The Crowood Press, Wiltshire, UK.

van Valen, L. (1973) A new evolutionary law. *Evolu. Theory.* **1**: 1–30.

Vansteenkiste, D., Tirry, L., Van Acker, J., and Stevens, M. (2004) Predispositions and symptoms of *Agrilus* borer attack in declining oak trees. *Annu. For. Sci.* **61**: 815–823.

Vayssier-Taussat, M., Albina, E., Citti, C., Cosson, J.-F., Jacques, M.-A., Lebrun, M.-H., *et al.* (2014) Shifting the paradigm from pathogens to pathobiome: new concepts in the light of meta-omics. *Front. Cell. Infect. Microbiol.* **4**: 29.

Vinatzer, B. a, Monteil, C.L., and Clarke, C.R. (2014) Harnessing Population Genomics to Understand How Bacterial Pathogens Emerge, Adapt to Crop Hosts, and Disseminate. *Annu. Rev. Phytopathol.* 1–25.

Walker, L., LeVine, H., and Jucker, M. (2006) Koch's postulates and infectious proteins. *Acta Neuropathol.* **112**: 1–4.

Wang, Z., Gerstein, M., and Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**: 57–63.

Westreich, S.T., Korf, I., Mills, D.A., and Lemay, D.G. (2016) SAMSA : a comprehensive metatranscriptome analysis pipeline. *BMC Bioinf.* **17**:399, 1-20.

Wilson, D.B. (2011) Microbial diversity of cellulose hydrolysis. *Curr. Opin. Microbiol.* **14**: 259–263.

Witzell, J. and Martín, J. A. (2008) Phenolic metabolites in the resistance of northern forest trees to pathogens — past experiences and future prospects. *Can. J. For. Res.* **38**: 2711–2727.

Woese, C.R. and Fox, G.E. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. U. S. A.* **74**: 5088–5090.

Yuan, Z.-C., Edlind, M.P., Liu, P., Saenkham, P., Banta, L.M., Wise, A.A., *et al.* (2007) The plant signal salicylic acid shuts down expression of the vir regulon and activates quormone-quenching genes in Agrobacterium. *Proc. Natl. Acad. Sci. U. S. A.* **104**: 11790–5.

Zhang, B., Cai, J., Duan, C.-Q., Reeves, M.J., and He, F. (2015) A review of polyphenolics in oak woods. *Int. J. Mol. Sci.* **16**: 6978–7014.

CHAPTER 2

The intergenic transcribed spacer region 1 as a molecular marker for identification and discrimination of *Enterobacteriaceae* associated with acute oak decline

Abstract

Both species of native oak trees in the UK (*Quercus robur* and *Quercus petraea*) are affected by a syndrome termed Acute Oak Decline (AOD). Black weeping patches on the stems of mature trees are the primary symptom of AOD and indicate the presence of necrotic lesions in the underlying tissue. Two newly described bacterial species, *Gibbsiella quercinecans* and *Brenneria goodwinii* are consistently isolated from necrotic lesions and are considered as potential pathogens of oak, but their identification against a background of other sporadic lesion-associated *Enterobacteriaceae* is a laborious and costly process. Here, the intergenic spacer region 1 (ITS1) was assessed for the rapid, inexpensive and accurate typing of bacterial strains belonging to the *Enterobacteriaceae*. This approach was initially validated using six *Enterobacteriaceae* strains isolated from oak trees before characterisation of a further twenty-eight strains. The number and length of ITS1 amplicons for each isolate varied significantly and was determined using agarose gel electrophoresis and polyacrylamide gel electrophoresis. Using these methods, the newly described AOD associated *Enterobacteriaceae*, *B. goodwinii* and *G. quercinecans* were successfully resolved to species level. The method was found to have equivalent sensitivity to the current standard method for strain identification (sequence analysis of the DNA gyrase B gene), but with reduced processing time and cost.

2.1 Introduction

Acute Oak Decline (AOD) is a syndrome partly derived from tissue necrosis of the inner bark of oak trees (Denman *et al.*, 2014). The decline is relatively new in Britain (reported cases of AOD have been observed for 20-30 years) and has been identified in both species of native oak, with the number of reported cases on the increase. The precise biotic and abiotic causes of AOD are currently unknown, but there are likely to be multiple factors involved. Networks of galleries produced by the larvae of the buprestid beetle *Agrilus biguttatus* have been discovered in trees with symptoms of AOD and are typically associated in and around areas of lesion formation (Brown *et al.*, 2014). The beetles themselves are rarely seen, but evidence of their occurrence is presented via larval galleries in the cambial zone of the oak tree. In addition, previous studies have described the consistent isolation of *Enterobacteriaceae* from necrotic lesions of affected oak trees in sampling sites across England (Brady *et al.*, 2010; Denman *et al.*, 2012). However, two newly-described bacterial species belonging to the *Enterobacteriaceae*, *Gibbsiella quercinecans* and *Brenneria goodwinii*, are consistently isolated from necrotic lesions on oak trees, with other species such as *Rahnella* spp. isolated from both healthy and necrotic tissues. Consequently, the almost exclusive association of *G. quercinecans* and *B. goodwinii* with the lesions of AOD affected trees has led to the hypothesis that these species play a central role in tissue necrosis (Denman and Webber, 2009).

The ribosomal RNA (rrn) operon (Fig. 2.1) and in particular, 16S rRNA, are conserved genomic regions commonly used for bacterial genotyping and taxonomy (Woese and Fox, 1977). However, due to the highly conserved nature of the 16S rRNA gene, resolution of closely related species within certain taxa such as the *Enterobacteriaceae* can be problematic (Janda and Abbott, 2007; Naum *et al.*, 2008). The dearth of sequence differences among closely related species contributes to a lack of phylogenetic power and limits its applicability for fine scale taxonomic resolution. The intergenic or internal transcribed spacer region 1 (ITS1) is part of the *rrn* operon, found between the genes for small (16/18S) and large (23/28S) rRNA subunits. The ITS1 region is more susceptible to mutations than other parts of the rRNA operon due to its non-functional role (Scheinert *et al.*, 1996). Consequently, it is an excellent candidate for molecular differentiation between both diverse and closely related species. ITS analysis, is a method most commonly applied to examine the extent of microbial diversity in environmental samples from soil and marine communities e.g. (Rappé *et al.*, 2002; Brown and Fuhrman, 2005), and it has been well described in medical microbiology for identifying species in clinical samples e.g. (Gurtler and Stanisich, 1996). However, its use in arboreal microbial analysis has not previously

been tested. Molecular studies typically use the 16S rRNA gene for taxonomic identification at the familial level, and although this is the gold standard of PCR based identification of taxonomic identity, it often lacks accuracy in discriminating at the species level (Gurtler and Stanisich, 1996). Conversely, the ITS1 region is highly variable in length and sequence in all bacteria and may thus be effective at distinguishing bacterial species and strains. It should be noted that many bacteria contain more than one copy of the *rrn* operon, which usually show strong sequence homogeneity (>98%) but may differ through insertions or deletions by a range of 2-301bp, with an average difference of 166bp (Klappenbach *et al.*, 2000; Larkin *et al.*, 2007). ITS1 amplicons will therefore exhibit substantial variation in size as amplicon range can vary from 150-1500bp with 85-90% of amplicons being between 150-600bp (Fisher and Triplett, 1999). Often the ITS1 region will code for tRNA genes, these are most frequently found in Gram-negative bacteria and give extra length to the sequence. However, if present they are usually more conserved than other parts of ITS1 (Daffonchio *et al.*, 1998). Typically, the ITS1 amplicon pattern resolved using agarose gels can be used to identify species, whereas the hypervariable indels (mutations) can be used to differentiate between strains (García-Martínez *et al.*, 1999).

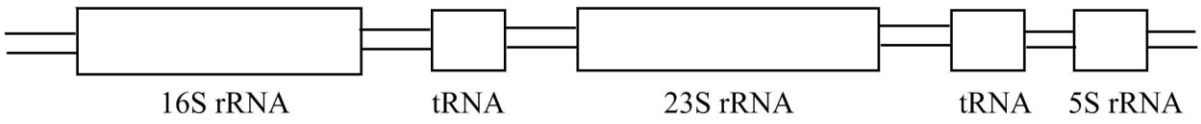


Figure 2.1 Example of a bacterial RNA operon (*rrn*). The *rrn* operon is frequently present in multiple copies within bacterial genomes and can vary substantially in size. It is arranged in a co-transcribed operon consisting of 16S rRNA or small subunit, 23S rRNA or large subunit, 5S rRNA and may contain tRNAs. The ITS1 region is found between 16S and 23S rRNA (Smit *et al.*, 2007).

Due to the conserved nature of 16S rRNA genes across the *Enterobacteriaceae*, *gyrB* is used as the phylogenetic marker gene of choice for identification and taxonomic resolution of strains isolated from the necrotic lesions of oaks affected by AOD. This approach has provided important insights into the ecology and epidemiology of *G. quercinecans* and *B. goodwinii* and their almost exclusive association with necrotic tissue in AOD affected trees. However, *gyrB* gene

PCR amplification, sequencing and phylogenetic analysis of hundreds/thousands of isolated strains from affected trees is a costly and laborious process, representing a barrier to elucidating the ecology and epidemiology of *B. goodwinii* and *G. quercinecans.* This is a pertinent issue, as there are numerous ongoing investigations into the role of bacteria in AOD and the need to confirm the identity and purity of strains is of paramount importance.

The aim of this study was therefore, to develop a rapid molecular diagnostic test for *B. goodwinii* and *G. quercinecans*, based on ITS1 profiling, to resolve species identity and purity of strains. We hypothesised that the ITS1 region represents a suitable genotypic marker for identification and differentiation of *B. goodwinii* and *G. quercinecans* strains associated with AOD and negates the requirement for gene sequencing, as multiple copies of the ITS1 region are present in the bacterial genome, with significant size variation in the length of the sequence.

Here, the ITS1 profiles of bacterial isolates from the necrotic lesions of oak trees were characterised and validated using a polyphasic analysis. Two DNA fingerprinting methods were tested and compared to verify and validate the use of ITS1 profiles in species identification and resolution. Amplified ITS1 PCR products were resolved via (i) 3% agarose gel electrophoresis and (ii) polyacrylamide gel electrophoresis. Finally, the ability to resolve isolated strains to species level was determined by a comparative phylogram of ITS1 profiles and the DNA gyrase B gene sequence of each strain.

2.2 Methods

2.2.1 Maintenance of bacterial strains

Enterobacteriaceae strains were isolated by Forest Research, from oak trees affected by AOD. Most strains had been previously identified to species level through DNA gyrase B (*gyrB*) sequencing and DNA-DNA hybridisation (Brady *et al.*, 2010; Denman *et al.*, 2012). The strains were stored in glycerol stocks at -80°C and maintained on nutrient agar (Oxoid) at 20°C.

2.2.2 ITS1 PCR amplification, agarose gel electrophoresis and sequencing

Each strain was sub-cultured from an individual colony five times, creating five sub-isolates per strain. Genomic DNA was prepared using the colony extraction method for ITS1 PCR reactions: one colony of bacterial cells was picked from an agar plate and added directly to the PCR assay tube (colony PCR). The ITS1 region of each sub-isolate was amplified using ITS1 specific oligonucleotide primers, designed for environmental bacterial communities ITSF (5'-GTCGTAACAAGGTAGCCGTA-3') and ITSReub (5'-GCCAAGGCATCCACC-3') (Cardinale *et al.*, 2004). PCR reactions were performed in 50 µL reaction volumes containing; a colony of each isolate, 1 x MyTaq Red Mix (Bioline), 50 µM each forward and reverse primer. The solution was made up to its final volume of 50 µL with molecular grade double distilled water. The thermal cycle consisted of initial denaturation at 95°C for 60s followed by 25 cycles of denaturation at 95°C for 15s, primer annealing at 55°C for 15s and elongation at 72°C for 10s. PCR amplification products were visualised using 3% agarose gel at 120 volts for 135 minutes and amplicon size was calculated using HyperLadder I (Bioline). ITS1 amplicons were excised from agarose gels and extracted directly using the QIAEX II get extraction kit (Qiagen) for all sequencing reactions. Purified PCR amplicons were sequenced by Macrogen Inc. These sequence data have been deposited to GenBank under accession numbers KJ418748 to KJ418834.

2.2.3 *gyrB* PCR amplification, gel electrophoresis and sequencing

The type II topoisomerase *gyrB* gene was selected as a quality control marker to positively identify bacterial species and verify the utility of ITS1 as a molecular marker for identification and discrimination of bacterial species. Oligonucleotide primers were *Enterobacteriaceae*

specific as described by (Brady *et al.*, 2010), gyrB01F (5'-TAARTTYGAYGAYAACTCYTAYAAAGT-3') and gyrB02R (5'-CMCCYTCCACCARGTAMAGT-3'). A separate forward primer was used for sequencing of *gyrB* PCR amplicons, gyrB07F (5'-GTVCGTTTCTGGCCVAG-3'). Genomic DNA was prepared using the boil prep method, where a colony of the isolate was picked from nutrient agar plates and suspended in 20 µl molecular grade water and incubated at 100°C for 5 mins before adding 1 µl to the PCR reaction mixture. PCR reactions were performed as described for ITS1 with an annealing temperature of 50°C. PCR amplification products were visualised via 1% agarose gel electrophoresis at 140 volts for 35 minutes. *gyrB* amplicons were excised from the gel and purified using the QIAEX II kit (Qiagen). Purified amplicons were sequenced at Macrogen Inc.

2.2.4 Polyacrylamide Gel Electrophoresis (PAGE)

The ITS1 region of thirty-four bacterial isolates from AOD affected trees was amplified as described for the agarose gel analysis. Electrophoresis was conducted using an Ingeny PhorU electrophoresis unit (Ingeny, Leiden, Netherlands). ITS1 PCR products (50 µl, prepared as described in section 3.2) were loaded into a 15% polyacrylamide gel and allowed to migrate through the gel for 16 hours at 100 volts. Amplicons were then visualised using a SYBR Gold stain (Invitrogen).

2.2.5 Phylogram analyses of ITS1 amplicon patterns and *gyrB* DNA sequences

Phylogenetic analyses of thirty-four isolates based on separation of their ITS1 amplicons from the PAGE analysis was conducted in Dendroscope (Huson and Scornavacca, 2012). The size (in base pairs) of ITS1 amplicons were scored individually using Bio-Rad Chemidoc software, which aligned matching amplicons in a presence absence test (with a sensitivity setting of 2) giving a binary output. The binary output was then altered; changing '0' to 'T' and '1' to 'C'. This altered binary data was aligned using ClustalW (Larkin *et al.*, 2007). The aligned data was passed onto Dendroscope which produced a neighbour joining dendrogram based on presence and absence of amplicons. The dendrogram was used to display ITS1 fingerprints from the polyacrylamide gel. *gyrB* sequence data was aligned using ClustalW, and a rectangular neighbour joining phylogram was generated using Dendroscope. Using Dendroscope, the ITS1 dendrogram and

gyrB phylogram, were combined into a tanglegram. The tanglegram directly compared the PAGE analysis against DNA gyrase B gene sequences for thirty-three strains.

2.3 Results

2.3.1 Validation of ITS1 typing as a tool for species resolution

Six bacterial strains (two x *G. quercinecans*, two x *B. goodwinii* and two x *Rahnella* spp.) were used to validate the veracity of the ITS1 amplicon typing approach. In order to confirm the purity of these strains, each strain was consecutively plated from a single colony five times and this subset (n=30 sub-isolates) was characterised by ITS1 PCR amplicon typing and *gyrB* gene sequencing to confirm that each isolate was correctly identified.

ITS1 specific PCR amplicons were generated for each of the thirty sub-isolates and resolved via 3% agarose gel electrophoresis (e.g. Fig. 2.2). The number of visible ITS1 amplicons for each strain tested varied from three to seven, with both strains of *G. quercinecans* possessing six ITS1 amplicons. The two *B. goodwinii* strains possessed different numbers of visible ITS1 amplicons (four and six respectively) and *Rahnella* spp. strains also possessed a different number of amplicons (four and seven) (Apendix II). These data demonstrate that ITS1 typing provides an effective method for the resolution of the six different *Enterobacteriaceae* strains tested at the species level. Furthermore, despite the presence of six amplicons for both *G. quercinecans* strains tested, differences in the length of these ITS1 amplicons provided strain level differentiation, and this was also the case for *B. goodwinii* and *Rahnella* spp. strains, which demonstrated the potential for strain specific resolution, where both the number and length of ITS1 amplicon varied between the two strains of each species.

Visible ITS1 amplicons for all sub-isolates were subsequently excised and purified from agarose gels, and where sufficient DNA was retrieved, each amplicon was sequenced to confirm its identity. ITS1 amplicon sequencing (n=87) and BLASTn searches (Altschul *et al.*, 1990) confirmed that all observed ITS1 amplification products were *bona fide* ITS1 sequences. In addition, each strain was identified via a confirmatory *gyrB* sequencing reaction (Appendix III). However, for one of the strains (*G. quercinecans* FRB98) agarose gel electrophoresis of five sub-isolates obtained from the same agar culture plate revealed two distinct ITS1 amplicon types (Fig. 2.2), where the presence of a contaminant is clearly visible, as three sub-isolates have three visible amplicons and two isolates have six amplicons. These data suggested the presence of a second bacterial strain in the culture of *G. quercinecans* FRB98 and *gyrB* sequencing confirmed the presence of two strains within the culture; *G. quercinecans* FRB98 (lanes 5 and 6, Fig. 2.2) and *Enterobacter cloacae* (lanes 2-4, Fig. 2.2), demonstrating an additional application of ITS1

ribotyping in strain maintenance and the identification of culture contaminants in laboratory strains. The pilot study therefore validated the use of ITS1 typing to separate members of the *Enterobacteriaceae* at both species and strain level, and this approach was subsequently tested and validated on three commonly used and independent laboratory methods for DNA fingerprinting.

2.3.2 Comparison of three independent methods for the resolution of ITS1 ribotypes in *Enterobacteriaceae* strains associated with AOD

2.4.2.1 3% agarose gel electrophoresis

Based on the results of the pilot study the ITS1 genomic region was proposed as an appropriate molecular marker for the resolution of members of the *Enterobacteriaceae*. The agarose verification method was subsequently extended to a wider selection of *Enterobacteriaceae* strains (n=34) (Supplementary Table 2.1). However, variations in ITS1 amplicon numbers for some strains were observed when the same strain was run on replicate agarose gels despite identical quantities of PCR amplified ITS1 fragments being loaded onto the gel. It is possible that in independent PCR reactions, certain amplicons are preferentially amplified due to PCR bias and agarose is not as sensitive a tool to visualise DNA fragments. Nevertheless, the sensitivity of detection using 3% agarose gel separation of ITS1 fragments did appear to provide a rudimentary resolution of strains and species, but the ITS1 amplicon differences observed between the same strain on replicate agarose gels suggest that a more sensitive method for ITS1 typing would be more appropriate. Consequently, polyacrylamide gel electrophoresis analysis of ITS1 PCR products was tested as an alternative method for the resolution of ITS1 ribotypes and compared with phylogenetic resolution against the *gyrB* gene phylogeny for each strain.

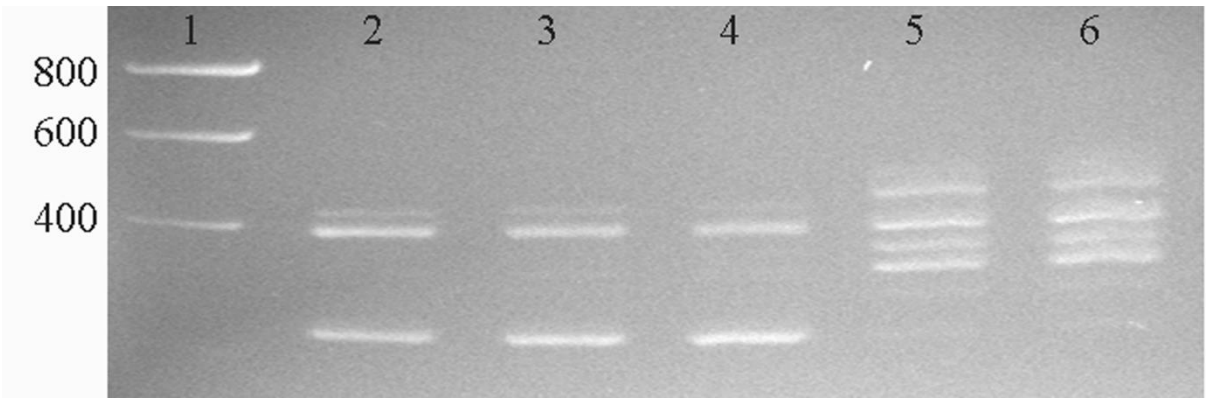


Figure 2.2. Agarose gel image of sub-isolates from *Gibbsiella quercinecans* FRB98 pure culture. Five sub-isolates (lanes 2–6) from a culture of *G. quercinecans* strain FRB98 revealed two distinct ITS1 amplicon patterns on a high-resolution 3% agarose gel, suggesting that the culture was impure. *gyrB* sequencing of sub-isolates in lanes 5 and 6 confirmed that these were *G. quercinecans*, whereas *gyrB* sequencing of sub-isolates in lanes 2–4 established that the sub-isolates were *Enterobacter cloacae*. Lane 1 contains HyperLadder I (Bioline). Ladder size is given in base pairs.

2.3.2.2 PAGE resolution of *Enterobacteriaceae* ITS1 ribotypes

The ITS1 ribotype profiles of the same thirty-four *Enterobacteriaceae* strains tested via agarose gel electrophoresis were analysed using PAGE to further test the utility of the ITS1 region as a marker for strain discrimination and provide greater resolution of the ITS1 amplicons (Fig. 2.2).

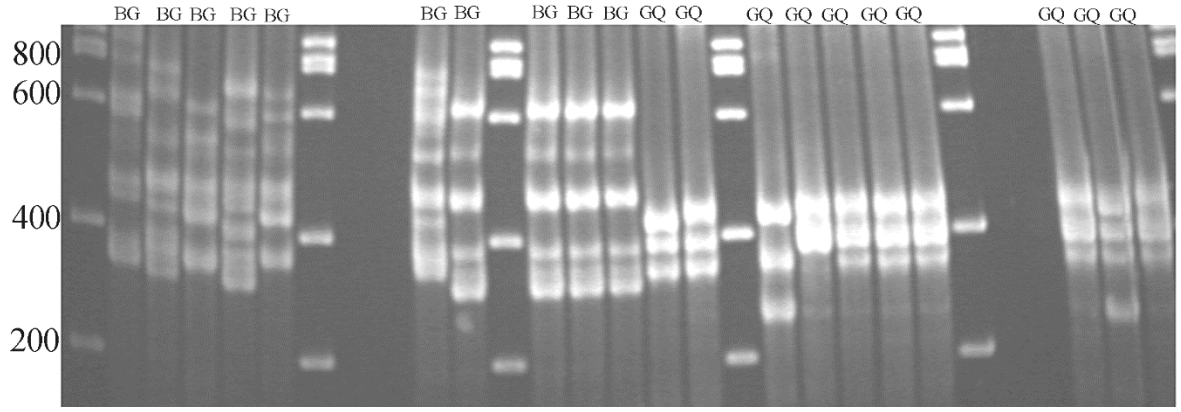


Figure 2.3. Polyacrylamide gel electrophoresis image of *Brenneria goodwinii* and *Gibbsiella quercinecans*. Polyacrylamide gel with ITS1 profiles for: *B. goodwinii*, lanes 2–6, 8–9 and 11–13; *G. quercinecans*, lanes 14–15, 17–21 and 23–25. Lanes 1,7,10,16 and 22 contain HyperLadder I (Bioline). Blank lanes are not included in the lane numbering. Ladder size is given in base pairs.

Visual inspection of ITS1 profiles for *G. quercinecans* and *B. goodwinii* (Fig. 2.3), suggests similarities between the ITS1 ribotypes of isolates of the same species. This discrimination was further demonstrated using a neighbour joining phylogenetic tree to infer relationships between species (Mount, 2008).

Phylogenetic resolution of the thirty-four isolates based on their ITS1 profile using PAGE successfully separated *G. quercinecans* and *B. goodwinii* strains into two lineages (Fig. 2.4). Twenty of the 34 strains were *G. quercinecans* and *B. goodwinii* isolates, and to further validate

the use of ITS1 as a molecular marker for the discrimination of *G. quercinecans* and *B. goodwinii*, a broad spectrum of related AOD strains (n=14) from the *Enterobacteriaceae* were also separated and compared using ITS1 ribotyping.

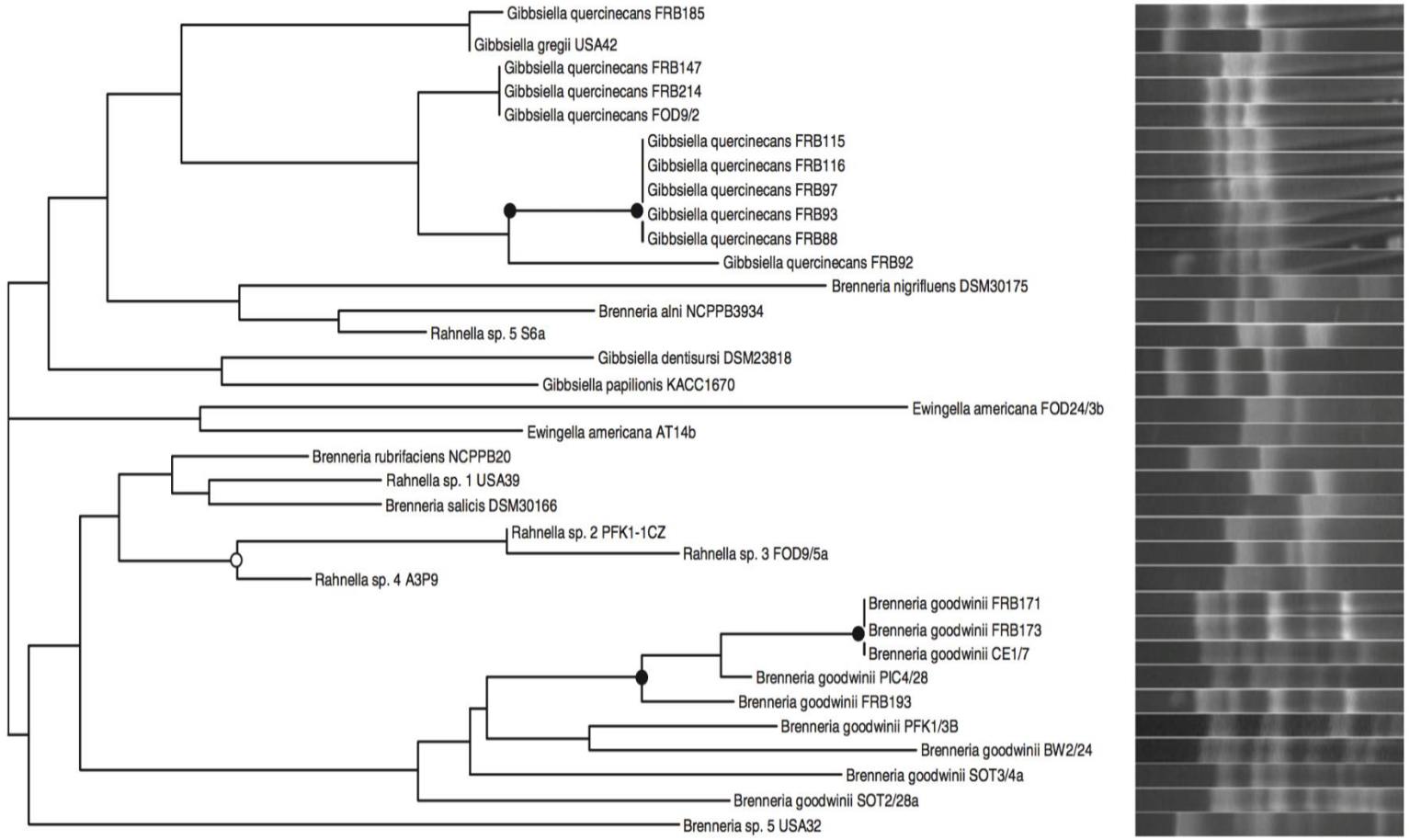


Figure 2.4. Neighbour-joining dendrogram of bacteria isolated from necrotic lesions of AOD affected trees and their ITS1 fingerprints resolved on a polyacrylamide gel. Nodes in which bootstrap value >95% are denoted as filled circles, and those between 75 and 95% as unfilled circles.

A presence/absence matching alignment was used to cluster the amplicons in Fig. 2.4, which demonstrates the relationship between isolates. One potential caveat to this approach is that similarly sized ITS1 amplicons can erroneously be grouped together, as the QuantityOne software (Bio-Rad) used in this study (or the similar amplicon matching open source software PyElph 1.4 (Pavel and Vasile, 2012)), cluster amplicons of a similar but not identical size. However, our data suggest that this does not result in the mis-identification of strains at the species level at least. These programs do have an adjustable sensitivity gauge, but ultimately require manual validation. An appropriately spaced molecular ladder (i.e. a long gap between each rung) can reduce bias by creating a larger area on the gel image.

The PAGE derived ITS1 dendrogams were compared against *gyrB* sequences using a tanglegram (Fig. 2.5). This contrasting tree reveals differences in resolution between *gyrB* sequencing and ITS1 ribotyping methods, but demonstrates the ability of both methods to resolve *G. quercinecans* and *B. goodwinii* at the species level. Furthermore, some strains could be resolved through their ITS1 amplicon pattern, for example *G. quercinecans* FRB92 has five amplicons whereas *G. quercinecans* FRB93 has four. PAGE successfully resolved ITS1 ribotypes to the species level for all *G. quercinecans* and *B. goodwinii* isolates, barring one exception; *G. gregii* USA42 clustered with *G. quercinecans* FRB185 and the latter strain has therefore not been successfully resolved according to species. However, it is clear from the PAGE profiles that *G. quercinecans* strain FRB185 has a different ITS1 PAGE profile to both *G. gregii* USA42 and also the other members of that species, but as it is currently a singleton strain for this ITS1 ribotype, it appears to be an outlier. Furthermore, *G. quercinecans* strain FRB185 also failed to cluster with the other members of this species via *gyrB* sequence analysis (Fig. 2.5), and clustered with *G. dentisursi* and *G. papillionis*, suggesting that the inability of both ITS1 ribotyping and *gyrB* sequencing to resolve FRB185 to species level is an issue specific to that strain, rather than a methodological issue. Nevertheless, *G. quercinecans* FRB185 occupies the same clade as the other *G. quercinecans* strains and it is likely that future typing of additional *G. quercinecans* strains will further resolve the intra-species and genus relationships between strains.

Discrimination of closely related bacterial strains using ITS1 amplicon length may sometimes be challenging as it is not possible to ensure each copy of the ITS1 region is amplified with equal concentration. Therefore, some ITS1 copies which may be present in the genome may not appear when visualised on a gel image or appear as a shadow. It should be noted that it is possible if not probable that not all ITS1 copies are resolved. However, resolved ITS1 fragments were visualised using PAGE consistently and reproducibly and the method was capable of validation

of the ITS1 maker and separation of the AOD isolates. On a cautionary note, there may be similarly weighted ITS1 fragments shared between unrelated bacteria causing relatedness between bacteria to be inferred when no such genotypic relationship exits (Jensen *et al.*, 1993). This is likely to be a rare event, but it should be noted as overlapping intergenic spacer size classes could lead to misinterpretation of results.

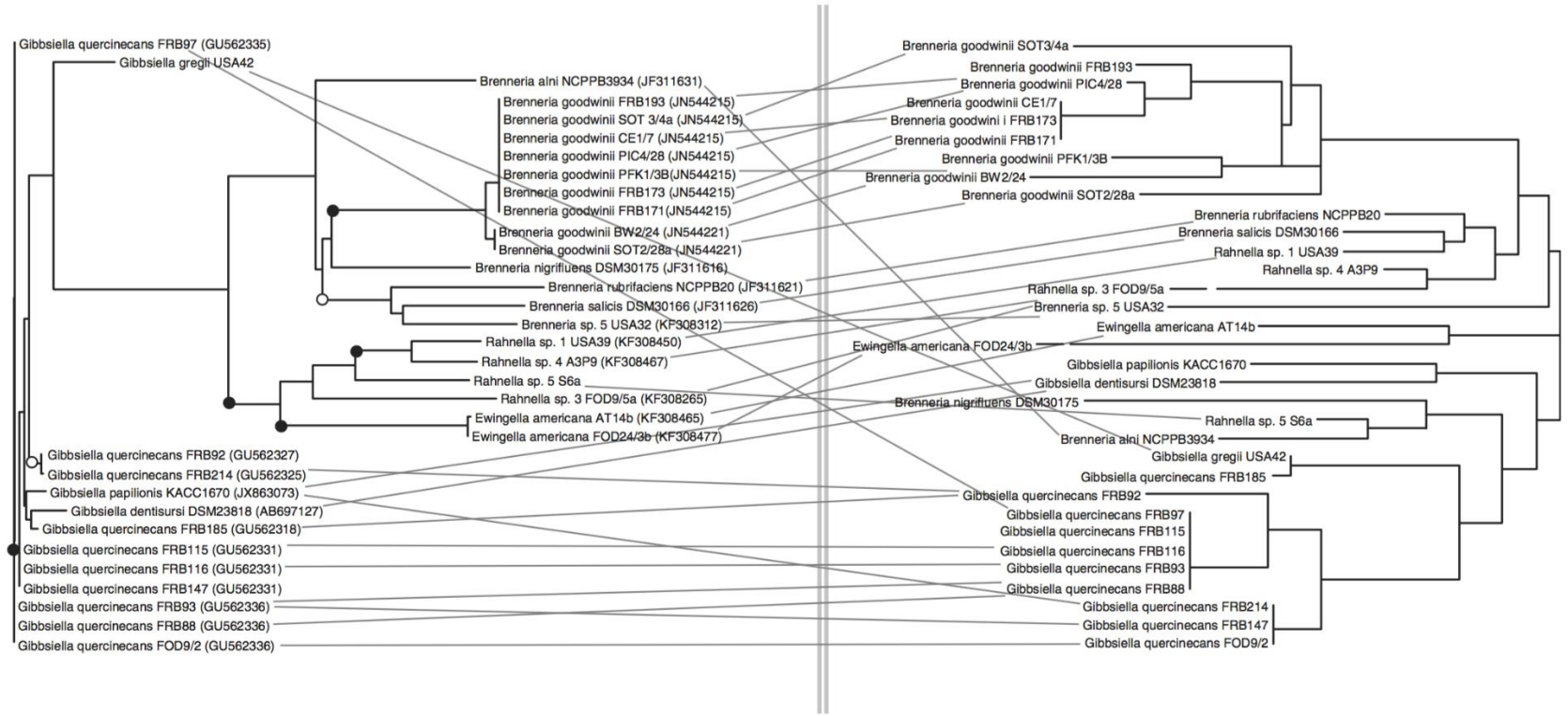


Figure 2.5. Tanglegram comparing *gyrB* phylogram (left) and polyacrylamide dendrogram (right). GenBank accession numbers are given in parentheses. Bootstrap values >95% are denoted as filled circles; values between 75% and 95% are denoted as clear circles.

2.3.2.3 *gyrB* phylograms vs. ITS1 typing

16S rRNA gene sequencing represents the most commonly used phylogenetic tool for estimating microbial diversity and would be improperly applied to the *Enterobacteriaceae* here, due to the conserved nature of the taxon at this locus (Mollet *et al.*, 1997). Therefore, to resolve interspecies taxonomic relatedness the *gyrB* gene is the marker of choice for taxonomic identification of *Enterobacteriaceae* strains isolated from AOD affected trees (Brady *et al.*, 2010; Denman *et al.*, 2012). Relationships between *Enterobacteriaceae* strains were revealed using the *gyrB* molecular marker to create a phylogenetic tree. This topology was compared to a neighbour joining dendrogram from the polyacrylamide gel generated via the ITS1 molecular marker using a tanglegram (Fig. 2.5). The tanglegram reveals that species level resolution of ITS1 is very similar to that of *gyrB*. Furthermore, strain separation may be possible using ITS1, especially in strains with a higher number of ITS1 copies, as this gives a greater prospect of variation. For example, *G. quercinecans* has fewer ITS1 copies than *B. goodwinii*, and this increases the resolution of ITS1 in *B. goodwinii* compared to *G. quercinecans*. Additionally, as Fig. 2.5 reveals, *B. goodwinii* FRB171 and *B. goodwinii* PFK1/3B are identical in their *gyrB* sequence, whereas the ITS1 copy number differs, and they are therefore separated into different branches of the tree. Therefore, as ITS1 copy number increases, the marker contains ever greater phylogenetic power to differentiate between closely related strains to a greater extent than that of *gyrB*. The tanglegram analysis conclusively demonstrates the suitability of ITS1 ribotyping for the resolution of *Enterobacteriaceae* strains.

2.4 Discussion

G. quercinecans and *B. goodwinii,* belonging to the family *Enterobacteriaceae,* are strongly suspected to play a leading role in AOD. One of the difficulties faced by microbiologists working in the field is culture-based differentiation of AOD bacterial isolates, which are morphologically similar and require DNA extraction, amplification of the *gyrB* gene, sequencing and phylogenetic analysis for taxonomic identification. This study demonstrates that ITS1 profiling can discriminate between species amongst the family *Enterobacteriaceae* to the same resolution as the most commonly used differentiation marker, DNA gyrase B (*gyrB*). The advantage that ITS1 ribotyping has over *gyrB*, is that it negates the requirement for a sequencing reaction and DNA sequence analysis, therefore reducing both cost and time. Polyacrylamide gel electrophoresis successfully separated ITS1 amplicons using intergenic heterogeneity in the ITS1 region using standard protocols. *gyrB* sequencing of individual AOD bacterial isolates validated the ITS1 amplicon separation and clustering analyses providing evidence that the method could be used for rapid and inexpensive environmental bacterial identification (Fig. 2.5).

Beyond the successful resolution of the two species of interest (*G. quercinecans* and *B. goodwinii*), our data suggest that ITS1 ribotyping analysis is not suitable for the discrimination of genus-level relationships. As described above, low fidelity of the ITS1 region provides the basis for this study. However, variability of the marker is such that inter-genus relationships are idiosyncratic and meaningful comparisons are limited. This is demonstrated via Figs. 2.3 and 2.4 where the *G. quercinecans* and *B. goodwinii* species are robustly separated into distinct clades. However, drawing any phylogenetic conclusions between genera for the isolates used in this study, is not supported via bootstrapping (Fig. 2.4). For example, on the PAGE phylogenetic tree, *Rahnella sp.*1 USA39 and *B. salicis* DSM30166 are placed on the same phylogenetic node, yet based on the random sampling of the bootstrapping analysis this node is not strongly supported, hence it could not be inferred that these isolates are closer than other inter-genus relationships on the tree. Therefore, our data has shown similar to previous studies (Gurtler and Stanisich, 1996), that ITS1 copy numbers are strain specific with inter-genus level relationships unclear.

In this study, ITS1 ribotyping has been tested and validated on arboreal *Enterobacteriaceae,* using two independent electrophoresis methods. However, it could equally be applied to other members of the *Enterobacteriaceae* including bacterial groups in clinical studies. There are reports in the literature on the inefficiency of the 16S rRNA gene or other conventional identification methods for screening bacterial isolates due to the sequence similarity between

closely related species (Fukushima *et al.*, 2002; Paradis *et al.*, 2005; Pavlovic *et al.*, 2011). For example, *Escherichia coli* and *Shigella spp.* pathovars are closely related with few disparate biochemical characteristics. Consequently, separation using biochemical tests or the 16S rRNA gene will often fail to resolve strains. Therefore, the method we propose here could thus be applied to other situations. Equally, the food industry requires diagnostic mechanisms for *Enterobacteriaceae* species, as these through a variety of species or strain specific mechanisms can spoil meat products (Doulgeraki *et al.*, 2011; Mofokeng *et al.*, 2011). Through building ITS1 profiles for individual *Enterobacteriaceae* species the method offers a rapid fingerprinting mechanism without the need for expensive and time consuming sequencing reactions, which is not only applicable in arboreal studies but across all disciplines of microbiology.

This is a simple, accurate, and inexpensive method for the rapid screening of *Enterobacteriaceae* samples and provides robust species specific profiles. The method allows environmental screening of a large number of samples providing species identification and differentiation data.

2.5 References

Altschul, S.F., Gish, W., Miller, W., Myers, E., and Lipman, D.J. (1990) Basic Local Alignment Search Tool. *J. Mol. Biol.* **215**: 403–410.

Brady, C., Denman, S., Kirk, S., Venter, S., Rodríguez-Palenzuela, P., and Coutinho, T. (2010) Description of *Gibbsiella quercinecans* gen. nov., sp. nov., associated with Acute Oak Decline. *Syst. Appl. Microbiol.* **33**: 444–50.

Brown, M. V. and Fuhrman, J.A. (2005) Marine bacterial microdiversity as revealed by internal transcribed spacer analysis. *Aquat. Microb. Ecol.* **41**: 15–23.

Brown, N., Inward, D.J.G., Jeger, M., and Denman, S. (2014) A review of *Agrilus biguttatus* in UK forests and its relationship with acute oak decline. *Forestry* 1–11.

Cardinale, M., Brusetti, L., Quatrini, P., Borin, S., Puglia, A.M., Rizzi, A., *et al.* (2004) Comparison of different primer sets for use in automated ribosomal intergenic spacer analysis of complex bacterial communities. *Appl. Environ. Microbiol.* **70**: 6147–6156.

Daffonchio, D., Borin, S., Frova, G., Manachini, P.L., and Sorlini, C. (1998) PCR fingerprinting of whole genomes: the spacers between the 16S and 23S rRNA genes and of intergenic tRNA gene regions reveal a different intraspecific genomic variability of *Bacillus cereus* and *Bacillus licheniformis* [corrected]. *Int. J. Syst. Bacteriol.* **48**: 107–116.

Denman, S., Brady, C., Kirk, S., Cleenwerck, I., Venter, S., Coutinho, T., and De Vos, P. (2012) *Brenneria goodwinii* sp. nov., associated with acute oak decline in the UK. *Int. J. Syst. Evol. Microbiol.* **62**: 2451–2456.

Denman, S., Brown, N., Kirk, S., Jeger, M., and Webber, J. (2014) A description of the symptoms of Acute Oak Decline in Britain and a comparative review on causes of similar disorders on oak in Europe. *Forestry* **87**: 535–551.

Denman, S. and Webber, J. (2009) Oak declines: new definitions and new episodes in Britain. *Q. J. For.* **103**: 285–290.

Doulgeraki, A.I., Paramithiotis, S., and Nychas, G.J.E. (2011) Characterization of the *Enterobacteriaceae* community that developed during storage of minced beef under aerobic or modified atmosphere packaging conditions. *Int. J. Food Microbiol.* **145**: 77–83.

Fisher, M.M. and Triplett, E.W. (1999) Automated Apporach for Ribosomal Intergenic Spacer Analysis of microbial diversity and its application to freshwater bacterial communities. *Appl. Environ. Microbiol.* **65**: 4630–4636.

Fukushima, M., Kakinuma, K., and Kawaguchi, R. (2002) Phylogenetic analysis of *Salmonella*, *Shigella*, and *Escherichia coli* strains on the basis of the *gyrB* gene sequence. *J. Clin. Microbiol.* **40**: 2779–2785.

García-Martínez, J., Acinas, S.G., Antón, A.I., and Rodríguez-Valera, F. (1999) Use of the 16S-23S ribosomal genes spacer region in studies of prokaryotic diversity. *J. Microbiol. Methods* **36**: 55–64.

Gurtler, V. and Stanisich, V. (1996) New approaches to typing and identification of bacteria using the 16S-23S rDNA spacer region. *Microbiology* **142**: 3–16.

Huson, D.H. and Scornavacca, C. (2012) Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.* **61**: 1061–1067.

Janda, J.M. and Abbott, S.L. (2007) 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. *J. Clin. Microbiol.* **45**: 2761–2764.

Jensen, M.A., Webster, J.A., and Straus, N. (1993) Rapid identification of bacteria on the basis of polymerase chain reaction-amplified ribosomal DNA spacer polymorphisms. *Appl. Environ. Microbiol.* **59**: 945–952.

Klappenbach, J. A., Dunbar, J.M., and Schmidt, T.M. (2000) rRNA operon copy number reflects ecological strategies of bacteria. *Appl. Environ. Microbiol.* **66**: 1328–1333.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., Mcgettigan, P.A., McWilliam, H., *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947–2948.

Mofokeng, L., Cawthorn, D.M., Witthuhn, R.C., Anelich, L.E.C.M., and Jooste, P.J. (2011) Characterization of cronobacter species (*Enterobacter sakazakii*) isolated from various south African food sources. *J. Food Saf.* **31**: 98–107.

Mollet, C., Drancourt, M., and Raoult, D. (1997) *rpoB* sequence analysis as a novel basis for bacterial identification. *Mol. Microbiol.* **26**: 1005–1011.

Mount, D.W. (2008) Choosing a method for phylogenetic prediction. *Cold Spring Harb. Protoc.* **3**:.

Naum, M., Brown, E.W., and Mason-Gamer, R.J. (2008) Is 16S rDNA a reliable phylogenetic marker to characterize relationships below the family level in the *Enterobacteriaceae*? *J. Mol. Evol.* **66**: 630–642.

Paradis, S., Boissinot, M., Paquette, N., Belanger, S.D., Martel, E.A., Boudreau, D.K., *et al.* (2005) Phylogeny of the *Enterobacteriaceae* based on genes encoding elongation factor Tu and F-ATPase beta-subunit. *Int. J. Syst. Evol. Microbiol.* **55**: 2013–2025.

Pavel, A.B. and Vasile, C.I. (2012) PyElph - a software tool for gel images analysis and phylogenetics. *BMC Bioinformatics* **13**: 9.

Pavlovic, M., Luze, A., Konrad, R., Berger, A., Sing, A., Busch, U., and Huber, I. (2011) Development of a duplex real-time PCR for differentiation between *E. coli* and *Shigella* spp. *J. Appl. Microbiol.* **110**: 1245–1251.

Rappé, M.S., Connon, S. A, Vergin, K.L., and Giovannoni, S.J. (2002) Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* **418**: 630–633.

Scheinert, P., Krausse, R., Ullmann, U., Söller, R., and Krupp, G. (1996) Molecular differentiation of bacteria by PCR amplification of the 16S–23S rRNA spacer. *J. Microbiol. Methods* **26**: 103–117.

Smit, S. Widmann, J. and Knight, R. (2007) Evolutionary rates vary among rRNA structural elements. Nucl. Acid Res. **35**: 3339-3354.

Woese, C.R. and Fox, G.E. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. U. S. A.* **74**: 5088–5090.

CHAPTER 3

Comparative genomic analysis of bacteria isolated from necrotic lesions of AOD affected trees

Abstract

Acute Oak Decline (AOD) is a recently described Decline-disease, which affects both species of British oak (*Quercus robur* and *Q. petraea*). The causative agents of AOD are not well characterised, however several bacterial species have been frequently isolated from necrotic lesions within affected oak tissue. Here, the whole genomes of *Gibbsiella quercinecans* FRB97 (T), FRB124, & N78 and *Brenneria goodwinii* FRB141 (T) & FRB171, isolated from necrotic lesions of AOD affected trees have been sequenced, annotated and compared to whole genomes of three closely related plant-associated bacteria, *B. alni* NCPPB 3934, *B. salicis* DSM30166 (phytopathogens) and *Ewingella americana* ATCC33852 (endophyte). Previous studies have consistently isolated *G. quercinecans* and *B. goodwinii* from necrotic lesions of AOD affected trees and described them as putative oak pathogens. Therefore, the aim of this chapter is to identify putative genome encoded virulence factors, which would enable the sequenced isolates to cause disease. Whole genome sequencing of the selected isolates was performed on the Illumina MiSeq benchtop sequencing platform. The bacterial isolates were phylogenetically positioned and annotated using resultant whole genome sequencing data. Phylogenetic data revealed that *B. goodwinii* and related tree pathogens of the genus *Brenneria* are positioned alongside the soft-rot *Enterobacteriaceae* group of necrotic bacterial pathogens. *G. quercinecans* occupies a separate clade alongside *E. americana*, *Rahnella spp.* and the wide-ranging opportunistic pathogen genus, *Serratia*. Whole genome annotations revealed that *G. quercinecans* FRB97 & FRB124 contain a type II secretion system, all sequenced *G. quercinecans* strains encode a type IV secretion system operon, along-with a repertoire of lignocellulolytic enzymes, including pectic enzymes, cellulases, and hemicellulases. Encoded virulence factors of *B. goodwinii* include, a T2SS operon in strain FRB171, which is partially absent in FRB141, however both sequenced *B. goodwinii* strains contain a T3SS operon and the associated pathogenic effectors, YopJ, DspA/E, DspF and multiple Hop homologs.

3.1 Introduction

Acute Oak Decline (AOD) is a recently described Decline-disease, affecting oak trees predominantly in the South-East of the United Kingdom. The Decline is spreading from East-West and has recently become established at the border between Wales and England. However, it remains most prevalent in the South-East and the Midlands of England (Brown *et al.*, 2016). The causative agents of AOD are uncertain and despite several substantive evidence trails welding putative causes together (e.g. biotic and abiotic factors), no complete definition is yet available, in part due to disparate theories on the causes of disease (Vuts *et al.*, 2015). Despite no complete description of the Decline, there are four diagnostic symptoms; 1) Cracked outer-bark plates with weeping patches 2) Evidence of larval galleries in the inner-bark or D-shaped exit holes on the outer-bark of the two-spotted oak buprestid, *Agrilus biguttatus*, 3) Necrotic lesions with black sticky exudate appearing on the outer bark and 4) inner-bark necrosis (Denman *et al.*, 2014). The assemblage of interacting causes aligns with classic descriptions of Decline-disease (Manion and Lachance, 1992), however this complexity makes definitive descriptions on the causative agents of AOD exceedingly complicated.

It has been suggested that bacterial pathogens are causative agents of AOD. Two recently described bacteria, *Gibbsiella quercinecans* (Brady *et al.*, 2010) and *Brenneria goodwinii* (Denman *et al.* 2012), have been consistently isolated from necrotic lesions of affected trees. Furthermore, there has been convincing evidence in the United Kingdom and across the European continent that bacteria, and in particular *G. quercinecans* and *B. goodwinii*, play a role in stem bleeding on oak (Scortichini *et al.*, 1993; Biosca *et al.*, 2003; Denman *et al.*, 2014). This suggests that the bacteria have the inherent genotypic capability to destruct tree tissue.

Putative bacterial pathogens can be systematically screened using whole genome sequencing (WGS). WGS is a global molecular analysis method, which can completely describe the biological potential of emerging bacterial plant pathogens (da Silva *et al.*, 2002; Green *et al.*, 2010; Peeters *et al.*, 2013). This is exemplified with the whole genome sequence of the potato pathogen, *Pectobacterium atrosepticum* which was published in 2004 (Bell *et al.*, 2004), and illuminated the organism's complexity, revealing a large number of genes coding for suppression and evasion of the host immune response (Toth *et al.*, 2015). Unravelling genomic sequences allows an idiosyncratic analysis of their biology, and provides an evidence-based approach to understand what makes a pathogen and ultimately proscribes the tools to control disease (Toth *et al.*, 2006).

This chapter presents the first genomic exploration of *G. quercinecans* and *B. goodwinii*, revealing numerous putative virulence genes and mechanisms, which may cause disease within an oak host. Using *in silico* tools, strains of *G. quercinecans* and *B. goodwinii* are compared against closely related bacterial phytopathogens. Furthermore, the sequence data provides a genomic baseline for *G. quercinecans* and *B. goodwinii*, revealing gene targets for preventative testing, and presenting data for use in further molecular studies of a contemporary threat to a keystone species within the United Kingdom.

3.2 Materials and Methods

3.2.1 Maintenance of bacterial strains

G. quercinecans FRB97, FRB124, & N78, *B. goodwinii* FRB141 & FRB171, and *E. americana* ATCC33852 were isolated by Forest Research from oak trees affected with AOD (Table 3.1). The strains were previously identified to species level through DNA gyrase B (*gyrB*) sequencing and DNA-DNA hybridisation (Brady *et al.* 2010; Denman *et al.* 2012). Isolates were stored in glycerol stocks at -80°C and maintained on nutrient agar (Oxoid) at 20°C.

3.2.2 Genome sequencing on Illumina MiSeq

Three strains of *G. quercinecans*, two strains of *B. goodwinii*, *Ewingella americana*, and two positive controls (*Brenneria alni* and *Brenneria salicis*, both plant pathogenic bacteria associated with alder and willow, respectively), were selected for whole genome sequencing (Table 1). A single colony of each strain was sampled from nutrient agar (Oxoid) and used as inoculum for liquid culture, which was grown overnight in nutrient broth (Oxoid) at 28°C, on a shaking incubator at 200 rpm. Total genomic DNA was isolated using the Genomic II extraction kit (Bioline) following the manufacturer's protocol. Extracted DNA was quantified using the Qubit flurometer (Life Technologies, Paisley, UK). DNA integrity was assessed using 2% agarose gel electrophoresis. Sequencing libraries were prepared using the Illumina Nextera XT DNA protocol (Illumina, Inc., Cambridge, UK). Briefly, samples were equalised for an input concentration of 1 ng/µl. DNA was fragmented, tagged ('tagment') and appended with adapters using an engineered transposome. The adapters were used as amplification targets for a limited cycle PCR. During the PCR, target DNA (insert) was amplified, and indexed sequences were added to both ends of the DNA, allowing paired end amplification of the insert. Finally, a further PCR was performed as per the manufacturer's instructions with the exception that 16 thermal cycles were completed as opposed to 12. Amplicon and insert size was assessed through agarose gel electrophoresis. Amplified DNA was purified using Agencourt AMPure XP beads (Beckman Coulter, Beverley, Massachusetts) and normalised with library normalisation additives. Samples were adjusted to a

concentration of 2 nM in 10 mM Tris-HCl and 0.1% Tween before being heat denatured and added to a single lane of the MiSeq Personal Sequencer (Illumina, Inc., Cambridge, UK).

3.2.3 Bioinformatics

General bioinformatic analyses were carried out on a locally installed Bio-Linux 8 workstation (Field *et al.* 2006). Applications requiring high computing power were undertaken on the High Performance Computing (HPC) Wales supercomputing network.

3.2.4 Post sequencing QC

Nextera XT adapter sequences were removed from raw FastQ files containing resultant sequencing reads, using Cutadapt v1.2.1, and the option –O 3, which specifies that a minimum of 3 base pairs have to match the adapter sequences before they were trimmed. Sequences were quality trimmed using Sickle v1.2 with a minimum quality score of 20. Reads of fewer than 10 base pairs (bp) were removed.

3.2.5 Bacterial genome assembly

Bacterial genome assemblers were systematically tested and scored according to most contiguous output and highest N_{50} score. The N_{50} metric is a commonly used genome assembly comparison statistic which uses contig length in bp to compare assemblies. The metric adds contig lengths together, starting with the largest, until the length equals half of the estimated chromosome length. The length of the contig which crosses half the chromosome length is presented as the N_{50} (Narzisi and Mishra, 2011). The best assemblies were then used for all further analyses (Table 3.1). *De novo* genome assemblies were generated using SPAdes v3.0 (Bankevich *et al.*, 2012) for all sequenced bacterial isolates, with k-mer values of 21, 33, 55, 77, 99,121, 143, 165, 187, 209, 231.

3.2.6 Post-assembly QC

Draft genome assembles were tested for contaminants using Blobtools v0.1 (Kumar *et al.*, 2013), producing a blobplot of each genome assembly (Figs. 3.1-3.8). Blobtools identifies contaminants

through examination of GC content, coverage, and taxonomic alignment of sequences to the complete non-redundant NCBI nucleotide database.

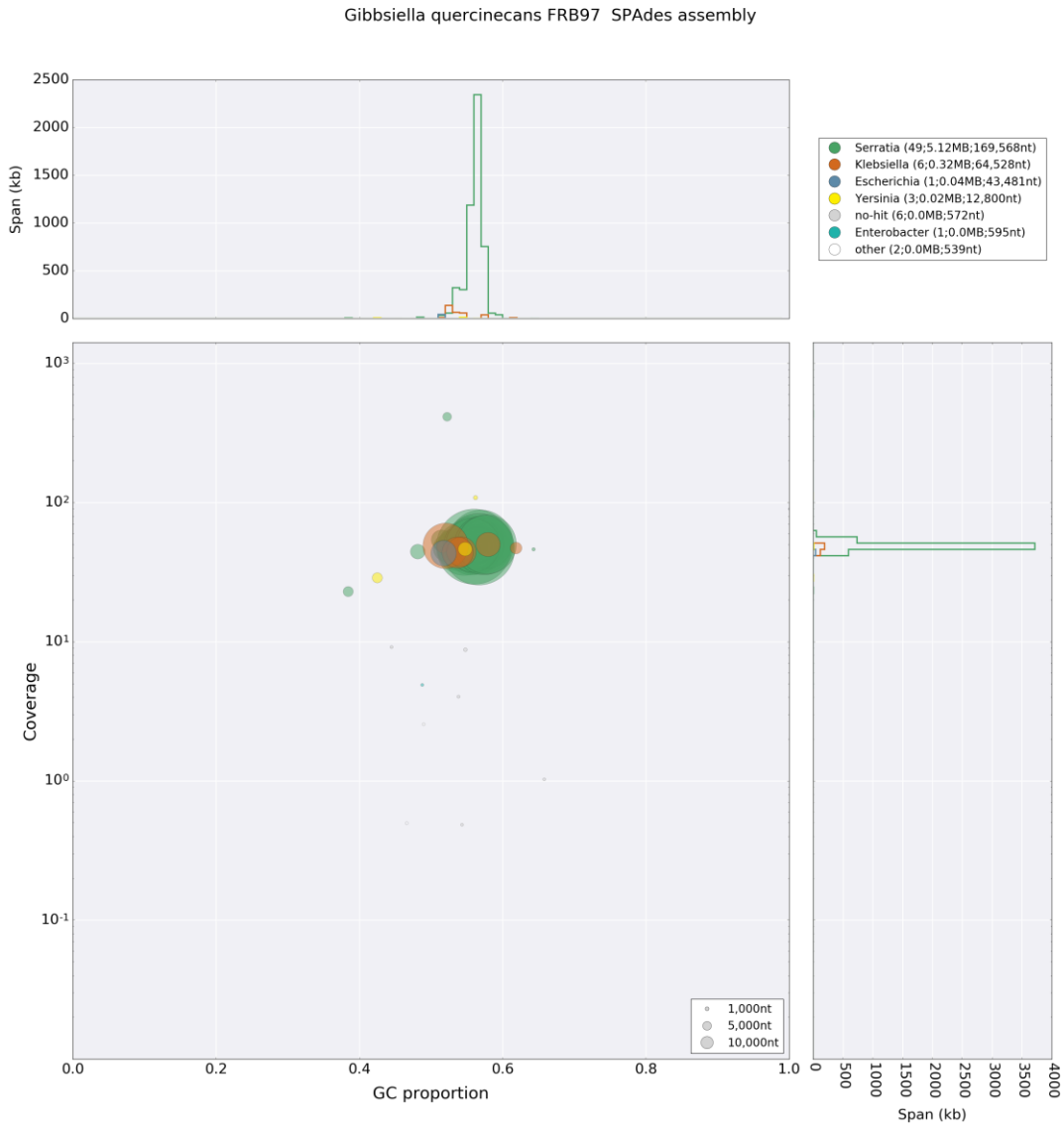


Figure 3.1. Blobplot of *Gibbsiella quercinecans* FRB97 SPAdes assembly. Taxonomic homology was measured between the SPAdes assembly of a sequenced *G. quercinecans* FRB97 isolate, and NCBI non-redundant, nucleotide database sequences. The taxonomy key shows the highest homology to the assembled contigs is from the genus *Serratia*. The plot also shows per contig coverage and GC proportion.

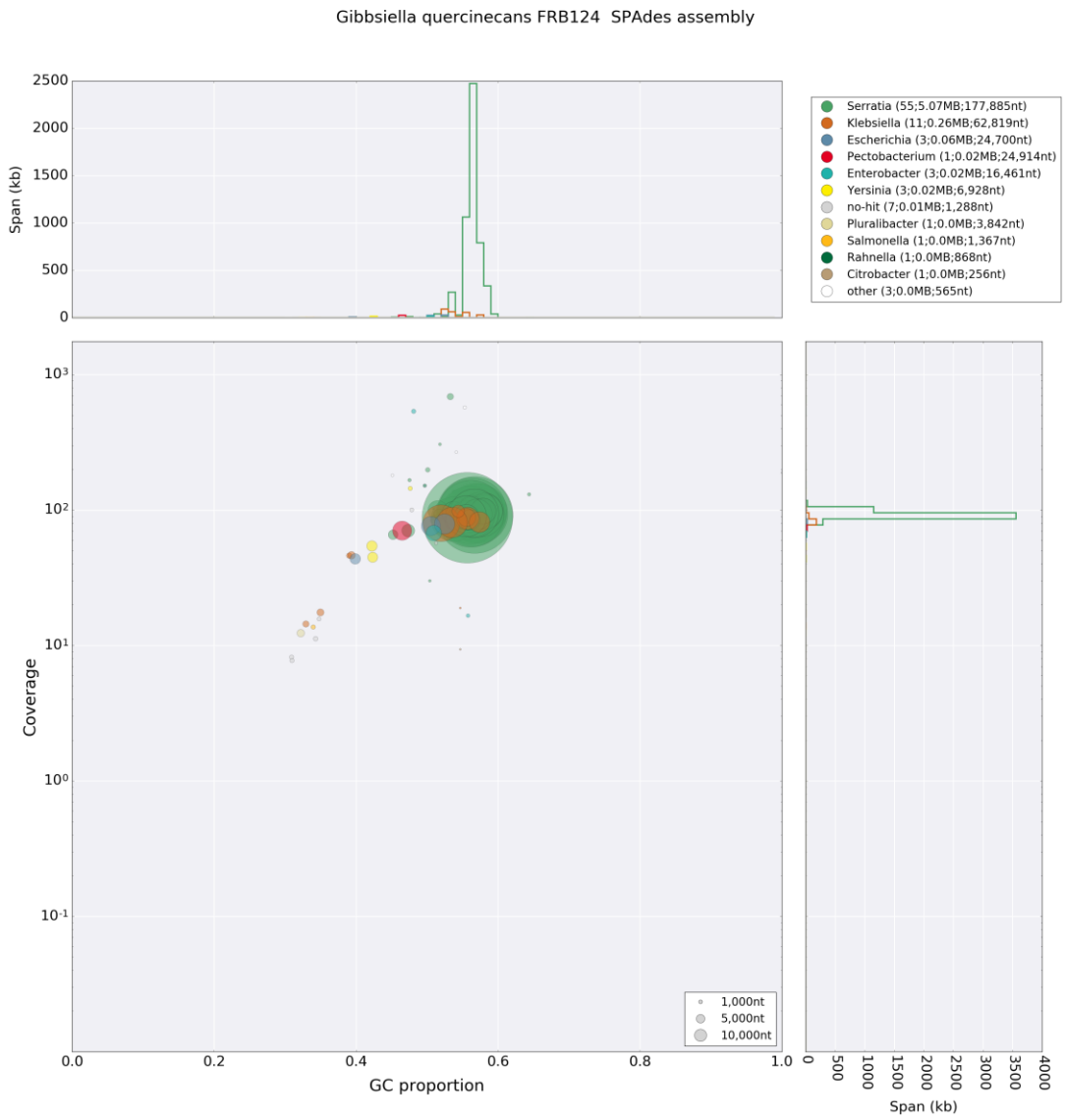


Figure 3.2. Blobplot of *Gibbsiella quercinecans* FRB124 SPAdes assembly. Taxonomic homology was measured between the SPAdes assembly of a sequenced *G. quercinecans* FRB124 isolate, and NCBI non-redundant, nucleotide database sequences. The taxonomy key shows the highest homology to the assembled contigs is from the genus *Serratia*. The plot also shows per contig coverage and GC proportion.

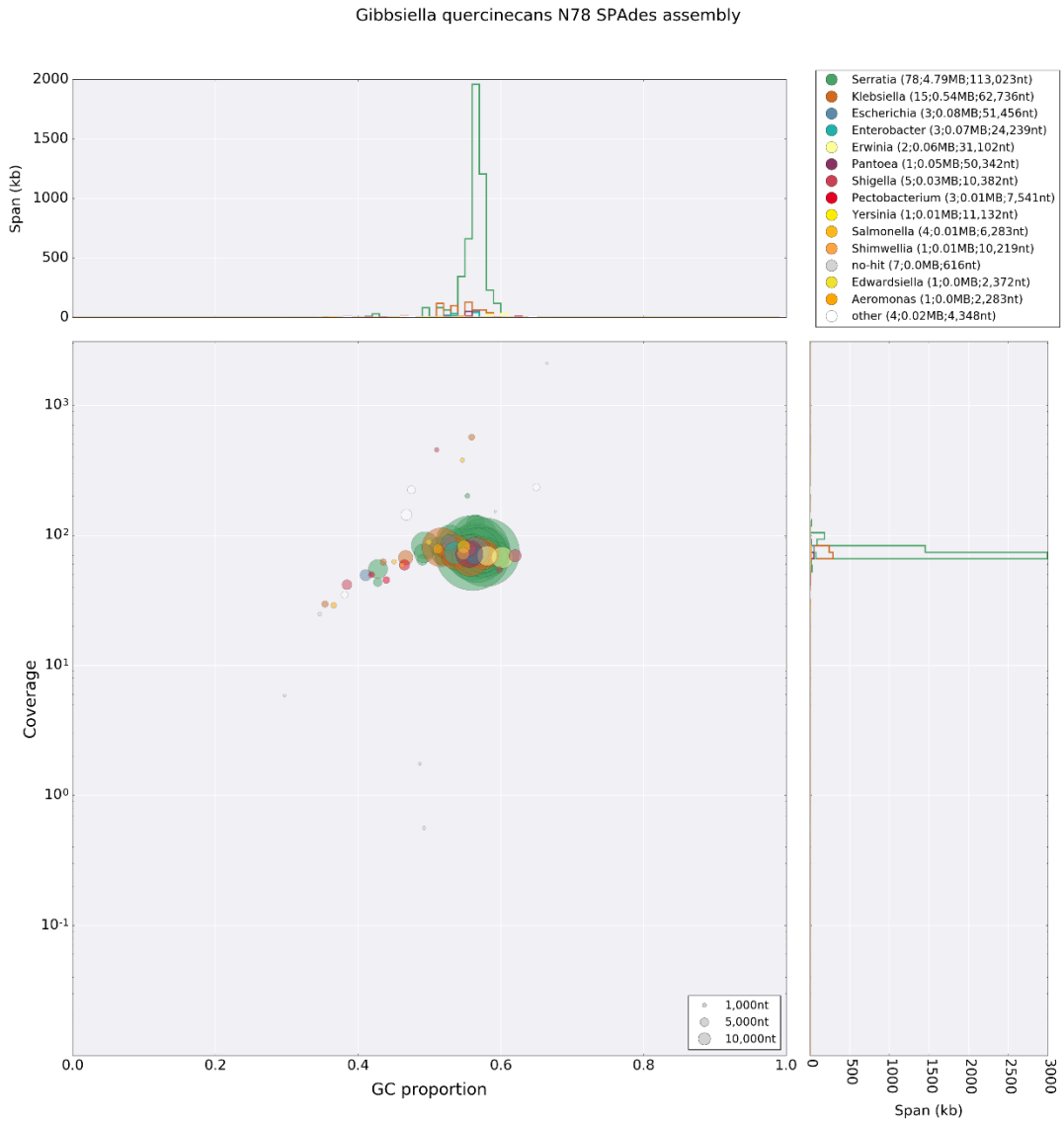


Figure 3.3. Blobplot of *Gibbsiella quercinecans* N78 SPAdes assembly. Taxonomic homology was measured between the SPAdes assembly of a sequenced *G. quercinecans* N78 isolate, and NCBI non-redundant, nucleotide database sequences. The taxonomy key shows the highest homology to the assembled contigs is from the genus *Serratia*. The plot also shows per contig coverage and GC proportion.

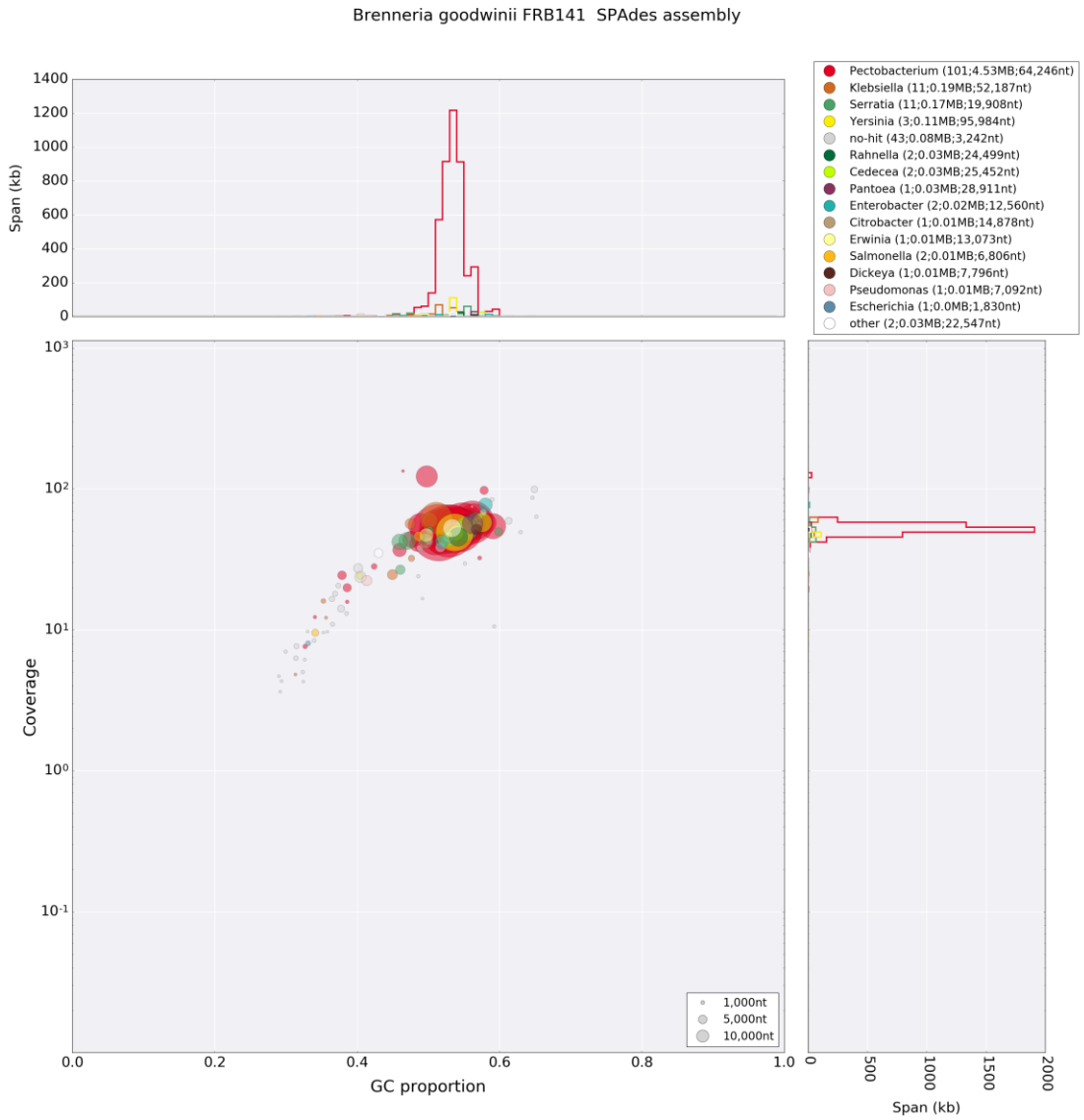


Figure 3.4. Blobplot of *Brenneria goodwinii* FRB141 SPAdes assembly. Taxonomic homology was measured between the SPAdes assembly of a sequenced *B. goodwinii* FRB141 isolate, and NCBI non-redundant, nucleotide database sequences. The taxonomy key shows the highest homology to the assembled contigs is from the genus *Pectobacterium*. The plot also shows per contig coverage and GC proportion.

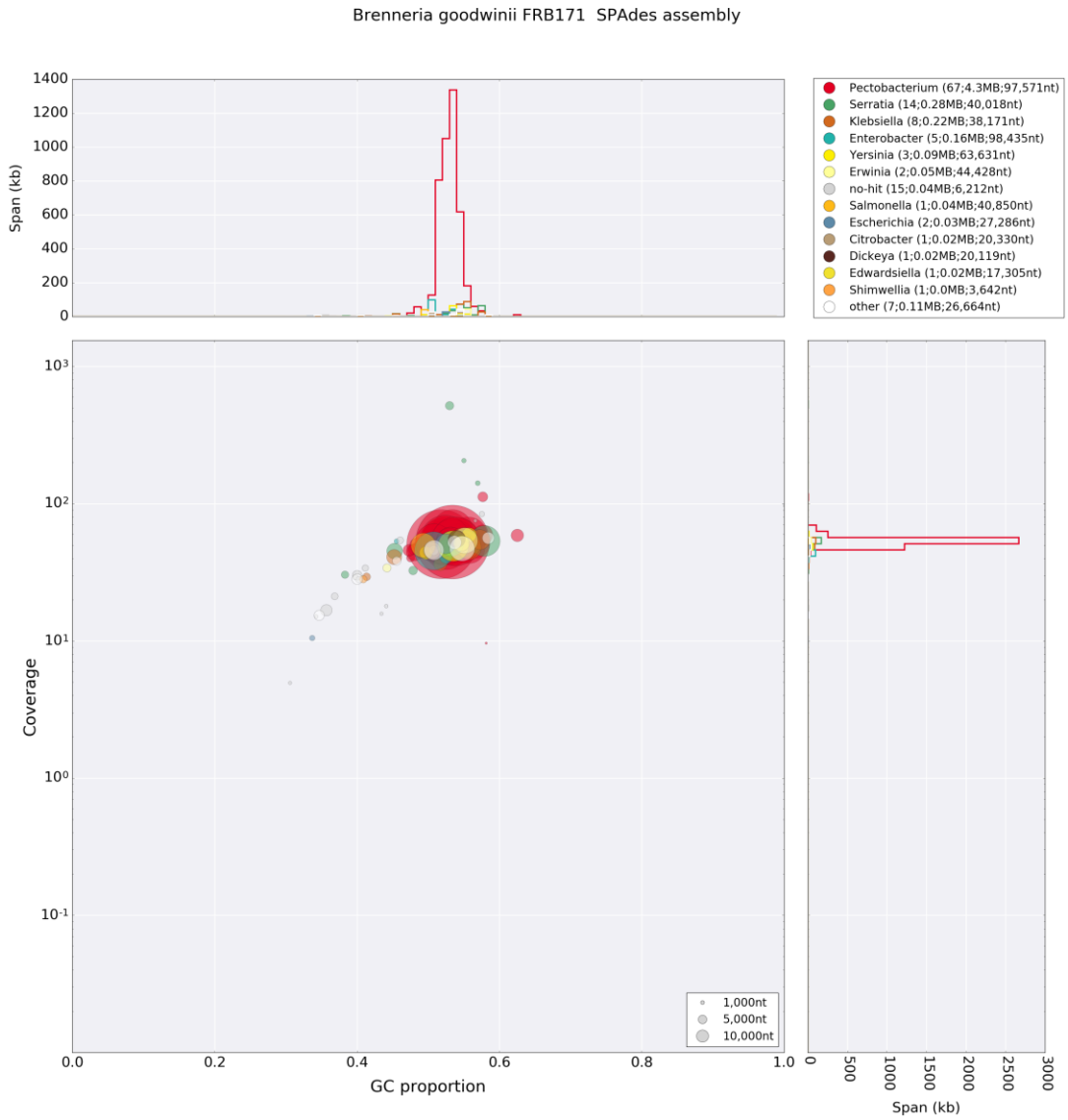


Figure 3.5. Blobplot of *Brenneria goodwinii* FRB171 SPAdes assembly. Taxonomic homology was measured between the SPAdes assembly of a sequenced *B. goodwinii* FRB171 isolate, and NCBI non-redundant, nucleotide database sequences. The taxonomy key shows the highest homology to the assembled contigs is from the genus *Pectobacterium*. The plot also shows per contig coverage and GC proportion.

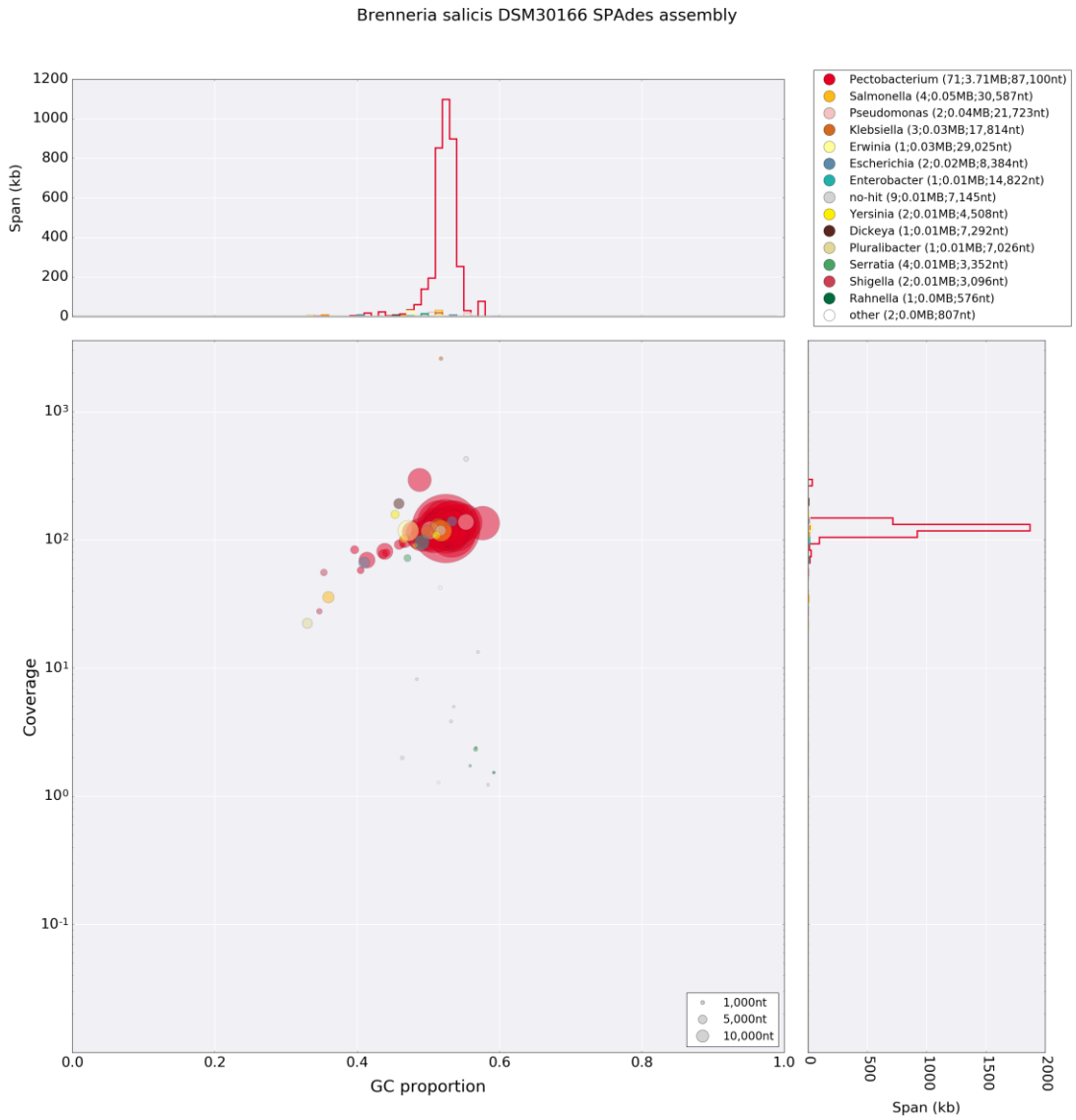


Figure 3.6. Blobplot of *Brenneria salicis* DSM30166 SPAdes assembly. Taxonomic homology was measured between the SPAdes assembly of a sequenced *B. salicis* DSM30166 isolate, and NCBI non-redundant, nucleotide database sequences. The taxonomy key shows the highest homology to the assembled contigs is from the genus *Pectobacterium*. The plot also shows per contig coverage and GC proportion.

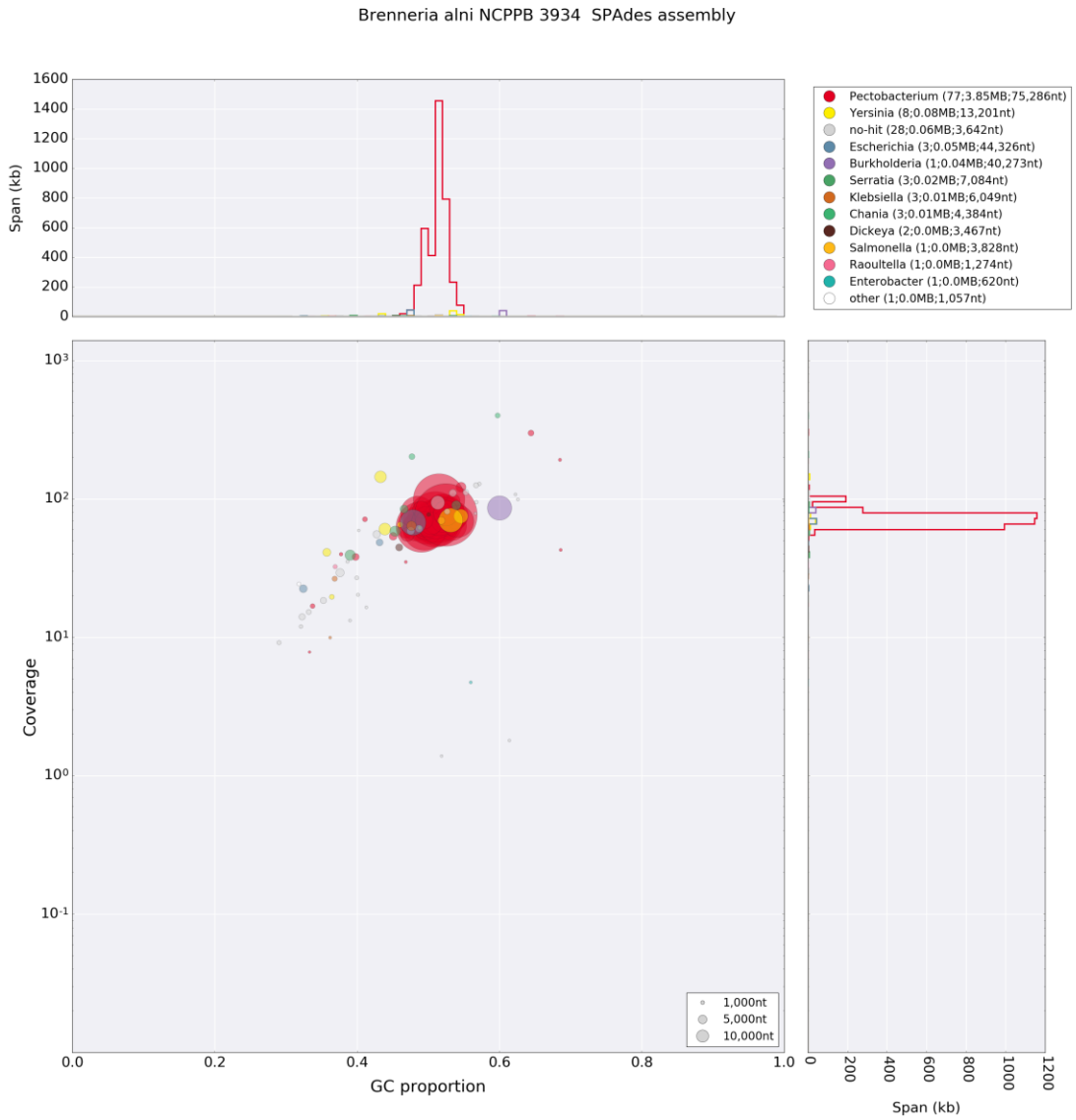


Figure 3.7. Blobplot of *Brenneria alni* NCPPB 3934 SPAdes assembly. Taxonomic homology was measured between the SPAdes assembly of a sequenced *B. alni* NCPPB 3934 isolate, and NCBI non-redundant, nucleotide database sequences. The taxonomy key shows the highest homology to the assembled contigs is from the genus *Pectobacterium*. The plot also shows per contig coverage and GC proportion.

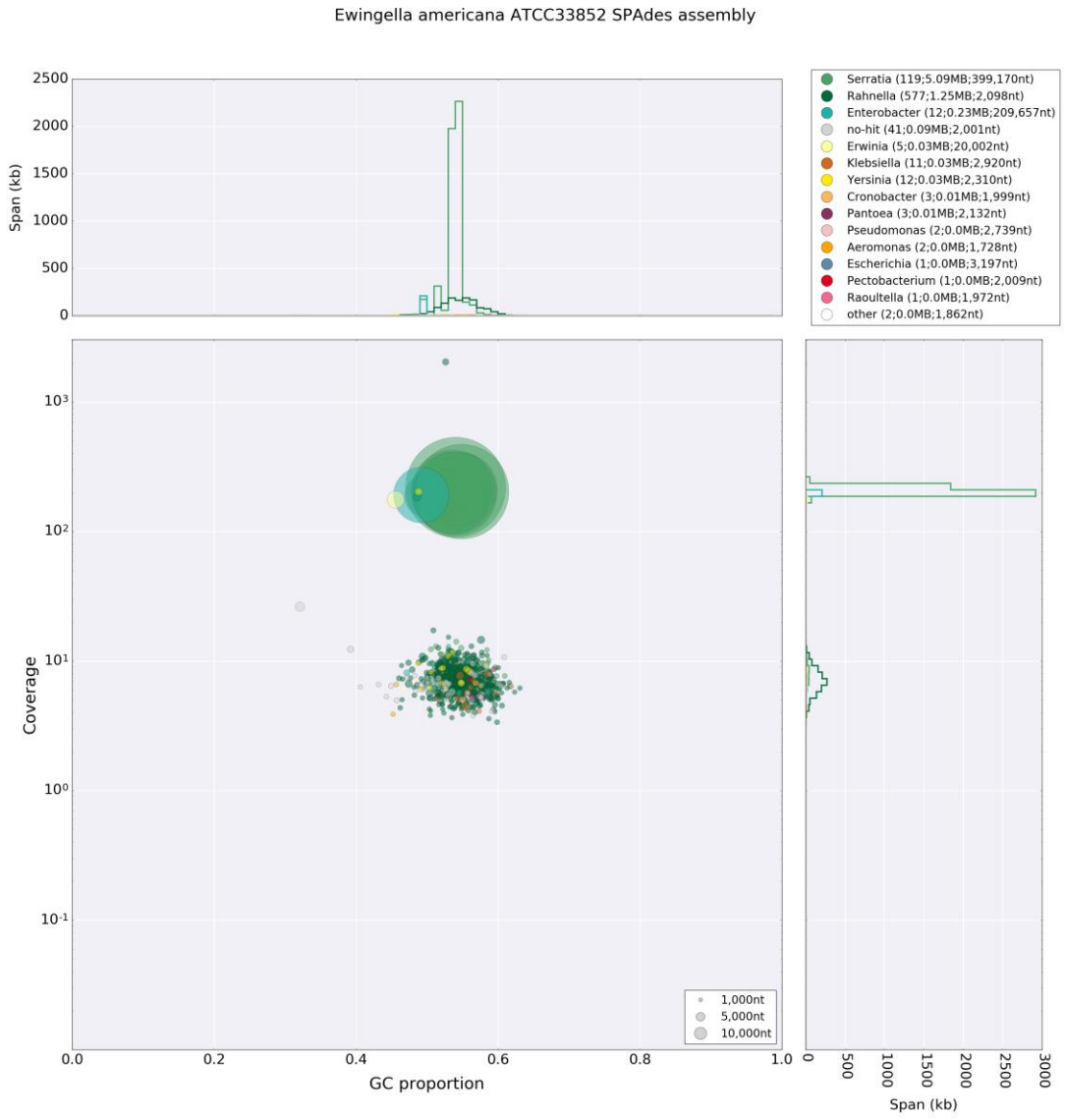


Figure 3.8. Blobplot of *Ewingella americana* ATCC33852 SPAdes assembly. Taxonomic homology between the SPAdes assembly of a sequenced *E. americana* ATCC33852 isolate, and NCBI non-redundant, nucleotide database sequences. The taxonomy key shows the highest homology to the assembled contigs is from the genus *Serratia*. The plot also shows per contig coverage and GC proportion.

3.2.7 Genome annotation

Annotations of the resultant SPAdes draft genome assemblies were automatically generated using the Prokka annotation pipeline v1.11 (Seemann 2014) and the RAST online annotation server (Aziz *et al.* 2008). Prokka annotations were used as input to search for CAZymes, which were automatically identified using the dbCAN online server (Yin *et al.* 2012). Type III, IV and VI secretion systems were annotated using the T346 hunter (Martínez-García *et al.* 2015). Type II secretion systems and assorted virulence factors were manually annotated using locally installed BLAST+ (v2.2.31) (Altschul *et al.*, 1990).

3.2.8 Phylogenomics

The bacterial isolates *G. quercinecans* FRB97, FRB124, & N78, *B. goodwinii* FRB141 & FRB171, *B. alni* NCPPB 3934, *B. salicis* DSM30166, and *E. americana* ATCC33852, were phylogenetically resolved using a whole genome comparison. This high resolution method phylogenetically positioned the isolates against species deciphered as closely related in previous multi-locus sequence analyses (Brady *et al.*, 2010; Denman *et al.*, 2012b). Additional species included within the phylogenetic analysis include, a selection of SRE species, the number one bacterial plant pathogen, based on scientific/economic importance, *Pseudomonas syringae* (Mansfield *et al.*, 2012), and a randomly selected Gram-positive bacterium, *Bacillus licheniformis*.

A total of fourteen genome sequences were downloaded from NCBI (accessed: 10/04/2016) in FastA format, annotated using Prokka, and used to generate a phylogenetic tree, via the microbial phylogenomic reconstruction package PhyloPhlAn v0.99 (Segata *et al.*, 2013), using default settings. PhyloPhlAn aligns homologous sequences, before selecting 400 universal sequences to build an unrooted, evolutionary distance-based phylogenetic tree. Evolutionary distance is defined by the degree of variation within the universal sequence alignment (Mount, 2008). The tree was plotted using the R Bioconductor package ggtree (v1.2.17) (Yu *et al.*)

3.3 Results and Discussion

3.3.1 Whole genome based phylogenetic resolution of *Gibbsiella quercinecans* and *Brenneria goodwinii*

Here, the *in silico* virulence potential was described for of two bacterial species, *B. goodwinii* and *G. quercinecans*, consistently isolated from necrotic tissue of trees affected by AOD, and a third, *E. americana* ATCC33852, which has been isolated on several occasions from trees affected by AOD. The whole genomes of three strains of *G. quercinecans*, two of *B. goodwinii,* and a single strain of *E. americana* were sequenced, assembled and annotated (Table 3.1). Previously, strains of *B. goodwinii, G. quercinecans*, and *E. americana* have been taxonomically placed alongside their nearest closest relative using multi-locus sequence analysis (MLSA) (Brady *et al.*, 2010, 2014b; Denman *et al.*, 2012). Whole genome sequencing offers higher resolution of taxa (around 1,000 times more information), through phylogenetic reconstruction of genome wide conserved genes (Vinatzer *et al.*, 2014), unlike the limited number of conserved genes typicaly used in MLSA analyses. Here, a comparative phylogenomic approach was used to reconstruct the evolutionary relationships of the sequenced genomes within the family *Enterobacteriaceae*.

Table 3.1. Genome metrics of Illumina sequenced and SPAdes assembled genomes from five bacteria isolated from necrotic lesions of AOD affected trees, two positive controls (+), and one negative control (-).

Organism	Origin	No. of contigs	N50 (G+C content %)	No. of genes (Coding density %)	Chromosome size (bp)	Sequencing output (Number of reads, Million)	Genome coverage
Gibbsiella quercinecans FRB97	Hoddesdon Park, UK (Brady *et al.*, 2010)	68	5548506 (55.9)	5087 (86.9)	5,504,004	1.7	49 X
Gibbsiella quercinecans FRB124	Outwood, UK (Brady *et al.*, 2010)	90	161,616 (55.6)	4852 (86.6)	5,469,793	2.4	92 X
Gibbsiella quercinecans N78	Burgos, Spain (Brady *et al.*, 2010)	129	96,420 (56.0)	5202 (86.4)	5,693,731	2.8	75 X
Brenneria goodwinii FRB141	Outwood, UK (Denman *et al.*,	185	61,230 (51.4)	4625 (85.8)	5,281,917	1.4	52 X

	2012a)(Denman *et al.*, 2012b)						
Brenneria goodwinii FRB171	Gorse Covert, UK (Denman *et al.*, 2012a)(Denman *et al.*, 2012b)	128	82,991 (53.4)	4881 (86.1)	5,377,922	1.6	52 X
Brenneria alni NCPPB 3934 (+)	Positive control. Isolated from Italian Alder (*Alnus cordata*) in Italy. Causative agent of bark canker (Surico *et al.*, 1996).	132	71,375 (52.4)	4013 (86.9)	4,127,267	1.7	73 X
Brenneria salicis DSM30166 (+)	Positive control. Isolated from willow (*Salix alba*	106	85,701 (86.4)	3781 (86.4)	3,929,937	3.9	136 X

	var. *caerulea*) in the UK. Causative agent of watermark disease (Maes *et al.*, 2009).						
Ewingella americana ATCC33852 (-)	Brockhampton, UK (Brady *et al.*, 2014b)	792	308,463 (53.9)	6592 (82.9)	6,777,808	3.8	111X

The microbial phylogenetic alignment program PhyloPhlan was used to detect a set of 400 conserved marker sequences from the draft whole genomes, alongside a further fourteen finished bacterial genomes. A phylogenomic comparison revealed the evolutionary distance between bacteria and concurred with previous phylogenetic approaches (Brady *et al.*, 2010; Denman *et al.*, 2012b). The nearest, whole genome sequenced relatives of *G. quercinecans* occupy a separate clade to *Brenneria spp.* and were positioned alongside the genera *Serratia*, *Erwinia* and *Rahnella* (Fig. 3.9). *B. goodwinii* is as expected, most closely aligned to *B. alni* NCPPB 3934 and *B. salicis* DSM30166, which are phytopathogens within the genus *Brenneria*, and the causative agents of bark canker in alder (Surico *et al.*, 1996) and watermark disease in willow (Maes *et al.*, 2009), and may have pathogenicity mechanisms which can also be found in *B. goodwinii.* Interestingly, *B. goodwinii* is within the same phylogenetic clade as the SRE group of opportunistic phytopathogens, which includes *Pectobacterium carotovorum* and *Dickeya dadantii* (Charkowski *et al.*, 2012). The putative endophyte *E. americana* ATCC33852, has been sporadically isolated from AOD affected oaks and is commonly found in the intestinal microbiome of snails and slugs. However, it has been linked to pathogenesis in mushrooms and can cause nosocomial infections in immunocompromised patients (Hassan *et al.*, 2012). *G. quercinecans* and *B. goodwinii* are closely related species within the family *Enterobacteriaceae*; their phylogenetic resolution is crucial for genomic comparisons between closely related species. Here, functional annotations from strains of *G. quercinecans* and *B. goodwinii* are compared to homologous genes in other species and evaluated with particular emphasis on homologous virulence factors in bacterial phytopathogens, highlighting the putative impact of these genes on AOD.

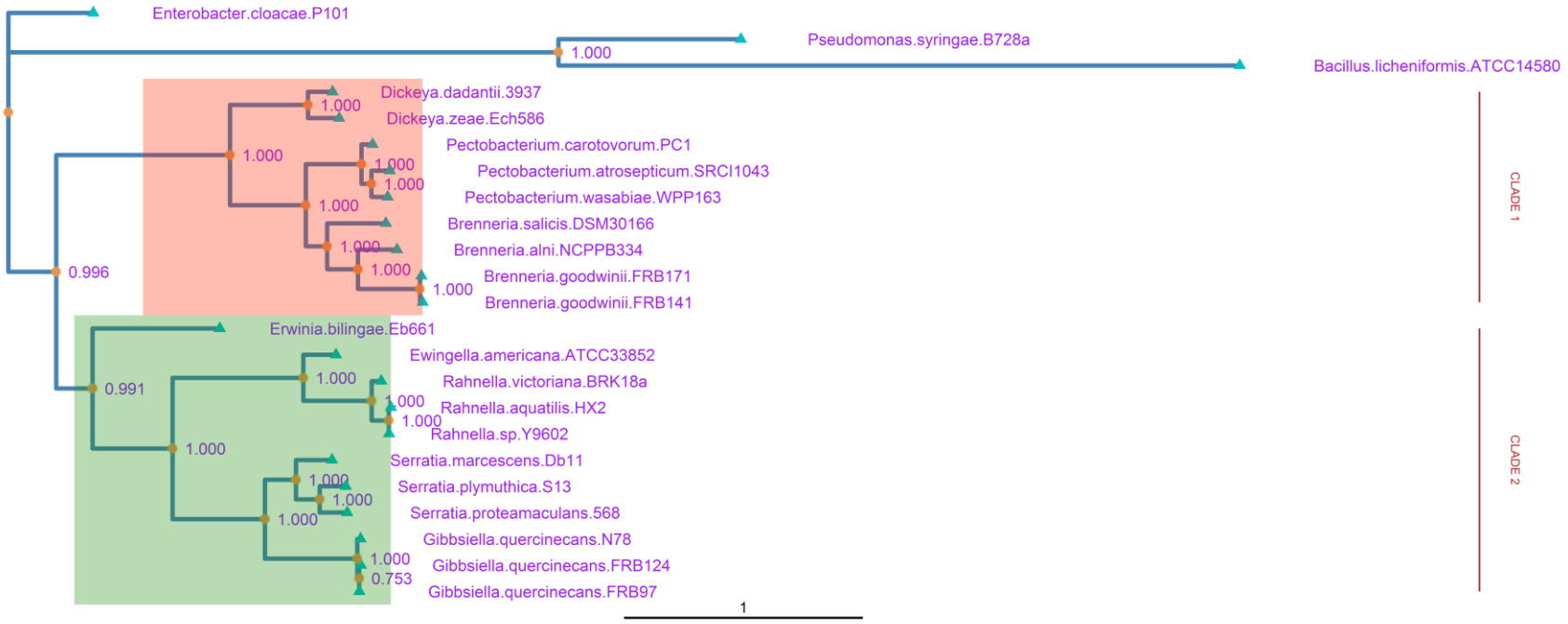


Figure 3.9. Phylogenetic tree based on evolutionary relationships between a set of 400 homologous marker sequences. The scale shows evolutionary distance between species. The tree is arbitrarily rooted. Numbers represent posterior probability of node position.

3.3.2 Putative virulence mechanisms encoded within *Gibbsiella quercinecans* and *Brenneria goodwinii*

3.3.2.1 Secretion systems

To enable macromolecules such as plant cell wall degrading enzymes (PCWDEs) to cross bacterial membranes, proteinaceous nanomachines known as secretion systems are employed to facilitate the process (Fig.3.10) (Costa *et al.*, 2015). Secretion systems are critical to the study of bacterial virulence, as they enable delivery of many toxins and effectors to recipient cells. Strains of *G. quercinecans* and *B. goodwinii* encode multiple secretion systems, the most prominent virulence enabling systems among closely related bacterial phytopathogens, are the type II, III, IV and VI secretion systems (Fig. 3.11). Secretion systems are key virulence determinants amongst the *Enterobacteriaceae* (Toth *et al.*, 2006; Mansfield *et al.*, 2012).

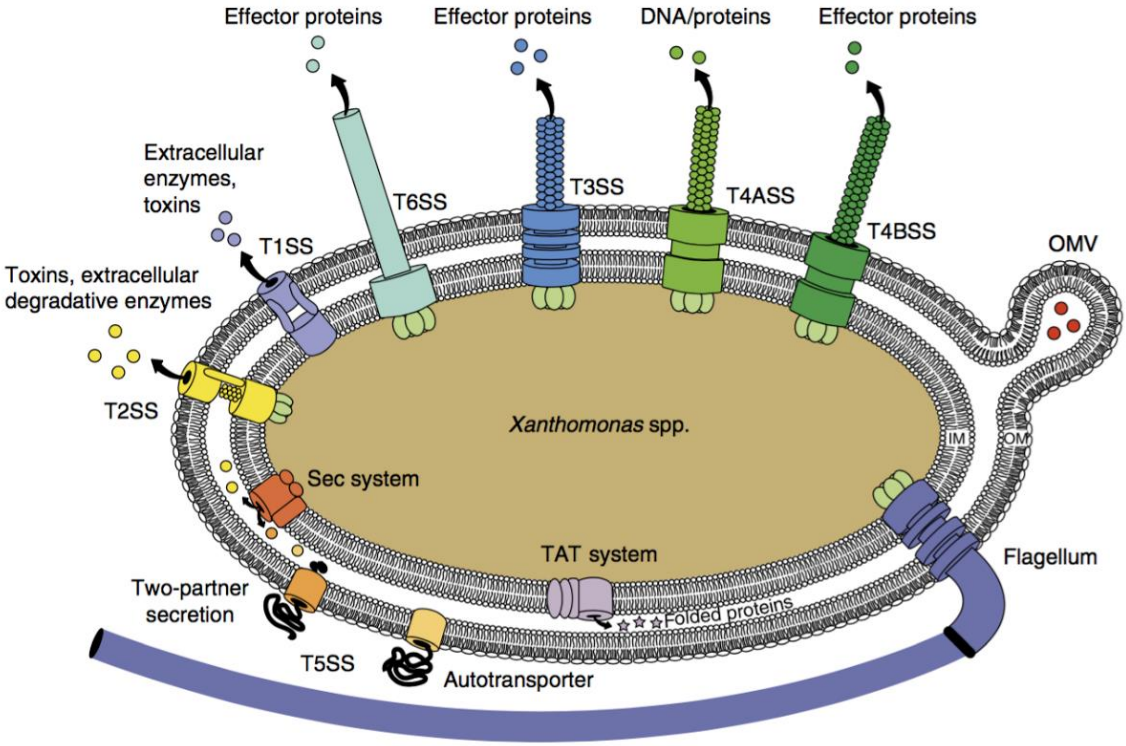


Figure 3.10. Secretion systems of *Xanthomonas spp.* Similar systems are found in the family *Enterobacteriaceae*, including Type I, II, III, IV & VI. Illustration from (Buttner and Bonas, 2010).

Pathogenicity determinants (including PCWDEs) require secretion system mediated transport into the extracellular surroundings or directly into host cells. The basic secretion apparatus consists of core components, which are necessary for secretion, and identification of these *in silico* is a strong indicator of functionality (Murdoch *et al.*, 2011). However, it should be noted that *in silico* validation is not evidence of secretion capable systems, as a large number of vestigial secretion system genes may be present without being functional (Ochman and Davalos, 2006).

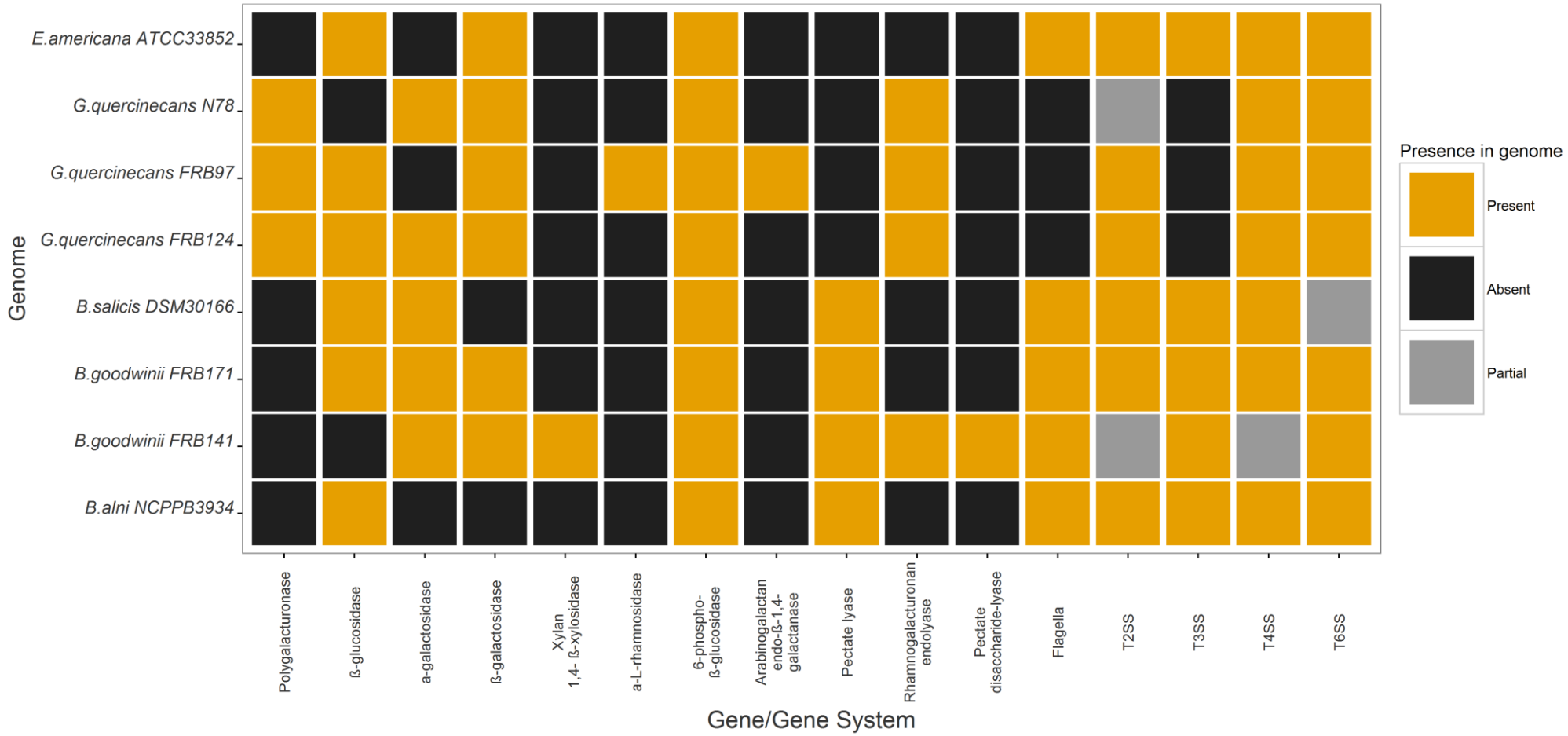


Figure 3.11. Presence/absence chart of secretion systems and PCWDEs in bacterial isolates recovered from necrotic lesions of AOD affected trees. Included in the chart are the phytopathogens, *B. alni* NCPPB 3934 and *B. salicis* DSM30166.

3.3.2.2 The general secretory pathway

Complete general secretory pathway (GSP) translocon operons were annotated *in silico* in *G. quercinecans* FRB97, FRB124, & N78, *B. goodwinii* FRB141 & FRB171, *B. alni* NCPPB 3934 and *B. salicis* DSM30166, and *E. americana* ATCC33852. The GSP of Gram-negative bacteria transports proteins from the cytoplasm, across the inner membrane to the periplasm. Cytosolic proteins are usually synthesised and regulated by signal peptides, which bind to the protein and are ultimately cleaved by signal peptidase upon passage across the cytoplasmic membrane (Palmer and Berks, 2012). The GSP has two possible translocons, which target proteins to the periplasm, these are: the secretion (Sec) translocase, which consists of three core proteins *secYEG*, which transports unfolded proteins, and the twin-arginine translocase (Tat), which also consists of three core proteins *tatABC* and transports folded proteins. The Sec translocase is an essential, universally conserved system across all prokaryotes. Despite being non-essential in most prokaryotes, the Tat translocase is essential for virulence in almost all animal and plant pathogens and is an exclusive inner membrane transport system for the T2SS (De Buck *et al.*, 2008; Costa *et al.*, 2015). Unlike other secretion systems, the T2SS requires periplasm-localised Sec or Tat secretion. The Sec translocase delivers proteins to several secretion systems, these are the Type I, II, IV and V secretion systems, with only the Type III and Type VI being strictly Sec-independent (Büttner and Bonas, 2002; Coulthurst, 2013). The Tat system only delivers proteins to the T2SS and its upregulation in gene expression studies may indicate T2SS activation (Voulhoux *et al.*, 2001). Whether the T2SS receives Sec or Tat targeted proteins depends on the N-terminal signal peptide of the protein substrate. There is no exclusivity to virulence protein periplasmic localisation, with both Sec and Tat translocons capable of facilitating first step virulence gene secretion. Furthermore, the same substrates are targeted to the periplasm by both Tat and Sec in strains of the plant pathogen *Ralstonia solanacearum* (González *et al.*, 2007). *R. solanacearum* is a necrotrophic phytopathogen which uses the Tat pathway for inner membrane secretion of virulence factors including PCWDEs. Similarly, a pectate lyase protein (*pnlH*) in *Dickeya dadantii* 3937 is targeted to the periplasm by the Tat translocase (Ferrandez and Condemine, 2008). However, the Sec pathway of *D. dadantii* EC16 secretes a pectate lyase protein (*pelW*) across the cytoplasmic membrane (He *et al.*, 1991). Evidence from investigations summarised here reveal the idiosyncratic nature of translocases, and their possible functionality within *G. quercinecans* and *B. goodwinii*.

3.3.2.3 Type II secretion system

Type II secretion is a two-step process and initially relies upon the GSP for substrate transport from the periplasm across the inner membrane (Desvaux *et al.*, 2004). The type II secretion apparatus is made up of twelve core proteins (namely T2S CDEFGHIJKLMNO), known as the secreton, which are required for a functional system (Cianciotto, 2005; Johnson *et al.*, 2006; Toth *et al.*, 2006). All twelve T2SS core genes were annotated in *E. americana* ATCC33852, *G. quercinecans* FRB97 and FRB124, but several were missing from strain N78 (Fig. 3.11). Similarly, *B. goodwinii* FRB141 had only three annotated core type II genes, whereas strain FRB171 contained a complete set of type II core genes located in a single operon (Cianciotto, 2005). Type II secretion systems are commonly found in γ-*proteobacteria*, especially the *Enterobacteriaceae* (Cianciotto, 2005), and are essential for virulence in the SRE as they are the only transport mechanism for PCWDEs (with the exception of some proteases, secreted via the type I autotransporter), to pass through the cell envelope into the extracellular space, thereby macerating plant host tissue and causing soft rot disease (Wandersman, 1989; Sandkvist, 2001; Liu *et al.*, 2008; Nykyri *et al.*, 2012). Type II secretion is a highly coordinated process within the SRE; the first step in the process is the release of quorum sensing signals into the extracellular environment, when cell density reaches a certain threshold; consequently, PCWDEs are synthesised, and targeted to the periplasm by the GSP before finally being secreted through the T2SS to the extracellular environment (Toth *et al.*, 2006). This process of intercellular communication leading to plant host destruction could be described as a recruitment-destruction (RD) pathway, where invading pathogens release small soluble molecules from the pre-macerated tissue before a synchronous mass release of PCWDEs. SRE induced tissue necrosis through the release of PCWDEs, including pectate lyase and cellulase are well characterised examples of the RD mechanism in *P. carotovorum* subsp. *carotovorum* and *D. dadantii* 3937 (Salmond G, 1994).

3.3.2.4 Type III secretion system

The type III secretion system is a notorious pathogenicity factor - essential to the virulence of prominent bacterial pathogens, such as *Yersinia* spp. and *Salmonella* spp. in animals, and *P. syringae*, *Erwinia amylovora* and *D. dadantii* in plants (Oh and Beer, 2005; Lovell *et al.*, 2009; Jahn *et al.*, 2011; Cordero-Alba *et al.*, 2012). *B. goodwinii* FRB141 & FRB171, *B. alni* NCPPB 3934 and *B. salicis* DSM30166, and *E. americana* ATCC33852, encode a single T3SS operon, consisting of nine core and various accessory genes. Membrane spanning T3SS proteins inject

effector proteins directly into the eukaryotic host cell cytoplasm, which elicits several pathogen-specific responses including, translocation of toxins (*E. amylovora*), evasion of ingestion by macrophages (*Yersinia spp.*), invasion and multiplication within host cells (*Salmonella* spp.) (Toth *et al.*, 2006). *B. goodwinii* FRB141 & FRB171 encode effector proteins (Fig. 3.11), which align with greatest homology to the *yopJ* effector gene of *Yersinia* spp., *hopM1* and two *hopPto* effectors of *P. syringae* pv. tomato. *P. syringae* is a canonical, economically important, hemibiotrophic plant pathogen with over fifty host specific pathovars, and a wide host range but mainly affecting herbaceous and woody plants (Green *et al.*, 2010). Pathogenesis of this primary pathogen is dependent on injection of T3SS effectors into the host cell, which is mediated by the hypersensitivity response and pathogenicity (*hrp*) pathway (Lindeberg *et al.*, 2012). *P. syringae* has an effector 'super-repertoire' encompassing 57 families, these effectors are designated either Hop or Avr and are generalists which can function on a wide range of plants. The T3SS is crucial in the pathology of *P. syringae* as knock-out mutants are avirulent (Alfano *et al.*, 2000). Empirical evidence suggests identification of a T3SS cluster within a putative pathogen is potentially critical to establishing the causality of disease (Arnold and Jackson, 2011). A T3SS cluster has been identified in *B. goodwinii* adding substantial evidence to its putative pathogenicity, which is further enhanced by the consistent isolation of the bacterium from necrotic AOD lesions. However, it should be noted that conserved, functional Hrp systems are found in the SRE but are considered a minor virulence component (Pérombelon, 2002). However, non-beneficial bacterial virulence genes are energetically expensive to maintain and are often rapidly purged from the genome (Pallen and Wren, 2007). The continued presence of a T3SS within many SRE genomes may reflect its importance in establishing a host niche, through manipulation of defences, prior to latent stage fulfilment of the RD pathway (Tampakaki *et al.*, 2010). Expression of the T3SS minor virulence factors prior to expression of major virulence factors is exemplified in the corn pathogen *Pantoea stewartii* subsp. *stewartii*; in this disease model, quorum sensing mediates pathogenicity, initially releasing T3SS *hrp* effectors, causing cell leakage and bacterial nutrient absorption, before secretion of its primary pathogenicity factor, a dense capsular biofilm, constituting the exopolysaccharide compound, stewartan (Coplin *et al.*, 1992; Pérombelon, 2002). Furthermore, there is increasing evidence of T3SS interaction with insect vectors, such as in *D. dadantii* 3937 (Costechareyre *et al.*, 2012) and in *P. stewartii* subsp. *stewartii* (Roper, 2011), providing a potentially significant advantage for encoding the T3SS. Alternatively, the T3SS may reflect a hemibiotrophic lifestyle of *B. goodwinii*, allowing interaction with the host rather than the instant killing behaviour of necrotrophic pathogens (Glazebrook, 2005). T3SS mediated manipulation may enable a parasitic relationship with the oak tree host, allowing successful

undetected colonisation due to manipulation of host defences (Abramovitch and Martin, 2004), and the acquisition of nutrients through induced low level cellular-leakage, before conditions become favourable to incite the RD pathway.

3.3.2.5 Type IV secretion system

The Type IV secretion system (T4SS) is unique amongst the macromolecular transportation systems in its diversity of delivery mechanisms (Alvarez-Martinez and Christie, 2009; Low *et al.*, 2014). The T4SS facilitates bacterial evolution through horizontal gene transfer (HGT), specifically, conjugation mediated transfer of large scale non-homologous gene systems; which is the most frequent method of bacterial HGT (Guglielmini *et al.*, 2013), with a broader range than transformation and transduction (Guglielmini *et al.*, 2014). The T4SS has evolved via a process of exaptation, to widen its diversity from its original conjugative role to include protein delivery functionality (Guglielmini *et al.*, 2013). Furthermore, the T4SS can release, and uptake DNA, from the extracellular environment (Voth *et al.*, 2012). A simple delineation of T4SS into conjugation (cT4SS) and protein secretion (pT4SS) subsystems is frequently used. This nomenclature scheme has been adopted by the specific annotation programs, T346Hunter (Martínez-García *et al.*, 2015), SecReT4 (Bi *et al.*, 2013), and MacSyFinder (Abby *et al.*, 2014), and for simplicity has been employed here. All bacterial strains described here had at least one T4SS, with the exception of *B. goodwinii* FRB141. Genes from incomplete T4SS and T2SS were identified in *B. goodwinii* FRB141, which encoded several genes from both systems but had many core genes missing. Identification of a T2SS and T4SS within strain FRB171 is perhaps evident of the dispensable nature of these systems within *B. goodwinii,* and unlike the T3SS is not necessary for survival within the oak lesion microbiome. *G. quercinecans* N78 encodes four T4SSs, three of which are cT4SS with the fourth having cT4SS and effector functionality, revealing substantial HGT potential and evidence of past HGT events. A single conjugation and effector T4SS was encoded within *E. americana* ATCC33852, *G. quercinecans* FRB97 & FRB124, *B. goodwinii* FRB171, and *B. salicis* DSM30166; *B. alni* NCPPB 3934 has a conjugation and effector cluster, along with a further cT4SS locus.

3.3.2.6 Type VI secretion system

The type VI secretion system (T6SS) is a recently described Gram-negative membrane transport system, which was identified simultaneously in *Vibrio cholera* (Pukatzki *et al.*, 2006) and

Pseudomonas aeruginosa (Mougous *et al.*, 2006). These initial discoveries revealed the role of T6SS in the one-step translocation of toxic effectors from the bacterial cytoplasm into eukaryotic hosts. Recently, there has been growing evidence that the T6SS is predominantly involved in interbacterial competition, targeting competitors and promoting survival in the polymicrobial environment (Russell *et al.* 2014). Subsequent analyses may refine this definition and reveal an evolutionary separation between anti-eukaryotic and anti-prokaryotic subsystems. Currently, however, the T6SS is organised in highly conserved polycistronic operons compromising thirteen core proteins, designated TssA-M (Shalom *et al.*, 2007). Two of these proteins known as VgrG/TssI (Valine-Glycine Repeat protein G) and Hcp/TssD (Haemolysin Co-Regulated protein) are part of the injectisome but can also function as effectors, and it has been suggested that they may play an additional role, in cell-cell signalling and biofilm formation (Murdoch *et al.*, 2011). Unlike other core T6SS genes, the VgrG and Hcp regions are diversity hot-spots, which appear evolutionarily distinct and differ substantially in GC content, indicating that they may be horizontally acquired (Shyntum *et al.*, 2014).

The T6SS penetrates host cells using the *vgrG* and *hcpD*injectisome to puncture target cells, a mechanism which is functionally equivalent to the contractile tail of a bacteriophage (Shneider *et al.*, 2013), and delivers toxic effector proteins into both prokaryotic (Hood *et al.*, 2010) and eukaryotic hosts (Boyer *et al.*, 2009). Many ecologically diverse *Enterobacteriaceae* species encode numerous T6SS loci (including pathogenic and epiphytic plant associated bacteria) (De Maayer *et al.*, 2011; Coulthurst, 2013), the sequenced strains of *G. quercinecans* (2-3 operons per strain), *B. goodwinii* (1 operon per strain), and *E. americana* (2 operons) follow this trend, encoding at least one T6SS. Due to its prevalence amongst the *Enterobacteriaceae* (more than an estimated 25% of species encode at least one locus), the T6SS confers a beneficial adaptation to a large number of bacteria in divergent environments, however from genome sequencing data of *G. quercinecans* and *B. goodwinii* strains, it cannot be inferred what benefit the T6SS offers (Ochman and Davalos, 2006; Coulthurst, 2013).

3.3.2.7 Plant cell wall degrading enzymes

Unlike the lipid based structures of mammalian cell walls, plant cell walls are composed of polysaccharides. PCWDEs equip phytopathogens with the capability to target these polysaccharides and utilise them as a nutrient source (Toth *et al.*, 2006). Animal pathogenic *Enterobacteriaceae* also produce PCWDEs, but these enzymes show reduced capacity to

macerate plant tissue (Manulis *et al.*, 1988). Canonical SRE plant pathogens discharge specific enzymes which macerate plant cell wall polysaccharides and release nutrients for bacterial growth (Toth *et al.*, 2003). These species depolymerise host tissues through the release of exoenzymes, including cellulases, proteases and multiple isoforms of pectinase (Beaulieu *et al.* 1993). Each pectinase isoform may specifically target a particular host, and encoding a range of isoforms allows the pathogen to attack various substrates. Cellulases degrade the primary and secondary plant cell walls, and along with proteases, are minor virulence factors in these pathogens. Pectinases are major pathogenicity factors, which break down the middle lamella (adhesive component which fastens components together), resulting in tissue collapse, cell damage and leakage (Barras *et al.*, 1994). Resultant degraded products are transported across the bacterial cell membrane and catabolised (Nasser *et al.*, 1992).

The CAZy database is a specific annotation tool, which identifies closely related carbohydrate active enzymes (CAZymes) and can be used to identify PCWDEs (Lombard *et al.*, 2014). CAZymes are enzymes involved in the assembly (glycosyltransferases) and breakdown (glycoside hydrolases, polysaccharide lyases and carbohydrate esterases) of complex carbohydrates (Levasseur *et al.*, 2013). CAZyme classes suspected of being PCWDEs can be further specified using enzyme commission (EC) classification, which provides a more detailed catalytic profile (Bairoch, 1999).

G. quercinecans has a notable polysaccharide lyase (PL) family enzyme, PL4 (EC 4.2.2.23), specifically, a pectate lyase, which degrades the plant cell wall component rhamnogalacturonan and was originally isolated in the phytopathogen, *Aspergillus aculeatus* (Fig. 3.11) (Kofod *et al.*, 1994). *B. goodwinii* FRB141 encoded this PL, and two pectate lyases, PL2 (EC 4.2.2.9) and PL3 (EC 4.2.2.2). Apart from *B. goodwinii* FRB141, sequenced *Brenneria spp.* encoded only PL3 enzymes. Pectin lyase enzymes were absent from *E. americana* ATCC33852.

Glycoside hydrolases (GH) are a superfamily of enzymes which hydrolyse glycosidic bonds, and have a significant role in the degradation of lignocellulosic biomass (Gibson *et al.*, 2011). The three sequenced strains of *G. quercinecans*, were the only isolates to encode GH family 28 polygalacturonase (EC 3.2.1.15; PehA), a homolog of which is secreted to the extracellular environment in *P. carotovorum* subs. *carotovorum*, and is a well characterised virulence-inducing pectinase (Abbott and Boraston, 2008). Of the sequenced genomes, only *B. goodwinii* FRB141 and *G. quercinecans* N78 did not encode the cellulose degrading GH family β-glucosidase, BglA (EC 3.2.1.21), which is a virulence factor in *Dickeya spp.* (Zhou *et al.*, 2015). A reducing α-

galactosidase (EC 3.2.1.22), acts against plant carbohydrate with glucose and galactose residues (melibiose), is encoded by *B. goodwinii* FRB141 & FRB171, *G. quercinecans* FRB124 & N78, and *E. americana* ATCC33852. This GH family 4 CAZyme is transported extracellularly by *D. dadantii* 3937 and catabolises the disaccharide, melibiose, for use as a carbon source (Hugouvieux-Cotte-Pattat and Charaoui-Boukerzaza, 2009). Both *B. salicis* DSM30166 and *B. alni* NCPPB 3934, lack a β-galactosidase (EC 3.2.1.23), which is encoded within all strains of *G. quercinecans*, *B. goodwinii,* and *E. americana* ATCC33852, and is homologous to a PCWDE in *P. carotovorum* subsp. *carotovorum* (Cui *et al.*, 2005). *G. quercinecans* FRB97 was the only isolate to encode a homolog of exo-1,4-β-D-xylosidase (EC 3.2.1.37), a hemicellulase, and part of the N-glycan degradative toolkit of the black rot phytopathogen *Xanthomonas campestris* pv. *campestris* (Dupoiron *et al.*, 2015). Furthermore, *G. quercinecans* FRB97 encodes a GH53, which is an arabinogalactan endo-β-1,4-galactanase (EC 3.2.1.89), and is a cellulolytic PCWDE found in diverse bacteria phyla, including *Cellulomonas spp.*, *Bacillus spp.*, and *Pseudomonas spp.* (Gilbert and Hazlewood, 1993; Levy *et al.*, 2002). Similarly, *B. goodwinii* FRB141 encodes a copy of GH78, α-L-rhamnosidase (EC 3.2.1.40), a PCWDE of the soft rot fungi *Xylaria polymorpha*, where it has glycosyl hydrolase and esterase activities, enabling the fungus to macerate lignocellulosic material (Nghi *et al.*, 2012). These data reveal the *in silico* virulence potential of *G. quercinecans* and *B. goodwinii* strains, not only to encode the mechanisms of plant cell wall destruction, but also their delivery systems, which are homologous to those encoded within canonical bacterial necrotrophic and hemibiotrophic phytopathogens (Charkowski *et al.*, 2012).

3.3.2.8 Regulation of virulence gene expression

Bacterial phytopathogens tightly control the release of PCWDEs and type III effectors through the relaying of environmental signals to an internal regulatory circuit (Cui *et al.*, 2005). This signalling response is central to the success of pectinolytic phytopathogens such as *P. carotovorum* subsp. *carotovorum*, which has a suite of characterised virulence regulators (Rodionov *et al.*, 2004). Homologous lignocellulosic degradation regulators from bacterial phytopathogens were identified in all strains of *G. quercinecans* and *B. goodwinii*, and *B. alni*, NCPPB 3934, *B. salicis* DSM30166, and interestingly the putative endophyte, *E. americana* ATCC33852, including ExpR, KdgR, and two-component signal transduction systems (TCSTS) for example, PhoP-PhoQ, and GacA-GacS, which are able to sample the external environment

and transcribe genes in response to fluctuating environmental conditions (Table 3.2) (Rio-Alvarez *et al.*, 2012).

Lignocellulosic regulators control virulence gene expression of the SRE and respond to environmental signals such as population cell density; if an appropriate environmental cue is provided, the repressor of pectinolysis, *kdgR*, will not be transcribed and degradative enzymes are released. Bacteria use quorum sensing signals known as acyl-homoserine lactones (AHLs) to relay information such as population density. AHLs released within necrotic *Pectobacterium spp.* tissue, bind to the *luxR* homologue *expR*,, which mediates activity of Type II secretion systems and no longer attaches to the global virulence and PCWDE repressor, *rsmA,* resulting in the overproduction of PCWDEs (Chatterjee *et al.*, 1995; Liu *et al.*, 2008). Contrastingly, virulence of *D. dadantii* is not substantially affected by quorum sensing signalling, but instead is mediated by intracellular transcriptional regulators reacting to environmental cues, particularly plant derived signal molecules, and specifically, pectate fragments which induce the production of 2-keto-3-deoxygluconate (KDG), which represses *kdgR*, and in a mechanism similar to *Pectobacterium spp.*, downregulates *rsmA*, (Nasser *et al.*, 1992; Reverchon and Nasser, 2013).

Phenolic compounds are also known to act as inducers of T3SS expression in *D. dadantii*, where they up-regulate expression of *hrpL* through the *rsmB-rsmA* pathway (Yang *et al.*, 2008) resulting in the release of PCWDEs by *Pectobacterium* and *Dickeya spp.*, and maceration of plant tissue via the RD pathway. *E. americana* ATCC33852 was isolated on several occasions from necrotic lesions of AOD affected trees, indicating that it can survive in this environment. The sporadic isolations in addition to the annotation of encoded lignocellulosic regulators, suggests that *E. americana* ATCC33852 may contribute to intercellular quorum sensing signalling, and form part of an interchangeable consortia of opportunistic *Enterobacteriaceae*, which do not directly contribute to pathogenesis but can participate in intercellular signalling and benefit through access to partially degraded sugars. However, as revealed in Fig. 3.11, *E. americana* ATCC33852, is unlikely to be a necrotrophic phytopathogen, as unlike strains of *G. quercinecans* and *B. goodwinii*, *E. americana* ATCC33852 does not have a degradative arsenal of PCWDEs to macerate host tissue. The combination of regulators, secretion systems, effectors, and degradative enzymes, in addition to the consistent isolation of *G. quercinecans* and *B. goodwinii*, is substantial cumulative evidence of their molecular potential to contribute to AOD. A similar body of cumulative evidence has previously revealed the mechanisms through which the SRE induce disease (Toth *et al.*, 2006). Genome encoded regulators and PCWDEs within strains of

G. quercinecans and *B. goodwinii*, reveal the RD pathway is a possible pathogenicity strategy in these bacteria.

Table 3.2. Virulence regulators encoded within *G. quercinecans* FRB97 (*Gq*97), FRB124 (*Gq*124), & N78 (*Gq*78), *B. goodwinii* FRB141 (*Bg*141) & FRB171 (*Bg*171), *B. alni* NCPPB 3934 (*Ba*), *B. salicis* DSM30166 (*Bs*), and *E. americana* ATCC33852 (*Ea*). Number in brackets denotes, number of gene copies. TCSTS – Two-component signal transduction system.

Virulence regulator	Genome	Predicted function	Reference
ompR	*Gq*97, *Gq*124, *Gq*78, *Bg*141, *Bg*171, *Ba*, *Bs*, *Ea*(4)	Master stress response activator in *Salmonella enterica*	(Duprey *et al.*, 2014)
phoP/pmrA	*Gq*97, *Gq*124, *Gq*78, *Bg*141, *Bg*171, *Ba*, *Ea*	TCSTS. Receptor component, induces expression of virulence genes in a wide variety of plant pathogens, and represses expression of virulence genes in *Dickeya spp.*, *Pectobacterium spp.*, and *Erwinia amylovora*	(Hyytiäinen *et al.*, 2003; Ellison and Miller, 2006; Costechareyre *et al.*, 2010)
phoQ/pmrB	*Gq*97, *Gq*124, *Gq*78, *Bg*141, *Bg*171, *Ba*, *Ea*	TCSTS. Sensor component. Detects low magnesium levels and signals	As above

		PhoP, and e.g. represses pectate lyase expression in *Dickeya spp.*	
uvrY/gacA	*Gq*97, *Gq*124, *Gq*78, *Bg*141, *Bg*171, *Ba*, *Bs*, *Ea*	TCSTS. Receptor component. Regulates expression of type I secretion of proteases in *Pectobacterium spp.* Represses *rsmA*, induces *rsmB*	(Mhedbi-Hajri *et al.*, 2011)
gacS/barA/pecS	*Gq*97, *Gq*124, *Gq*78, *Bg*141, *Bg*171, *Ba*, *Bs*, *Ea*	TCSTS. Sensor component. Prevents premature expression of virulence genes in many plant pathogens, thus avoiding detection by plant immune system	As above
virR/expR/luxR	*Gq*97, *Gq*124, *Gq*78, *Bg*141, *Bg*171, *Ba*, *Bs*, *Ea*	Quorum sensing molecule receptor in plant pathogenic bacteria	(von Bodman *et al.*, 2003)

rsmA/csrA	*Gq*97, *Gq*124, *Gq*78, *Bg*141, *Bg*171, *Ba*, *Bs*, *Ea*	Post-transcriptional repressor of virulence gene expression in the SRE	(Broberg *et al.*, 2014)
rsmB/csrB	*Gq*97, *Gq*124, *Gq*78, *Bg*141, *Bg*171, *Ba*, *Bs*, *Ea*	RNA post-transcriptional inhibitor of virulence repressor *rsmA* in *Pectobacterium spp.*	(Vakulskas *et al.*, 2015)
slyA/marR/pecS	*Gq*97, *Gq*124, *Gq*78, *Bg*141, *Bg*171, *Ba*, *Bs*, *Ea*	Positive regulator for many virulence genes in plant and animal bacterial pathogens	(Haque *et al.*, 2009)
pecT/hexA	*Gq*97, *Gq*124, *Gq*78, *Bg*141, *Bg*171, *Ba*, *Bs*, *Ea*	Positive regulator of PCWDEs in many plant pathogens	(Harris *et al.*, 1998; Cui *et al.*, 2005)
rpoS	*Gq*97, *Gq*124, *Gq*78, *Bg*141, *Bg*171, *Ba*, *Bs*, *Ea*(2)	Alternative sigma factor which enables bacterial survival under osmotic and oxidative stress, and starvation	(Andersson *et al.*, 1999)

kdgR	*Gq*97(4), *Gq*124(4), *Gq*78(3), *Bg*141(4), *Bg*171(4), *Ba*, *Bs*(2), *Ea*(3)	Key repressor of pectinolysis in the SRE	(Nasser *et al.*, 1992)
fliA	*Bg*141, *Bg*171, *Ba*, *Bs*, *Ea*	Encodes alternative transcription factor. Induces assembly of flagella and chemotaxis genes in many *Enterobacteriaceae* species, and in some species the T3SS	(Jahn *et al.*, 2008)
fur	*Gq*97, *Gq*124, *Gq*78, *Bg*141, *Bg*171, *Ba*, *Bs*, *Ea*	Key repressor of iron uptake in many bacteria	(Escolar *et al.*, 1999)

3.3.2.9 Motility

The presence of flagella in *B. goodwinii* has been documented microscopically (Denman *et al.*, 2012), and is now confirmed *in silico* by whole genome sequences, as numerous flagellar genes (including a complete cluster of core genes), were identified within all *Brenneria* isolates. Additionally, sequenced *B. goodwinii* strains had a high number of chemotactic genes. Contrastingly, *G. quercinecans* strains had no documented flagella (Brady *et al.*, 2010), a result now confirmed *in silico* with few annotated flagellar genes, and fewer annotated chemotactic associated genes. Flagella can be a key pathogenicity factor, as in *D. dadantii* and *P. carotovorum*, where flagella are involved in disease progression (Hossain *et al.*, 2005; Antúnez-Lamas *et al.*, 2009). In order to utilise flagella efficiently bacteria respond to environmental cues, such as the chemotactic response to AHLs. A signal is transferred from membrane receptors to the flagella regulators *flhC*, *flhD* and *fliA*, and bacteria then swim or swarm (depending on the environment) towards an attractant or away from a repellent (Lux and Shi, 2004; Bowden *et al.*, 2013). This chemotactic signal cascade is enhanced by the density of methyl-accepting chemotaxis proteins (MCPs) on the bacterial membranes. The number of MCPs is markedly higher in *Enterobacteriaceae* pathogens associated with plants than those associated with animals, e.g. *E. coli* typically encodes five compared to more than thirty in *Dickeya* and *Pectobacterium* spp. This is reflected in the significant role of chemotaxis and flagella in *D. dadantii,* where the bacterial wild type strain is able to penetrate *Arabidopsis* leaves (Antúnez-Lamas *et al.*, 2009), whereas strains with mutagenized chemotactic signal transduction systems and flagellar motor genes have impaired swimming ability and are unable to penetrate leaves.

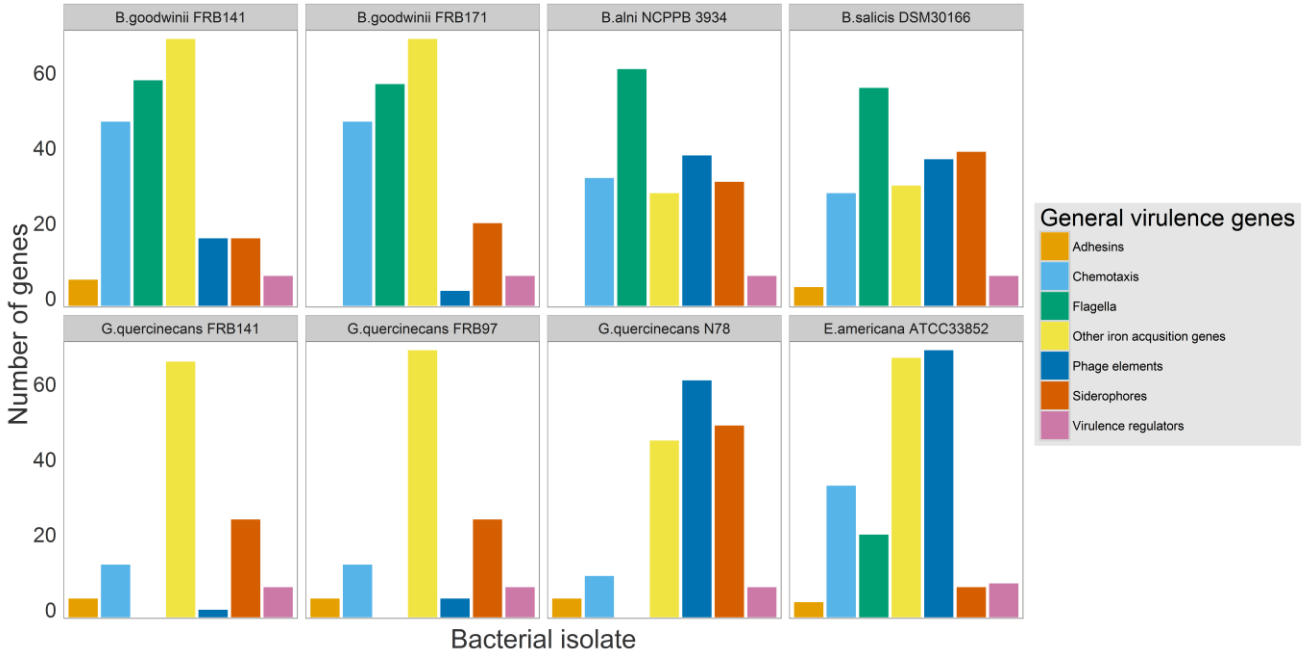


Figure 3.12. Functional annotation of virulence genes. General virulence genes encoded within five bacterial strains, isolated from necrotic lesions of AOD affected trees, and the phytopathogens *B. alni* NCPPB 3934 and *B. salicis* DSM30166 and a putative endophyte *E. americana* ATCC33852.

3.3.2.10 Iron acquisition

Iron is an essential element for bacterial survival, it has a vital role in the catalytic activity of proteins and its absence is often a limiting factor for growth (Ratledge and Dover, 2000). Competition for this environmentally scarce element has fuelled a host-pathogen, evolutionary arms race (Kieu *et al.*, 2012). At the heart of bacterial iron acquisition is the ferric uptake regulator (Fur), which represses or induces uptake systems depending on the intracellular concentration of ferrous iron (Fones and Preston, 2013). For example, Fur controls the direct uptake of iron through cell wall binding proteins such as the TonB-dependent uptake system, and directs the synthesis and release of diffusible iron chelators such as siderophores, which sequester iron from the environment (Noinaj *et al.*, 2010). The Fur gene, siderophores and other iron acquisition genes are encoded within all sequenced strains of *G. quercinecans*, *Brenneria spp.*, and *E. americana* (Table 3.12). *G. quercinecans* FRB97 was found to have more siderophores, n=26 than *B. goodwinii* FRB141, n=16 and fewer cell wall iron-capturing enzymes n=47 compared to n=60 in *B. goodwinii* FRB141, more TonB associated proteins n=8, compared to n=4 in *B. goodwinii* FRB141. Notably, *G. quercinecans* N78 had a higher number of siderophores than other sequenced *G. quercinecans* strains, (n=58, compared to n=26 & n=26), and all strains had substantially more siderophores than the putative endophyte, *E. americana* ATCC33852 (n=8). The coding of an increased number of siderophores, is the key horizontal acquired mechanism of increased virulence in a hyper-virulent variant of *Klebsiella pneumoniae*, which has recently acquired a siderophore encoding genomic island, resulting in a significant increase in morbidity and mortality of infection within healthy human hosts (Shon *et al.*, 2013).

For plant pathogens, acquiring iron from the host is an invasion strategy which enables and enhances infection (Franza *et al.*, 2002). Under conditions of iron deficiency, the plant pathogen *D. dadantii* promotes infection by releasing the siderophores, achromobactin and chrysobactin (Franza *et al.*, 2005). Within an iron-starved *Arabidopsis thaliana* host, disease progression of *D. dadantii* is severely diminished and major virulence factors were not expressed, indicating a correlation between plant iron status and infection (Kieu *et al.*, 2012). Notably, during the course of *D. dadantii* infection, iron is visibly reduced from cellular compartments and plant cell walls,

as bacterial siderophores successfully scavenge host iron (Aznar, Patrit, *et al.*, 2014). The siderophore-mediated iron acquisition process of *D. dadantii* is coincident with pectate lyase secretion, which degrades plant cell walls and releases host iron, a process which is within the *D. dadantii* RD pathway. Siderophore mediated virulence is a key factor within plant (Dellagi *et al.*, 2009) and human (Russo *et al.*, 2014) infections, however siderophores are ubiquitous in bacteria (with the notable exception of *Lactobacilli*), and some plants have been known to use bacterial siderophores symbiotically to enhance their own supply of soluble iron (Archibald, 1983; Aznar, Chen, *et al.*, 2014). Therefore, the presence of siderophores is linked to bacterial virulence but requires phenotypic or transcriptomic data for confirmation.

3.4 Conclusions

These data reveal *G. quercinecans* and *B. goodwinii* encode the virulence capability to contribute to the destruction of plant tissue within a polymicrobial infection site. The central aim of the study was to present the first genomic catalogue of putative virulence genes encoded within bacteria consistently isolated from necrotic lesions of AOD affected trees. The resultant genomic itinerary reveals putative bacterial pathogens at the molecular level and explores their potential to establish infection on susceptible oak trees. Comparative genomic analysis identified virulence genes homologous to those in closely related bacterial phytopathogens and provided many gene targets for further characterisation. Overall, the chapter provides evidence of the genome wide virulence potential of *G. quercinecans* and *B. goodwinii* recovered from necrotic oak lesions and adds to the growing body of evidence implicating *G. quercinecans* and *B. goodwinii* as contributing pathogens within AOD.

3.5 References

Abbott, D.W. and Boraston, A.B. (2008) Structural biology of pectin degradation by *Enterobacteriaceae*. *Microbiol. Mol. Biol. Rev.* **72**: 301–316.

Abby, S.S., Neron, B., Menager, H., Touchon, M., and Rocha, E.P.C. (2014) MacSyFinder: A program to mine genomes for molecular systems with an application to CRISPR-Cas systems. *PLoS One* **9**: e110726.

Abramovitch, R.B. and Martin, G.B. (2004) Strategies used by bacterial pathogens to suppress plant defenses. *Curr. Opin. Plant Biol.* **7**: 356–364.

Alfano, J.R., Charkowski, A.O., Deng, W.L., Badel, J.L., Petnicki-Ocwieja, T., van Dijk, K., and Collmer, A. (2000) The *Pseudomonas syringae* Hrp pathogenicity island has a tripartite mosaic structure composed of a cluster of type III secretion genes bounded by exchangeable effector and conserved effector loci that contribute to parasitic fitness and pathogenicity in plants. *Proc. Natl. Acad. Sci. U. S. A.* **97**: 4856–4861.

Altschul, S.F., Gish, W., Miller, W., Myers, E., and Lipman, D.J. (1990) Basic Local Alignment Search Tool. *J. Mol. Biol.* **215**: 403–410.

Alvarez-Martinez, C.E. and Christie, P.J. (2009) Biological diversity of prokaryotic type IV secretion systems. *Microbiol. Mol. Biol. Rev.* **73**: 775–808.

Andersson, R.A., Kõiv, V., Norman-Setterblad, C., and Pirhonen, M. (1999) Role of RpoS in virulence and stress tolerance of the plant pathogen *Erwinia carotovora* subsp. *carotovora*. *Microbiology* **145**: 3547–3556.

Antúnez-Lamas, M., Cabrera-Ordóñez, E., López-Solanilla, E., Raposo, R., Trelles-Salazar, O., Rodríguez-Moreno, A., and Rodríguez-Palenzuela, P. (2009) Role of motility and chemotaxis in the pathogenesis of *Dickeya dadantii* 3937 (ex *Erwinia chrysanthemi* 3937). *Microbiology* **155**: 434–42.

Archibald, F. (1983) *Lactobacillus plantarum*, an organism not requiring iron. *FEMS Microbiol. Lett.* **19**: 29–32.

Arnold, D.L. and Jackson, R.W. (2011) Bacterial genomes: evolution of pathogenicity. *Curr. Opin. Plant Biol.* **14**: 385–91.

Aznar, A., Chen, N.W.G., Rigault, M., Riache, N., Joseph, D., Desmaële, D., *et al.* (2014) Scavenging iron: a novel mechanism of plant immunity activation by microbial siderophores. *Plant Physiol.* **164**: 2167–83.

Aznar, A., Patrit, O., Berger, A., and Dellagi, A. (2014) Alterations of iron distribution in *Arabidopsis* tissues infected by *Dickeya dadantii*. *Mol. Plant Pathol.* 1–18.

Bairoch, A. (1999) The ENZYME data bank in 1999. *Nucleic Acids Res.* **27**: 310–311.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A, Dvorkin, M., Kulikov, A.S., *et al.* (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**: 455–77.

Barras, F., van Gijsegem, F., and Chatterjee, A. (1994) Extracellular enzymes and pathogenesis of soft-rot *Erwinia*. *Annu. Rev. Phytopathol.* **32**: 201–34.

Bell, K.S., Sebaihia, M., Pritchard, L., Holden, M.T.G., Hyman, L.J., Holeva, M.C., *et al.* (2004) Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora* subsp. *atroseptica* and

characterization of virulence factors. *Proc. Natl. Acad. Sci. U. S. A.* **101**: 11105–10.

Bi, D., Liu, L., Tai, C., Deng, Z., Rajakumar, K., and Ou, H.Y. (2013) SecReT4: A web-based bacterial type IV secretion system resource. *Nucleic Acids Res.* **41**: 660–665.

Biosca, E.G., González, R., López-López, M.J., Soria, S., Montón, C., Pérez-Laorga, E., and López, M.M. (2003) Isolation and Characterization of *Brenneria quercina*, Causal Agent for Bark Canker and Drippy Nut of *Quercus* spp. in Spain. *Phytopathology* **93**: 485–92.

von Bodman, S.B., Bauer, W.D., and Coplin, D.L. (2003) Quorum sensing in plant-pathogenic bacteria. *Annu. Rev. Phytopathol.* **41**: 455–482.

Bowden, S.D., Hale, N., Chung, J.C.S., Hodgkinson, J.T., Spring, D.R., and Welch, M. (2013) Surface swarming motility by *Pectobacterium atrosepticum* is a latent phenotype that requires O antigen and is regulated by quorum sensing. *Microbiology* **159**: 2375–85.

Boyer, F., Fichant, G., Berthod, J., Vandenbrouck, Y., and Attree, I. (2009) Dissecting the bacterial type VI secretion system by a genome wide *in silico* analysis: what can be learned from available microbial genomic resources? *BMC Genomics* **10**: 104.

Brady, C., Denman, S., Kirk, S., Venter, S., Rodríguez-Palenzuela, P., and Coutinho, T. (2010) Description of *Gibbsiella quercinecans* gen. nov., sp. nov., associated with Acute Oak Decline. *Syst. Appl. Microbiol.* **33**: 444–50.

Brady, C., Hunter, G., Kirk, S., Arnold, D., and Denman, S. (2014) *Rahnella victoriana* sp. nov., *Rahnella* bruchi sp. nov., *Rahnella woolbedingensis* sp. nov., classification of *Rahnella* genomospecies 2 and 3 as *Rahnella variigena* sp. nov. and *Rahnella inusitata* sp. nov., respectively and emended description of the genus. *Syst. Appl. Microbiol.* **37**: 545–552.

Broberg, M., Lee, G.W., Nykyri, J., Lee, Y.H., Pirhonen, M., and Palva, E.T. (2014) The global response regulator ExpA controls virulence gene expression through RsmA-mediated and RsmA-independent pathways in *Pectobacterium wasabiae* SCC3193. *Appl. Environ. Microbiol.* **80**: 1972–1984.

Brown, N., Jeger, M., Kirk, S., Xu, X., and Denman, S. (2016) Spatial and temporal patterns in symptom expression within eight woodlands affected by Acute Oak Decline. *For. Ecol. Manage.* **360**: 97–109.

De Buck, E., Lammertyn, E., and Anne, J. (2008) The importance of the twin-arginine translocation pathway for bacterial virulence. *Trends Microbiol.* **16**: 442–453.

Buttner, D. and Bonas, U. (2010) Regulation and secretion of *Xanthomonas* virulence factors. *FEMS Microbiol. Rev.* **34**: 107–133.

Büttner, D. and Bonas, U. (2002) Port of entry--the type III secretion translocon. *Trends Microbiol.* **10**: 186–192.

Charkowski, A., Blanco, C., Condemine, G., Expert, D., Franza, T., Hayes, C., *et al.* (2012) The role of secretion systems and small molecules in soft-rot *Enterobacteriaceae* pathogenicity. *Annu. Rev. Phytopathol.* **50**: 425–49.

Chatterjee, A., Cui, Y., Liu, Y., Dumenyo, C.K., and Chatterjee, A.K. (1995) Inactivation of *rsmA* leads to overproduction of extracellular pectinases, cellulases, and proteases in *Erwinia carotovora* subsp. *carotovora* in the absence of the starvation/cell density-sensing signal, N-(3-oxohexanoyl)-L-homoserine lactone. *Appl. Environ. Microbiol.* **61**: 1959–1967.

Cianciotto, N.P. (2005) Type II secretion: a protein secretion system for all seasons. *Trends Microbiol.* **13**: 581–8.

Coplin, D.L., Frederick, R.D., Majerczak, D.R., and Tuttle, L.D. (1992) Characterization of a gene cluster that specifies pathogenicity in *Erwinia stewartii*. *Mol. Plant-Microbe Interact.* **5**: 81–88.

Cordero-Alba, M., Bernal-Bayard, J., and Ramos-Morales, F. (2012) SrfJ, a *Salmonella* type III secretion system effector regulated by PhoP, RcsB, and IolR. *J. Bacteriol.* **194**: 4226–4236.

Costa, T.R.D., Felisberto-Rodrigues, C., Meir, A., Prevost, M.S., Redzej, A., Trokter, M., and Waksman, G. (2015) Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nat. Rev. Microbiol.* **13**: 343–359.

Costechareyre, D., Balmand, S., Condemine, G., and Rahbé, Y. (2012) *Dickeya dadantii*, a plant pathogenic bacterium producing cyt-like entomotoxins, causes septicemia in the pea aphid *Acyrthosiphon pisum*. *PLoS One* **7**: e30702.

Costechareyre, D., Dridi, B., Rahbé, Y., and Condemine, G. (2010) Cyt toxin expression reveals an inverse regulation of insect and plant virulence factors of *Dickeya dadantii*. *Environ. Microbiol.* **12**: 3290–3301.

Coulthurst, S.J. (2013) The Type VI secretion system - a widespread and versatile cell targeting system. *Res. Microbiol.* **164**: 640–54.

Cui, Y., Chatterjee, A., Hasegawa, H., Dixit, V., Leigh, N., and Chatterjee, A.K. (2005) ExpR, a LuxR homolog of *Erwinia carotovora* subsp. *carotovora*, activates transcription of *rsmA* , which specifies a global regulatory RNA-binding protein. *J. Bacteriol.* **187**: 4792–4803.

Dellagi, A., Segond, D., Rigault, M., Fagard, M., Simon, C., Saindrenan, P., and Expert, D. (2009) Microbial siderophores exert a subtle role in *Arabidopsis* during infection by manipulating the immune response and the iron status. *Plant Physiol.* **150**: 1687–1696.

Denman, S., Brady, C., Kirk, S., Cleenwerck, I., Venter, S., Coutinho, T., and De Vos, P. (2012) *Brenneria goodwinii* sp. nov., associated with acute oak decline in the UK. *Int. J. Syst. Evol. Microbiol.* **62**: 2451–2456.

Denman, S., Brown, N., Kirk, S., Jeger, M., and Webber, J. (2014) A description of the symptoms of Acute Oak Decline in Britain and a comparative review on causes of similar disorders on oak in Europe. *Forestry* **87**: 535–551.

Desvaux, M., Parham, N.J., Scott-Tucker, A., and Henderson, I.R. (2004) The general secretory pathway: A general misnomer? *Trends Microbiol.* **12**: 306–309.

Dupoiron, S., Zischek, C., Ligat, L., Carbonne, J., Boulanger, A., Duge De Bernonville, T., *et al.* (2015) The N-Glycan cluster from *Xanthomonas campestris* pv. *campestris*: A toolbox for sequential plant N-Glycan processing a toolbox for sequential plant N-Glycan processing. *J. Biol. Chem.* **290**: 6022–6036.

Duprey, A., Reverchon, S., and Nasser, W. (2014) Bacterial virulence and Fis: adapting regulatory networks to the host environment. *Trends Microbiol.* **22**: 92–9.

Ellison, D.W. and Miller, V.L. (2006) Regulation of virulence by members of the MarR/SlyA family. *Curr. Opin. Microbiol.* **9**: 153–159.

Escolar, L., Pérez-Martín, J., and De Lorenzo, V. (1999) Opening the iron box: Transcriptional metalloregulation by the fur protein. *J. Bacteriol.* **181**: 6223–6229.

Ferrandez, Y. and Condemine, G. (2008) Novel mechanism of outer membrane targeting of proteins in Gram-negative bacteria. *Mol. Microbiol.* **69**: 1349–1357.

Fones, H. and Preston, G.M. (2013) The impact of transition metals on bacterial plant disease. *FEMS*

Microbiol. Rev. **37**: 495–519.

Franza, T., Mahé, B., and Expert, D. (2005) Erwinia chrysanthemi requires a second iron transport route dependent of the siderophore achromobactin for extracellular growth and plant infection. *Mol. Microbiol.* **55**: 261–275.

Franza, T., Michaud-Soret, I., Piquerel, P., and Expert, D. (2002) Coupling of iron assimilation and pectinolysis in Erwinia chrysanthemi 3937. *Mol. Plant. Microbe. Interact.* **15**: 1181–1191.

Gibson, D.M., King, B.C., Hayes, M.L., and Bergstrom, G.C. (2011) Plant pathogens as a source of diverse enzymes for lignocellulose digestion. *Curr. Opin. Microbiol.* **14**: 264–270.

Gilbert, H.J. and Hazlewood, G.P. (1993) Bacterial cellulases and xylanases. *J. Gen. Microbiol.* **139**: 187–194.

Glazebrook, J. (2005) Contrasting Mechanisms of Defense Against Biotrophic and Necrotrophic Pathogens. *Annu. Rev. Phytopathol.* **43**: 205–227.

González, E.T., Brown, D.G., Swanson, J.K., and Allen, C. (2007) Using the Ralstonia solanacearum tat secretome to identify bacterial wilt virulence factors. *Appl. Environ. Microbiol.* **73**: 3779–3786.

Green, S., Studholme, D.J., Laue, B.E., Dorati, F., Lovell, H., Arnold, D., *et al.* (2010) Comparative genome analysis provides insights into the evolution and adaptation of *Pseudomonas syringae* pv. *aesculi* on *Aesculus hippocastanum*. *PLoS One* **5**: e10224.

Guglielmini, J., De La Cruz, F., and Rocha, E.P.C. (2013) Evolution of conjugation and type IV secretion systems. *Mol. Biol. Evol.* **30**: 315–331.

Guglielmini, J., Néron, B., Abby, S.S., Garcillán-Barcia, M.P., La Cruz, D.F., and Rocha, E.P.C. (2014) Key components of the eight classes of type IV secretion systems involved in bacterial conjugation or protein secretion. *Nucleic Acids Res.* **42**: 5715–5727.

Haque, M.M., Kabir, M.S., Aini, L.Q., Hirata, H., and Tsuyumu, S. (2009) SlyA, a MarR family transcriptional regulator, is essential for virulence in *Dickeya dadantii* 3937. *J. Bacteriol.* **191**: 5409–5418.

Harris, S.J., Shih, Y.L., Bentley, S.D., and Salmond, G.P. (1998) The hexA gene of *Erwinia carotovora* encodes a LysR homologue and regulates motility and the expression of multiple virulence determinants. *Mol. Microbiol.* **28**: 705–17.

Hassan, S., Amer, S., Mittal, C., and Sharma, R. (2012) *Ewingella americana*: An emerging true pathogen. *Case Rep. Infect. Dis.* **2012**: 1–2.

He, S.Y., Schoedel, C., Chatterjee, A.K., and Collmer, A. (1991) Extracellular secretion of pectate lyase by the *Erwinia chrysanthemi* out pathway is dependent upon Sec-mediated export across the inner membrane. *J. Bacteriol.* **173**: 4310–4317.

Hood, R.D., Singh, P., Hsu, F., Güvener, T., Carl, M. a, Trinidad, R.R.S., *et al.* (2010) A type VI secretion system of *Pseudomonas aeruginosa* targets a toxin to bacteria. *Cell Host Microbe* **7**: 25–37.

Hossain, M.M., Shibata, S., Aizawa, S.-I., and Tsuyumu, S. (2005) Motility is an important determinant for pathogenesis of *Erwinia carotovora* subsp. *carotovora*. *Physiol. Mol. Plant Pathol.* **66**: 134–143.

Hugouvieux-Cotte-Pattat, N. and Charaoui-Boukerzaza, S. (2009) Catabolism of raffinose, sucrose, and melibiose in *Erwinia chrysanthemi* 3937. *J. Bacteriol.* **191**: 6960–6967.

Hyytiäinen, H., Sjöblom, S., Palomäki, T., Tuikkala, A., and Palva, E.T. (2003) The PmrA-PmrB two-

component system responding to acidic pH and iron controls virulence in the plant pathogen *Erwinia carotovora* ssp. *carotovora*. *Mol. Microbiol.* **50**: 795–807.

Jahn, C.E., Selimi, D. a, Barak, J.D., and Charkowski, A.O. (2011) The *Dickeya dadantii* biofilm matrix consists of cellulose nanofibres, and is an emergent property dependent upon the type III secretion system and the cellulose synthesis operon. *Microbiology* **157**: 2733–44.

Jahn, C.E., Willis, D.K., and Charkowski, A.O. (2008) The flagellar sigma factor fliA is required for *Dickeya dadantii* virulence. *Mol. Plant. Microbe. Interact.* **21**: 1431–1442.

Johnson, T.L., Abendroth, J., Hol, W.G.J., and Sandkvist, M. (2006) Type II secretion: from structure to function. *FEMS Microbiol. Lett.* **255**: 175–86.

Kieu, N.P., Aznar, A., Segond, D., Rigault, M., Simond-Côte, E., Kunz, C., *et al.* (2012) Iron deficiency affects plant defence responses and confers resistance to *Dickeya dadantii* and *Botrytis cinerea*. *Mol. Plant Pathol.* **13**: 816–27.

Kofod, L. V., Kauppinen, S., Christgau, S., Andersen, L.N., Heldt-Hansen, H.P., Dorreich, K., and Dalboge, H. (1994) Cloning and characterization of two structurally and functionally divergent rhamnogalacturonases from *Aspergillus aculeatus*. *J. Biol. Chem.* **269**: 29182–29189.

Kumar, S., Jones, M., Koutsovoulos, G., Clarke, M., and Blaxter, M. (2013) Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. *Front. Genet.* **4**: 237.

Levasseur, A., Drula, E., Lombard, V., Coutinho, P.M., and Henrissat, B. (2013) Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol. Biofuels* **6**: 41.

Levy, I., Shani, Z., and Shoseyov, O. (2002) Modification of polysaccharides and plant cell wall by endo-1,4-beta-glucanase and cellulose-binding domains. *Biomol. Eng.* **19**: 17–30.

Lindeberg, M., Cunnac, S., and Collmer, A. (2012) Pseudomonas syringae type III effector repertoires: Last words in endless arguments. *Trends Microbiol.* **20**: 199–208.

Liu, H., Coulthurst, S.J., Pritchard, L., Hedley, P.E., Ravensdale, M., Humphris, S., *et al.* (2008) Quorum sensing coordinates brute force and stealth modes of infection in the plant pathogen Pectobacterium atrosepticum. *PLoS Pathog.* **4**: e10000093.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**: D490-5.

Lovell, H.C., Mansfield, J.W., Godfrey, S. a C., Jackson, R.W., Hancock, J.T., and Arnold, D.L. (2009) Bacterial evolution by genomic island transfer occurs via DNA transformation in planta. *Curr. Biol.* **19**: 1586–90.

Low, H.H., Gubellini, F., Rivera-Calzada, A., Braun, N., Connery, S., Dujeancourt, A., *et al.* (2014) Structure of a type IV secretion system. *Nature* **508**: 550–3.

Lux, R. and Shi, W. (2004) Chemotaxis-guided movements in bacteria. *Crit. Rev. Oral Biol. Med.* **15**: 207–220.

De Maayer, P., Venter, S.N., Kamber, T., Duffy, B., Coutinho, T. a, and Smits, T.H.M. (2011) Comparative genomics of the Type VI secretion systems of *Pantoea* and *Erwinia* species reveals the presence of putative effector islands that may be translocated by the VgrG and Hcp proteins. *BMC Genomics* **12**: 576.

Maes, M., Huvenne, H., and Messens, E. (2009) Brenneria salicis, the bacterium causing watermark

disease in willow, resides as an endophyte in wood. *Environ. Microbiol.* **11**: 1453–62.

Manion, P. and Lachance, D. (1992) Forest decline concepts: An overview. In, *Forest Decline Concepts*. The American Phytopathological Society, pp. 181–190.

Mansfield, J., Genin, S., Magori, S., Citovsky, V., Sriariyanum, M., Ronald, P., *et al.* (2012) Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol. Plant Pathol.* **13**: 614–29.

Manulis, S., Kobayashi, D., and Keen, N. (1988) Molecular cloning and sequencing of a pectate lyase gene from *Yersinia pseudotuberculosis*. *J. Bacteriol.* **170**: 1825–1830.

Martínez-García, P.M., Ramos, C., and Rodríguez-Palenzuela, P. (2015) T346Hunter: A novel web-based tool for the prediction of type III, type IV and type VI secretion systems in bacterial genomes. *PLoS One* **10**: e0119317.

Mhedbi-Hajri, N., Malfatti, P., Pédron, J., Gaubert, S., Reverchon, S., and Van Gijsegem, F. (2011) PecS is an important player in the regulatory network governing the coordinated expression of virulence genes during the interaction between *Dickeya dadantii* 3937 and plants. *Environ. Microbiol.* **13**: 2901–14.

Mougous, J.D., Cuff, M.E., Raunser, S., Shen, A., Zhou, M., Gifford, C.A., *et al.* (2006) A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. *Science* **312**: 1526–1530.

Mount, D.W. (2008) Choosing a method for phylogenetic prediction. *Cold Spring Harb. Protoc.* **3**:.

Murdoch, S.L., Trunk, K., English, G., Fritsch, M.J., Pourkarimi, E., and Coulthurst, S.J. (2011) The opportunistic pathogen *Serratia marcescens* utilizes type VI secretion to target bacterial competitors. *J. Bacteriol.* **193**: 6057–69.

Narzisi, G. and Mishra, B. (2011) Comparing *de novo* genome assembly: the long and short of it. *PLoS One* **6**: e19175.

Nasser, W., Reverchon, S., and Robert-Baudouy, J. (1992) Purification and functional characterization of the KdgR protein, a major repressor of pectinolysis genes of *Erwinia chrysanthemi*. *Mol. Microbiol.* **6**: 257–265.

Nghi, D.H., Bittner, B., Kellner, H., Jehmlich, N., Ullrich, R., Pecyna, M.J., *et al.* (2012) The wood rot ascomycete *Xylaria polymorpha* produces a novel GH78 glycoside hydrolase that exhibits alpha-L-rhamnosidase and feruloyl esterase activities and releases hydroxycinnamic acids from lignocelluloses. *Appl. Environ. Microbiol.* **78**: 4893–4901.

Noinaj, N., Guillier, M., Barnard, T.J., and Buchanan, S.K. (2010) TonB-dependent transporters: regulation, structure, and function. *Annu. Rev. Microbiol.* **64**: 43–60.

Nykyri, J., Niemi, O., Koskinen, P., Nokso-Koivisto, J., Pasanen, M., Broberg, M., *et al.* (2012) Revised phylogeny and novel horizontally acquired virulence determinants of the model soft rot phytopathogen *Pectobacterium wasabiae* SCC3193. *PLoS Pathog.* **8**: e1003013.

Ochman, H. and Davalos, L.M. (2006) The nature and dynamics of bacterial genomes. *Science* **311**: 1730–3.

Oh, C.-S. and Beer, S. V (2005) Molecular genetics of *Erwinia amylovora* involved in the development of fire blight. *FEMS Microbiol. Lett.* **253**: 185–92.

Pallen, M.J. and Wren, B.W. (2007) Bacterial pathogenomics. *Nature* **449**: 835–842.

Palmer, T. and Berks, B.C. (2012) The twin-arginine translocation (Tat) protein export pathway. *Nat Rev Micro* **10**: 483–496.

Peeters, N., Guidot, A., Vailleau, F., and Valls, M. (2013) *Ralstonia solanacearum*, a widespread bacterial plant pathogen in the post-genomic era. *Mol. Plant Pathol.* **14**: 651–662.

Pérombelon, M.C.M. (2002) Potato diseases caused by soft rot erwinias: An overview of pathogenesis. *Plant Pathol.* **51**: 1–12.

Pukatzki, S., Ma, A.T., Sturtevant, D., Krastins, B., Sarracino, D., Nelson, W.C., *et al.* (2006) Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the *Dictyostelium* host model system. *Proc. Natl. Acad. Sci. U. S. A.* **103**: 1528–33.

Ratledge, C. and Dover, L.G. (2000) Iron metabolism in pathogenic bacteria. *Annu. Rev. Microbiol.* **54**: 881–941.

Reverchon, S. and Nasser, W. (2013) *Dickeya* ecology, environment sensing and regulation of virulence programme. *Environ. Microbiol. Rep.* **5**: 622–636.

Rio-Alvarez, I., Rodríguez-Herva, J.J., Cuartas-Lanza, R., Toth, I., Pritchard, L., Rodríguez-Palenzuela, P., and López-Solanilla, E. (2012) Genome-wide analysis of the response of *Dickeya dadantii* 3937 to plant antimicrobial peptides. *Mol. Plant. Microbe. Interact.* **25**: 523–33.

Rodionov, D.A., Gelfand, M.S., and Hugouvieux-Cotte-Pattat, N. (2004) Comparative genomics of the KdgR regulon in *Erwinia chrysanthemi* 3937 and other gamma-proteobacteria. *Microbiology* **150**: 3571–3590.

Roper, M.C. (2011) *Pantoea stewartii* subsp. *stewartii*: Lessons learned from a xylem-dwelling pathogen of sweet corn. *Mol. Plant Pathol.* **12**: 628–637.

Russell, A.B., Peterson, S.B., and Mougous, J.D. (2014) Type VI secretion system effectors: poisons with a purpose. *Nat. Rev. Microbiol.* **12**: 137–48.

Russo, T.A., Olson, R., MacDonald, U., Metzger, D., Maltese, L.M., Drake, E.J., and Gulick, A.M. (2014) Aerobactin mediates virulence and accounts for increased siderophore production under iron-limiting conditions by hypervirulent (hypermucoviscous) *Klebsiella pneumoniae*. *Infect. Immun.* **82**: 2356–2367.

Salmond G (1994) Secretion of extracellular virulence factors by plant pathogenic bacteria. *Annu. Rev. Phytopathol.* **32**: 181–200.

Sandkvist, M. (2001) Type II Secretion and Pathogenesis. **69**: 3523–3535.

Scortichini, M., Stead, D.E., and Pia Rossi, M. (1993) Oak decline: Aerobic bacteria associated with declining Quercus cerris in Central Italy. *Eur. J. For. Pathol.* **23**: 120–127.

Segata, N., Börnigen, D., Morgan, X.C., and Huttenhower, C. (2013) PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat. Commun.* **4**: 2304.

Shalom, G., Shaw, J.G., and Thomas, M.S. (2007) *In vivo* expression technology identifies a type VI secretion system locus in *Burkholderia pseudomallei* that is induced upon invasion of macrophages. *Microbiology* **153**: 2689–99.

Shneider, M.M., Buth, S.A., Ho, B.T., Basler, M., Mekalanos, J.J., and Leiman, P.G. (2013) PAAR-repeat proteins sharpen and diversify the type VI secretion system spike. *Nature* **500**: 350–353.

Shon, A.S., Bajwa, R.P.S., and Russo, T.A. (2013) Hypervirulent (hypermucoviscous) *Klebsiella pneumoniae*. *Virulence* **4**: 107–118.

Shyntum, D.Y., Venter, S.N., Moleleki, L.N., Toth, I., and Coutinho, T.A. (2014) Comparative genomics of type VI secretion systems in strains of *Pantoea ananatis* from different environments. *BMC Genomics* **15**: 163.

da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R., Quaggio, R.B., *et al.* (2002) Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* **417**: 459–63.

Surico, G., Mugnai, L., Pastorelli, R., Giovannetti, L., and Stead, D.E. (1996) *Erwinia alni*, a new species causing bark cankers of alder (Alnus miller) species. *Int. J. Syst. Bacteriol.* **46**: 720–726.

Tampakaki, A.P., Skandalis, N., Gazi, A.D., Bastaki, M.N., Sarris, P.F., Charova, S.N., *et al.* (2010) Playing the harp: Evolution of our understanding of hrp/hrc genes. *Annu. Rev. Phytopathol.* **48**: 347–370.

Toth, I., Humphris, S., Campbell, E., and Pritchard, L. (2015) Why genomics research on *Pectobacterium* and *Dickeya* makes a difference. *Am. J. Potato Res.* 218–222.

Toth, I.K., Bell, K.S., Holeva, M.C., and Birch, P.R.J. (2003) Soft rot *erwiniae*: From genes to genomes. *Mol. Plant Pathol.* **4**: 17–30.

Toth, I.K., Pritchard, L., and Birch, P.R.J. (2006) Comparative genomics reveals what makes an enterobacterial plant pathogen. *Annu. Rev. Phytopathol.* **44**: 305–36.

Vakulskas, C. a., Potts, A.H., Babitzke, P., Ahmer, B.M.M., and Romeo, T. (2015) Regulation of bacterial virulence by Csr (Rsm) systems. *Microbiol. Mol. Biol. Rev.* **79**: 193–224.

Vinatzer, B. a, Monteil, C.L., and Clarke, C.R. (2014) Harnessing population genomics to understand how bacterial pathogens emerge, adapt to crop hosts, and disseminate. *Annu. Rev. Phytopathol.* 1–25.

Voth, D.E., Broederdorf, L.J., and Graham, J.G. (2012) Bacterial Type IV secretion systems: versatile virulence machines. *Future Microbiol.* **7**: 241–257.

Voulhoux, R., Ball, G., Ize, B., Vasil, M.L., Lazdunski, A., Wu, L.F., and Filloux, A. (2001) Involvement of the twin-arginine translocation system in protein secretion via the type II pathway. *EMBO J.* **20**: 6735–6741.

Vuts, J., Woodcock, C.M., Sumner, M.E., Caulfield, J.C., Reed, K., Inward, D.J.G., *et al.* (2015) Responses of the two-spotted oak buprestid, *Agrilus biguttatus* (Coleoptera: Buprestidae), to host tree volatiles. *Pest Manag. Sci.* **72**: 845-851.

Wandersman, C. (1989) Secretion, processing and activation of bacterial extracellular proteases. *Mol. Microbiol.* **3**: 1825–1831.

Yang, S., Peng, Q., San Francisco, M., Wang, Y., Zeng, Q., and Yang, C.-H. (2008) Type III secretion system genes of *Dickeya dadantii* 3937 are induced by plant phenolic acids. *PLoS One* **3**: e2973.

Yu, G., Smith, D., Zhu, H., Guan, Y., and Lam, T. ggtree: an R package for visualization and annotation of phylogenetic tree with different types of meta-data. *submitted*.

Zhou, J., Cheng, Y., Lv, M., Liao, L., Chen, Y., Gu, Y., *et al.* (2015) The complete genome sequence of *Dickeya zeae* EC1 reveals substantial divergence from other *Dickeya* strains and species. *BMC Genomics* **16**: 571.

CHAPTER 4

A comparison of the whole genome sequences of *Gibbsiella quercinecans* FRB97 and *Brenneria goodwinii* FRB141 using second and third generation sequencing technologies

Abstract

An organism's biology is described within its genome. Accurate genomic reconstruction using sequencing technologies is crucial to correctly understand an organism's lifestyle. Here, second and third generation sequencing methods were used to construct and compare the whole genome sequences of two bacteria, *Gibbsiella quercinecans* FRB97 and *Brenneria goodwinii* FRB141, consistently isolated from the lesions of trees affected by Acute Oak Decline (AOD), Second generation Illumina MiSeq sequencing produced draft assemblies with 68 and 185 contigs, for *G. quercinecans* FRB97 and *B. goodwinii* FRB141, whereas third generation Pacific Biosciences (PacBio) RSII sequencing produced complete, highly accurate, finished genomes. Sequencing output was substantially lower for the PacBio with 0.2 and 0.3 Mbp of data, compared to 1.7 and 1.4 Mbp of data for the Illumina MiSeq, for *G. quercinecans* FRB97 and *B. goodwinii* FRB141. However, due to the long PacBio read length and error correction tools, resultant PacBio genome assemblies provided greater fidelity and finished genomes, increasing the accuracy of biological inferences from the resultant data. Second generation sequencing produced great advances in our understanding of the biological world, however, it is hindered by erroneous sequences and short read length, these failings are corrected with the highly accurate, long read length of third generation sequencing platforms.

4.1 Introduction

Whole genome sequencing technology has improved relentlessly in terms of quantity of output and accuracy of base calling, from the publication of the first whole genome shotgun sequencing study over twenty years ago (Fleischmann *et al.*, 1995), through to the development of high-throughput sequencing more than a decade ago, to the emerging methods of single molecule, long read sequencing (Loman and Pallen, 2015). The major breakthrough for whole genome sequencing came in 1995, with the sequence of the first free living organism, *Haemophilus influenzae* using Sanger sequencing (Fleischmann *et al.*, 1995). The technological breakthrough was due to the use of shotgun sequencing for the first time, as previous sequencing efforts, such as that of the much smaller genome of bacteriophage φX174, sequenced cloned restriction enzyme fragments (Sanger *et al.*, 1977). This pioneering work, a coordinated effort by research groups across the United States, bypassed the prerequisite of genome mapping from previous sequencing efforts and randomly cloned DNA fragments, prior to sequencing. The introduction of the Roche 454 GS20 high-throughput sequencing platform in 2005, heralded the advent of next generation sequencing (NGS) (Loman *et al.*, 2012). This technological advance produced affordable, high quantity output, and greatly reduced the lab intensive, expensive methods of first generation sequencing (Kisand and Lettieri, 2013). For over a decade, NGS has fuelled the large scale application of whole genome sequencing, combining a revolutionary, in demand product with market forces, which has driven continuous technological advances and wider application of NGS tools (Loman *et al.*, 2012).

NGS sequencing methods produce a large volume of short reads containing non-random errors and typically rely on de Bruijn graph based assemblers (Pop, 2009). Resultant assemblies are usually broken into dozens to hundreds of contigs, as the short reads cannot span low complexity regions and produce fragmented draft assemblies (Chin *et al.*, 2013). Unfortunately, mis-assemblies are an integral part of NGS sequencing of which, collapsed repeats, sequence rearrangements and inversions are the classic signatures (Phillippy *et al.*, 2008). Within the last few years, novel platforms with substantially improved capacity to increase sequence length have advanced the technology available for genome assembly and ushered in third generation of whole genome sequencers. Single molecule, real time (SMRT) whole genome sequencing produced by the Pacific Biosciences RSII instrument can produce complete *de novo* reference genomes with minimal errors (>99.999% accuracy) (Chin *et al.*, 2013). There are four advantages in using the Pacific Biosciences sequencing platform: 1) errors are random, enabling a highly accurate

consensus sequence to be produced; 2) typically, PCR bias and amplification errors produced in NGS platforms are eliminated as DNA is read directly from the template; 3) GC content produces less sequencing bias; and 4) read length is greatly increased (~14,000 bp with C4 sequencing chemistry), enabling resolution of repeats (Koren *et al.*, 2013; Korlach, 2014). Improved sequence accuracy allows biological insights based on this data to be asserted with confidence. Therefore, to create highly accurate genomes of *G. quercinecans* and *B. goodwinii*, associated with AOD, the type strains of these species have been sequenced using Pacific Biosciences SMRT sequencing.

Here, an *in silico* comparison, of genome assembly and annotation is described, between the second generation Illumina MiSeq and the third generation Pacific Biosciences RSII sequencing platforms. The chapter explores the impact of the choice of sequencing technologies on the accuracy of genome assemblies and the impact on resultant gene annotations of *Gibbsiella quercinecans* FRB97 and *Brenneria goodwinii* FRB141.

4.2 Methods

4.2.1 Bacterial culture and DNA extraction

Bacterial strains sequenced by the Illumina MiSeq were cultured and processed as described in section 3.2.2.

The type strains of *G. quercinecans* FRB97 and *B. goodwinii* FRB141 were selected for Single Molecule Real-Time (SMRT) sequencing on the Pacific Biosciences RSII platform. A single colony of each strain was picked from nutrient agar (Oxoid) and used as inoculum for overnight culture in nutrient broth (Oxoid), at 28°C, and shaken at 100 rpm. Total genomic DNA was extracted using the Gentra Puregene Yeast/Bact. kit (Qiagen) and quantified using the Qubit fluorometer (Life Technologies, Paisley, UK). DNA integrity was assessed through 1% agarose gel electrophoresis. Extracted total DNA had a weight of 10 μg.

4.2.2 Genome sequencing using Illumina MiSeq

Genomic DNA sequence libraries were prepared and sequenced as described in section 3.2.2.

4.2.3 Genome sequencing on Pacific Biosciences RSII

Genomic DNA sequence libraries were prepared and sequenced by DUGSIM at Duke University, NC, USA, using 6 SMRT cells per library and a 180-minute movie.

4.2.4 Bioinformatic analysis

General bioinformatic analyses were carried out on a locally installed Bio-Linux 8 workstation (Field *et al.*, 2006) and applications requiring high computing power were undertaken on the HPC Wales supercomputing network.

4.2.5 Genome assembly of *G. quercinecans* FRB97 and *B. goodwinii* FRB141

De novo assembly was performed on an Amazon EC2 image using the hierarchical genome assembly 3 (HGAP3) workflow, incorporating the CELERA assembler. The genomes were finished using Quiver consensus polisher, giving mean confidence (QV) values of 49 for both

genomes. Genome assemblies have been deposited in NCBI with accession numbers CP014136 (*G. quercinecans* FRB97) and CP014137 (*B. goodwinii* FRB141).

4.2.6 Genome annotation

Whole genome general annotations were automatically generated using the Prokka annotation pipeline v1.11 (Seemann, 2014) and the RAST online annotation server (Aziz *et al.*, 2008). Prokka annotations were used as input to search for CAZymes, which were identified using the dbCAN online server (Yin *et al.*, 2012). Type III, IV and VI secretion systems were annotated using the T346 hunter (Martínez-García *et al.*, 2015). Whole genomes were displayed using the genome visualisation aesthetic Circos (Krzywinski *et al.*, 2009). Circos input data was generated using the file formatting tools BEDTools v2.17.0 (Quinlan and Hall, 2010) and SAMtools v0.1.19 (Li *et al.*, 2009).

4.2.7 Comparison of *G. quercinecans* FRB97 and *B. goodwinii* FRB141 sequencing results from Illumina and PacBio platforms

Draft genomes created from assemblies of llumina MiSeq sequences from *G. quercinecans* FRB97 and *B. goodwinii* FRB141 isolates, were aligned against the finished reference assemblies of the same strains from PacBio sequences using BLAST+ (v2.2.31) (Altschul *et al.*, 1990) and visualised in the genome browser Artemis (Carver *et al.*, 2012).

4.3 Results

4.3.1 Genome assembly of *G. quercinecans* FRB97 and *B. goodwinii* FRB141using PacBio SMRT sequencing

The sequencing output produced average coverage for *G. quercinecans* FRB97 and *B. goodwinii* FRB141 genomes of, 400 and 393 times, respectively (Table 4.1). The resultant assemblies were circular contigs, with N_{50} values of 5,548,506 and 5,395,301 representing the complete chromosomes of *G. quercinecans* FRB97 and *B. goodwinii* FRB141 (Table 4.1). *G. quercinecans* FRB97 and *B. goodwinii* FRB141 have GC contents of 56% and 53%, respectively (Figs. 4.1 & 4.2).

Table 4.1. Genome sequencing, assembly and annotation metrics of *G. quercinecans* FRB97 and *B. goodwinii* FRB141. Mbp = mega base pairs.

Organism	Sequencing platform	Assembler	Contigs	N50 (G+C content %)	No. of genes (Coding density %)	Chromosome size (bp)	Sequencing output (Number of reads, Million)	Genome coverage
Gibbsiella quercinecans FRB97	Pacific Biosciences RSII	HGAP3	1	5548506 (55.9)	5126 (87.2)	5548506	0.3	400 X
Gibbsiella quercinecans FRB97	Illumina MiSeq	SPAdes	68	160,581 (56.0)	5087 (86.9)	5504004	1.7	48.9 X
Brenneria goodwinii FRB141	Pacific Biosciences RSII	HGAP3	1	5395301 (53.1)	4905 (86.3)	5395301	0.2	393 X
Brenneria goodwinii FRB141	Illumina MiSeq	SPAdes	185	61,230 (51.4)	4625 (85.8)	5281917	1.4	52 X

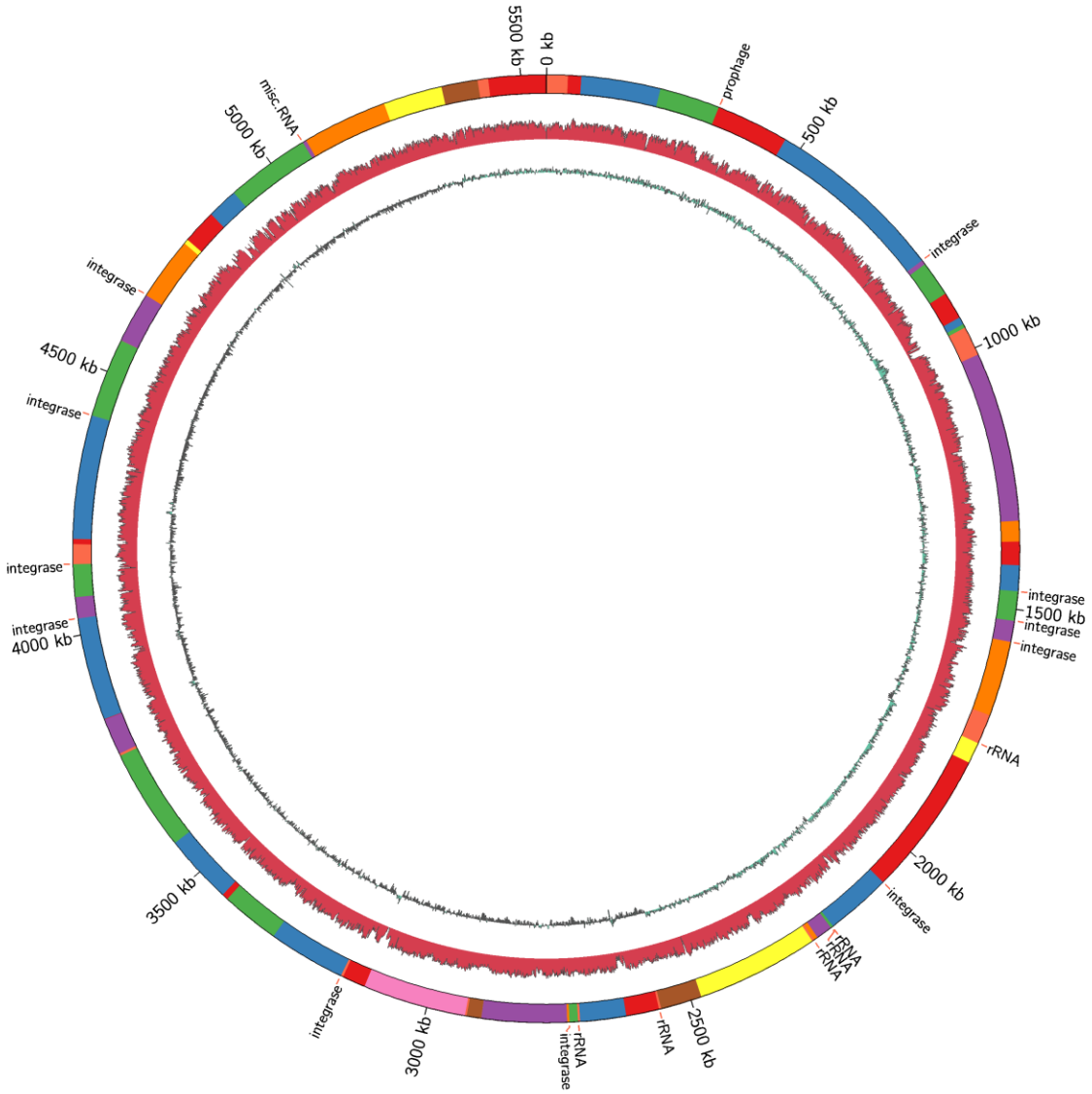


Figure 4.1. Circular representation of the *G. quercinecans* FRB97 PacBio genome assembly. The PacBio complete assembly (chromosome) is represented by the outermost circular. Contigs from the *G. quercinecans* FRB97 SPAdes draft assembly are overlain on the chromosome. Contigs are demarcated by a new colour, each colour represents a separate contig. Gaps are black. Labels signify features which force contig breaks, where annotations exist, they are found upstream, downstream and spanning contig breaks. The red middle circle represents variation in G+C content and the innermost circle represents G+C skew, across the PacBio genome assembly.

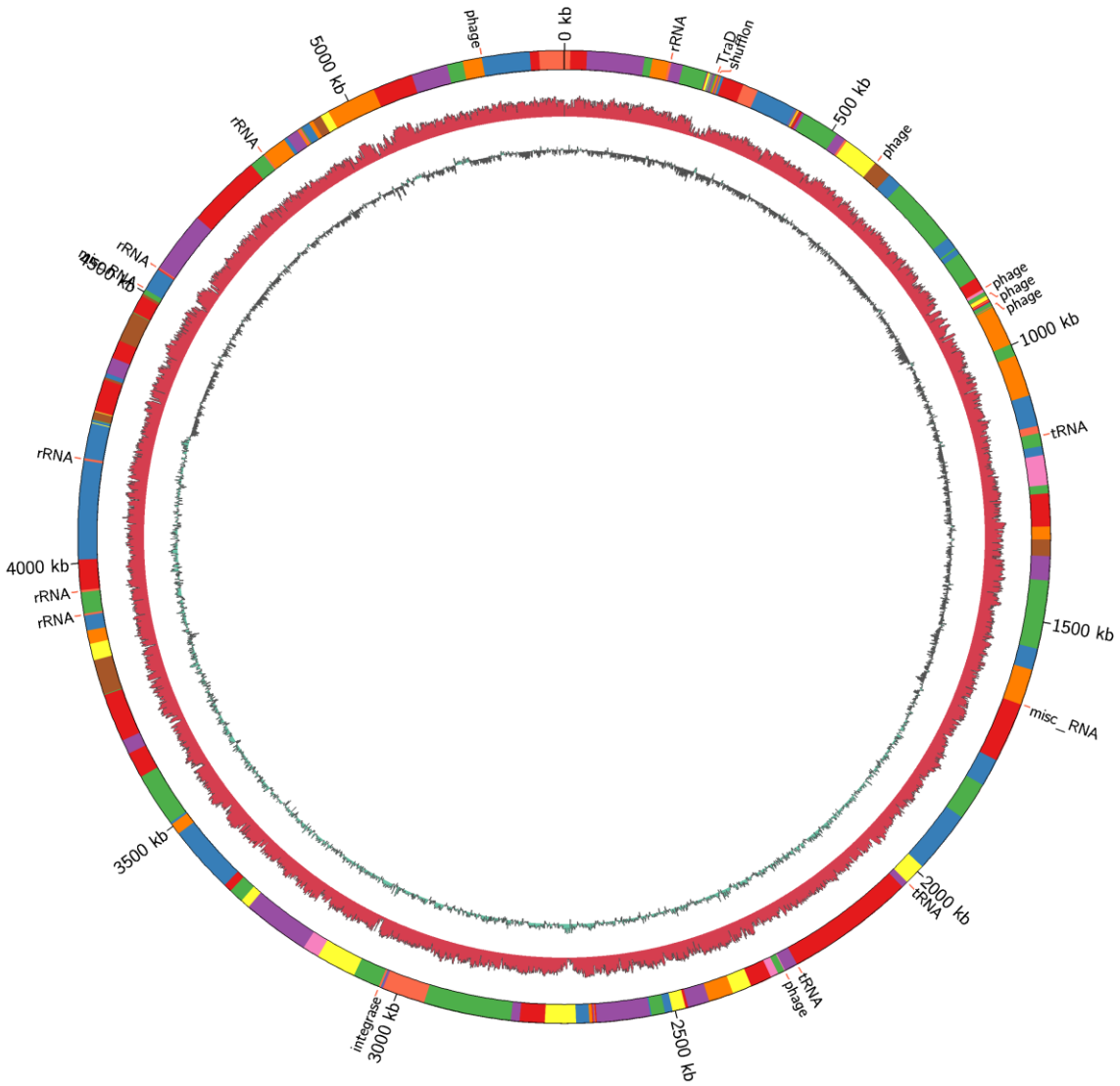
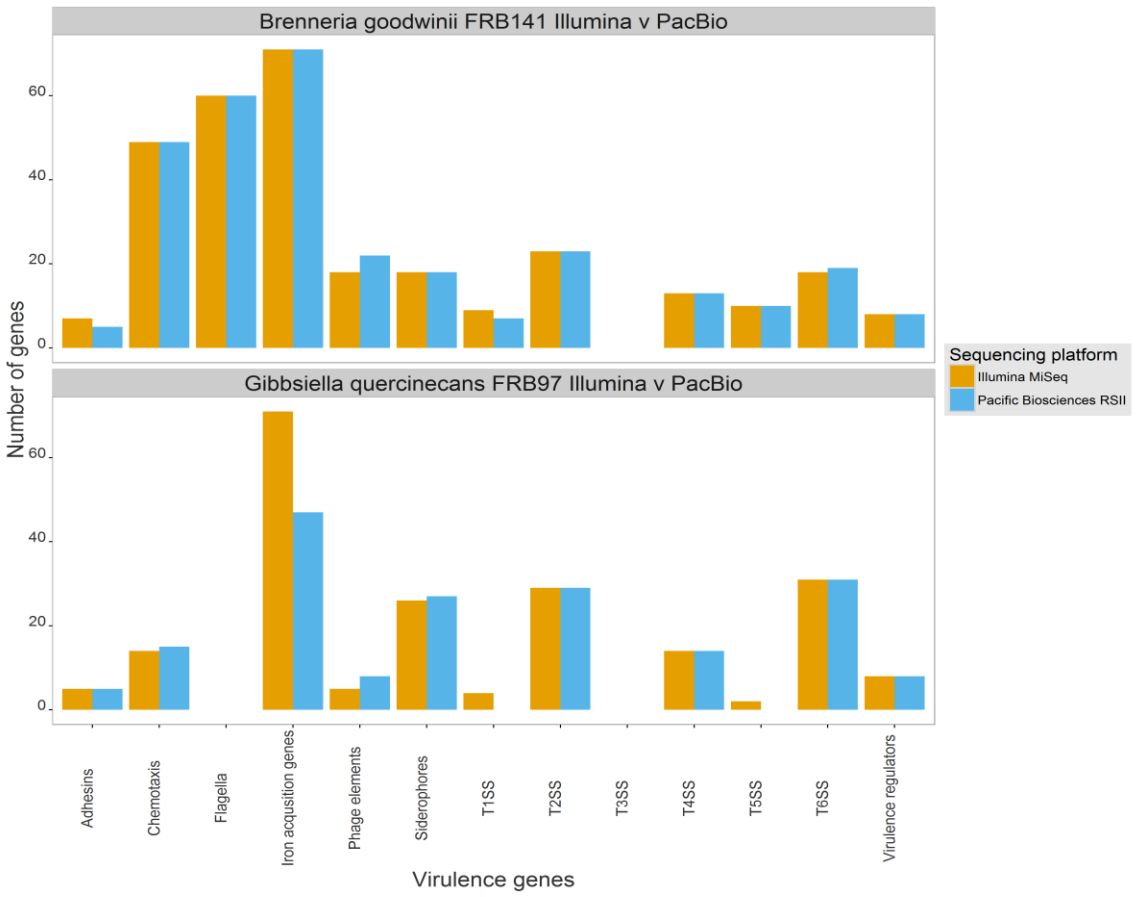


Figure 4.2. Circular representation of the *B. goodwinii* FRB141 PacBio genome assembly. The PacBio complete assembly (chromosome) is represented by the outermost circular. Contigs from the *B. goodwinii* FRB141 SPAdes draft assembly are overlain on the chromosome. Contigs are demarcated by a new colour, each colour represents a separate contig. Gaps are black. Labels signify features which force contig breaks, where annotations exist, they are found upstream, downstream and spanning contig breaks. The red middle circle represents variation in G+C content and the innermost circle represents G+C skew, across the PacBio genome assembly.

4.3.2 Genome alignment

Second generation Illumina MiSeq and third generation PacBio RSII sequencing platforms were compared using whole genome alignments. Resultant BLAST alignments allowed observations of the causative agents of contig breaks, which were revealed by overlaying reference genome, gene annotations against contig break points (Figs. 4.1 & 4.2). There were many causes of contig breaks; however, there were several common annotations at the break loci, with some featuring more commonly in the particular draft genomes, for example, contigs spanning transposases (i.e. the transposase gene was encoded at the 3' end of one contig and 5' end of the subsequent contig) produced many breaks in *G. quercinecans* FRB97 (Appendix IV), whereas there were no contig spanning transposable elements in the *B. goodwinii* FRB141 genome (Appendix V), which had several phage spanning genes across its 185 contigs. The *B. goodwinii* FRB141 draft assembly did have inherently more breaks than the Illumina *G. quercinecans* FRB97 draft assembly, and was littered with short broken genome regions, embedded with short hypothetical proteins and marked by areas of phage insertion, producing stunted contigs.

a.



b.

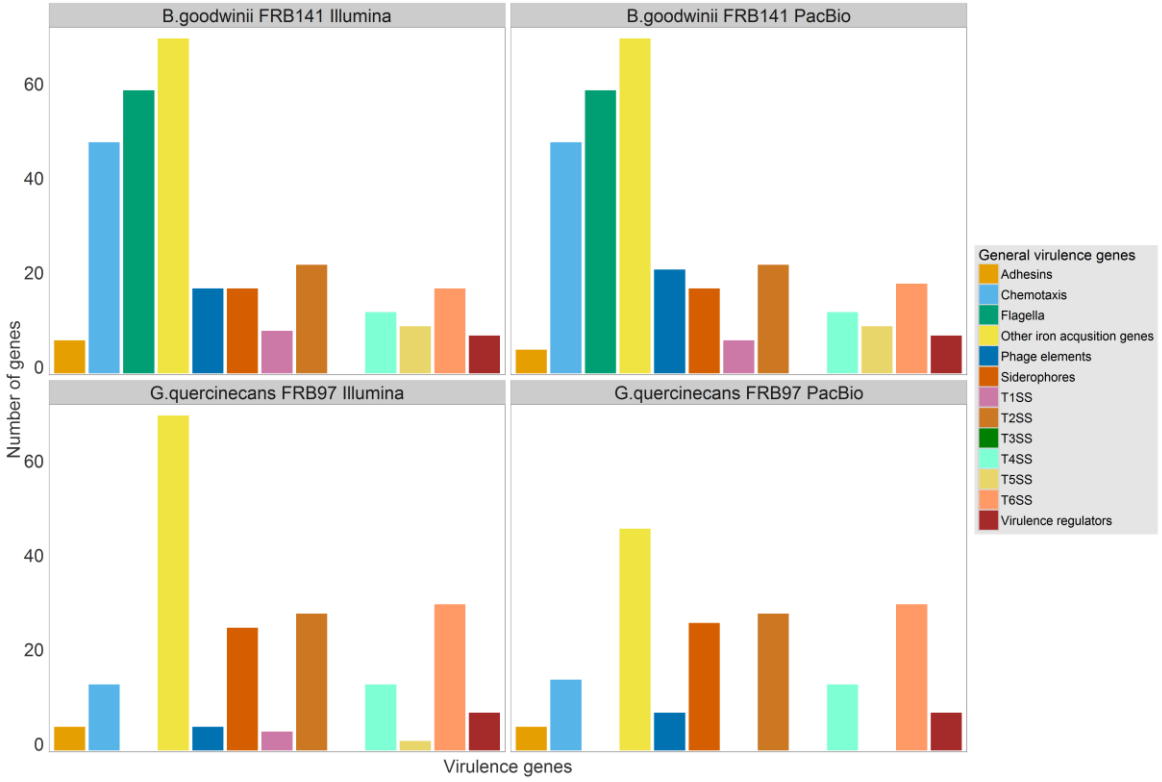


Figure 4.3. Comparison of virulence gene annotation using Illumina MiSeq and PacBio sequencing platforms. **a,** (top). Comparison of virulence genes, based on genome assemblies of the Illumina MiSeq and PacBio sequencing output. **b**, (bottom). Genome annotations of the *G. quercinecans* FRB97 and *B. goodwinii* FRB141, draft (Illumina) and finished (PacBio) assemblies.

4.4 Discussion

4.4.1 Genome assembly

Draft genome sequences of *G. quercinecans* FRB97 and *B. goodwinii* FRB141 were generated using the Illumina MiSeq NGS platform and assembled using the de Bruijn graph based assembler, SPAdes (Table 4.1). The aim of de Bruijn graph based assembly is to pair reads, through homologous alignment, at specific vertices and generate the most contiguous representation of the sequenced chromosome (Compeau *et al.*, 2011). The SPAdes assembler was unable to find a complete (Hamiltonian) path between each vertex of overlapping reads, and contiguity was lost, breaking the assemblies into 68 (*G. quercinecans* FRB97) and 185 (*B. goodwinii* FRB141) separate contigs (Pop, 2009). Discontinuous assemblies are a limitation of NGS, as unbridgeable gaps are created by low-complexity regions such as the multi-copy rRNA operon, which can range in size between 5-7 kbp (Koren *et al.*, 2013). The short read length produced by NGS platforms, such as the 250 bp reads from the Illumina MiSeq used in this chapter, is too short to traverse large, low complexity genomic regions. However, the third generation PacBio sequencing platform has the required read length to span repeat regions in prokaryotic genomes. Therefore, to create high-quality, finished reference genomes, the PacBio RSII sequencing platform was used to sequence the whole genomes of *G. quercinecans* FRB97 and *B. goodwinii* FRB141. Resultant genome assemblies produced a single contig for each species, which were error corrected to 99.99% accuracy, as part of the HGAP3 assembly process (Koren and Phillippy, 2015). The SPAdes draft genome assemblies were aligned against the reference genome using BLAST+, which revealed substantial contiguity and a broadly accurate representation of the genome (Figs. 4.1 and 4.2). The aligned contig ends were examined for classic signatures of contig break features such as the rRNA operon and mobile elements (Figs. 4.1 and 4.2). The *G. quercinecans* FRB97 draft genome was broken often by mobile elements such as transposases, which have long repetitive regions (Koren and Phillippy, 2015). Contrastingly, the *B. goodwinii* FRB141 draft assembly had few obvious causes for breaks and was characterised by unannotated hypothetical genes at contig break points (Fig. 4.2). This may be due to the lower sequence output for *B. goodwinii* FRB141 compared to *G. quercinecans* FRB97, which may have been insufficient to accurately represent certain genome regions, such as long genes. This is exemplified by a type I secreted agglutinin encoded within the *B. goodwinii* FRB141 PacBio genome, which is 24970 bp and breaks the Illumina assembly into several contigs, due to the inability of the sequencing reads to transverse its entire length (Supplementary

Table 4.2). Lower sequencing coverage, non-random errors and short read length of the Illumina MiSeq sequencing platform, resulted in assembly discontiguity. Erroneous or missing genomic regions produces a downstream continuum of annotation errors during both structural and then functional annotation, as structurally annotated coding domains are called at incorrect loci, producing falsely annotated gene function and false inferences of an organism's genomic potential.

4.4.2 Genome annotation

Draft and finished genome assemblies of *G. querciencans* FRB97 and *B. goodwinii* FRB141 were annotated using Prokka. Prokka is an automated annotation system, which can rapidly annotate finished or draft prokaryotic genomes. Ideally, Prokka annotated genomes should subsequently be manually validated using specific annotation tools as described in Chapter 2. However, the wide applicability of Prokka, covering all gene categories allows insights on the influence of assembly contiguity on annotations (Fig. 4.3a). Overall, the structural annotation reflected favourably on the *G. quercinecans* FRB97 SPAdes assembly, with only 39 coding domains missed, compared to 280 coding domains missing from the *B. goodwinii* FRB141 draft genome annotation. A possible reason for the missing genes is the disparity in the estimated genome size of the *B. goodwinii* FRB141 draft assembly, which has 113384 missing base pairs (bp), compared with *G. quercinecans* FRB97 which has only 44502 bp missing (Table 4.1), suggesting that many genes may be on the genomic regions absent from the draft assembly. Surprisingly, resultant Prokka annotations of the virulence genes revealed greater synchronicity between the *B. goodwinii* FRB141 genomes, with few differences and a larger discrepancy in the *G. quercinecans* FRB97 genomes, this is exemplified in the iron acquisition genes with twenty four more genes found in the draft genome than the finished version (Fig. 4.3b). These misannotations in the draft genome may be due to wrongly called genes during the structural annotation, where a misaligned region causes a frameshift or an uncaptured gap, and produces misalignment of a gene start site and subsequent wide-ranging incorrect functional annotation (Richardson and Watson, 2013).

4.5 Conclusions

Pacific Biosciences RS II third generation sequencing offers improved accuracy and increased data than the second generation Illumina MiSeq platform. Third generation sequencing technology provided highly accurate finished genomes for the *G. quercinecans* FRB97 and *B. goodwinii* FRB141 AOD associated bacterial genomes. Comprehensive inferences of organismal biology can be inferred with increased confidence, and future work, such as RNA sequencing can reliably align expressed transcripts against these genomes.

4.6 References

Altschul, S.F., Gish, W., Miller, W., Myers, E., and Lipman, D.J. (1990) Basic Local Alignment Search Tool. *J. Mol. Biol.* **215**: 403–410.

Aziz, R.K., Bartels, D., Best, A. A, DeJongh, M., Disz, T., Edwards, R. a, *et al.* (2008) The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**: 75.

Carver, T., Harris, S.R., Berriman, M., Parkhill, J., and McQuillan, J.A. (2012) Artemis: An integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* **28**: 464–469.

Chin, C.-S., Alexander, D.H., Marks, P., Klammer, A. A, Drake, J., Heiner, C., *et al.* (2013) Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **10**: 563–9.

Compeau, P.E.C., Pevzner, P.A., and Tesler, G. (2011) How to apply de Bruijn graphs to genome assembly. *Nat. Biotechnol.* **29**: 987–991.

Field, D., Tiwari, B., Booth, T., Houten, S., Swan, D., Bertrand, N., and Thurston, M. (2006) Open software for biologists: from famine to feast. *Nat. Biotechnol.* **24**: 801–803.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Ewen, F., Kerlavage, A.R., *et al.* (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science.* **269**: 496–512.

Kisand, V. and Lettieri, T. (2013) Genome sequencing of bacteria: sequencing, *de novo* assembly and rapid analysis using open source tools. *BMC Genomics* **14**: 211.

Koren, S., Harhay, G.P., Smith, T.P., Bono, J.L., Harhay, D.M., McVey, S.D., *et al.* (2013) Reducing assembly complexity of microbial genomes with single-molecule sequencing. *Genome Biol.* **14**: R101.

Koren, S. and Phillippy, A.M. (2015) One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. *Curr. Opin. Microbiol.* **23**: 110–120.

Korlach, J. (2014) Returning to more finished genomes. *Genomics Data* 17–19.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., *et al.* (2009) Circos: an information esthetic for comparative genomics. *Genome Res* **19**: 1639–1645.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Loman, N.J., Constantinidou, C., Chan, J.Z.M., Halachev, M., Sergeant, M., Penn, C.W., *et al.* (2012) High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat. Rev. Microbiol.* **10**: 599–606.

Loman, N.J. and Pallen, M.J. (2015) Twenty years of bacterial genome sequencing. *Nat. Rev. Microbiol.* 1–9.

Martínez-García, P.M., Ramos, C., and Rodríguez-Palenzuela, P. (2015) T346Hunter: A novel web-based tool for the prediction of type III, type IV and type VI secretion systems in bacterial genomes. *PLoS One* **10**: e0119317.

Phillippy, A.M., Schatz, M.C., and Pop, M. (2008) Genome assembly forensics: finding the elusive mis-assembly. *Genome Biol.* **9**: R55.

Pop, M. (2009) Genome assembly reborn: recent computational challenges. *Brief. Bioinform.* **10**: 354–66.

Quinlan, A.R. and Hall, I.M. (2010) BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.

Richardson, E.J. and Watson, M. (2013) The automatic annotation of bacterial genomes. *Brief. Bioinform.* **14**: 1–12.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74**: 5463–7.

Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* **30**: 2068-2069 1–2.

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012) DbCAN: A web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **40**: 445–451.

CHAPTER 5

Transcriptome analysis reveals differential gene expression of *Gibbsiella quercinecans* FRB97 and *Brenneria goodwinii* FRB141 in axenic and co-cultures containing oak tree tissue

Abstract

Comparing expression of bacterial genomes in the presence of changing substrates provides insights into regulatory alterations which precede phenotypic adjustments in contrasting environments. RNA-sequencing (RNA-seq) was applied to characterise the transcriptomes of *Gibbsiella quercinecans* FRB97 (*Gq*) and *Brenneria goodwinii* FRB141 (*Bg*), that are associated with lesion formation in the vascular tissue of trees affected by Acute Oak Decline (AOD). *Gq* and *Bg* were cultured in axenic and co-cultures containing milled oak sapwood and phloem, at early and mid-exponential growth phase. Resultant data provided a comprehensive genome-wide overview of expression in response to tree tissue, and allowed an investigation of putative virulence gene upregulation in *Gq* and *Bg,* which have been implicated as members of a pathogen complex in oak affected by AOD. Results reveal substantial upregulation of *Gq* virulence genes in phloem cultures, particularly at 2 hours post inoculation. This response indicates that *Gq* can utilise phloem components as a nutrient source. The functional shift in *Gq* in sapwood compared to control was minimal with only 8 differentially expressed genes, however, when in co-culture with *Bg* in sapwood, *Gq* differentially regulated 390 genes, including many putative virulence genes, such as GH78 and GH31. Differential gene expression of *Bg* was substantially higher in axenic cultures with sapwood, and to a lesser extent in phloem tissue, than a nutrient broth control (no oak tissue) at 2 hours post inoculation. In co-culture, *Bg* had a reduced functional differential compared to the control, indicating an altered response to tree tissue. An analysis of differentially expressed genes revealed that *Bg* has a general dampening of differential gene expression in response to tree tissue when in co-culture with *Gq*, but an examination of the differentially expressed genes revealed a putative cooperative, antagonistic role in response to oak sapwood and phloem, as virulence genes such as GH51, HrpN (a T3SS harpin effector), and the flagella apparatus were upregulated. This putative cooperation is a possible cause of necrosis in AOD affected trees, as *Gq* and *Bg* cross-regulate key phytopathogenic virulence genes. This study reveals upregulation of virulence-encoded genes in response to the presence of oak sapwood and phloem, and the possible induction of virulence in co-culture, in replication of environmental polymicrobial behaviour.

5.1 Introduction

Gibbsiella quercinecans FRB97 (*Gq*) and *Brenneria goodwinii* FRB141 (*Bg*) were isolated from necrotic lesions of trees affected by AOD (Brady *et al.*, 2010; Denman *et al.*, 2012a). Their nearest closest relatives are generalists which can switch from an asymptomatic epiphytic, endophytic or saphrophytic lifestyle to become plant pathogens (e.g. *Dickeya* spp. and *Pectobacterium spp.*) and others are opportunistic human pathogens (e.g. *Serratia* spp.). These canonical bacterial phytopathogens have the ability to transition between several environments, for example *D. solani* can transition from an endophytic or epiphytic lifestyle; on the plant surface or the intercellular spaces, to pathogenicity in response to specific environmental cues such as an increase in temperature or low oxygen availability (Potrykus *et al.*, 2014). However, it is unclear from their genome sequence alone what allows this versatility, as subtle genetic variations, such as upregulating a particular gene, can produce distinct ecological phenotypes (Joshi *et al.*, 2015). Transcriptomic sequencing using RNA-seq technologies enables fine scale analysis of regulatory changes across *Gq* and *Bg* genomes, when cultured in divergent temporal and physiological conditions, providing key insights of possible adaptation to an oak host.

RNA sequencing (RNA-seq) is the global analysis of gene expression and is currently the most informative technique for the study of intricate genome wide changes in gene expression as a response to a changing environment. Similar to microarrays, RNA-seq is a high throughput technology, but has the advantages of low background signal, low sample input, single base pair resolution technology, with no requirement for *a priori* coding domain knowledge (Wang *et al.*, 2009). Next generation (NGS) sequencing methods typically use the Illumina HiSeq platform due to its high-throughput, for transcriptome experiments. Transcriptome analyses can encapsulate physiological and temporal changes in bacterial genome regulation in response to environmental changes. A generalist bacterium can survive in multiple environments and vary its lifestyle through modification of its gene expression, transcriptomics can provide valuable insights of the switch to a pathogenic lifestyle (Brankatschk *et al.*, 2014). To understand this transition, differentially expressed genes (DEG) are identified and key genes associated with particular phenotypes can be inferred. To gain these insights, resultant data (reads) from a sequencing platform is first aligned to genes in an annotated genome and then the number of reads per gene is counted. Most read count programs are designed for eukaryotic features such as splice variants which skew the analysis (Magoc *et al.*, 2013). Additionally, unlike eukaryotes, it has been discovered from a study of 220 prokaryotic genomes that 29% of genes overlap with a segment of neighbouring genes (Kingsford *et al.*, 2007). Therefore, dedicated bacterial specific read count

programs such as EDGE-pro (Magoc *et al.*, 2013) or Rockhopper (McClure *et al.*, 2013) are required for a biologically relevant approach to read count data. Analysis of the global transcriptome can provide broad and subtle insights of how bacteria may alternate from an innocuous microbiota component to a pathogenic phase, and facilitate the destruction of seemingly indefatigable oak tree tissue.

Soft-rot bacteria are wide-ranging necrotrophic phytopathogens, which cause disease prior to colonisation. Using brute force tactics, they release degradative enzymes to break down plant cell tissue and ingest degraded nutrients (Toth and Birch, 2005). Hemibiotrophic pathogens such as *Pseudomonas syringae* are host-specific, stealth pathogens which typically inject effectors to manipulate and colonise the host, living relatively harmoniously before causing serious latent infection (Lindeberg *et al.*, 2012).

Oak trees are sizeable organisms with long lifetimes and require a robust morphology to withstand pressure from bacterial pathogens (Eyles *et al.*, 2010). They are tough lignocellulose-based organisms comprising mostly polysaccharides (cellulose, hemicellulose, and lignin) and small quantities of structural proteins (de Lima *et al.*, 2016). Cellulose is composed of long unbranched glucose homopolymers joined by hydrogen bonds which are formed into microfibrils and make up over half of wood weight (Klemm *et al.*, 2005). Hemicellulose is a branched heteropolymer composed of xylans, with various concentrations of pentose and hexose including: xylose, arabinose, and glucuronic acid. Lignin is a phenolic heteropolymer consisting of three phenylpropanoid units, which are entwined around the microfibrils and provide the firm structure of wood (Scheller and Ulvskov, 2010). Pectin is a complex polysaccharide which is found in highest abundance in the middle lamella between cells and reduces in concentration through the cell wall to the plasma membrane (Sriamornsak, 2003). Through the adhesive properties of pectin, cells are embedded alongside cellulose microfibrils and hemicellulose, providing the tree with structural stability. Pectin consists of homogalacturans, xylogalacturan, and rhamnogalacturan I and II, with a backbone of D-galacturonic acid (GalA) units, crosslinked by α-(-4) glycosidic linkages and strengthened by rhamnose, arabinose, galactose and xylose side chains. Consequently, soft-rot bacterial phytopathogens must produce a repertoire of plant cell wall degrading enzymes (PCWDEs) which target the many isoforms of pectin (Barras *et al.*, 1994), cellulose, hemicellulose and minor components such as proteins, macerating the pecto-cellulosic matrix of plant tissue and causing cell death through osmotic shock (Costechareyre *et al.*, 2012).

This chapter explores transcriptomic gene expression of axenic and co-cultures of *Gq* and *Bg*, in media containing milled oak sapwood and phloem. *Gq* and *Bg* are hypothesised to be necrotic pathogens of oak, therefore the aim of the study is to understand regulatory changes in the bacterial genome which may enable virulence against oak tree tissue and gain further insights of the role of these bacteria in AOD.

5.2 Materials and methods

5.2.1 Strains, growth medium and conditions

Environmental strains of *Gibbsiella quercinecans* FRB97 and *Brenneria goodwinii* FRB141 were obtained by Forest Research from AOD affected trees. Isolates were maintained on nutrient agar (Oxoid) at room temperature. To simulate growth on sapwood and phloem, cells were cultured in nutrient broth (Oxoid) containing 1% (w/v) milled sapwood (NBS), nutrient broth with 1% (w/v) milled phloem (NBP) and a control consisting of nutrient broth only (NB). Cell cultures were incubated for 24 Hours Post Inoculation (HPI). This 24-hour culture represents lag to stationary phase of the growth curve (see growth curve; Supplementary Fig. 5.1). Initially, a 10 ml starter culture from a single colony was incubated to stationary phase at 28°C on a shaking incubator at 100 rpm. A 1% inoculum was taken from a cell starter culture, incubated overnight, at 28°C, centrifuged at 10,000 rpm in a 1.5 ml Eppendorf tube, resuspended and added to three replicate culture flasks containing 150 ml volumes of NB, NBS, and NBP (Fig. 5.1). The flasks were incubated at 28°C and 105 rpm, for 24 h, with cell suspensions collected 2, 6, 12, and 24 HPI. At each time point 25 ml of liquid was collected in a 50 ml Falcon tube and centrifuged for 5 mins at 3000 rpm. The supernatant was discarded and pelleted cells were frozen in liquid nitrogen.

Figure 5.1. Culture media for RNA-seq test environments. Experimental design for RNA-seq analysis of *Gq* and *Bg* in axenic and co-culture with milled oak tissue. Culture media of Nutrient Broth (NB) + phloem (left), NB + sapwood (centre), and NB (right). A total of 27 culture flasks were used, with 3 independent replicates of the three treatments (1) *Bg*, (2) *Gq*, and (3) co-cultures of *Bg* + *Gq* (n=9) for each of the three medium types.

5.2.2 RNA extraction from bacterial cultures

Total RNA was extracted from pelleted cells using the RNeasy Mini Kit (Qiagen), according to manufacturer's instructions. Genomic DNA was removed from extracted RNA samples using a TURBO DNA-free DNase kit (Ambion).

5.2.3 qPCR detection of residual DNA

Residual DNA quantities were measured using general bacterial 16S rRNA gene primers (Table 5.1) Duplicate qPCR assays were performed on the 7900HT Fast Real-Time PCR System (Applied Biosystems). Each reaction was performed in a 20 µl final volume, containing 60 ng DNA, 10 µl of 2x QuantiFast SYBR Green PCR Master Mix (Qiagen), 1 mM (final concentration) forward and reverse primer and dd H_2O. Cycling conditions were 95°C for 5 min, followed by 45 cycles of 95°C for 10 seconds, and 60°C for 30 seconds, with fluorescence detection in the combined annealing and extension step. DNA amplification levels were recorded to measure the presence of residual DNA contaminants.

Table 5.1. qPCR primers and probes used for quality control (pA/pH) and testing of differential virulence gene expression (remaining primers/probes). Oligonucleotide dual labelled probes were designed for the TssD (Hcp) gene of *Gq* and the FliA gene of *Bg*. Whole gene (WG) primers were designed for use as controls. The remaining primer assays used SYBR Green as a dye. All primers and probes were custom designed, with the exception of pA and pH, designed by (Edwards *et al.*, 1989).

Primers/Probe	Sequence (5'-3')	Label (5'-3')	Specificity	Melting temperature (°C)	Product size (bp)
pA	AGAGTTTGATCCTGGCTCAG	-	General bacterial	55	~1534
pH	AAGGAGGTGATCCAGCCGCA	-			
HcpF	TCCTGAAAGTCGACGGTGTGTCC	-	*Gq*	57	~293
HcpR	TGTACAGTGGTCACCAGCAC	-			
HcpTaq	AGATTCCAATCACACGGGCTGGACCG	Fam – BHQ1		67	-
HcpWGF	ATGGCTATTGATATGTTCCTGAAG	-		55	~483
HcpWGR	AATCCGAAGCGCCAAAAGAAACTAC	-			
FliAF	CCATGAAGCATTGCGGCTAC	-	*Bg*	59	~146
FliAR	GCTGAACGGCATAGGTGGTA	-			
FliATaq	TCGTCTTCCTGCGAGCGTTGA	Hex – BHQ3		68	-
FliAWGF	GCTATGTCCCTTTAGTACG	-		55	~722
FliAWGR	CATCAGTTGATGCAATGGATTGGC	-			

SecYF	TTACTTTCATCTGCCTGATCCC	-	*Gq* & *Bg*	62	~131
SecYR	CATCAGAGTTTGCACTTGAGC	-			
SecYWGF	GTGCTAAAGGCGGASTYG	-		*55*	~1093
SecYWGR	GTAATGACSCGYYTWACCYTG	-			

5.2.4 Quality control of RNA extracted from *Gibbsiella quercinecans* FRB97 *and Brenneria goodwinii* FRB141

DNase treated total RNA was quality controlled using 1% agarose gel electrophoresis and visualised under UV light. Samples were tested for contaminants using a NanoDrop 3300 spectrophotometer (Thermo Scientific) and RNA concentration was measured using a Qubit fluorometer hsRNA kit (Life Technologies) (Table 5.2).

Table 5.2. Total and depleted RNA extractions of *Gibbsiella quercinecans* FRB97 and *Brenneria goodwinii* FRB141 in nutrient broth, nutrient broth and 1% sapwood, and nutrient broth and 1% phloem. HPI = hours post inoculation. *Gq* = *G. quercinecans* FRB97, *Bg* = *B. goodwinii* FRB141. The depletion efficiency measuring total RNA before and after depletion was measured, and represented here as mRNA recovery. mRNA recovery (%) is expected at around 1% of the depleted total RNA concentration, concentations above this threshold indicate post-depletion overspill of total RNA.

	Organism	Medium	HPI	Total RNA volume (µl)	Depleted RNA concentration (ng/µl)	mRNA Recovery (%)
Sample 1	*Gq*	Broth	2	750	8.46	1.1
Sample 2	*Gq*	Sapwood	2	750	9.12	1.2
Sample 3	*Gq*	Phloem	2	750	12.8	1.7
Sample 4	*Gq*	Broth	6	750	14.5	1.9
Sample 5	*Gq*	Sapwood	6	750	24.8	3.3
Sample 6	*Gq*	Phloem	6	750	19.8	2.6
Sample 7	*Bg*	Broth	2	750	4.4	0.6
Sample 8	*Bg*	Sapwood	2	750	5.72	0.8
Sample 9	*Bg*	Phloem	2	750	15.1	2.0
Sample 10	*Bg*	Broth	6	750	8.28	1.1
Sample 11	*Bg*	Sapwood	6	750	11.3	1.5
Sample 12	*Bg*	Phloem	6	750	22.2	3.0
Sample 13	*Gq* & *Bg*	Broth	2	750	6.08	0.8
Sample 14	*Gq* & *Bg*	Sapwood	2	750	19.1	2.5
Sample 15	*Gq* & *Bg*	Phloem	2	750	22.4	3.0
Sample 16	*Gq* & *Bg*	Broth	6	750	16	2.1
Sample 17	*Gq* & *Bg*	Sapwood	6	750	34	4.5
Sample 18	*Gq* & *Bg*	Phloem	6	750	32.8	4.4

5.2.5 Quantification of virulence gene expression via qPCR

Three virulence genes were tested for differential gene expression across four time points in a qPCR assay. The assay was used to select the two most appropriate time points to characterise the gene expression profiles of *Gq* and *Bg* during its growth cycle by RNA-seq transcriptome sequencing. Identification of the time points, with the highest transcriptional activity were measured using qPCR gene expression assays. The assays measured transcript expression levels, which were considered as analogous to expression across the genome. The secreted type VI gene TssD (Hcp) (N.B. the TssA-TssM nomenclature was suggested to unite the T6SS naming structure (Shalom *et al.*, 2007) and will be used here, with important common names in brackets) was selected to represent metabolic activity in *Gq*, the virulence regulating sigma factor FliA (Jahn *et al.*, 2008) was selected from *Bg* and the inner membrane translocation regulator SecY was selected in both genomes.

Oligonucleotide primers and hydrolysis probes for TaqMan assays, were custom designed from the annotated genes for Hcp and FliA (Table 5.1). Duplicate qPCR assays were performed on the 7900HT Fast Real-Time PCR System (Applied Biosystems). Each reaction had a 20 µl total volume, with 25 ng RNA, 10 µl of Precision One- Step qRT-PCR Mastermix (PrimerDesign, Southampton, UK), 6 pmols forward and reverse primer, 3 pmols probe and RNase/DNase free H_2O. Cycling conditions were; reverse transcription at 55°C for 10 mins, 95°C for 8 mins, followed by 40 cycles of 95°C for 10 seconds.

A consensus sequence of the SecY homologs from *Bg* and *Gq*, was used to design qPCR primers (Table 5.1). Each reaction had a 20 µl total volume, with 60 ng DNA, 10 µl of 2x QuantiFast SYBR Green PCR Master Mix (Qiagen), 1mM (final concentration) forward and reverse primer and RNase/DNase H_2O Cycling conditions were; 95°C for 5 min, followed by 45 cycles of 95°C for 10 seconds, and 60°C for 30 seconds. A melt curve was included at the end of each SYBR Green SecY qPCR assay to confirm the presence of a single amplicon.

Whole gene primers designed from each of the virulence genes were used to create qPCR standard dilution curves which measured virulence gene expression in *Gq* and *Bg* (Table 5.1). The whole genes of *tssD*, *fliA*, and *secY* were amplified using end-point PCR, with specific custom designed oligonucleotide primers primer sets, HcpWG, FliAWG, and SecYWG (Table 5.1). Each PCR reaction had a 50 µl total volume, including a colony of each *Gq* (Hcp and SecY) or *Bg* (FliA and SecY), 1 X MyTaq Red Mix (Bioline), 50 μmol^{-1} each forward and reverse primer. The solution was made up to its final volume of 50 µl with molecular grade double-distilled water.

The thermal cycle consisted of initial denaturation at 95°C for 60 s followed by 25 cycles of denaturation at 95°C for 15 s, primer annealing at 60°C for 15 s and elongation at 72°C for 10 s. Amplicons were excised from a 1% agarose (Bioline) gel and purified using the QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's instructions. Purified DNA concentration was measured using the Qubit dsDNA BR Assay Kit (Life Technologies). The DNA concentration (molecules/µl) added to qPCR standard dilution curves was calculated using equation (1):

$$\left(\frac{X}{(b \times 660)}\right) \times (6.022 \times 10^{23}) \quad (1)$$

Where X is the concentration of DNA in g/µl and b is the length of the PCR product in base pairs.

Each reaction was compared against a standard curve ranging from 3 x 10^8 to 3 x 10^1 gene copies, which revealed relative transcript abundance for each virulence gene.

5.2.6 RNA depletion

Total RNA was pooled from three biological replicates in equimolar quantities giving a total quantity of 750 ng. Total rRNA was depleted to enrich mRNA (transcripts) using the RiboZero rRNA depletion kit (Illumina). The protocol was performed according to manufacturer's instructions. Post-depletion mRNA concentrations were measured using a Qubit flurometer (Invitrogen) (Table 5.2). Remaining traces of rRNA was measured by the Centre for Genomic Research (CGR) (University of Liverpool, UK), using the Agilent 2100 BioAnalyzer.

5.2.7 Transcriptome profiling using the Illumina HiSeq RNA sequencing platform

Library preparation, transcriptomic sequencing, and post-sequencing QC of 18 depleted RNA samples was performed by Centre for Genomic Research (CGR), University of Liverpool, UK. One lane of the Illumina HiSeq 2500 sequencing platform with v4 chemistry reagents was used to sequence the cDNA libraries.

5.2.8 RNA-seq QC

Illumina adapter sequences were removed from raw FastQ files containing the sequencing reads using Cutadapt v1.2.1, using the option –O 3, which specifies that at least 3 base pairs have to match the adapter sequences before they were trimmed. Sequences were quality trimmed using Sickle v1.2 with a minimum quality score of 20. Reads shorter than 10 bp were removed. RNA-seq QC was performed by Centre for Genomic Research (CGR), University of Liverpool, UK.

5.2.9 Transcript counts and statistical analysis

Raw quality trimmed reads were aligned against Prokka (Seemann, 2014) annotated *Gq* and *Bg* coding domains using the bacteria specific, gene expression estimation tool EDGE-pro v1.3.1 (Magoc *et al.*, 2013). Output files with normalised values in reads per kilobase of gene per million reads mapped (RPKM) were reformatted for input to the differential expression estimation program GFOLD v1.1.4 (Feng *et al.*, 2012). The GFOLD algorithm calculates differential gene expression (DEG) between single replicate RNA-seq data, using the posterior estimate of raw fold change between samples, producing a value known as the GFOLD which is analogous to fold change in multi-replicate studies. Results were plotted using the R package ggplot2.

5.3 Results and Discussion

5.3.1 qPCR analysis of *Gibbsiella quercinecans* FRB97 and *Brenneria goodwinii* FRB141 virulence marker gene expression in oak tissue over 24 hours

To understand physiological and temporal changes in gene expression in *Gq* and *Bg*, qPCR assays were designed to measure gene expression at four time points (2, 6, 12, and 24 hours post inoculation), representing lag to stationary phase of growth (Supplementary Fig. 5.1) in axenic and co-cultures, in NBS and NBP. The aim of this experiment was to identify the most appropriate time points for transcriptome analysis using RNA-seq. Representative genes were selected due to their known contribution to virulence in closely related pathogens. TssD (Hcp), along with TssI (VgrG) forms part of the T6SS injectosome and can also act as an effector on eukaryotic hosts (Coulthurst, 2013). *tssD* is secreted constitutively in the closely related nosocomial pathogen *S. marcescens* Db10, where it enables the bacteria to persist in polymicrobial environments (Murdoch *et al.*, 2011). Environmental persistence in the polymicrobial community is enabled via the secretion of antibacterial effectors, which travel through the *tssD* and *tssI* outer membrane spanning channel and into target cells. There were four TssD genes found within the *Gq* genome and 2-3 T6SS operons. A TssD gene was chosen for qPCR expression analysis here, as it was located within the most complete T6SS operon.

The gene selected to represent expression in *Bg* is the alternative sigma factor FliA, which controls flagella filament synthesis, chemotaxis machinery, and motor switch complex genes in the model organism *E. coli*, and participates in the surface swarming motility of *P. atrosepticum*. *fliA* is the third regulator in the flagella transcription cascade and is required for functioning of the flagellar apparatus, which is part of the regulon of the master virulence controller $FlhD_4C_2$ (Bowden *et al.*, 2013). Notably, *flhD*$_4$*C*$_2$binds to the virulence repressor *pecT*, thereby allowing expression of virulence regulators including *gacA* and *rsmB*, which control induction of PCWDE expression in *P. carotovorum*.

The final representative gene encoded within *Gq* and *Bg* - SecY is the central component of the SecYEG translocon, an inner membrane transport system which is instrumental to soft-rot disease in closely related bacterial phytopathogens (Korotkov *et al.*, 2012). *secY* is part of the inner membrane transport system which forms a two-part translocation structure in combination with the T2SS system to secrete PCWDEs from the cytoplasm to the extracellular environment (He *et al.*, 1991).

qPCR assays revealed that gene expression was highest at 6 HPI for *tssD* (an average of 226123 absolute transcript copies at 2 HPI, 3499143 at 6 HPI, 274795 at 12 HPI, and 40716 at 24 HPI), 2 HPI for *fliA* (an average of 55266 absolute transcript copies at 2 HPI, 8661 at 6 HPI, and 2631 at 12 HPI), and high expression at 2 and 6 HPI for *secY* (an average of 2043390 absolute transcript copies at 2 HPI and 504169 at 6 HPI) (Fig. 5.2). These data reveal that 2 HPI and 6 HPI produced the highest gene expression within the three genes, a result which was taken as analogous of expression across the genome with particular inference of virulence gene expression. Therefore, 2 HPI and 6 HPI were selected for RNA-seq expression profiling of *Gq* and *Bg* on oak tissue.

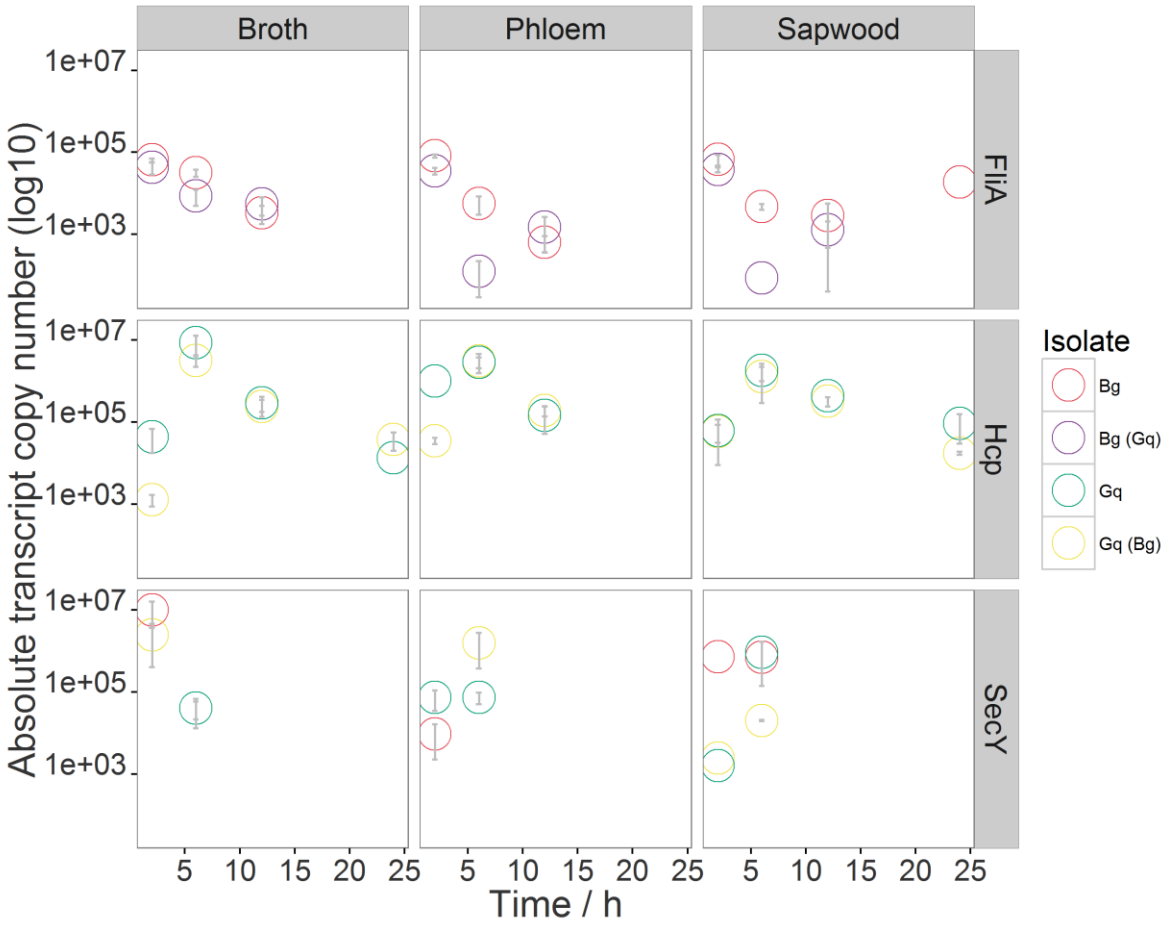


Figure 5.2. qPCR expression analysis of 3 virulence marker genes in *Gibbsiella quercinecans* FRB97 and *Brenneria goodwinii* FRB141. Top: gene expression profile for the *Bg* gene FliA cultured in NB (broth), NBP (phloem), and NBS (sapwood) at 2, 6, 12, 24 HPI. Middle: gene expression profile for the *Gq* gene TssD (Hcp) under the same growth conditions and time points. Bottom: gene expression profile for SecY in both *Gq* & *Bg* under the same growth conditions and time points. Red = *Bg*, Green = *Bg* co-cultured with *Gq*, Blue = *Gq*, Purple = *Gq* co-cultured with *Bg*. Parentheses indicate when co-culture species were also present in the culture flask, but their expression was not being measured in that particular reaction.

5.3.2 Total RNA depletion and transcriptome sequencing

Total RNA was extracted from cultures at 2 HPI and 6 HPI (Fig. 5.3). To maximise genome-wide expression data from the RNA sequencing, total RNA was depleted to remove rRNA and enrich mRNA. Within extracted total RNA, rRNA is expected to represent over 90% of the total molecules, The depletion process has varying efficiency, dependent on the kit used and the biology of the bacterial species, the quantity of rRNA can be sustantially reduced to leave a high proportion of mRNA in the depleted sample (O'Neil *et al.*, 2013).

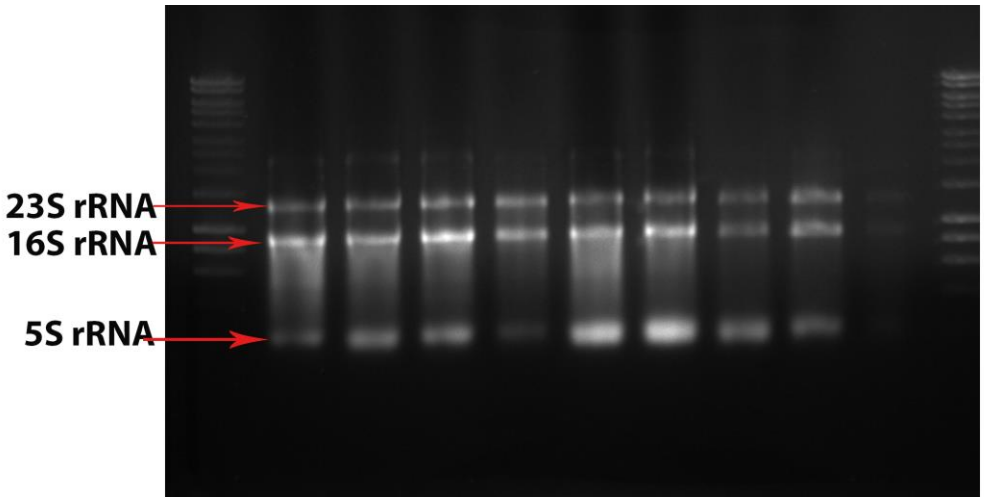


Figure 5.3. Total RNA extracted from *Gibbsiella quercinecans* FRB97 and *Brenneria goodwinii* FRB141 cells at 6 hours post inoculation. Lanes 1 and 11 contain a molecular size ladder (HyperLadder I, Bioline), lanes 2-5 contain RNA extracted from *G. quercinecans* FRB97, lanes 6-10 contain RNA extracted from *B. goodwinii* FRB141. rRNA subunits are highlighted; the smear between the extracted subunits is mRNA.

Resultant depleted RNA was used as input for transcriptome sequencing. Sequencing output provided extensive genome coverage, as between 20 and 35 million reads were obtained per sample (Fig. 5.4) and this was in excess of the estimated 5-10 million reads required for sufficient genome-wide coverage of a typical bacterial transcriptome (Haas *et al.*, 2012).

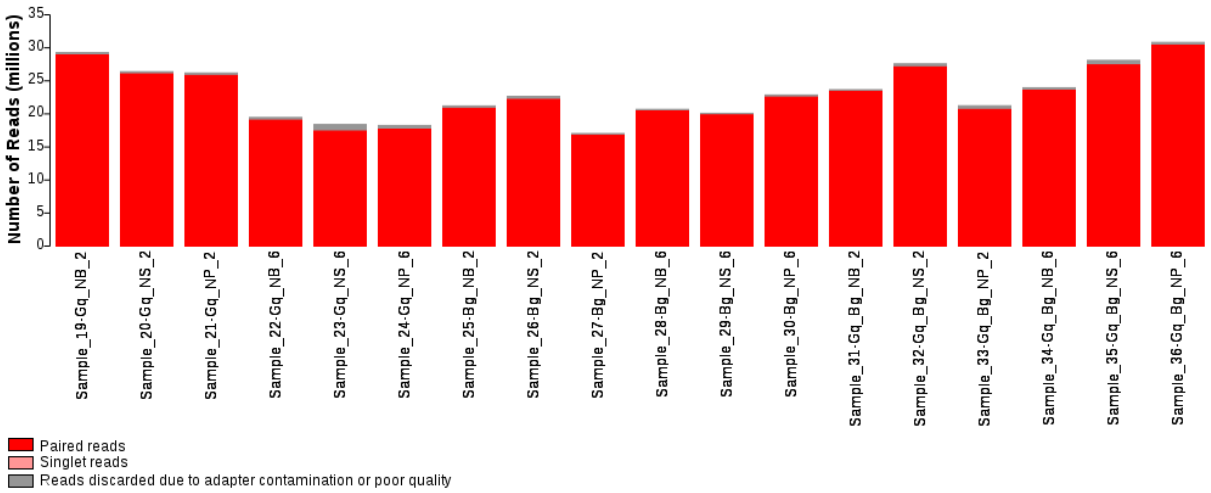


Figure 5.4. RNA-seq transcript output. Raw paired end read counts (scale in millions) from: six *Gq* libraries (lanes 1-6), six *Bg* libraries (lanes 7-12), and six combined *Gq* and *Bg* libraries (lanes 13-18). Red = paired reads; Pink = singlet reads; Grey = low quality reads. Figure provided by Centre for Genomic Research (CGR), University of Liverpool, UK.

5.3.3 A comparison of RNA-seq expression profiles from axenic and co-cultured *Gibbsiella quercinecans* FRB97 and *Brenneria goodwinii* FRB141 at 2 HPI and 6 HPI with media containing oak sapwood and phloem

The gene expression profiles of *Gq* and *Bg* were measured in axenic conditions and in co-culture, and pairwise comparisons of differential gene expression were made between NB and NBS, NB and NBP, NBS and NBP at 2 HPI and 6 HPI. There was a total of twelve test environments for each bacterium. A GFOLD comparison of differential expression revealed genes which had GFOLD values less than -1.5 and greater than 1.5, these genes were designated as being significantly differentially expressed (Feng *et al.*, 2012). This chapter compares gene expression of *Gq* and *Bg*, cultured within oak tissue containing medium, therefore the metabolic shifts described here were induced in response to oak tissue.

5.3.3.1 RNA-seq analysis of qPCR virulence marker genes

RNA-seq data revealed high gene expression of *fliA* in axenic *Bg* culture at 2 HPI, and differential upregulation in co-culture with *Gq*, in NBS and NBP cultures at 2 HPI only, with gene expression being suppressed with the addition of *Gq* in NB. *tssD* was highly expressed at 6 HPI, concurring with the qPCR data (Fig. 5.2), and was differentially upregulated at 2

HPI in NBS and NBP compared to NB. TssD is part of the T6SS injectisome and can have anti-bacterial functionality (Murdoch *et al.*, 2011). Within the *Gq* transcriptome *tssD* was upregulated in NBS and NBP, suggesting that it is part of a wider virulence transcription cascade, and may respond to eukaryotic stimuli. *secY* was highly expressed at 2 HPI in NB and NBS, but not NBP in axenic and co-cultures cultures of *Gq* and *Bg*. Furthermore, *secY* was differentially upregulated only in axenic *Gq* culture in NBS culture at 6 HPI, suggesting that *secY* is constitutively expressed at basal levels but responds to sapwood tissue at a later time point. This data may reflect the increased role of *secY* as the bacteria progress into stationary growth phase, where secondary metabolite transport is increased and *secY*, as a key secretor and translocator of bacterial proteins has an important role (Mori *et al.*, 2010). Transcriptomic expression data of *tssD*, *fliA*, and *secY* data broadly correlates with the qPCR data, however, small variations may be explained by the high sensitivity of qPCR (Geiss *et al.*, 2008; Wang *et al.*, 2009).

5.3.4 Differentially expressed genes in *Gibbsiella quercinecans* FRB97

5.3.4.1 Overview of *Gibbsiella quercinecans* FRB97 differential gene expression in all test environments

Differential gene expression of axenic *Gq* was high in NBP but limited in NBS. Co-cultures of *Gq* and *Bg* substantially increased differential gene expression of *Gq* in NBS. Generally, *Gq* had substantially higher differential gene expression at 2 HPI than 6 HPI. In NBS compared to NB at 2 HPI, *Gq* had only eight differentially expressed genes (DEG). The identification of only eight DEGs indicates that compared to other environments, there was minimal functional adjustment at 2 HPI (Fig. 5.5). However, the same conditions of NB compared to NBS at 6 HPI revealed a substantially increased number of DEG. The functional adjustment of *Gq* to sapwood in all test environments is minimal, compared to the substantially higher functional adjustment of *Gq* in response to phloem.

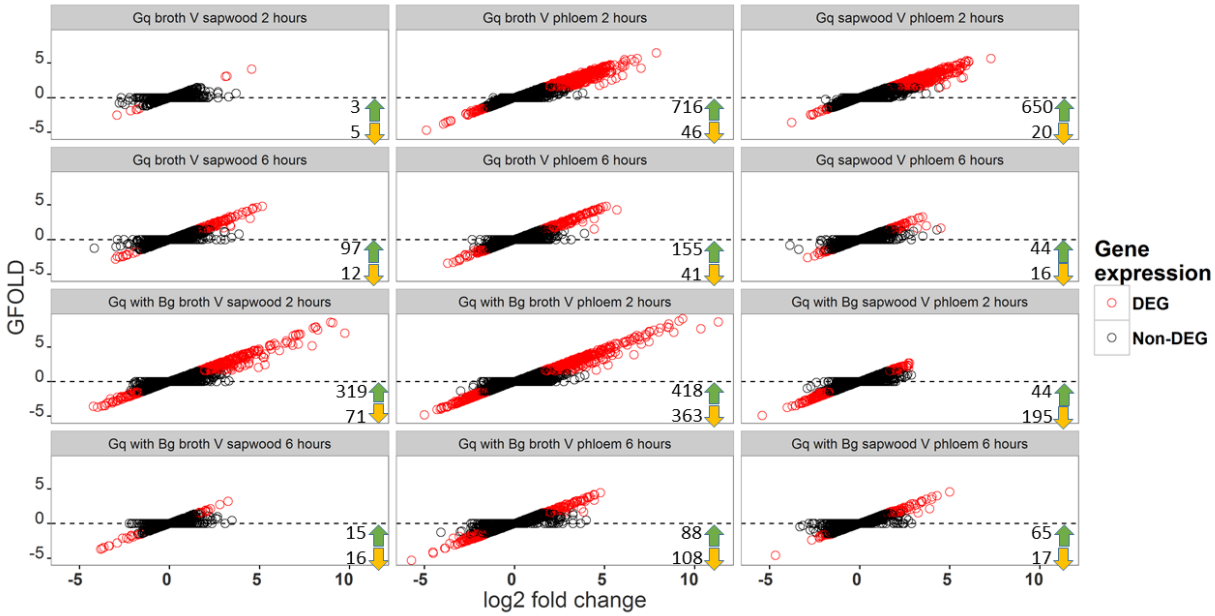


Figure 5.5. Differential gene expression estimates calculated using the GFOLD algorithm for *Gibbsiella quercinecans* FRB97 in 12 environments. Test environments were, broth, sapwood and phloem at 2 and 6 hours, and in these same conditions but in co-culture with *Bg*. Total upregulated (green arrow) and downregulated (yellow arrow) DEG for each environment are numerically listed within each box (environment). Red = DEG, Black = non-DEG.

The increased differential gene expression of *Gq* in NBP was most dramatic at 2 HPI, with a total of 762 DEG. Of these 716 were upregulated, compared to 46 downregulated genes. A similar pattern was observed in NBS compared to NPS, with 670 genes differentially regulated, with 20 downregulated and 650 upregulated, clearly indicating a functional adjustment between NB, NBS and NBP growth conditions (Fig. 5.5). The axenic *Gq* functional expression profile was consistent across the three expression profiles at 2 hours. Growth in phloem has the greatest impact on the functional expression profile; phloem induces upregulation of hundreds of genes compared to NB and NBS, whereas NBS has a limited differential to NB. A further indicator of the impact of phloem is revealed through expression of virulence genes (Fig. 5.6), which were upregulated in NBP compared to NB, and upregulated in NBP compared to NBS.

To understand the functional importance of phase shifts in metabolic activity the functional profiles of differentially regulated transcriptome genes were examined revealing an upsurge

in virulence genes in response to phloem tissue, with a clear link to plant pathogenesis (Toth *et al.*, 2006).

5.3.4.2 Differential expression of PCWDEs in *Gibbsiella quercinecans* FRB97

Expression of PCWDEs across the twelve test environments revealed nine differentially expressed PCWDEs (Fig. 5.6) (Toth *et al.*, 2003; Gibson *et al.*, 2011). Of the differentially expressed PCWDEs, a β-glucosidase GH1 (3.2.1.86) and a polygalacturonase GH28 (EC 3.2.1.15) were differentially expressed in NBP at 6 HPI compared to NB. The upregulation of polygalacturonase, a pectic enzyme is significant, as this is a major soft-rot virulence factor of *P. carotovorum* subsp. *carotovorum* and *Xanthomonas spp.* (Hugouvieux-Cotte-Pattat *et al.*, 1996; Ryan *et al.*, 2011). Expression of polygalacturonase relies on quorum-sensing signalling in *P. carotovorum* subsp. *carotovorum* for mass production and T2SS mediated secretion to the extracellular environment. Polygalacturonase expression is regulated by the global virulence repressor RsmA in *P. carotovorum* subsp. *carotovorum*, and although this gene is found in *Gq*, it was not differentially repressed along with the upregulation of polygalacturonase, indicating greater complexity to its regulation (Chatterjee *et al.*, 1995). Axenic culture of *Gq* at 6 HPI and co-culture at 2 HPI revealed differential upregulation of GH28 and very high normalised read counts (RPKM). High expression of GH28 at 6 HPI may be induced from several stimuli, including pectin fragments from phloem tissue, increase in cell density, and the relaying of quorum sensing signals. Expression of polygalacturonase was high at 6 hours in NB, suggesting constitutive expression of the enzyme after 6 hours, however polygalacturonase was induced to higher expression when in co-culture with *Bg* at 2 HPI. Increased expression of PCWDEs in co-culture with *Bg* occurred in many environments (Fig. 5.6), suggesting that *Bg* stimulates *Gq* PCWDEs, including polygalacturonase. The upregulated β-glucosidase (GH1) at 6 HPI, was encoded within an *asc* operon, which has a homolog in *P. carotovorum* subsp. *carotovorum* (An *et al.*, 2005). The *asc* operon consists of an interacting set of genes involved in sugar transport and metabolism, including the cellulose degrading 6-phospho-β-glucosidase (GH1), an outer membrane porin and the bacterial phosphotransferase system (PTS). The operon is a component of the virulence arsenal of *P. carotovorum* subsp. *carotovorum* and can hydrolyse aromatic glycosides, such as arbutin and salicin on the plant host. The degraded polysaccharides are then internalised and metabolised by the PTS system enzymes A-C. However, within these test environments expression of some PCWDEs was low. For

example, the polysaccharide lyase family (PL) 4 enzyme is differentially expressed in NBP compared to NB, but this was compared to very low expression of the gene at other time points (Fig. 5.6), therefore due to low-level expression the reliability of this result requires further experimental validation (Feng *et al.*, 2012).

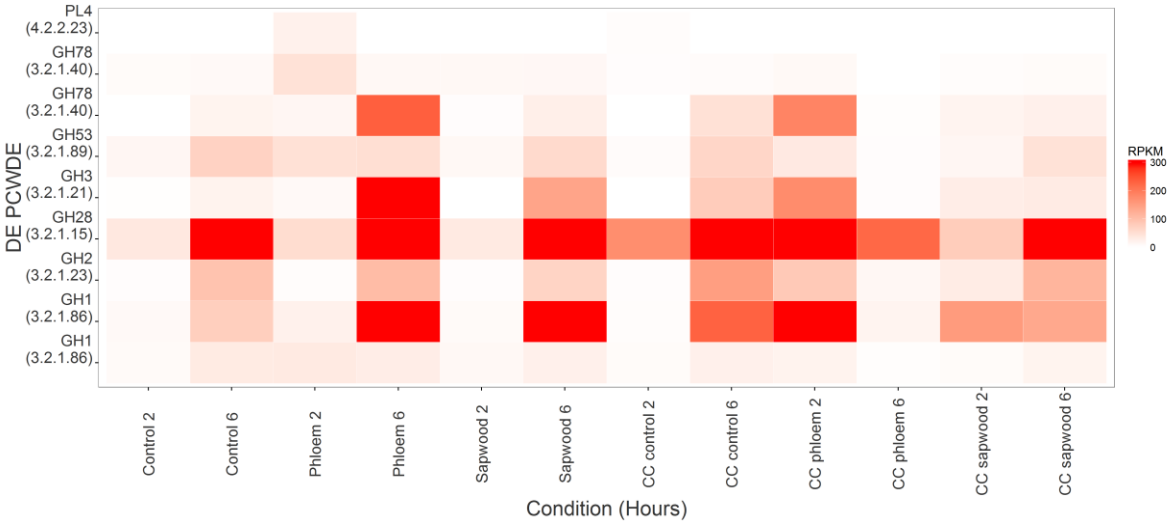


Figure 5.6. Heatmap of expressed plant cell wall degrading enzyme transcripts in *Gibbsiella quercinecans* FRB97. Reads per kilobase of gene per million reads Mapped (RPKM) of PCWDEs in *Gq*. PCWDEs displayed on the y-axis were differentially expressed across a treatment. Increasing red intensity is equivalent to an increase in gene expression to an imposed maximum of 300 RPKM. DE = Differentially Expressed. CC = Co-Culture.

5.3.4.3 Differential expression of virulence regulators in *Gibbsiella quercinecans* FRB97

Many characterised lignocellulosic regulators were differentially expressed in *Gq* (Fig. 5.7). As described in chapter 3, homologous regulators may have antithetical functions within closely related bacteria (Vakulskas *et al.*, 2015). This makes predictions of uncharacterised regulatory networks problematic, however within transcriptomic studies, the function of virulence regulators may be predicted through correlations with expression of their transcribed products. Notable virulence regulators which were differentially expressed across *Gq* sample environments include, *rsmB*, *slyB*, four *pecT* loci, two *luxR* loci, *kdgR*, *marR*, both components of the *phoP-phoQ* two-component signal transduction system (TCSTS). Across *Gq* test environments, there were five distinct patterns of significant differential expression amongst the virulence regulators (Fig. 5.7), these were; 1) Regulators which were significantly differentially expressed in NBP compared to NB and NBS at 2 HPI: *marR*, *slyB*

and three *pecT* homologs. 2) Significantly differentially expressed in NBP at 2 HPI in axenic cultures and co-culture with *Bg*: *pecT*, *phoQ*, and *kdgR*. 3) Significantly differentially downregulated in NBP: *virK*. 4) Significantly differentially expressed in response to co-culture with *Bg* and oak tissue: *rsmB*. 5) Other regulators, which were differentially expressed but had no clear stimuli: *phoP* and two *luxR* loci.

Within axenic *Gq* cultures, in NBS and NBP at 2 HPI, *marR* was significantly differentially upregulated and *slyB* which is within the MarR family of transcriptional regulators (Haque *et al.*, 2009), was differentially downregulated. Homologs of MarR and SlyB have positive and negative regulatory roles, and are involved in the response to oxidative stress, antibiotic resistance, sensing of aromatic compounds and virulence (Ellison and Miller, 2006). Within *Gq*, differential expression of *marR* and *slyB* may have been a response to expected oxidative stress, as the regulators react to the presence of tree tissue in the environment, inciting the stress response and possibly production of virulence genes, as has been described in *E. coli* MarR homologs (Grove, 2013). Three *pecT* homologs were significantly differentially upregulated in axenic cultures at 2 HPI, PecT is a PCWDE repressor in *D. dadantii* 3937 (Jahn *et al.*, 2008), however positive expression of *pecT* here suggests they are positive regulators of PCWDE in *Gq*.

Within axenic and co-cultures of NBS and NBP at 2 HPI, *phoQ* and *kdgR* were differentially expressed. A homology based interpretation of *phoQ* upregulation is problematic, as *phoP-phoQ* is the master inducer of virulence genes in *Salmonella*, *Xanthomonas* and other Gram-negative bacteria, but represses virulence genes in plant pathogenic *Enterobacteriaceae* such as *Erwinia amylovora*, *Pectobacterium* spp. and *Dickeya* spp. PhoP-PhoQ (PhoPQ) is a two-component signal transduction system (TCSTS), which senses environmental levels of Mg^{2+}, antimicrobial peptides, and mildly acidic pH. The membrane bound sensor histidine kinase, *phoQ*, represses or activates virulence genes, through the cytoplasmic response regulator, *phoP*, for example, the negative regulation of PCWDEs in *P. carotovorum* subsp. *carotovorum* in response to extracellular iron (Hyytiäinen *et al.*, 2003). Transcriptional profiles from *Gq*, suggest that PhoPQ is a positive regulator of virulence, as the PhoPQ upregulated test environments correspond to increased virulence gene expression.

The downregulation of *slyB* at 2 HPI provides additional evidence of PhoPQ as a positive virulence regulator, as *slyB* within the SRE is a negative regulator of virulence (Ellison and Miller, 2006). There are several cognate genes upregulated in the PhoPQ circulatory

response, such as magnesium and nickel transport genes, the arn operon and LPS components. Homologs of these genes were downregulated in *Gq* in NBP at 2 and 6 HPI. This is paradoxical, as the *Gq* expression phenotype in this environment is of virulence gene expression, however switching off one set of virulence genes whilst increasing expression of others may be a sophisticated mechanism to concentrate resources in response to subtle environmental cues. Interestingly, separate PhoPQ loci were downregulated in axenic *Gq* cultures at 2 and 6 HPI, a potentially pivotal finding as separate repertoires of PCWDEs were differentially expressed at each time point, indicating that repression of each PhoPQ locus is central to upregulation of a specific set of PCWDEs. However, downregulation of the PhoPQ subsidiary gene VirK provides caveats to this hypothesis (Rio-Alvarez *et al.*, 2012), as upregulation of a *virK* homolog increases antimicrobial resistance in *D. dadantii* 3937 and virulence protein export in *Shigella flexneri* and enteroinvasive/enteroaggregative *E. coli* (Nakata *et al.*, 1992; Tapia-Pastrana *et al.*, 2012), and is also a component regulator of the T3SS signalling cascade, stimulating synthesis of *hrpG*, in the economically important plant pathogen *Ralstonia solanacearum* (Valls *et al.*, 2006; Mansfield *et al.*, 2012). Therefore, downregulation of *virK* in NBP (the sample environment with the highest expression of virulence genes), suggests either an antithetical role within *Gq* or a possible nuanced role for the gene, perhaps in the upregulation of specific virulence factors.

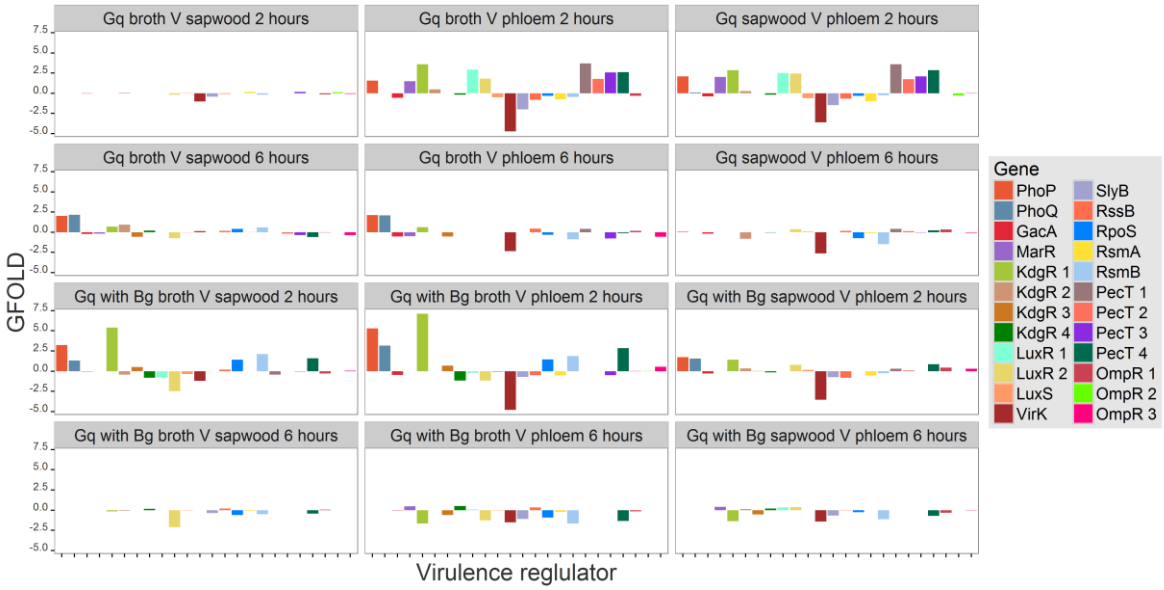


Figure 5.7. GFOLD comparison of virulence gene regulators in *Gibbsiella quercinecans* FRB97. *Gq* = *G. quercinecans* FRB97, *Bg* = *B. goodwinii* FRB141.

In addition to *phoQ*, *kdgR,* a repressor of pectinolysis within the SRE (Barras *et al.*, 1994), was upregulated in axenic cultures of *Gq*, in NBS and NBP at 2 and 6 HPI. The PCWDE-mediated destruction of pectin-rich plant cell tissue is adopted as a survival strategy under nutrient deplete conditions in the soft-rot *Enterobacteriaceae* (SRE) (Charkowski *et al.*, 2012). One of the cues for upregulation of the pectinolysis pathway is the availability of partially degraded pectin, particularly PCWDE substrate by-products such as oligogalacturonates and rhamnose. The SRE are opportunists and the switch to pathogenicity in *D. dadantii* 3937 is initiated through the uptake of oligogalacturonates through the porin *kdgM*, which unleashes a transcriptional cascade of PCWDEs and translocation of these through the T2SS (Blot *et al.*, 2002). Two of these key virulence membrane porins (*kdgM*) were upregulated in *Gq* in NBP at 2 HPI. The first *kdgM* is located on an upregulated operon which includes a carbohydrate specific outer membrane porin, a β-glucoside specific phosphotransferase (PTS) gene, an oligogalacturonide transporter, and a GH105, putative PCWDE (EC 3.2.1.172). The second *kdgM* channel is situated between two upregulated *pecT* (LysR homolog) family transcriptional regulators and alongside the PCWDE, rhamnogalacturonate lyase (rhiE; EC 4.2.2.23) a polysaccharide lyase family 4 (PL4) CAZyme. Downstream of the porin are six upregulated T2SS or Out system genes, part of the T2SS operon found within *Gq*. This secretion system transports rhamnogalacturonate lyase to the extracellular environment in *D. dadantii* 3937 (Laatu and Condemine, 2003), and its upregulation here in addition to porins and PCWDEs, is symptomatic of characterised necrotrophic bacterial phytopathogens. Global control over pectin catabolism in plant pathogenic bacteria is regulated by the repressor *kdgR* (Rodionov *et al.*, 2004). *kdgR* binds to the KdgR box, upstream of the coding domain at the promoter region and prevents transcription. However, upon uptake of pectin polymers, three monomers are formed, galacturonate, glucoronate, and DKI, which produce a common intermediate 2-keto-3-deoxygluconate (KDG), this prevents *kdgR* binding and allows expression of around 13 virulence gene operons and 50 genes, which comprise the KdgR regulon (Hugouvieux-Cotte-Pattat *et al.*, 1996). Despite being known as a classical repressor, KdgR is also known to positively regulate at least two genes in *D. dadantii* 3937 (Rodionov *et al.*, 2004). This offers a possible explanation for the upregulation of *kdgR*, and the pectinolysis pathway genes in *Gq* in NBP at 2 HPI. However, the fold change for this regulator (3.6) is less than that in NBS compared to NB with *Bg* at 2 HPI (5.4), or that of NBP compared to NB with *Bg* at 2 HPI (7.1). In the co-culture with *Bg* there is no upregulation of the Out system, or the porin KdgM, indicating that another regulator is hierarchically superior to KdgR in the pectinolysis

signalling cascade. This may be explained by the availability of pectin fragments, which are higher in milled phloem than sapwood, as a similar upregulation of some pectinolysis genes was found in the NBP co-culture at 2 hours, whereas it is predominantly at the 6 HPI time points that virulence genes are differentially expressed in NBS cultures (Fig. 5.6).

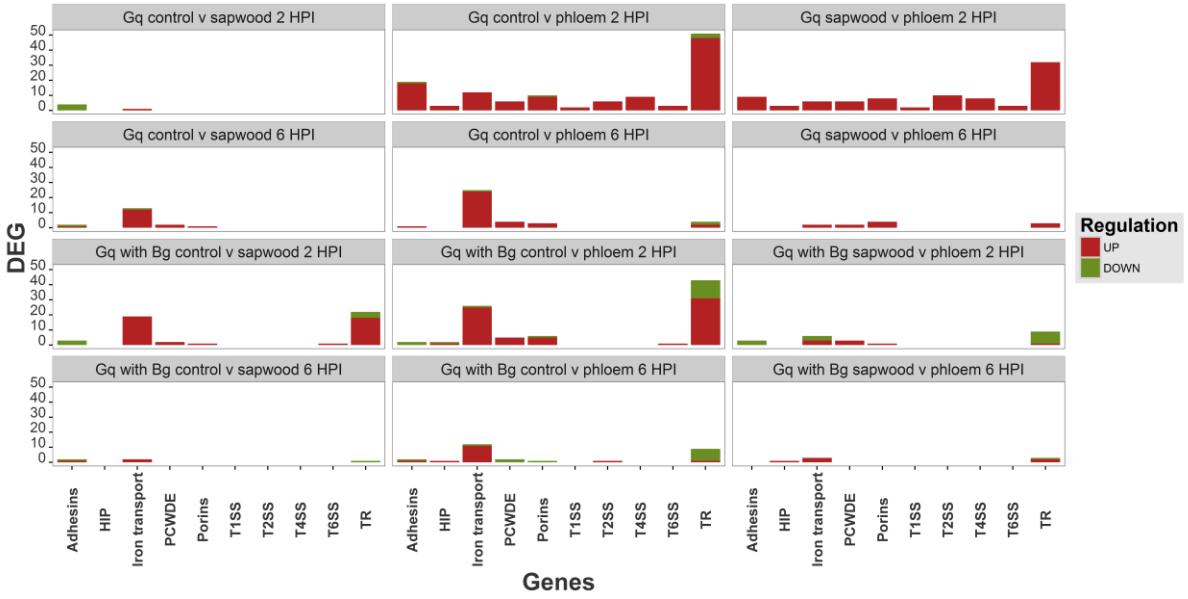


Figure 5.8. Virulence gene expression profiles of *Gibbsiella quercinecans* FRB97 in axenic culture, and co-culture with *Brenneria goodwinii* FRB141. DEG = differentially expressed genes. HIP = host interacting protein, TR = transcription regulator. Red = Upregulated, Green = Down regulated.

A possible interacting component with KdgR, is the post-transcription regulator *rsmB*. *rsmB* degrades transcripts of the virulence gene repressor *rsmA*, thereby allowing the transcription of virulence genes. *rsmA* was not significantly differentially expressed in any of the test environments, whereas *rsmB* was significantly differentially upregulated when *Gq* was in co-culture with *Bg* and in the NBS and NBP v NB environments at 2 HPI, and was significant differentially downregulated in co-culture with *Bg* in NBP v NB at 6 HPI. This suggests that the degradation of the PCWDE transcriptional repressor *rsmA*, is induced through the addition of *Bg* but only in the presence of sapwood and phloem, implying that *Bg* stimulates PCWDE production in *Gq,* when oak tissue is present.

5.3.4.4 Differential expression of adhesins in *Gibbsiella quercinecans* FRB97

As described, the largest functional adjustment of virulence associated genes was in NBP compared to NB and NBS (Fig. 5.8). Upregulated adhesins in NBP at 2 HPI include the Fim gene cluster, a virulence factor of *X. campestris* pv. *vesicatoria* which produces bundle forming fimbriae and allows cell aggregation (Kalkkinen *et al.*, 1997), and type IV pili, which permit adherence of *P. syringae* in the phyllosphere (Hirano and Upper, 2000). Adhesin mediated virulence is used by the plant pathogen *Xylella fastidiosa* to create occlusions in the sapwood (xylem) of host plants (notably causing Pierce's disease of grapevine) (Clifford *et al.*, 2013). Through the release of adhesins including type I pili, fimbriae, and hemagglutinins, the *X. fastidiosa* pathogen can successfully colonise its host. A similar upregulation of adhesins occurs in early NBP expression profiles of *Gq*, including upregulation of type IV pili and the FimA fimbriae gene cluster. An adhesin dependent release of PCWDEs through the Out system (also upregulated in this profile) of *Gq* may allow cells to aggregate and co-ordinate the release of PCWDEs, which would amplify destruction and increase nutrient release. This co-ordinated virulence tactic is employed by *D. dadantii* to attach to the host, aggregate and release virulence enzymes (Rojas *et al.*, 2002).

5.3.4.5 Differential expression of iron acquisition genes in *G. quercinecans* FRB97

A co-ordinated approach to virulence is essential for successful colonisation of the host. However, this requires careful metalloregulation of the cell to maintain sufficient iron levels and prevent the formation of free radicals through the Fenton reaction (Reverchon and Nasser, 2013). As discussed in chapter 3, iron is an essential cofactor in virulence enzymes, however bacteria require soluble iron, which has low bioavailability in the plant host. To acquire iron from the host, bacterial phytopathogens release siderophores such as enterobactin, into the extracellular environment and use TolC secreted haem-binding proteins such as the hemophore HasA to acquire haem from host haemoglobin (Ratledge and Dover, 2000). Complete siderophore operons, including HasA and TolC were extensively upregulated in *Gq* NBP cultures, this expression was especially prevalent at 6 HPI in axenic culture and at 2 HPI in co-culture. At these time points there was a high release of PCWDEs, a pattern mirrored in *D. dadantii*, which releases siderophores and PCWDEs to increase the import of nutrients and increase the degree of virulence (Reverchon and Nasser, 2013).

5.3.4.6 Differential expression of the type IV secretion system in *Gibbsiella quercinecans* FRB97

Nine T4SS genes were upregulated in NBP compared to NB, and eight upregulated in NBP compared to NBS in axenic culture at 2 HPI (Fig. 5.8). Upregulated accessory T4SS proteins suggest a possible horizontal origin of the T4SS, with upregulation of a downstream phage like regulator, an upstream transposase and integrase, and the relaxosome gene MobC which enables conjugal mobilisation through the dissolution of oriT DNA (Zhang and Meyer, 1997). Notably, the upregulated MobC in *Gq* has a homolog in *Agrobacterium tumefaciens*, namely the relaxase protein VirD2, which also functions as an effector protein, and forms a nucleoprotein complex with crown gall inducing oncogenic T-DNA, with *virD2* escorting the T-DNA through the pilus structure and into the host cytoplasm. Further T4SS accessory genes upregulated in axenic *Gq* at 2 HPI were, TraC, a gene required for maturation of F pilin to F pilus which is involved in contact to recipient cells (Schandel *et al.*, 1992), and TraG a gene not involved in construction of the nanomachine but essential for targeted substrate translocation.

The type IV secretion system of *Gq* is a type IVA based subgroup (VirB/D), with an operon cluster of 13 genes, including 11 out of 12 core components. A typical type IVA system is a conjugation/effector group and can interact with the host through pilin mediated attachment and delivery of effectors into the host cytoplasm (Zechner *et al.*, 2012; Martínez-García *et al.*, 2015). Despite being found sporadically across the SRE, the T4SS is not thought to be involved in virulence. However, *P. atrosepticum* contains a type IVA (VirB/D) operon, which is structurally similar to that of *Gq*, it was revealed that a mutation in the *virB4* gene resulted in reduced virulence in the plant host (Bell *et al.*, 2004). Despite, the widespread residency of type IVA systems within SRE pathogens their role is cryptic (Toth *et al.*, 2006; Charkowski *et al.*, 2012). Therefore, the upregulation of T4SS in NBP at 2 HPI in axenic culture suggests a possible phloem induced role in virulence, possibly through attachment to the phloem tissue and translocation of effectors.

5.3.5 *Brenneria goodwinii* FRB141 DEG

5.3.5.1 Overview of *Brenneria goodwinii* FRB141 differential gene expression in all test environments

DEG from *Bg* in NBS compared to NB at 2 HPI was strikingly different to *Gq* in the same comparison. There were 872 differentially expressed genes, with 821 genes upregulated and 51 genes downregulated (Fig. 5.9). This metabolic transition of *Bg* in NBS, is a comprehensive response to different environments. This is the largest differential regulation of genes in *Bg*, which significantly, includes the T3SS virulence operon alongside its effector *dsbA*, the exopolygalacturonate lyase *pelW* and a repertoire of variable substrate PCWDEs. A large but reduced repertoire of genes was upregulated in NBP compared to NB, with a total of 489 DEG, including 418 upregulated and 71 downregulated.

In co-culture with *Gq*, there was a reduced response to NBS and NBP with 153 DEG in NBS compared to NB and 295 DEG in NBP compared to NB at 2HPI. A comparison of the co-culture profile revealed a similar number of differentially expressed virulence genes but differences in the differentially expressed gene systems. The comparatively low number of DEG across the 6 HPI test environments indicates a reduced differential compared to NB at this time point. Furthermore, in both axenic and co-culture test environments, there is a substantially increased functional adjustment at 2 HPI compared to 6 HPI (Fig. 5.9).

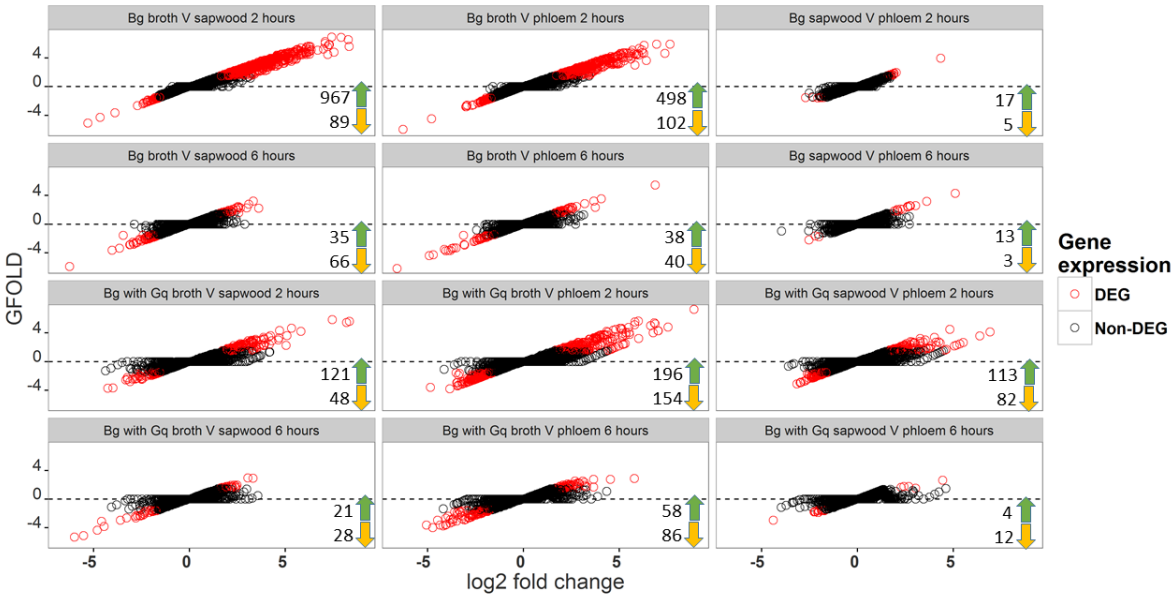


Figure 5.9. Differential gene expression estimates calculated using the GFOLD algorithm for *Brenneria goodwinii* FRB141 in 12 environments. Test environments were, broth, sapwood and phloem at 2 and 6 hours, and in these same conditions but with *Gq*. Total upregulated (green arrow) and downregulated (yellow arrow) DEG for each environment are numerically listed within each box (environment). Red = DEG, Black = non-DEG.

5.3.5.2 Differential expression of PCWDEs in *Brenneria goodwinii* FRB141

Differential expression of PCWDEs by *Bg* had a clear pattern of upregulation in co-culture, including the GH families 1, 3, 51, and 78 (Fig. 5.10). This suggests that *Gq* stimulated the expression of PCWDEs within *Bg*, when oak tissue was present. As described in chapter 3, the core proteins required for the T2SS secretion system function were not identified in *Bg*. This is paradoxical as there were numerous PCWDE homologues from closely related phytopathogens encoded within the genome, some of which were secreted in this experiment (Fig. 5.10). Furthermore, the pectinolysis pathway, mediated by *kdgR*, was upregulated alongside the *kdgM* porin in axenic NBS compared to NB culture at 2 HPI. Mysteriously, the auto-transporters required for transport of PCWDEs across the inner membrane, i.e. the Sec and Tat translocases are encoded but downregulated compared to NB, and substantially downregulated in NBS expression profiles. The subtle interactions governing the expression of pectinolysis pathway genes within *Bg* requires fine-scale *in vitro* analysis, with the possibility of a bespoke PCWDE secretion method.

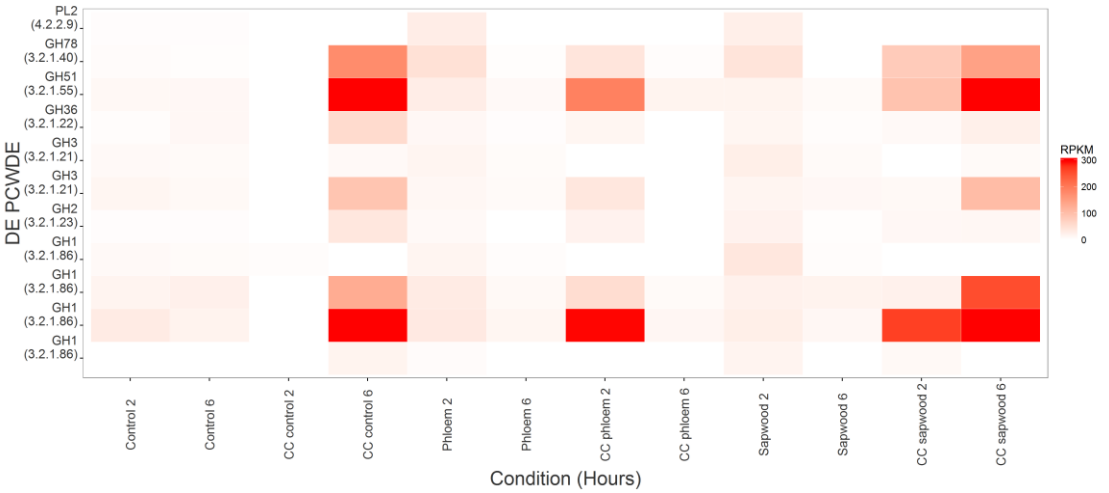


Figure 5.10. Heatmap of expressed plant cell wall degrading enzyme transcripts in *Brenneria goodwinii* FRB141. Reads per kilobase of gene per million reads Mapped (RPKM) of PCWDEs in *Bg*. PCWDEs displayed on the y axis were differentially expressed in any expression profile. Increasing red intensity is equivalent to increase in gene expression to an imposed maximum of 300 RPKM. DE = Differentially Expressed. CC = Co-Culture.

5.3.5.3 Differential expression of virulence regulators in *Brenneria goodwinii* FRB141

There were several significantly differentially expressed regulators across the *Bg* test environments which have virulence homologs in bacterial pathogens. These virulence regulators were: PhoPQ, *rsmA*, *lexA*, *luxR*, *kdgR* and *virF*. Both components of the PhoPQ TCSTS, were downregulated in NBP at 2 HPI, within axenic and co-cultures of *Bg*. This suggests that PhoPQ is directly responding to phloem tissue and may be repressing the T3SS signalling cascade and other virulence genes, which are upregulated in response to sapwood, where there is no significant differential downregulation of PhoPQ. However, there were some virulence genes upregulated within the NBP test environment, suggesting that PhoPQ is a virulence repressor, but it has a low hierarchical position, and may be supplanted by superior regulators.

An *rsmA* homolog is significantly differentially downregulated in the most virulent profile for *Bg*, axenic culture of NBS at 2 HPI. As described above, *rsmA* binds to RNA virulence transcripts, promoting their degradation, its effect is nullified in the presence of the regulatory RNA *rsmB*, which degrades *rsmA* and allows virulence genes to mature (Hyytiäinen *et al.*, 2001). In *P. carotovorum* subsp. *carotovorum* it would be expected that upregulation of *rsmA* would reduce virulence gene expression, as would downregulation of *rsmB*. This empirical evidence in addition to downregulation of *rsmA* in the most virulence test profile suggests that RsmA is critical to the virulence of *Bg*.

The stress response repressor *lexA* is significantly differentially upregulated within axenic NBS culture at 2 HPI. *lexA* binds to the SOS box, preventing transcription of DNA repair proteins, which can be competitively removed from the binding site by a DNA damage regulator, RecA (Yamaguchi and Inouye, 2011). However, the LexA repressor has a functional spectrum beyond the stress response, including the inducement of horizontal gene transfer and the upregulation of pathogenicity island encoded virulence factors (Žgur-Bertok,

2013). Within *Bg*, LexA may function as a virulence repressor, which is stimulated by sapwood tissue, as is unlikely that *Bg* would induce the stress response only in NBS at 2 HPI.

The *luxR* transcriptional regulator is significantly differentially upregulated in *Bg* axenic NBS culture at 2 HPI, and NBP co-culture at 2 HPI, but downregulated in NBP co-culture at 6 HPI. This suggests that *luxR* is stimulated in axenic NBS culture, but repressed in NBS by *Gq*, and induced in NBP by *Gq*. This finding corresponds with the increased virulence profile of *Gq* in NBP at 2 HPI, and suggests that *Gq* induces *luxR* expression within *Bg*, to coincide with the release of virulence genes by *Gq*. The function of two *luxR* loci in *Gq* was unclear, however the *Bg* expression data suggests, that LuxR is a component of a polymicrobial virulence response to phloem tissue. This result links with the function of the LuxR homolog, ExpA in *P. carotovorum* subsp. *carotovorum*, where *expR* activates production of the RNA binding protein *rsmA*, but is repressed by AHL leading to the inhibition of *rsmA* and the maturation of virulence proteins (Cui *et al.*, 2005). There is no obvious link in this study between *rsmA* and *luxR*, however there may be an analogous interaction. A similar pattern of transcription expression to *luxR* was produced by the xylose catabolism promoter, *xylR*, which was significantly upregulated in axenic NBS cultures and in NBP co-cultures. This suggests that *Gq* has a cross-species regulatory effect on *Bg*, inducing upregulation of *xylR* and *luxR*. A possible function for upregulation of *xylR* in *Bg* is positive regulation xylose catabolism and transport, as a *xylR* homolog in *E. coli* displays this functionality (Desai and Rao, 2010).

Within *Bg* test environments, the downregulation of *kdgR* correlates with the upregulation of virulence genes, thereby appearing to follow its characterised role of virulence gene repression (Nasser *et al.*, 1992). There are four homologs of the pectinolysis repressor *kdgR* encoded within *Bg*, only one was significantly differentially expressed, this downregulation corresponds with an upregulation in PCWDEs. However, as described above, *kdgR* is upregulated in the most virulent profiles of *Gq* and as Figure 5.11 reveals, a *kdgR* homolog is downregulated in the most virulent profiles of *Bg*. This may be due to it superiority in the signalling hierarchy within these two bacteria, where its pectinolysis repression is supplanted by a more potent regulator in *Gq*, but this ambiguity proves the requirement for individual analysis of all virulence regulators in bacteria.

The regulator VirF is a plasmid encoded virulence factor which mediates expression of T3SS in *Shigella flexneri*, and activates *yop* and *ysc* T3SS effectors in *Yersinia enterocolitica* causing

dysentery and a range of conditions in human hosts (Gemski *et al.*, 1980; Dohlich *et al.*, 2014). *virF* is one of the highest upregulated genes within axenic *Bg* cultures at 2 HPI (Fig. 5.11), and is differentially downregulated in profiles with no upregulation of the T3SS. Differential upregulation of *virF*, a characterised inducer of T3SS genes, in *Bg* at 2 HPI coincides with that of the T3SS apparatus genes, suggesting a direct role for VirF within the *Bg* T3SS pathosystem.

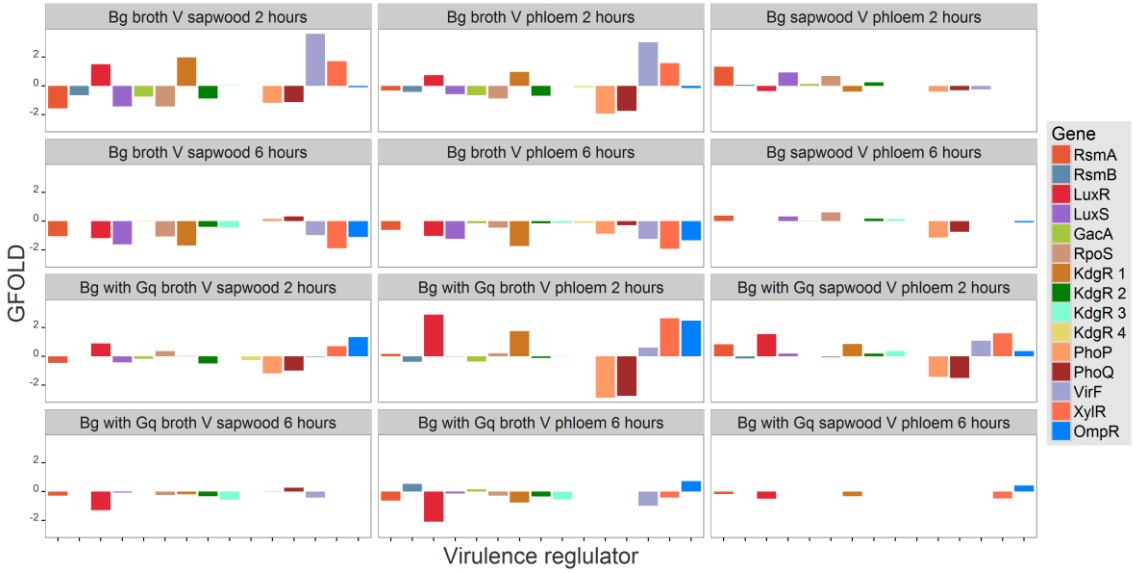


Figure 5.11. GFOLD comparison of virulence gene regulators in *Brenneria goodwinii* FRB141. *Bg* = *B. goodwinii* FRB141, *Gq* = *G. quercinecans* FRB97.

5.3.5.4 Differential expression of the NF-T3SS in *Brenneria goodwinii* FRB141

The *Bg* non-flagellar (NF) T3SS is a key virulence factor and the central component of virulence in the canonical bacterial phytopathogen *P. syringae* (Arnold and Jackson, 2011). Through the pilus mediated secretion of effectors from the bacterial cytoplasm to the host cytosol (a function which distinguishes the NF-T3SS from the flagellar T3SS), *P. syringae* can evade host detection and cause pathogenesis in a narrow (pathovar specific) host range (Pallen *et al.*, 2005; Lindeberg *et al.*, 2012). *Bg* is more distantly related to *P. syringae* (Family: *Pseudomonadaceae*) than to the SRE (Family: *Enterobacteriaceae*), which are also known to encode NF-T3SS. Unlike *P. syringae*, the SRE are broad range pathogens which rely on translocation of PCWDEs as their principal virulence mechanism; some virulent species of SRE such as *P. wasabiae* have no encoded T3SS (Nykyri *et al.*, 2012). Therefore, it is unclear

if the T3SS are primary virulence components as in *P. syringae* or secondary to PCWDEs, as in the SRE. The T3SS was upregulated in NBS and NBP at 2 HPI (all nine core T3SS genes are upregulated in NBS, with five upregulated in NBP), which are the only expression profiles with a T3SS upregulation phenotype (Fig. 5.12). Crucially a type III effector with a homolog in the pectinolytic, wide host range Pseudomonad and a close relative of *P. syringae*, *P. viridiflava* effector, *hopPtoM* (Araki *et al.*, 2006; Sarris *et al.*, 2012) and its protein chaperone *shcM*, were upregulated in axenic NBS cultures at 2 hours. The effector and chaperone are located at a separate locus to the secretion system machinery and nearby several upregulated transposes and integrases, suggesting horizontal transfer of the effector and chaperone. Intriguingly, a homolog of the *hopPtoM* upregulated effector, *hrpW* from *P. syringae* pv. tomato, can elicit the plant hypersensitivity response and has pectate lyase functionality (Charkowski *et al.*, 1998; Collmer *et al.*, 2002). Additionally, a homolog of the *dspA/E* (Disease SPecific), T3SS effector from the fireblight pathogen *Erwinia amylovora*, was also upregulated at 2 HPI in axenic NBS and NBP cultures (Oh and Beer, 2005). This is a cytotoxic effector, which resides in the *E. amylovora* HEE (Hrp effector and elicitors) gene cluster. Furthermore, the DspA/E gene controls a multifaceted virulence regulon in the SRE bacterium *P. atrosepticum*, including the Out secretion system and many virulence regulators (Coulthurst *et al.*, 2008). Similar to *hopPtoM*, *dspA/E* is located on a separate operon to the core T3SS gene cluster, and is upregulated in the same expression phenotype as the core T3SS cluster, suggesting it could have a role as an effector or in virulence regulation. This suggests a possible role for the T3SS in pathogenesis, similar to *P. syringae* which releases effectors upon contact with the host (Tampakaki *et al.*, 2010). Despite the lack of T3SS secreton upregulation in co-culture test environments, a key harpin virulence effector protein *hrpN*, was induced by *Gq* in NBS and NBP at 2 HPI. A HrpN homolog in *D. dadantii* 3937 is induced by phenolic plant components (Yang *et al.*, 2008). However, *Bg* requires co-culture with *Gq*, and oak tissue components to express *hrpN*. This result indicates that a specific virulence phenotype is produced within *Bg*, when in co-culture with *Gq*, a virulence phenotype which differs to the axenic response to oak tissue. The differentially expressed T3SS virulence genes in *Bg*, are homologous to key virulence enzymes in bacterial phytopathogens which may have evolved new functions, and therefore require further testing for complete functional characterisation (Vinatzer *et al.*, 2014).

A shared characteristic between the T3SS of the fireblight pathogen *E. amylovora*, the SRE member *D. dadantii*, *Gq* and *Bg*, is a putative insect vector. The buprestid beetle *Agrilus*

biguttatus has a close association with AOD (being present in more than 90% of confirmed cases), and is suspected of playing a key role in Decline-disease pathogenicity, through vectoring the bacteria between hosts (Denman *et al.*, 2014). This is particularly relevant to *Gq* and *Bg* as their range and mode of transmission between hosts remains unclear (Brown *et al.*, 2016). Insect vectors which transmit pathogens between hosts are described for some SRE species, but this occurrence is not well characterised (Charkowski *et al.*, 2012). The T3SS of *E. amyovora* encodes three T3SS, two of which have no reported effect on plant virulence and were suggested to interact with their insect vector during transmission (Zhao *et al.*, 2009). The T3SS of *D. dadantii* injects cytotoxic effectors into the pea aphid, *Acyrthosiphon pisum*, which harbour the bacteria on plant leaves, potentially vectoring them between hosts. The role of the T3SS in *Bg* is currently unknown, but perhaps it has a similar function to the T3SS of *D. dadantii* and mediates transmission of the bacterium through an insect vector.

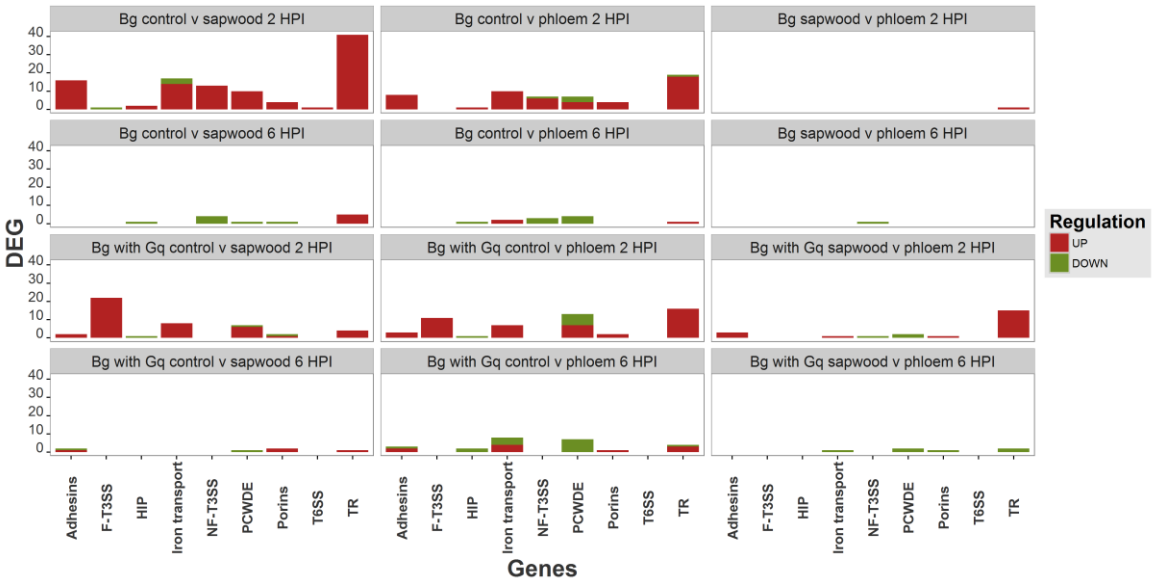


Figure 5.12. Virulence gene expression profiles of *Brenneria goodwinii* FRB97 in axenic culture, and co-culture with *Gibbsiella quercinecans* FRB97. DEG = differentially expressed genes. HIP = host interacting protein, TR = transcription regulator. Red = Upregulated, Green = Down regulated.

5.3.5.5 Differential expression of flagellar genes in *Brenneria goodwinii* FRB141

The flagellum and type IV pili are motility organelles which enable cellular locomotion in bacteria (Lux and Shi, 2004). The flagellum is the main mechanism of bacterial motility and is evolutionary derived from the effector-translocation, NF-T3SS (Pallen and Matzke, 2006). This large modular system consists up to 50 genes but with a core of just twenty required for function. A large flagellar cluster of 39 genes is encoded within *Bg*. Only two expression profiles contained upregulation of this cluster, with twenty-two upregulated genes in NBS compared to NB and eleven from NBP compared to NB, both of which were in co-culture with *Gq* at 2 HPI (Fig. 5.10). The axenic cultures of *Bg* had no upregulated flagellar genes, indicating they require the dual trigger of suitable host tissue and chemotactic inducement from other bacteria. This suggests synchronicity and a possible mutualistic relationship between the species, which may use this strategy to enhance the disease-causing process (Venturi and da Silva, 2012).

5.3.6 Polymicrobial symbiosis

The co-ordination of virulence through interspecies signalling enhances pathogenicity and is part of a polymicrobial infection, this is increasingly recognised as an intrinsic component of the natural environment, with the focus of research into bacterial phytopathogens shifting from single species, to recognising that they do not exist in isolation but rather as multispecies communities (Hosni *et al.*, 2011). For example, the olive knot pathogen *Pseudomonas savastanoi* pv. *savastanoi* has enhanced virulence when in co-culture with the endophyte *Erwinia toletana,* which releases N-acylhomoserine lactone (AHL) allowing induction of virulence factors at the most appropriate moments (da Silva *et al.*, 2014a). The inducement of the nf-T3SS is driven by media constituents in axenic culture, but the activation of the F-T3SS in the same media is stimulated by *Gq*, could be viewed in the light of polymicrobial infection research. *Gq* and *Bg* co-exist in the oak lesion microbiome, therefore their relationship may be antagonistic with the associated release of toxin/antitoxin system and competing T6SS secretion systems (Coulthurst, 2013). However, bacteria can have co-operative relationships, and act synergistically to recruit other species to an infection site and release exoenzymes as a quorum (Venturi and da Silva, 2012). This symbiosis can extend to primary insect pathogens involved in pathogenesis, such as egg hatching induced by mice gut microbiota assisting virulence of the parasitic nematode *Trichuris muris* (Hayes *et al.*, 2010). *Gq* and *Bg* may follow a polymicrobial infection model, where each individual is not able to

act as a primary pathogen but through acting co-operatively, the bacteria are able to degrade tree tissue and acquire nutrients. The nature of the polymicrobial environment in bacterial phytopathology is only beginning to be understood, and does not conform to the classical microbiological approach of proving pathology through Koch's postulates (Falkow, 2004). Instead, a contemporary holistic approach such as that first proposed by Falkow in 1988 (Falkow, 1988), of incorporating molecular techniques to examine the fundamental role of interacting genes as opposed to single causative organisms may benefit understanding the necrotic lesion microbiota of AOD affected trees.

5.4 Conclusions

There are several conclusions arising from this work; 1) Both *Gq* and *Bg* had higher DEG at 2 HPI than 6 HPI, 2) *Gq* had substantial gene upregulation in axenic NBP culture but not NBS, particularly at 2 HPI, 3) *Gq* had a substantial genome wide response to sapwood when in co-culture with *Bg*, 4) *Gq* is induced to express virulence genes within NBS when in co-culture with *Bg*, 5) *Bg* had substantial DEG in NBS at 2 HPI, but also in NBP, albeit with a reduced DEG profile, 6) In axenic NBS culture, *Bg* expressed notorious T3SS plant pathogenic effectors, 7) *Bg* had a reduced DEG profile in NBS and NBP when in co-culture with *Gq,* but was induced by *Gq* to produce specific virulence genes including PCWDEs and a T3SS harpin.

Both *Gq* and *Bg* had substantial metabolic shifts in gene regulation within specific expression profiles. Gene expression *in vitro* can be extrapolated to the polymicrobial environment of a necrotic oak lesion from which *Gq* and *Bg* were originally isolated. The high number of DEG in *Gq* when cultured in NBP and the pathogenic potential of many of these genes, reveal that *Gq* may have the prerequisite tools to break down phloem tissue, suggesting that within the inner-bark of an oak lesion, *Gq* contribute to tissue necrosis. *Bg* has a substantial increase in DEG when sapwood and phloem were added to the culture media, including expression of many virulence genes, suggesting that *Bg* can reside within the sapwood, or the phloem. Intriguingly, the number of DEG within *Gq* and *Bg* co-cultures suggest that *Gq* has genome wide DEG, including virulence genes when in co-culture with *Bg* and that *Bg* has a substantially reduced response to tree tissue within the experimental medium when in co-culture with *Gq*. This suggests that bacterial interaction within necrotic oak lesions of AOD affected trees, contribute to the pathogenic phenotype of the lesion, and the presence of *Gq* without *Bg*, and vice-versa will produce a markedly different pathogen phenotype.

To confirm the findings of this study, the resultant data requires validation. The gold standard in transcriptomic research is to verify key differentially expressed genes via qPCR assays. Therefore, validation of differentially expressed virulence genes, which may be involved in the maceration of oak tree tissue within a necrotic lesion, should be confirmed using these methods.

5.5 References

An, C.L., Lim, W.J., Hong, S.Y., Shin, E.C., Kim, M.K., Lee, J.R., *et al.* (2005) Structural and biochemical analysis of the asc operon encoding 6-phospho-βglucosidase in *Pectobacterium carotovorum* subsp. *carotovorum* LY34. *Res. Microbiol.* **156**: 145–153.

Araki, H., Tian, D., Goss, E.M., Jakob, K., Halldorsdottir, S.S., Kreitman, M., and Bergelson, J. (2006) Presence/absence polymorphism for alternative pathogenicity islands in *Pseudomonas viridiflava*, a pathogen of Arabidopsis. *Pnas* **103**: 5887–5892.

Arnold, D.L. and Jackson, R.W. (2011) Bacterial genomes: evolution of pathogenicity. *Curr. Opin. Plant Biol.* **14**: 385–91.

Barras, F., van Gijsegem, F., and Chatterjee, A. (1994) Extracellular enzymes and pathogenesis of soft-rot *Erwinia*. *Annu. Rev. Phytopathol.* **32**: 201–34.

Bell, K.S., Sebaihia, M., Pritchard, L., Holden, M.T.G., Hyman, L.J., Holeva, M.C., *et al.* (2004) Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora* subsp. *atroseptica* and characterization of virulence factors. *Proc. Natl. Acad. Sci. U. S. A.* **101**: 11105–10.

Blot, N., Berrier, C., Hugouvieux-Cotte-Pattat, N., Ghazi, A., and Condemine, G. (2002) The oligogalacturonate-specific porin KdgM of *Erwinia chrysanthemi* belongs to a new porin family. *J. Biol. Chem.* **277**: 7936–7944.

Bowden, S.D., Hale, N., Chung, J.C.S., Hodgkinson, J.T., Spring, D.R., and Welch, M. (2013) Surface swarming motility by *Pectobacterium atrosepticum* is a latent phenotype that requires O antigen and is regulated by quorum sensing. *Microbiology* **159**: 2375–85.

Brady, C., Denman, S., Kirk, S., Venter, S., Rodríguez-Palenzuela, P., and Coutinho, T. (2010) Description of *Gibbsiella quercinecans* gen. nov., sp. nov., associated with Acute Oak Decline. *Syst. Appl. Microbiol.* **33**: 444–50.

Brankatschk, K., Kamber, T., Pothier, J.F., Duffy, B., and Smits, T.H.M. (2014) Transcriptional profile of *Salmonella enterica* subsp. *enterica* serovar Weltevreden during alfalfa sprout colonization. *Microb. Biotechnol.* **7**: 528–544.

Brown, N., Jeger, M., Kirk, S., Xu, X., and Denman, S. (2016) Spatial and temporal patterns in symptom expression within eight woodlands affected by Acute Oak Decline. *For. Ecol. Manage.* **360**: 97–109.

Charkowski, A., Blanco, C., Condemine, G., Expert, D., Franza, T., Hayes, C., *et al.* (2012) The role of secretion systems and small molecules in soft-rot *Enterobacteriaceae* pathogenicity. *Annu. Rev. Phytopathol.* **50**: 425–49.

Charkowski, A.O., Alfano, J.R., Preston, G., Yuan, J., He, S.Y., and 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. *J. Bacteriol.* **180**: 5211–5217.

Chatterjee, A., Cui, Y., Liu, Y., Dumenyo, C.K., and Chatterjee, A.K. (1995) Inactivation of *rsmA* leads to overproduction of extracellular pectinases, cellulases, and proteases in *Erwinia carotovora* subsp. *carotovora* in the absence of the starvation/cell density-sensing signal, N-(3-oxohexanoyl)-L-homoserine lactone. *Appl. Environ. Microbiol.* **61**: 1959–1967.

Clifford, J.C., Rapicavoli, J.N., and Roper, M.C. (2013) A rhamnose-rich O-antigen mediates adhesion, virulence, and host colonization for the xylem-limited phytopathogen *Xylella fastidiosa*. *Mol.*

Plant. Microbe. Interact. **26**: 676–85.

Collmer, A., Lindeberg, M., Petnicki-Ocwieja, T., Schneider, D.J., and Alfano, J.R. (2002) Genomic mining type III secretion system effectors in *Pseudomonas syringae* yields new picks for all TTSS prospectors. *Trends Microbiol.* **10**: 462–469.

Costechareyre, D., Balmand, S., Condemine, G., and Rahbé, Y. (2012) *Dickeya dadantii*, a plant pathogenic bacterium producing cyt-like entomotoxins, causes septicemia in the pea aphid *Acyrthosiphon pisum*. *PLoS One* **7**: e30702.

Coulthurst, S.J. (2013) The Type VI secretion system - a widespread and versatile cell targeting system. *Res. Microbiol.* **164**: 640–54.

Coulthurst, S.J., Lilley, K.S., Hedley, P.E., Liu, H., Toth, I.K., and Salmond, G.P.C. (2008) DsbA plays a critical and multifaceted role in the production of secreted virulence factors by the phytopathogen *Erwinia carotovora* subsp. *atroseptica*. *J. Biol. Chem.* **283**: 23739–53.

Cui, Y., Chatterjee, A., Hasegawa, H., Dixit, V., Leigh, N., and Chatterjee, A.K. (2005) ExpR, a LuxR Homolog of *Erwinia carotovora* subsp. *carotovora*, Activates Transcription of *rsmA* , Which Specifies a Global Regulatory RNA-Binding Protein. *J. Bacteriol.* **187**: 4792–4803.

Denman, S., Brady, C., Kirk, S., Cleenwerck, I., Venter, S., Coutinho, T., and De Vos, P. (2012) *Brenneria goodwinii* sp. nov., associated with acute oak decline in the UK. *Int. J. Syst. Evol. Microbiol.* **62**: 2451–2456.

Denman, S., Brown, N., Kirk, S., Jeger, M., and Webber, J. (2014) A description of the symptoms of Acute Oak Decline in Britain and a comparative review on causes of similar disorders on oak in Europe. *Forestry* **87**: 535–551.

Desai, T.A. and Rao, C. V. (2010) Regulation of arabinose and xylose metabolism in *Escherichia coli*. *Appl. Environ. Microbiol.* **76**: 1524–1532.

Dohlich, K., Zumsteg, A.B., Goosmann, C., and Kolbe, M. (2014) A substrate-fusion protein is trapped inside the Type III secretion system channel in *Shigella flexneri*. *PLoS Pathog.* **10**: e1003881.

Edwards, U., Rogall, T., Blöcker, H., Emde, M., and Böttger, E.C. (1989) Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Res.* **17**: 7843–7853.

Ellison, D.W. and Miller, V.L. (2006) Regulation of virulence by members of the MarR/SlyA family. *Curr. Opin. Microbiol.* **9**: 153–159.

Eyles, A., Bonello, P., Ganley, R., and Mohammed, C. (2010) Induced resistance to pests and pathogens in trees. *New Phytol.* **185**: 893–908.

Falkow, S. (2004) Molecular Koch's postulates applied to bacterial pathogenicity--a personal recollection 15 years later. *Nat. Rev. Microbiol.* **2**: 67–72.

Falkow, S. (1988) Molecular Koch's postulates applied to microbial pathogenicity. *Rev. Infect. Dis.* **10**: 7–10.

Feng, J., Meyer, C.A., Wang, Q., Liu, J.S., Liu, X.S., and Zhang, Y. (2012) GFOLD: A generalized fold change for ranking differentially expressed genes from RNA-seq data. *Bioinformatics* **28**: 2782–2788.

Geiss, G.K., Bumgarner, R.E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D.L., *et al.* (2008) Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat. Biotechnol.* **26**: 317–25.

Gemski, P., Lazere, J.R., and Casey, T. (1980) Plasmid associated with pathogenicity and calcium dependency of *Yersinia enterocolitica*. *Infect. Immun.* **27**: 682–685.

Gibson, D.M., King, B.C., Hayes, M.L., and Bergstrom, G.C. (2011) Plant pathogens as a source of diverse enzymes for lignocellulose digestion. *Curr. Opin. Microbiol.* **14**: 264–270.

Grove, A. (2013) MarR family transcription factors. *Curr. Biol.* **23**: R142–R143.

Haas, B.J., Chin, M., Nusbaum, C., Birren, B.W., and Livny, J. (2012) How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes? *BMC Genomics* **13**: 734.

Haque, M.M., Kabir, M.S., Aini, L.Q., Hirata, H., and Tsuyumu, S. (2009) SlyA, a MarR family transcriptional regulator, is essential for virulence in *Dickeya dadantii* 3937. *J. Bacteriol.* **191**: 5409–5418.

Hayes, K., Bancroft, A., Goldrick, M., Portsmouth, C., Roberts, I., and Grencis, R. (2010) Exploitation of the intestinal microflora by the parasitic nematode *Trichuris muris*. *Science.* **328**: 1391–1394.

He, S.Y., Schoedel, C., Chatterjee, A.K., and Collmer, A. (1991) Extracellular secretion of pectate lyase by the *Erwinia chrysanthemi* out pathway is dependent upon Sec-mediated export across the inner membrane. *J. Bacteriol.* **173**: 4310–4317.

Hirano, S.S. and Upper, C.D. (2000) Bacteria in the leaf ecosystem with emphasis on *Pseudomonas syringae*-a pathogen, ice nucleus, and epiphyte. *Microbiol. Mol. Biol. Rev.* **64**: 624–53.

Hosni, T., Moretti, C., Devescovi, G., Suarez-Moreno, Z.R., Fatmi, M.B., Guarnaccia, C., *et al.* (2011) Sharing of quorum-sensing signals and role of interspecies communities in a bacterial plant disease. *ISME J.* **5**: 1857–1870.

Hugouvieux-Cotte-Pattat, N., Condemine, G., Nasser, W., and Reverchon, S. (1996) Regulation of pectinolysis in *Erwinia chrysanthemi*. *Annu. Rev. Microbiol.* **50**: 213–257.

Hyytiäinen, H., Montesano, M., and Palva, E.T. (2001) Global regulators ExpA (GacA) and KdgR modulate extracellular enzyme gene expression through the RsmA-rsmB system in *Erwinia carotovora* subsp. *carotovora*. *Mol. Plant. Microbe. Interact.* **14**: 931–938.

Hyytiäinen, H., Sjöblom, S., Palomäki, T., Tuikkala, A., and Palva, E.T. (2003) The PmrA-PmrB two-component system responding to acidic pH and iron controls virulence in the plant pathogen *Erwinia carotovora* ssp. *carotovora*. *Mol. Microbiol.* **50**: 795–807.

Jahn, C.E., Willis, D.K., and Charkowski, A.O. (2008) The flagellar sigma factor *fliA* is required for *Dickeya dadantii* virulence. *Mol. Plant. Microbe. Interact.* **21**: 1431–1442.

Joshi, J.R., Burdman, S., Lipsky, A., Yariv, S., and Yedidia, I. (2015) Plant phenolic acids affect the virulence of *Pectobacterium aroidearum* and *P. carotovorum* ssp. *brasiliense* via quorum sensing regulation. *Mol. Plant Pathol.* **17**:, 487-500.

Kalkkinen, N., Haahtela, K., Wengelnik, K., and Bonas, U. (1997) Characterization of the *fimA* gene encoding bundle-forming fimbriae of the plant pathogen *Xanthomonas campestris* pv. *vesicatoria*. *J. Bactiol.* **179**: 1280–1290.

Kingsford, C., Delcher, A.L., and Salzberg, S.L. (2007) A unified model explaining the offsets of overlapping and near-overlapping prokaryotic genes. *Mol. Biol. Evol.* **24**: 2091–2098.

Klemm, D., Heublein, B., Fink, H.P., and Bohn, A. (2005) Cellulose: Fascinating biopolymer and sustainable raw material. *Angew. Chemie - Int. Ed.* **44**: 3358–3393.

Korotkov, K. V., Sandkvist, M., and Hol, W.G.J. (2012) The type II secretion system: biogenesis, molecular architecture and mechanism. *Nat. Rev. Microbiol.* **10**: 336–351.

Laatu, M. and Condemine, G. (2003) Rhamnogalacturonate lyase rhiE is secreted by the out system in *Erwinia chrysanthemi*. *J. Bacteriol.* **185**: 1642–1649.

de Lima, E.A., Machado, C.B., Zanphorlin, L.M., Ward, R.J., Sato, H.H., and Ruller, R. (2016) GH53 endo-beta-1,4-galactanase from a newly isolated *Bacillus licheniformis* CBMAI 1609 as an enzymatic cocktail supplement for biomass saccharification. *Appl. Biochem. Biotechnol.* **179**: 415-426.

Lindeberg, M., Cunnac, S., and Collmer, A. (2012) *Pseudomonas syringae* type III effector repertoires: Last words in endless arguments. *Trends Microbiol.* **20**: 199–208.

Lux, R. and Shi, W. (2004) Chemotaxis-guided movements in bacteria. *Crit. Rev. Oral Biol. Med.* **15**: 207–220.

Magoc, T., Wood, D., and Salzberg, S.L. (2013) EDGE-pro: Estimated degree of gene expression in prokaryotic genomes. *Evol. Bioinform. Online* **9**: 127–36.

Mansfield, J., Genin, S., Magori, S., Citovsky, V., Sriariyanum, M., Ronald, P., *et al.* (2012) Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol. Plant Pathol.* **13**: 614–29.

Martínez-García, P.M., Ramos, C., and Rodríguez-Palenzuela, P. (2015) T346Hunter: A novel web-based tool for the prediction of type III, type IV and type VI secretion systems in bacterial genomes. *PLoS One* **10**: e0119317.

McClure, R., Balasubramanian, D., Sun, Y., Bobrovskyy, M., Sumby, P., Genco, C. a., *et al.* (2013) Computational analysis of bacterial RNA-Seq data. *Nucleic Acids Res.* **41**: 1–16.

Mori, T., Ishitani, R., Tsukazaki, T., Nureki, O., and Sugita, Y. (2010) Molecular mechanisms underlying the early stage of protein translocation through the Sec translocon. *Biochemistry* **49**: 945–950.

Murdoch, S.L., Trunk, K., English, G., Fritsch, M.J., Pourkarimi, E., and Coulthurst, S.J. (2011) The opportunistic pathogen *Serratia marcescens* utilizes type VI secretion to target bacterial competitors. *J. Bacteriol.* **193**: 6057–69.

Nakata, N., Sasakawa, C., Okada, N., Tobe, T., Fukuda, I., Suzuki, T., *et al.* (1992) Identification and characterization of *virK*, a virulence-associated large plasmid gene essential for intercellular spreading of *Shigella flexneri*. *Mol. Microbiol.* **6**: 2387–2395.

Nasser, W., Reverchon, S., and Robert-Baudouy, J. (1992) Purification and functional characterization of the KdgR protein, a major repressor of pectinolysis genes of *Erwinia chrysanthemi*. *Mol. Microbiol.* **6**: 257–265.

Nykyri, J., Niemi, O., Koskinen, P., Nokso-Koivisto, J., Pasanen, M., Broberg, M., *et al.* (2012) Revised phylogeny and novel horizontally acquired virulence determinants of the model soft rot phytopathogen *Pectobacterium wasabiae* SCC3193. *PLoS Pathog.* **8**: e1003013.

O'Neil, D., Glowatz, H., and Schlumpberge, M. (2013) Ribosomal RNA depletion for efficient use of RNA-seq capacity. *Curr. Protoc. Mol. Biol.* 1–8.

Oh, C.-S. and Beer, S. V (2005) Molecular genetics of *Erwinia amylovora* involved in the development of fire blight. *FEMS Microbiol. Lett.* **253**: 185–92.

Pallen, M.J., Beatson, S. A., and Bailey, C.M. (2005) Bioinformatics, genomics and evolution of non-flagellar type-III secretion systems: a Darwinian perpective. *FEMS Microbiol. Rev.* **29**: 201–229.

Pallen, M.J. and Matzke, N.J. (2006) From the origin of species to the origin of bacterial flagella. *Nat. Rev. Microbiol.* **4**: 784–790.

Potrykus, M., Golanowska, M., Hugouvieux-Cotte-Pattat, N., and Lojkowska, E. (2014) Regulators involved in *Dickeya solani* virulence, genetic conservation, and functional variability. *Mol. Plant. Microbe. Interact.* **27**: 700–711.

Ratledge, C. and Dover, L.G. (2000) Iron metabolism in pathogenic bacteria. *Annu. Rev. Microbiol.* **54**: 881–941.

Reverchon, S. and Nasser, W. (2013) *Dickeya* ecology, environment sensing and regulation of virulence programme. *Environ. Microbiol. Rep.* **5**: 622–636.

Rio-Alvarez, I., Rodríguez-Herva, J.J., Cuartas-Lanza, R., Toth, I., Pritchard, L., Rodríguez-Palenzuela, P., and López-Solanilla, E. (2012) Genome-wide analysis of the response of *Dickeya dadantii* 3937 to plant antimicrobial peptides. *Mol. Plant. Microbe. Interact.* **25**: 523–33.

Rodionov, D.A., Gelfand, M.S., and Hugouvieux-Cotte-Pattat, N. (2004) Comparative genomics of the KdgR regulon in *Erwinia chrysanthemi* 3937 and other gamma-proteobacteria. *Microbiology* **150**: 3571–3590.

Rojas, C.M., Ham, J.H., Deng, W.-L., Doyle, J.J., and 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. *Proc. Natl. Acad. Sci. U. S. A.* **99**: 13142–7.

Ryan, R.P., Vorhölter, F.-J., Potnis, N., Jones, J.B., Van Sluys, M.-A., Bogdanove, A.J., and Dow, J.M. (2011) Pathogenomics of *Xanthomonas*: understanding bacterium-plant interactions. *Nat. Rev. Microbiol.* **9**: 344–55.

Sarris, P.F., Trantas, E.A., Mpalantinaki, E., Ververidis, F., and Goumas, D.E. (2012) *Pseudomonas viridiflava*, a multi host plant pathogen with significant genetic variation at the molecular level. *PLoS One* **7**: e36090.

Schandel, K.A., Muller, M.M., and Webster, R.E. (1992) Localization of TraC, a protein involved in assembly of the F conjugative pilus. *J. Bacteriol.* **174**: 3800–3806.

Scheller, H.V. and Ulvskov, P. (2010) Hemicelluloses. *Annu. Rev. Plant Biol.* **61**: 263–289.

Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* **30**: 2068-2069.

Shalom, G., Shaw, J.G., and Thomas, M.S. (2007) *In vivo* expression technology identifies a type VI secretion system locus in *Burkholderia pseudomallei* that is induced upon invasion of macrophages. *Microbiology* **153**: 2689–99.

da Silva, D.P., Castaneda-Ojeda, M.P., Moretti, C., Buonaurio, R., Ramos, C., and Venturi, V. (2014) Bacterial multispecies studies and microbiome analysis of a plant disease. *Microbiol. (United Kingdom)* **160**: 556–566.

Sriamornsak, P. (2003) Chemistry of pectin and its pharmaceutical uses : A review. *Silpakorn Univ. J. Soc. Sci. Humanit. Arts* **3**: 206–228.

Tampakaki, A.P., Skandalis, N., Gazi, A.D., Bastaki, M.N., Sarris, P.F., Charova, S.N., *et al.* (2010) Playing the harp: Evolution of our understanding of hrp/hrc genes. *Annu. Rev. Phytopathol.* **48**: 347–370.

Tapia-Pastrana, G., Chavez-Dueñas, L., Lanz-Mendoz, H., Teter, K., and Navarro-García, F. (2012) VirK is a periplasmic protein required for efficient secretion of plasmid-encoded toxin from enteroaggregative *Escherichia coli*. *Infect. Immun.* **80**: 2276–2285.

Toth, I.K., Bell, K.S., Holeva, M.C., and Birch, P.R.J. (2003) Soft rot erwiniae: From genes to

genomes. *Mol. Plant Pathol.* **4**: 17–30.

Toth, I.K. and Birch, P.R.J. (2005) Rotting softly and stealthily. *Curr. Opin. Plant Biol.* **8**: 424–429.

Toth, I.K., Pritchard, L., and Birch, P.R.J. (2006) Comparative genomics reveals what makes an enterobacterial plant pathogen. *Annu. Rev. Phytopathol.* **44**: 305–36.

Vakulskas, C. a., Potts, A.H., Babitzke, P., Ahmer, B.M.M., and Romeo, T. (2015) Regulation of bacterial virulence by Csr (Rsm) systems. *Microbiol. Mol. Biol. Rev.* **79**: 193–224.

Valls, M., Genin, S., and Boucher, C. (2006) Integrated regulation of the type III secretion system and other virulence determinants in *Ralstonia solanacearum*. *PLoS Pathog.* **2**: 0798–0807.

Venturi, V. and da Silva, D.P. (2012) Incoming pathogens team up with harmless "resident" bacteria. *Trends Microbiol.* **20**: 160–164.

Vinatzer, B. A, Monteil, C.L., and Clarke, C.R. (2014) Harnessing population genomics to understand how bacterial pathogens emerge, adapt to crop hosts, and disseminate. *Annu. Rev. Phytopathol.* **52**: 19–43.

Wang, Z., Gerstein, M., and Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**: 57–63.

Yamaguchi, Y. and Inouye, M. (2011) Regulation of growth and death in *Escherichia coli* by toxin–antitoxin systems. *Nat. Rev. Microbiol.* **9**: 779–790.

Yang, S., Peng, Q., San Francisco, M., Wang, Y., Zeng, Q., and Yang, C.-H. (2008) Type III secretion system genes of *Dickeya dadantii* 3937 are induced by plant phenolic acids. *PLoS One* **3**: e2973.

Zechner, E.L., Lang, S., and Schildbach, J.F. (2012) Assembly and mechanisms of bacterial type IV secretion machines. *Philos. Trans. R. Soc. B Biol. Sci.* **367**: 1073–1087.

Žgur-Bertok, D. (2013) DNA damage repair and bacterial pathogens. *PLoS Pathog.* **9**: e1003711.

Zhang, S. and Meyer, R. (1997) The relaxosome protein MobC promotes conjugal plasmid mobilization by extending DNA strand separation to the nick site at the origin of transfer. *Mol. Microbiol.* **25**: 509–516.

Zhao, Y., Sundin, G.W.S.W., and Wang, D. (2009) Construction and analysis of pathogenicity island deletion mutants of *Erwinia amylovora*. *Can. J. Microbiol.* **55**: 457–464.

CHAPTER 6

Metagenomic and metatranscriptomic analysis of AOD lesions reveals the *in situ* abundance and function of *Gibbsiella quercinecans* and *Brenneria goodwinii*

Abstract

Metagenomic and metatranscriptomic methodologies provide molecular access to the hidden world of microbiomes. These technologies allow sequencing of environmental DNA and RNA, providing data on microbiome community membership, functional potential, and activity. Here, established and custom bioinformatic techniques were applied to reveal the role of *Brenneria goodwinii* and *Gibbsiella quercinecans* in AOD lesion microbiomes. Within sequenced AOD lesion metagenomes, *G. quercinecans* and *B. goodwinii* had substantial relative functional abundance, which varied between microbiomes, with *B. goodwinii* having an average of 2225 coding domains, representing almost half of its total genomic coding capacity. *G. quercinecans* had a substantial but lesser functional abundance across lesion environments than *B. goodwinii*, with an average of 858 homologous coding domains in active oak lesions, around a fifth of its total coding capacity. Aligned coding domains from a healthy metagenome to *G. quercinecans* and *B. goodwinii* were almost completely absent, with a single coding domain, functionally annotated as a highly conserved ribosomal gene in *G. quercinecans*, and two coding domains, annotated as a hypothetical gene and a sulphate ABC transport permease commonly found across the *Enterobacteriaceae*, aligned to *B. goodwinii*. Metatranscriptome alignment from two active oak lesions, revealed substantial, genome-wide functional activity within *G. quercinecans* and *B. goodwinii*, including expression of key phtopathogenic virulence enzymes, such as the virulence regulator *virF*, the PCWDEs α-L-rhamnosidase and pectate disaccharide lyase in *B. goodwinii*, and an avirulence effector, virulence regulators *kdgR*, and *virK*, and the T6SS effector *Hcp* in *G. quercinecans*. Furthermore, virulence genes expressed *in vitro* described in chapter 5, correlate with many expressed transcripts recovered *in planta,* such as α-L-rhamnosidase, *virF, rsmA* in *B. goodwinii* and rhamnogalacturonate lysase, *kdgR*, *virK, phoP* and *hcp* in *G. quercinecans*. The data described here reveals high relative functional abundance and functional activity of *G. quercinecans* and *B. goodwinii* within necrotic AOD lesions.

6.1 Introduction

6.1.1 Metagenomics

Environmental genomics or metagenomics uncovers the taxonomic and functional diversity of a polymicrobial community within a defined environment (Wooley *et al.*, 2010). Metagenomics provides a cultivation-independent approach to identify the key organisms and enzymes which are metabolic drivers within an ecosystem (Handelsman *et al.*, 1998; Thomas *et al.*, 2012). Before the advent of metagenomics, cultivation based approaches and 16S rRNA community profiling returned only a fraction of the microbial diversity and presented minimal inference of community function (Thomas *et al.*, 2012), until two pioneering studies went beyond targeted sequencing of individual genomes and revealed the potential of shotgun sequencing, with surveys of community structure and function in the Sargasso Sea (Venter *et al.*, 2004) and an environmental acidic biofilm (Tyson *et al.*, 2004). These proof of principle studies used novel bioinformatic methods to illuminate the huge microbial diversity within previously hidden worlds (Gilbert and Dupont, 2011). A wealth of information was provided, encouraging the subsequent widespread uptake of metagenomics, which has produced wide-ranging characterisation of the microbiome, including phylogenetic composition, known and novel functional types, adaptation to specific environments and the distribution of gene families across ecosystems (Knight *et al.*, 2012).

There are numerous bioinformatic tools to analyse metagenomic data and in comparison to other 'omic' technologies, the methodology could be regarded as mature (McMurdie and Holmes, 2014). Metagenomic sequence reads can either be assembled or directly catalogued using the raw sequencing output (reads) to produce taxonomic and functional profiles of the community. Typically, the first step after sequencing and quality control is assembly; common assemblers include RAY-meta (Boisvert *et al.*, 2012) and MetaVelvet (Namiki *et al.*, 2012). Assembled contigs can be taxonomically and functionally annotated using many freely available programs such as Prokka (Seemann, 2014), which is a customisable Unix based annotation program that predicts coding domains and functional annotations through hidden markov models and sequence homology. An alternative annotation method is MG-RAST, which is an open source interactive metagenomic annotation pipeline, providing an excellent starting point for characterisation of assembled data and contextualisation of results (Meyer *et al.*, 2008). Using standalone BLAST (Basic Local Alignment Search Tool) (Altschul *et al.*, 1990), assembled and structurally annotated metagenome datasets can be queried against locally installed databases,

this bioinformatically ancient algorithm, remains one of the leading methods to characterise the relative abundance and functional profile of bacteria within a metagenome. These methods allow the taxonomic and functional goals of metagenomics to be addresses, i.e. (1) what is in the community? and (2) what functional potential do they have (Sharpton, 2014)?

6.1.2 Metatranscriptomics

RNA-sequencing (RNA-seq) of transcripts recovered from the natural environment is a metatranscriptomic method, which enables the study of gene expression within a complex microbiome, providing a real-time snap shot of microbial activity and function that cannot be captured through metagenomic gene inventories (Moran *et al.*, 2013; Reck *et al.*, 2015). Metagenomic analysis provides insights of potential community activity, the gap between potential and actual activity can be bridged through metatranscriptomics (Vieites *et al.*, 2009; Mason *et al.*, 2012). Due to the inherent difficulties in laboratory processing of bacterial transcripts, metatranscriptomic methodologies are technically challenging, as bacterial RNA degrades rapidly, with an estimated half-life of around 5 minutes in *E. coli*, independent of growth rate (Moran *et al.*, 2013). The rapid turnover of mRNA molecules is catalysed by an assortment of RNases which constitute the RNA degradosome (Deutscher, 2006). Therefore, the extraction of total RNA and subsequent ribodepletion to remove ribosomal RNA (rRNA) is time sensitive, adding substantial pressure on fieldwork and laboratory processing. Cellular total RNA composition is predominantly rRNA, which constitutes around 96% of total cellular RNA, with mRNA making up around ~4% (Gilbert, 2010). Furthermore, it has been estimated that transcript abundance within active environmental bacterial isolates is substantially lower than those grown *in vitro*, with ~1800 transcripts in laboratory grown isolates compared to ~200 in their environmental equivalents (Alberti *et al.*, 2014). This adds a further difficulty to RNA-seq, as samples should be depleted of rRNA prior to sequencing, to prevent rRNA overwhelming the sample and limiting results. There are standardised protocols for the processing of inherently unstable mRNA, however workflows for the analysis of resultant sequencing output lack standardisation, which results in ad-hoc methods with low reproducibility of resultant data between labs (Westreich *et al.*, 2016).

In this chapter, metagenomic and metatranscriptomic methods are used to identify the taxonomic and functional relationships between *Gibbsiella quercinecans*, *Brenneria goodwinii*, and AOD lesion microbiomes. Whole genome sequencing and annotation of *G. quercinecans*

and *B. goodwinii* was described in chapter 3 and chapter 4, and *in vitro* alignment of expressed virulence genes against the genomes of *G. quercinecans* FRB97 and *B. goodwinii* FRB141 was described in chapter 5. Here, necrotic lesion community coding domains and expressed transcripts were aligned against the genomes of *Gibbsiella quercinecans* FRB97 and *Brenneria goodwinii* FRB141, revealing the relative functional presence and activity of each species within necrotic lesions from AOD affected trees.

6.2 Methods

6.2.1 Sampling

Sampling of oak tree microbiomes was carried out by Dr Emma Ransom-Jones and Dr James McDonald, as follows. In total, six active lesions, one asymptomatic zone of a symptomatic tree and one healthy tree samples were collected for DNA extraction and metagenome analysis. Samples AH, AES, AMS, AMA and ALS were collected from Attingham Park (OS Eastings 356033, Northings 310372), November 2013; samples RES, RMS and RLS were collected from Runs Wood (OS Eastings 563207, Northings 310858), November 2013. Two lesion samples were collected from a tree with active AOD bleeds for RNA extraction. For RNA sampling, two separate lesions from a single tree were sampled from Attingham Park (OS Eastings 356033, Northings 310372), June 2013.

6.2.2 DNA and RNA extraction

Metagenomic DNA and metatranscriptomic RNA was extracted from lesion and heathy oak tissue samples by Dr Emma Ransom-Jones and Dr James McDonald.

For metagenomic DNA extractions, sections of inner-bark, sapwood and phloem were frozen in liquid nitrogen and ground in a pestle and mortar to homogenise the tissue. DNA from replicate 0.6 g samples of homogenized tissue were extracted in triplicate using the Power Soil DNA Isolation kit (MoBio). DNA extracts for each sample were subsequently pooled prior to oak DNA depletion using the NEBNext Microbiome DNA Enrichment Kit (New England Biolabs), with the exception of the non-symptomatic samples which were not enriched (AMA and AH). Samples were purified using Agencourt AMPure XP (Beckman Coulter) beads according to the manufacturer's protocol. DNA was quantified using a Qubit fluorometer (Thermo Fisher).

RNA was extracted by removing the outer bark and swabs of the lesion fluid were collected in addition to tissue from the active margins of the lesion, and gallery tissue. Swabs and tissue samples were immediately frozen in liquid nitrogen and transported back to the laboratory in a vessel containing liquid nitrogen. Samples were stored at -80°C prior to processing. Prior to RNA extraction, samples frozen in liquid nitrogen were ground in a pestle and mortar to homogenise the tissue. RNA was extracted using the PowerSoil Total RNA Isolation Kit (MoBio) according

to the manufacturer's instructions. RNA was quantified using a Qubit florometer (Thermo Fisher).

6.2.3 Metagenomic and metatranscriptomic sequencing

Metagenomic and metatranscriptomic sequencing and post-sequencing quality control was carried out by the Centre for Genomic Research (CGR), University of Liverpool.

Metagenome sequencing libraries were prepared from samples using the Nextera XT Library preparation kit (Illumina), and sequenced using 2x100bp paired-end sequencing on the Illumina HiSeq platform.

Metatranscriptome libraries were prepared using the strand-specific ScriptSeq preparation kit (Illumina, Inc), and sequenced using 2x100bp paired-end sequencing on the Illumina HiSeq platform.

Sequencing output for all reads were trimmed using first Cutadapt v1.2.1 (Martin, 2011) and additionally Sickle v1.2.00 (Joshi and Fass, 2011).

6.2.4 Bioinformatic analysis

General bioinformatic analyses were carried out on a locally installed Bio-Linux 8 workstation (Field *et al.*, 2006) and applications requiring high computing power were undertaken on the HPC Wales supercomputing network.

6.2.5 Metagenome analysis

Metagenomic reads were assembled using RAY-meta (Boisvert *et al.*, 2012). Assemblies were annotated using Prokka (Seemann, 2014). Metagenomic genes were aligned against whole genome annotations of *G. quercinecans* FRB97 and *B. goodwinii* FRB141 genes using BLASTx (Altschul *et al.*, 1990). Metagenomic protein sequences with greater than or equal to 97% homology for at least 50 amino acids within *G. quercinecans* FRB97 and *B. goodwinii* FRB141 were considered a match (Table 6.1). A custom Perl script was designed to extract matches.

6.2.6 Metatranscriptome analysis

Complete metatranscriptomic RNA from two AOD lesions, extracted from the same oak tree, were sequenced and *in silico* depleted of rRNA by the Centre for Genomic Research (CGR), University of Liverpool. Metatranscriptome samples were combined and aligned to the *G. quercinecans* FRB97 and *B. goodwinii* FRB141 whole genomes using Bowtie2 v2.02 (Langmead *et al.*, 2009), using local mode to maximise alignment score. Aligned reads were converted from SAM to BAM format and indexed using SAMtools v1.2 (Li *et al.*, 2009). A gene was considered as expressed if 3 or more transcripts aligned to the gene and had greater than 20% coverage of the gene (as described in (Versluis *et al.*, 2015)). A custom Perl script was designed to extract metatranscriptome hits. Aligned transcripts were visualised in Artemis (Carver *et al.*, 2012).

6.3 Results/Discussion

6.3.1 Extraction and alignment of oak lesion metagenomes to *G. quercinecans* and *B. goodwinii*

Using whole genome sequencing, the *in silico* virulence gene catalogues of *G. quercinecans* FRB97 and *B. goodwinii* FRB141 were identified and compared against canonical bacterial phytopathogens. The results of this work were reported in Chapters 3 and 4. The *in silico* genomic annotation of virulence factors demonstrates putative pathogenicity but not disease causation. To uncover the functional importance of *G. quercinecans* and *B. goodwinii* in active necrotic lesions, recovered metagenomic DNA was aligned against *G. quercinecans* and *B. goodwinii*, thereby revealing their relative functional abundance within the metagenome. Resultant data revealed extensive functional homology between *G. quercinecans* and *B. goodwinii*, and *in silico* reconstructed AOD lesion microbiomes from various sampling sites (Table 6.1). Symptomatic trees had an average of 858 homologous coding domains with *G. quercinecans* FRB97 (Fig. 6.1), and 2225 homologous coding domains with *B. goodwinii* FRB141 (Fig. 6.2). The asymptomatic zone of an affected tree had 526 homologous coding domains with *G. quercinecans* FRB97 and 3929 with *B. goodwinii* FRB141, suggesting that affected trees have a putative pathogenic microbial component outside the necrotic zone. The healthy tree had one homologous coding domain with *G. quercinecans* FRB97, which was annotated as a ribosomal protein, and two homologous coding domains with *B. goodwinii* FRB141, annotated as a hypothetical protein and a sulphate ABC transport permease found in many *Enterobacteriaceae*. This suggest that *G. quercinecans* and *B. goodwinii* are absent in AOD asymptomatic trees. However, an extensive spatial and temporal metagenomic sampling survey, to validate the environmental abundance and natural reservoirs of *G. quercinecans* and *B. goodwinii* is required to confirm this result.

Relative functional abundance measured through homologous alignment of *G. quercinecans* and *B. goodwinii*, to the lesion metagenomes varied between extraction sites. Metagenomes were extracted on trees in various stages of Decline, these were designated as early-stage, mid-stage or late-stage and from two sampling sites, Attingham and Runs Wood. *G. quercinecans* had its highest relative functional abundance in the Attingham early-stage, symptomatic (AES) metagenome. Contrastingly, second highest functional abundance of *G. quercinecans* was in Runs Wood late-stage symptomatic (RLS) metagenome. This suggests a randomised distribution of *G. quercinecans* at different stages of AOD, but assignment of trees with specific stages of Decline is ambiguous and not thought to provide definitive conclusions of tree health (S.

Denman, personal communication). As described in chapter 5, *G. quercinecans* expressed most virulence genes within phloem culture at 2 hours; early stage virulence gene expression may correlate with its relative functional abundance in the AES metagenome, or it may be that similar to the SRE, *G. quercinecans* expresses virulence genes independent of disese state and induced by pectin availability. (Flego *et al.*, 1997). However, biological correlations are limited as the exact stage of Decline is qualitatively expressed and is substantially removed from the sterile conditions of the *in vitro* transcriptomic study. The highest homology matches from AOD affected lesions to *B. goodwinii*, were in the AMS, ALS, RES, and RLS lesion microbiomes, suggesting a randomised distribution to the tree disease state, with early, late, and mid all represented, however as previously mentioned the health state descriptions are not currently measurable and may misinterpret the stage of Decline.

Two additional bacterial genomes were measured for relative abundance using the homologous coding domain alignment method. The selected bacteria, *Paenibacillus polymyxa* SC2 and *Bacillus licheniformis* ATCC14580, have been identified in DNA extracts in low relative abundance within symptomatic and healthy oak tree metagenomes. This finding was revealed in a subsequent unpublished metagenomic study (M. Broberg, personal communication). The low relative abundance of these species is in contrast to the high relative abundance of *G. quercinecans* and *B. goodwinii*. *P. polymyxa* SC2 and *B. licheniformis* ATCC14580 were therefore used to test alignment efficiency. There was minimal alignment between metagenomic coding domains and *P. polymyxa* SC2 (Appendix VII). The second control *B. licheniformis* ATCC14580 was completely absent from the metagenome (Appendix VIII), verifying the relative abundce of *G. quercinecans* and *B. goodwinii*, but also that the method could be improved, with increased sensitivity to include low abundance species.

Table 6.1. Alignment of AOD lesion metagenome coding domains to the genomes of *Gibbsiella quercinecans* FRB97 and *Brenneria goodwinii* FRB141. Numbers represent homologous coding domain alignments between assembled metagenomic libraries from AOD symptomatic trees, a non-symptomatic tree, and an asymptomatic zone of a symptomatic tree and assembled bacterial coding domains of *B. goodwinii* FRB141 (T) and *G. quercinecans* FRB97 (T), and two controls *P. polymyxa* SC2 and *B. licheniformis* ATCC14580. AES – Attingham early symptomatic, AMS – Attingham midstage symptomatic, ALS – Attingham late symptomatic, AH – Attingham healthy, AMA – Attingham midstage asymptomatic, RES – Runs Wood early symptomatic, RMS – Runs Wood midstage symptomatic, RLS – Runs Wood late symptomatic.

Metagenome	Total genes	*G. quercinecans* FRB97	*B. goodwinii* FRB141	*Paenibacillus polymyxa SC2*	*Bacillus licheniformis ATCC14580*
AES	216140	2507	56	1	0
AMS	192147	1493	3838	6	0
ALS	227496	72	1001	0	0
AH	58146	1	2	0	0
AMA	105148	526	3929	7	0
RES	211149	61	4181	1	0
RMS	195109	23	428	5	0
RLS	196863	993	3846	8	0

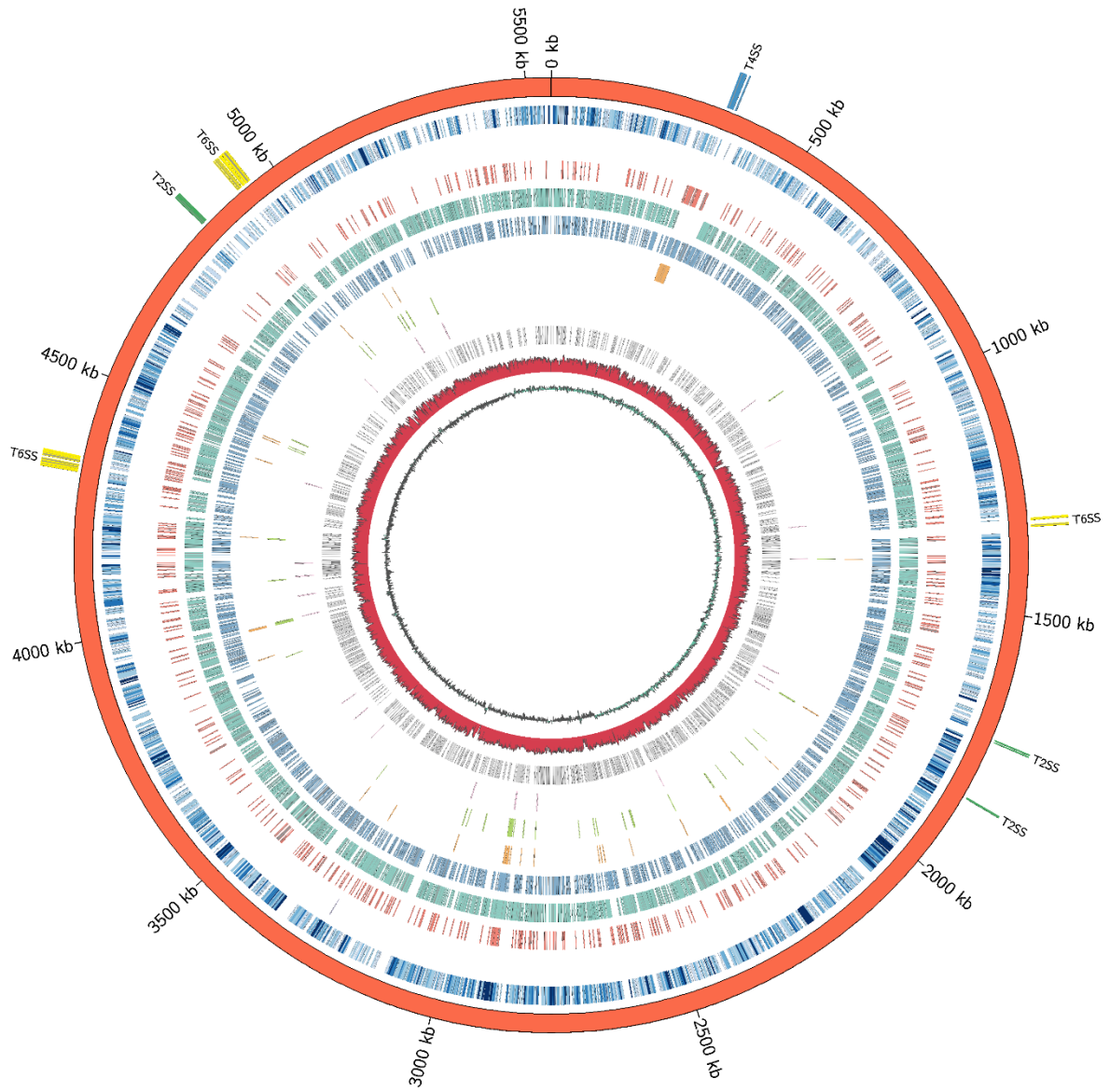


Figure 6.1. AOD metagenome coding domains and metatranscripts aligned against the *Gibbsiella quercinecans* FRB97 genome. Encoded secretion system genes are depicted at their genomic loci. The outermost circle represents a circular reconstruction of the *G. quercinecans* genome. The following circles represent, from outside to inside: 1) metatranscriptome heatmap, with increased blue saturation representing increasing read alignments, 2) Attingham healthy (AH) aligned metagenome coding domains, 3) Attingham mid asymptomatic (AMA) aligned metagenome coding domains, 4) Attingham early symptomatic (AES) aligned metagenome coding domains, 5) Attingham mid-stage symptomatic (AMS) aligned metagenome coding domains, 6) Attingham late symptomatic (ALS) aligned metagenome coding domains, 7) Runs Wood early symptomatic (RES) aligned metagenome coding domains, 8) Runs Wood mid symptomatic (RMS) aligned metagenome coding domains, 9) Runs Wood late symptomatic (RLS) aligned metagenome coding domains, 10) G+C content across the *G. quercinecans* genome, 11) G+C skew across the *G. quercinecans* genome.

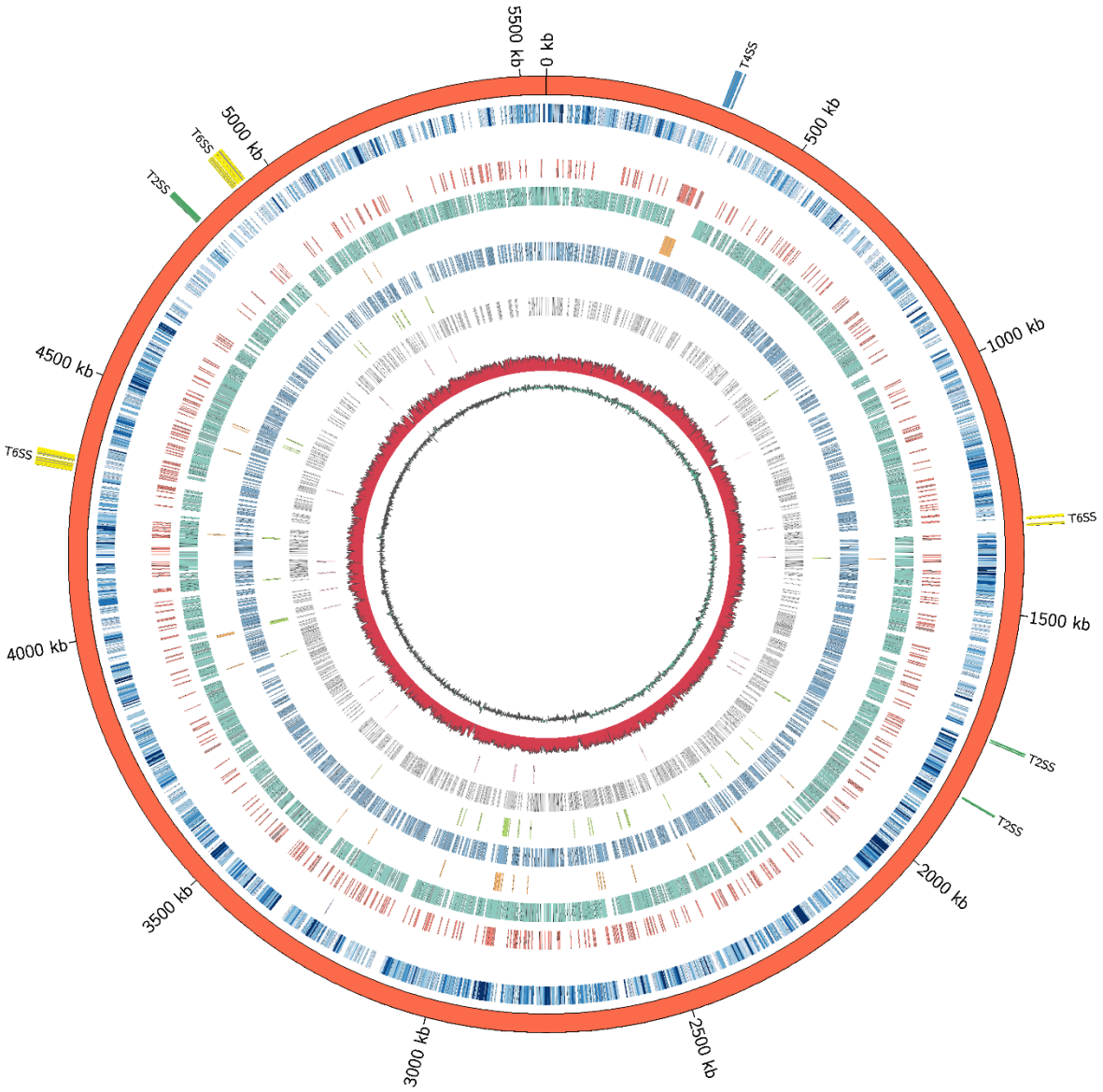


Figure 6.2. AOD metagenome coding domains and metatranscripts aligned against the *Brenneria goodwinii* FRB141 genome. Encoded secretion system genes are depicted at their genomic loci. The outermost circle represents a circular reconstruction of *B. goodwinii* genome. The following circles represent, from outside to inside: 1) metatranscriptome heatmap, with increased blue saturation representing increasing read alignment, 2) Attingham healthy (AH) aligned metagenome coding domains, 3) Attingham mid asymptomatic (AMA) aligned metagenome coding domains, 4) Attingham early symptomatic (AES) aligned metagenome coding domains, 5) Attingham mid-stage symptomatic (AMS) aligned metagenome coding domains, 6) Attingham late symptomatic (ALS) aligned metagenome coding domains, 7) Runs Wood early symptomatic (RES) aligned metagenome coding domains, 8) Runs Wood

mid symptomatic (RMS) aligned metagenome coding domains, 9) Runs Wood late symptomatic (RLS) aligned metagenome coding domains, 0) G+C content across the *B. goodwinii* genome, 11) G+C skew across the *B. goodwinii* genome.

6.3.2 Extraction and alignment of lesion metatranscriptomic sequences to the genomes of *Gibbsiella quercinecans* FRB97 and *Brenneria goodwinii* FRB141

Two metatranscriptome datasets were recovered from two AOD lesions present on the same tree. These expression datasets were combined *in silico* and aligned against the genomes of *G. quercinecans* FRB97 and *B. goodwinii* FRB141, allowing the analysis of transcript alignment against coding domains and particularly, virulence genes. The alignment method used two parameters to measure gene expression; (1) number of transcript hits against a gene (Fig. 6.3) and (2) the percentage of the total length of the gene covered by the transcripts (Fig. 6.4). Gene expression analysis (for definition of expression as applied here see section 6.2.6) revealed the presence of key virulence factors from both species in the metatranscriptome of AOD lesions.

G. quercinecans FRB97 expressed polygalacturonase (EC 3.2.1.15), a widely studied group of pectinases utilised by many phytopathogens, which are secreted upon host contact and split the cellulose/hemicellulose chains of the primary cell wall, releasing nutrients for absorption (Juge, 2006), rhamnogalacturonate lyase (EC 4.2.2.23) a well characterised PCWDE in *D. dadantii* which degrades the pectin backbone (Laatu and Condemine, 2003), and cellobiose degrading β-glucosidases (EC 3.2.1.21 & EC 3.2.1.86), which hydrolyse the cellulose intermediate cellobiose to produce glucose (Zamocky *et al.*, 2006). This data is critical to the understanding of putative pathogenicity of *G. quercinecans* within AOD, as the expressed PCWDEs within a necrotic lesion are likely to damage the host.

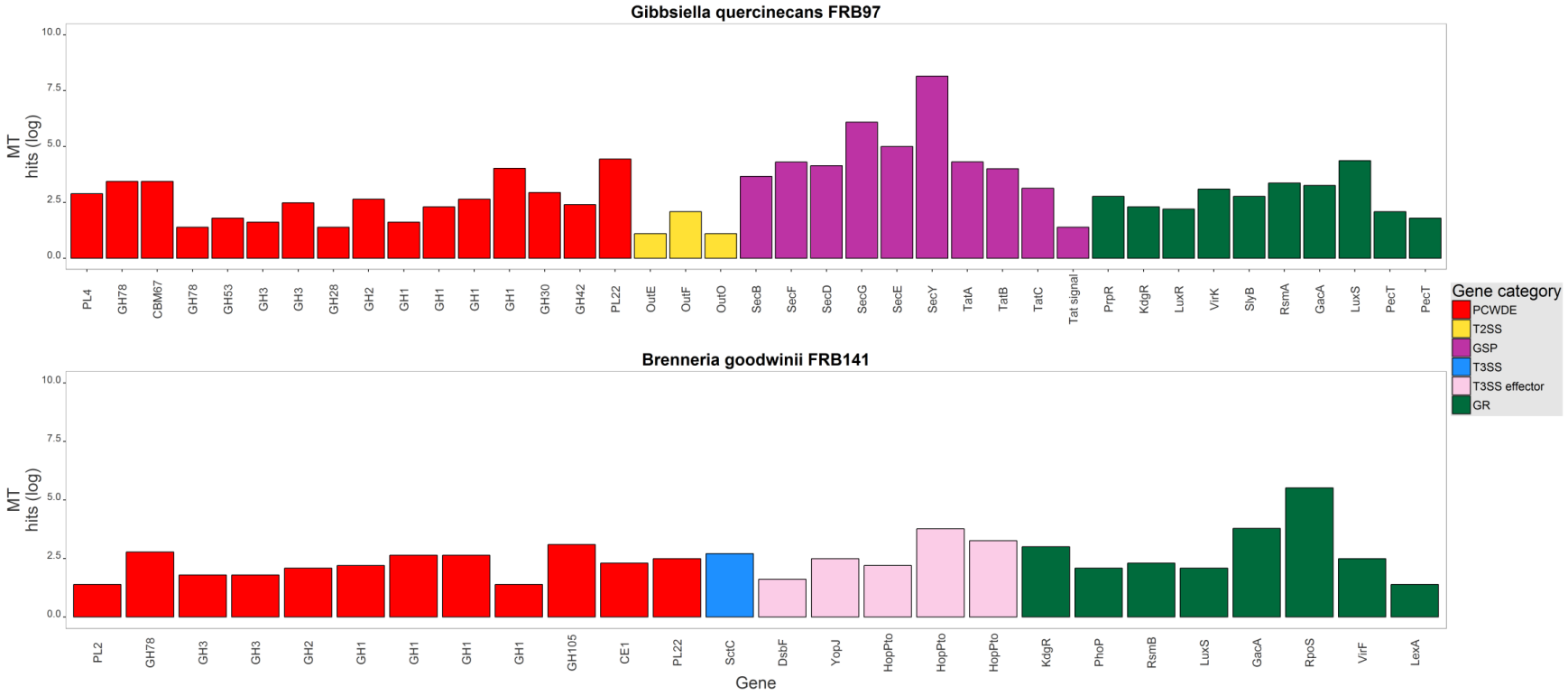


Figure 6.3. Metatranscripts aligned against virulence genes encoded within *Gibbsiella quercinecans* FRB97 and *Brenneria goodwinii* FRB141 genomes. MT – metatranscriptome. Gene categories are represented by the following colours, red - plant cell wall degrading enzymes (PCWDEs), purple - general secretory pathway (GSP), yellow - type II secretion system (T2SS), blue - type III secretion system (T3SS), pink - type III secretion system effectors (T3SS effectors), and green - global regulators (GR).

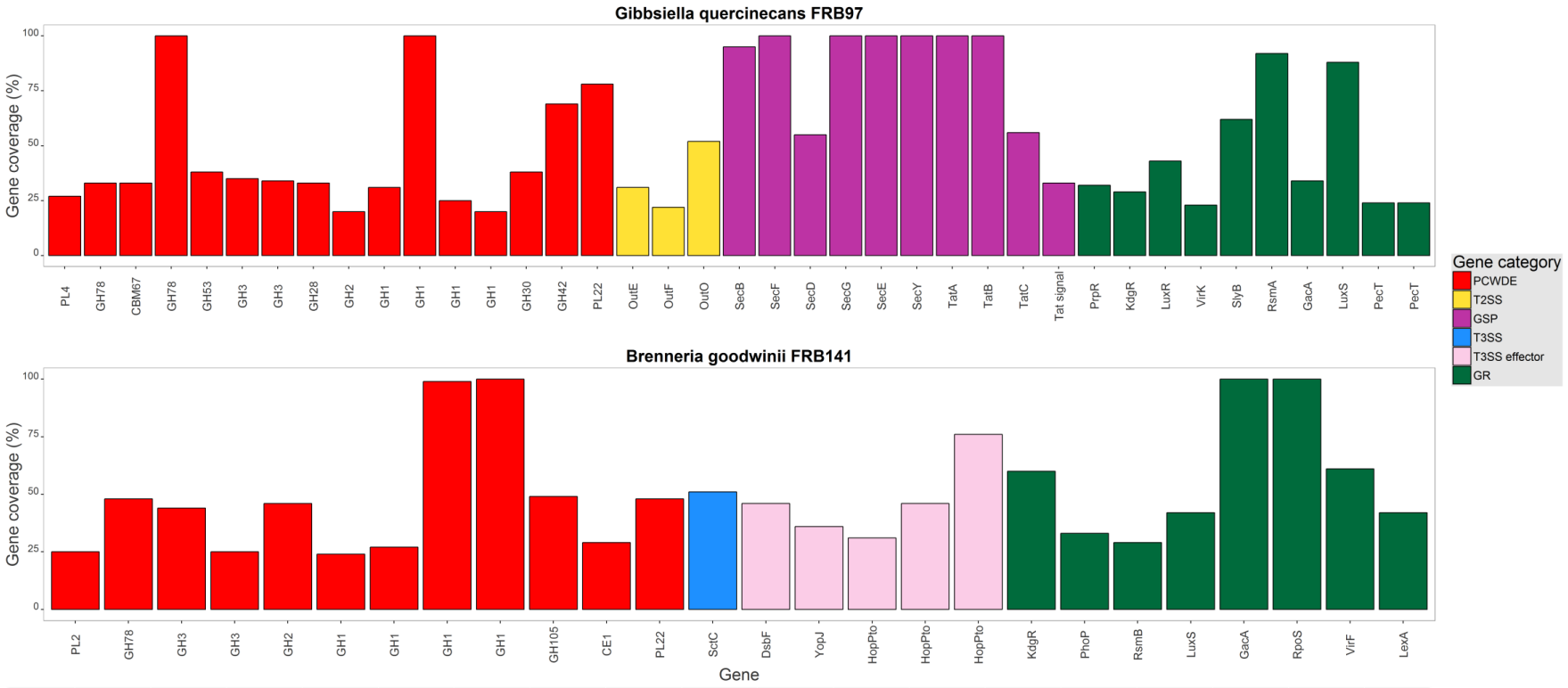


Figure 6.4. Gene coverage of aligned transcripts against virulence genes encoded within *Gibbsiella quercinecans* FRB97 and *Brenneria goodwinii* FRB141 genomes. Gene categories are represented by the following colours, red - plant cell wall degrading enzymes (PCWDEs), purple - general secretory pathway (GSP), yellow - type II secretion system (T2SS), blue - type III secretion system (T3SS), pink - type III secretion system effectors (T3SS effectors), and green - global regulators (GR).

B. goodwinii FRB141 expressed the key virulence PCWDE pectate disaccharide-lyase (EC 4.2.2.9), a pectinase found in *D. dadantii* (Shevchik *et al.*, 1999) and oligogalacturonide lyase (EC 4.2.2.6), which hydrolyses polygalacturonic acid, the by-product of pectin lyase (Moran *et al.*, 1968), and further virulence enzymes including 6-phospho-β-glucosidase (EC 3.2.1.86) and a β-galactosidase (EC 3.2.1.23) with activity against the hemicellulose component xyloglucan (Benko *et al.*, 2008). Despite expressing numerous PCWDEs, only a partial operon for T2SS within the *B. goodwinii* FRB141 genome was identified, therefore the mechanism for PCWDE secretion is unknown, but it is possibly due to alternative strain hosting the T2SS and strain FRB141 having an attenuated virulence potential, for example *B. goodwinii* FRB171 which as chapter 3 reveals, has a complete T2SS operon. One possible alternative translocation mechanism is through the T6SS operon which is most frequently involved in inter-bacterial competition, and can secrete antibacterial glycoside hydrolases, but it is also known to target eukaryotes and knock-out infection models produce attenuated virulence in *P. aeruginosa* and *Vibrio cholera* (Russell *et al.*, 2014). Intriguingly the Sec translocon (one of two possible inner membrane export systems for which the outer membrane T2SS is dependent), was highly expressed in both *G. quercinecans* FRB97, as could be expected due to the presence of a complete T2SS operon, but also in *B. goodwinii* FRB141, thereby providing a mysterious cargo and purpose. Key phytopathogenic enzymes - T3SS effectors, encoded within *B. goodwinii* such as *yopJ* and *hopPto*, were significantly expressed (Fig. 6.3), however the relatively low expression of these enzymes may refelect their role in host manipulation, prior to latent release of degradative enzymes, therefore the large scale expression of T3SS effectors may have occurred prior to the onset of a necrotrophic state (Tampakaki *et al.*, 2010). These data reveal evidence of a contribution to virulence for *G. quercinecans* FRB97 and *B. goodwinii* FRB141, within AOD affected oak trees, with virulence mechanisms comparable to that of their close relatives the SRE, with a similar opportunistic, hemibiotrophic lifestyle and when appropriate conditions arise they participate in the recruitment-destruction (RD) mode of attack.

6.3.3 Correlations between transcriptomic and metatranscriptomic datasets

Cominbing datasets between related experiments increases the validity and accuracy of conclusions. Here, datasets from *in vitro* transcriptomics reported in chapter 5 and *in planta* metatranscriptomic datasets were combined to increase the likelihood of uncovering key virulence factors involved in oak tissue degradation. Virulence regulators are important indicatiors of bacterial activity and reveal internal reactions to external stimuli (Caes and de

Lorenzo., 2005). Differential expression of regulators in response to oak tissue was described in section 5.3.4.4 for *G. quercinecans* and in section 5.3.53 for *B. goodwinii*. Within *G. quercinecans* the virulence regulator *virK,* which is the principal regulator of intracellular spreading in *S. flexneri*, was downregulated at 2 HPI when *G. quercinecans* was cultured axenically in phloem, indicating its involment in repression of virulence. Similarly, a key regulator of virulence within *Yersinia*, *virF* which exacerbates expression of T3SS effectors was downregulated in the most virulent profiles of *B. goodwinii,* including axenic culture at 2 HPI in sapwood. Both *virK* in *G. quercinecans* and *virF* in *B. goodwinii* were recovered from active oak lesions and aligned against the respective genomes. Annotated regulators of pectionlysis, including *slyB*, *pecT*, and *kdgR* in *G. quercinecans*, and *rsmA* in *B. goodwinii* were differentially expressed *in vitro* and significantly expressed *in planta*, indicating that they are involved in the regulation of virulence both *in vitro* and *in planta*. Across *in vitro* and *in planta* datasets expression of necrotic enzymes and T3SS effectors was recorded. Within both datasets expressed PCWDEs included pectate lyase (EC 4.2.2.2) in *G. quercinecans*, and pectate disaccharide lyase (EC 4.2.2.9) in *B. goodwinii*. These synergistic results show consistency in the data and further increase the weight of evidence suggesting that 1) *G. quercinecans* and *B. goodwinii* are active in a necrotic lesion of an AOD affected tree, 2) *G. quercinecans* and *B. goodwinii* release virulence enzymes in response to tree tissue and in the lesion environment and, 3) *G. quercinecans* and *B. goodwinii* actively macerate tree tissue.

6.4 Conclusions

Metagenomics and metatranscriptomics are information rich technologies which enable large scale insights on fundamental aspects of the microbial world. This chapter reveals substantial relative abundance of *G. quercinecans* and *B. goodwinii* within necrotic AOD lesions and their near absence from a healthy oak microbiome. Additionally, metatranscriptomics uncovered the active expression of a substantial proportion of the *G. quercinecans* and *B. goodwinii* genome within necrotic lesions, including numerous virulence genes. Furthermore, many *in planta* expressed virulence genes correlated with *in vitro* cultures of *G. quercinecans* and *B. goodwinii*, which were upregulated in the presence of oak tissue. This chapter uncovers the substantial presence and activity of *G. quercinecans* and *B. goodwinii* in active oak lesions, indicating their central importance to AOD and their likely contribution to tissue necrosis.

6.5 References

Alberti, A., Belser, C., Engelen, S., Bertrand, L., and Orvain, C. (2014) Comparison of library preparation methods reveals their impact on interpretation of metatranscriptomic data. *BMC Genomics*. **15** 1–13.

Altschul, S.F., Gish, W., Miller, W., Myers, E., and Lipman, D.J. (1990) Basic Local Alignment Search Tool. *J. Mol. Biol.* **215**: 403–410.

Benko, Z., Siika-aho, M., Viikari, L., and Reczey, K. (2008) Evaluation of the role of xyloglucanase in the enzymatic hydrolysis of lignocellulosic substrates. *Enzyme Microb. Technol.* **43**: 109–114.

Boisvert, S., Raymond, F., Godzaridis, E., Laviolette, F., and Corbeil, J. (2012) Ray Meta: scalable *de novo* metagenome assembly and profiling. *Genome Biol.* **13**: R122.

Cases, I. de Lorenzo, V. Promoters in the environment: Transcriptional regulation in its natural context. *Nat. Rev. Microbiol.* **3**: 105-118.

Carver, T., Harris, S.R., Berriman, M., Parkhill, J., and McQuillan, J.A. (2012) Artemis: An integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* **28**: 464–469.

Deutscher, M.P. (2006) Degradation of RNA in bacteria: Comparison of mRNA and stable RNA. *Nucleic Acids Res.* **34**: 659–666.

Field, D., Tiwari, B., Booth, T., Houten, S., Swan, D., Bertrand, N., and Thurston, M. (2006) Open software for biologists: from famine to feast. *Nat. Biotechnol.* **24**: 801–803.

Flego, D., Pirhonen, M., Saarilahti, H., Palva, T.K., and Palva, E.T. (1997) Control of virulence gene expression by plant calcium in the phytopathogen *Erwinia carotovora*. *Mol. Microbiol.* **25**: 831–838.

Gilbert, J.A. and Dupont, C.L. (2011) Microbial metagenomics: beyond the genome. *Ann. Rev. Mar. Sci.* **3**: 347–371.

Gilbert, J. A. (2010) Beyond the infinite - tracking bacterial gene expression. *Microbiol. Today* **37**: 82–85.

Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., and Goodman, R.M. (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.* **5**: R245–R249.

Joshi, N. and Fass, J. (2011) Sickle. A sliding-window, adaptive, quality-based trimming tool for FastQ files. (Version 1.33) [Software]. Available at https://github.com/najoshi/sickle.

Juge, N. (2006) Plant protein inhibitors of cell wall degrading enzymes. *Trends Plant Sci.* **11**: 359–367.

Knight, R., Jansson, J., Field, D., Fierer, N., Desai, N., Fuhrman, J. A., *et al.* (2012) Unlocking the potential of metagenomics through replicated experimental design. *Nat. Biotechnol.* **30**: 513–520.

Laatu, M. and Condemine, G. (2003) Rhamnogalacturonate lyase rhiE is secreted by the out system in *Erwinia chrysanthemi*. *J. Bacteriol.* **185**: 1642–1649.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**: R25.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**: 10.

Mason, O.U., Hazen, T.C., Borglin, S., Chain, P.S., Dubinsky, E. A, Fortney, J.L., *et al.* (2012) Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill. *ISME.* **6**: 1715–1727.

McMurdie, P.J. and Holmes, S. (2014) Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput. Biol.* **10**: e1003531.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E., Kubal, M., *et al.* (2008) The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**: 386.

Moran, F., Seiichi, N., and Starr, M. (1968) Oligogalacturonide trans-Eliminase of *Erwinia carotovora*. *Arch. Biochem. Biophys.* **125**: 734–741.

Moran, M.A., Satinsky, B., Gifford, S.M., Luo, H., Rivers, A., Chan, L.-K., *et al.* (2013) Sizing up metatranscriptomics. *ISME J.* **7**: 237–243.

Namiki, T., Hachiya, T., Tanaka, H., and Sakakibara, Y. (2012) MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Res.* **40**: e155.

Reck, M., Tomasch, J., Deng, Z., Jarek, M., Husemann, P., and Wagner-Döbler, I. (2015) Stool metatranscriptomics: A technical guideline for mRNA stabilisation and isolation. *BMC Genomics* **16**: 494.

Russell, A.B., Peterson, S.B., and Mougous, J.D. (2014) Type VI secretion system effectors: poisons with a purpose. *Nat. Rev. Microbiol.* **12**: 137–48.

Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* **30**: 2068-2069.

Sharpton, T.J. (2014) An introduction to the analysis of shotgun metagenomic data. *Front. Plant Sci.* **5**: 1–14.

Shevchik, V.E., Kester, H.C.M., Benen, J. A E., Visser, J., Robert-Baudouy, J., and Hugouvieux-Cotte-Pattat, N. (1999) Characterization of the exopolygalacturonate lyase PelX of *Erwinia chrysanthemi* 3937. *J. Bacteriol.* **181**: 1652–1663.

Tampakaki, A.P., Skandalis, N., Gazi, A.D., Bastaki, M.N., Sarris, P.F., Charova, S.N., *et al.* (2010) Playing the Harp: Evolution of our understanding of hrp/hrc genes. *Annu. Rev. Phytopathol.* **48**: 347–370.

Thomas, T., Gilbert, J., and Meyer, F. (2012) Metagenomics - a guide from sampling to data analysis. *Microb. Inform. Exp.* **2**: 3.

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., *et al.* (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**: 37–43.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J. A, *et al.* (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**: 66–74.

Versluis, D., Leimena, M.M., Garcia, J.R., and Andrea, M.M.D. (2015) Mining microbial metatranscriptomes for expression of antibiotic resistance genes under natural conditions. *Scient. Rep.* **5**: 1–10.

Vieites, J.M., Guazzaroni, M.-E., Beloqui, A., Golyshin, P.N., and Ferrer, M. (2009) Metagenomics approaches in systems microbiology. *FEMS Microbiol. Rev.* **33**: 236–55.

Westreich, S.T., Korf, I., Mills, D.A., and Lemay, D.G. (2016) SAMSA : a comprehensive metatranscriptome analysis pipeline. *BMC Bioinf.* **17**: 1-20.

Wooley, J.C., Godzik, A., and Friedberg, I. (2010) A primer on metagenomics. *PLoS Comput. Biol.* **6**: 1–13.

Zamocky, M., Ludwig, R., Peterbauer, C., Hallberg, B.M., Divne, C., Nicholls, P., and Haltrich, D. (2006) Cellobiose dehydrogenase--a flavocytochrome from wood-degrading, phytopathogenic and saprotropic fungi. *Curr. Protein Pept. Sci.* **7**: 255–280.

CHAPTER 7

Discussion

7.1 Introduction

Oak woodlands are central to the British ecosystem, and are an intrinsic component of the United Kingdom's national identity. The obdurate structure of oak enables the tree to survive for hundreds of years. The oak tissue components, lignin, cellulose, hemicellulose, pectin and tannin, form a recalcitrant structure, preventing transgression of invading pathogens. However, British oak trees are threatened by a recently described Decline-disease, acute oak decline (AOD). There has been a previous outbreak of AOD (Day, 1927), however the current episode has persisted for a longer period, increasing in incidence and range, as it spreads from the East of England to the Welsh border (Brown *et al.*, 2016). The aetiology of AOD follows the classic Decline-disease spiral (Fig. 1.6), with many interacting factors and no clear primary pathogenic cause. Two species of bacteria have been consistently isolated from necrotic lesions on affected oaks, *Gibbsiella quercinecans* and *Brenneria goodwinii* (Brady *et al.*, 2010; Denman *et al.*, 2012b). Upon isolation, molecular phylogenetic approaches, using the 16S rRNA and DNA gyrase B (*gyrB*) genes positioned *G. quercinecans* next to *Serratia* (an opportunistic pathogen) but outside the *Serratia* genus, and *B. goodwinii* as a novel species within the genus *Brenneria* (a genus in which all of the described members are plant pathogens). Therefore, a novel genus and species, *Gibbsiella quercinecans*, and novel species, *B. goodwinii* were created to accommodate these bacteria. Despite the 100% co-occurrence of *G. quercinecans* and *B. goodwinii* isolates from necrotic lesions on AOD affected trees, and the lack of isolations of *G. quercinecans* and *B. goodwinii* from healthy oak trees, no previous studies had provided definitive evidence of their pathogenic contribution to AOD. Therefore, a genomic analysis of *G. quercinecans* and *B.*

goodwinii was undertaken to analyse their contribution to AOD, through characterisation of their pathogenicity potential and putative mechanisms of virulence.

7.2 The intergenic transcribed spacer region 1 as a molecular marker for identification and discrimination of *Enterobacteriaceae* associated with acute oak decline

Identification of closely related strains of *G. quercinecans*, *B. goodwinii* and other *Enterobacteriaceae* species, isolated from necrotic lesions of AOD affected trees, is a time consuming process and typically requires marker gene sequencing and analysis. Rapid identification of bacterial isolates from AOD affected trees facilitates characterisation of a putative AOD site, and enables resources to be diverted to front line research. However, differentiation and continuous verification of *Enterobacteriaceae* isolates can be achieved in house using PCR amplification and agarose gel visualisation, to distinguish between isolates within a few hours. Here, ITS1, an intergenic region located between the 5S rRNA and 16S rRNA, and within the rRNA operon was tested as a marker gene for the rapid typing of *Enterobacteriaceae* associated with AOD (Jensen *et al.*, 1993). Bacterial genomes often contain multiple copies of the rRNA operon and therefore have multiple copies of the ITS1 region, and as this region is non-coding (with the exception of tRNAs), prone to mutation, duplications and insertions. This heterogeneity imbues the region with the property of a high resolving power (Klappenbach *et al.*, 2000). This resolving power is easily visualised as the *Enterobacteriaceae* typically have several variable size ITS1 regions, which range between 150-1500 bp, allowing species to easily be discerned based on ITS1 copy number and length using an agarose gel image. The ITS1 molecular marker allowed the rapid and cost effective differentiation of a substantial number of species isolated from necrotic lesions of AOD affected trees and continues to be utilised for rapid strain verification purposes. In addition, a published paper describes application of ITS1 in AOD research and more widely for diagnostic purposes (Appendix I). Furthermore, the ITS1 molecular marker facilitated the genome sequencing efforts described in chapter 3, by providing insights on variability between strains, verifying the purity of cultures, and positively identifying strains during the quality control process for DNA extraction and genomic DNA library preparation.

7.3 Comparative genomics of bacteria isolated from necrotic lesions of AOD affected trees

Gibbsiella quercinecans and *Brenneria goodwinii* have been consistently isolated from necrotic lesions of AOD affected trees. To understand their role within AOD, three strains of *G. quercinecans*, two strains of *Brenneria goodwinii*, and *Ewingella americana* ATCC33852 which has been infrequently isolated from necrotic lesions of AOD affected trees, were selected for whole genome sequencing. Whole genome sequencing (WGS) is a molecular analysis tool which can characterise an organism's DNA, thereby underpinning the complete molecular analysis of an organism (Goldman and Landweber, 2016). Here, WGS techniques combined with the annotation and identification of virulence genes, allowed comparison of the genomes of canonical bacterial phytopathogens with *G. quercinecans* and *B. goodwinii*. Chapter 3 represents the first whole genome analysis of members of the genera *Gibbsiella* and *Brenneria*. Functional annotations revealed an encoded T2SS in *G. quercinecans* FRB97 & FRB124, a T3SS in *B. goodwinii* FRB141 & FRB171 (there was no T3SS in any *G. quercinecans* strain), a T4SS in all strains of *G. quercinecans* and *B. goodwinii* FRB171 and at least one copy of the T6SS operon was encoded in all strains of *G. quercinecans* and *B. goodwinii*. Products of the T2SS and T3SS, PCWDEs and T3SS effectors were encoded at variable levels across the sequenced isolates, with major PCWDEs such as pectate lyase encoded within *B. goodwinii* FRB141 & FRB171, pectate disaccharide lyase in *B. goodwinii* FRB141, rhamnogalacturonan endolyase and polygalacturonase in all strains of *G. quercinecans*, and T3SS effectors such as pectinolytic Hop effectors in *B. goodwinii* FRB141 & FRB171. Furthermore, there was substantial strain level variation within each genus, including a partial T2SS core subunit within *G. quercinecans* N78 and *B. goodwinii* FRB141, and a partial T4SS operon in *B. goodwinii* FRB141. The presence of these secretion systems and secreted virulence enzymes within *G. quercinecans* and *B. goodwinii* is an important contribution to the understanding of putative phytopathogens, furthermore the putative endophyte *E. americana* ATCC33852, encoded core secretion subunits for T2-T6SS, but crucially lacks a repertoire of major PCWDEs and T3SS effectors, which are key components of the virulence arsenal of typical bacterial phytopathogens.

The identification of virulence genes is central to evidence of putative pathogenicity provided within this thesis, as genes homologous to those described here, are the principal virulence factors in the top ten bacterial phytopathogens (Mansfield *et al.*, 2012). Subsequent research within this thesis (chapter 5 and chapter 6), builds on data from chapter 3 to understand the relationship of encoded virulence factors of *G. quercinecans* and *B. goodwinii*, and the oak host. Furthermore, subsequent wider studies of *G. quercinecans* and *B. goodwinii* can build upon these data as a genomic and transcriptomic baseline to expand the knowledge on bacterial phytopathogens within AOD and other Decline-disease.

7.4 A comparison of the whole genome sequences of *Gibbsiella quercinecans* FRB97 and *Brenneria goodwinii* FRB141 using second and third generation sequencing technologies

Whole genome sequencing technology has advanced significantly since the advent of next generation sequencing, with the Roche 454 platform in 2005 (Loman and Pallen, 2015). The development of the Roche 454 and continuous improvements to its output, have been driven by the demand for rapid and accurate sequencing. The Roche 454 was superseded by Illumina sequencing platforms, which offer higher-throughput data and relatively low error rates compared to the Roche 454. Recently, the development of third generation or single molecule sequencing has further progressed the field. Third generation sequencing has been driven by Pacific Biosciences (PacBio) and in particular the RSII sequencing platform, which has long read length (>20 kb per read) and produces random errors, allowing the easy correction of errors during assembly, producing highly accurate and contiguous assemblies. However, the third generation Oxford Nanopore, MinIon sequencing instrument, has the potential to overtake PacBio as the sequencer of choice due its potential unlimited read length (i.e. limited only by the quality of input DNA), portability (sequencing stations can be set up in the field) and sequencing speed (with DNA extraction to sequence visualisation in only a few hours) (Garalde *et al.*, 2016).

Chapter 4 compares Illumina and PacBio sequencing output and assemblies, illustrating the higher accuracy of the PacBio and therefore greater confidence of resultant biological inferences. PacBio sequencing produced the first finished assemblies of *G. quercinecans* FRB97 (T) and *B.*

goodwinii FRB141 (T), with single contigs for each bacterium, this was opposed to the 68 (*G. quercinecans* FRB97) and 185 (*B. goodwinii* FRB141) contigs produced as a result of Illumina MiSeq sequencing. Subsequent work in chapter 5 and chapter 6, involving transcriptomic, metagenomic and metatranscriptomic sequencing outputs, were aligned with confidence as the bacterial genome assemblies described in chapter 4 were highly accurate. Furthermore, finished genome assemblies of *G. quercinecans* FRB97 and *B. goodwinii* FRB141, can be used as reference genomes in subsequent studies, for example, comparative and population genomic analyses.

7.5. Transcriptome analysis reveals differential gene expression of *Gibbsiella quercinecans* FRB97 and *Brenneria goodwinii* FRB141 in axenic and co-cultures containing milled oak tree sapwood and phloem

The role of *G. quercinecans* and *B. goodwinii* within the AOD Decline-disease is currently unknown. Chapter 3 describes the genome encoded virulence factors of these bacteria, providing evidence of virulence potential within strains of *G. quercinecans* and *B. goodwinii*. Gene expression studies add to existing evidence of pathogenicity by describing the expression of virulence genes in response to contact with the oak host. Chapter 4 provides highly accurate reference genome assemblies, which can be used for the mapping of transcriptome reads in expression studies. To gain insights into the role of *G. quercinecans* FRB97 and *B. goodwinii* FRB141 in necrotic lesions of AOD affected trees, bacterial isolates were cultured in medium containing oak sapwood and phloem, subsequently expressed transcripts were aligned against the highly accurate reference genomes. This experiment was designed to study differential expression of virulence genes in response to the addition of oak tissue. The principal virulence mechanism of canonical bacterial phytopathogens is the release of PCWDEs and T3SS effectors; expression of these enzymes was examined *in vitro*, in chapter 5.

Within each bacterium, there was an outstanding response to oak tissue; *G. quercinecans* FRB97 had a distinct functional upregulation of virulence genes in response to phloem tissue at 2 HPI, *B. goodwinii* FRB141 had a clear functional shift with upregulation of virulence genes in sapwood tissue at 2 HPI. Additionally, there were several minor but notable trends including, genome

wide upregulation, including virulence genes in *G. quercinecans* FRB97 when in co-culture with *B. goodwinii* FRB141 in media containing sapwood tissue. The core T3SS secreton and T3SS effectors of *B. goodwinii* FRB141 were differentially upregulated in axenic culture containing sapwood media at 2 HPI. A separate set of virulence genes including a pectinolytic T3SS effector was upregulated in co-culture with *G. quercinecans* FRB97.

Chapter 5 provides clear evidence of virulence gene expression by *G. quercinecans* and *B. goodwinii* in response to oak tissue. Furthermore, *in vitro* polymicrobial interactions between *G. quercinecans* and *B. goodwinii*, including inducement of specific virulence genes may reflect environmental interactions within necrotic lesions of AOD affected trees. To build on the findings described in chapter 5, introduced virulence gene mutations to PCWDEs and T3SS effectors would demonstrate specific evidence of contribution to virulence of these genes and gene systems in an appropriate model organism. Moreover, expression data recovered from *in planta* cultures of wild type and quorum sensing mutants may reveal evidence of virulence symbiosis between *G. quercinecans* and *B. goodwinii*, and possibly further lesion *Enterobacteriaceae* strains, inferring that quorum sensing mediated release of virulence enzymes is induced by *Enterobacteriaceae* community members *in vitro*, and varies according to the lesion microbiota. Evidence of AOD polymicrobial infection is a mechanism analogous to olive knot disease caused by *Pseudomonas savastanoi* pv. *savastanoi*, whose growth is facilitated by olive knot *Enterobacteriaceae* community members, *Pantoea agglomerans, Erwinia tolenta,* and *E. oleae* (Marchi *et al.*, 2006; Buonaurio *et al.*, 2015).

7.6 Metagenomic and metatranscriptomic analysis of DNA and RNA recovered from AOD lesions and aligned against *Gibbsiella quercinecans* FRB97 and *Brenneria goodwinii* FRB141

Comparative metagenomic sequencing of necrotic lesions from AOD affected trees and a healthy tree revealed a shift in the functional abundance of the microbiome towards a *G. quercinecans* and *B. goodwinii* dominated metagenome in trees with AOD. Homologous alignment of metagenomic coding domains and *G. quercinecans* FRB97 and *B. goodwinii* FRB141 coding domains, revealed the near absence of these bacteria in healthy trees, and high abundance in AOD affected trees. Furthermore, determining functional abundance within the metagenome

provides gene catalogues, which reveal the functional potential of a microbiome and correlates with actual activity (Vieites *et al.*, 2009; Mason *et al.*, 2012).

Two metatranscriptomes recovered from active lesions on an AOD affected tree were aligned to the genomes of *G. quercinecans* FRB97 and *B. goodwinii* FRB141, revealing genome wide functional activity of the isolates. Furthermore, expressed genes include encoded phytopathogenic virulence genes including PCWDEs in *G. quercinecans* and *B. goodwinii,* and T3SS effectors in *B. goodwinii*. This homologous alignment of RNA transcripts concurs with metagenomic data recovered from the lesion sites, accumulating evidence of the activity of the bacteria to cause disease and their activity within an active lesion.

These data reveal the *in planta* prominence of *G. quercinecans* and *B. goodwinii* within necrotic lesions of AOD affected trees. Metagenomic data clearly indicates that within an active lesion, *G. quercinecans* and *B. goodwinii* have a substantial role in the microbiome. This data was enhanced by the recovery of metatranscripts with substantial alignment to *G. quercinecans* FRB97 and *B. goodwinii* FRB141. Overall, *in planta* sequencing of metagenomes and metatranscriptomes indicate a central role of *G. quercinecans* and *B. goodwinii* within necrotic lesions of AOD affected trees.

To support the findings of this study, expanded metagenomic and metatranscriptomic sampling regimes from healthy and AOD affected trees over successive time periods would allow a temporal overview of seasonal variation and contribution of *G. quercinecans* and *B. goodwinii* to AOD Decline-disease.

7.7 Future work

The overarching aim of AOD research is prevention or eradication. Within that aim are several objectives, which include characterisation of the pathogen(s) as described in this thesis. The aim of this thesis was to characterise two putative bacterial pathogens, *G. quercinecans* and *B. goodwinii* at the molecular level. It is clear from chapter 3 and chapter 4, that *G. quercinecans* and *B. goodwinii* have the genome encoded potential to cause disease within an oak lesion. However, the complexity of Decline-disease, dictates there is no single primary pathogen, and

therefore multiple interacting causes are responsible, requiring a holistic approach and an understanding of the role of bacteria within the Decline spiral (Fig. 1.6). This requires identification of microorganisms associated with oak lesions, and an understanding of how putative pathogens regulate release of their virulence factors, i.e. what genes do they express and when are they expressed? Some of these questions have been addressed in chapters 5 and 6 of this thesis, however regulatory changes within a natural environment may vary temporally and spatially, requiring extended analysis over several years. The end result of this study would be a detailed conceptual model describing bacterial regulatory changes and the environmental signals inducing this change, within the necrotic or pre-necrotic external environment.

The genomes of several strains of *G. quercinecans* and *B. goodwinii* have been described in chapter 3, with the number and identity of virulence genes varying between strains. The effect of this variation on pathogenicity has not been determined. Using a combination of *in silico* and *in planta* methods, a wide selection of *G. quercinecans* and *B. goodwinii* strains could be compared in order to evaluate the functional importance and activity of encoded virulence genes. A parallel study using plant pathogenicity tests on, firstly, model organisms, and secondly, oak tissue or oak saplings, would determine contrasting virulence phenotypes between the strains. Key virulence genes such as regulators or direct virulence factors identified as central to pathogenicity using evidence from comparative genomic, transcriptomic and metatranscriptomic studies within the most pathogenic strains, would be systematically removed using site directed mutagenesis or chromosomal inactivation (Datsenko and Wanner, 2000), creating mutant strains which would subsequently be inoculated onto the plant models to characterise the effect on pathogenicity. Resultant data would expose the pathogenic lifestyle of these bacteria and unravel the key pathogenicity mechanisms.

Chapter 3 reveals encoded T3SS core subunits and associated effectors within *B. goodwinii* FRB141 & FRB171; these were transcribed *in vitro* and *in planta*, as described in chapter 5 and chapter 6. The T3SS is a remarkable system, as it can directly translocate effectors into eukaryotic cells and is a notorious virulence weapon of bacterial phytopathogens (van Dijk *et al.*, 2002). Therefore, the T3SS of *B. goodwinii* merits a thorough examination. The *B. goodwinii* core T3SS subunit and a harpin effector (HrpN) are homologous to those encoded within *P. carotovorum* subsp. *carotovorum*. *B. goodwinii* T3SS effectors have homologs in bacterial

phytopathogens including, *Dickeya zeae* (DspF/AvrF, chaperone of AvrE), *Erwinia amylovora* (Two DspA/E genes, homologs of the AvrA, avirulence effector of *P. syringae*), *P. syringae* (YopJ/HopZ) and *P. viridiflava* (HopPtoM) (van Dijk *et al.*, 2002; Pallen *et al.*, 2005). Using site directed mutagenesis techniques, and a defined hypersensitivity response (HR) to *B. goodwinii* in a plant model, the T3SS core subunit could be invalidated, through the removal of T3SS core genes. Mutant and wild-type HR could be compared, and effectors sequentially removed to understand their individual contribution to virulence.

Studies of the mammalian gut microbiota have substantial levels of research activity and funding, which is unmatched within ecological studies of plant microbiota. Mammalian microbiota research has uncovered substantial complexity governing host invasion by a variety of bacterial pathogens (Bäumler and Sperandio, 2016). Transference of knowledge from mammalian microbiome studies to bacterial involvement within Decline-disease of plant species would enable analogous insights of host-pathogen reactions and polymicrobial infection, as similar subtle interactions are likely. A shift in gut microbiota from the clostridia and *Bacteroidetes* groups, to Proteobacteria, predominantly *Enterobacteriaceae*, has been described in humans and mice affected by colitis, a phenomena known as dysbiosis. Dysbiosis has implications for the application of Koch's postulates, as dysbiosis preceded by poor host health is often a pre-requisite for successful pathogen colonisation. Therefore, a re-appraisal of the postulates, including caveats for the pre-requisites of disease such as community membership and polymicrobial interactions would improve their accuracy. For AOD, an analogous shift from Gram-positive to Gram-negative bacteria (predominantly *Enterobacteriaceae*) has been described within bacterial communities from healthy to AOD affected oak trees (Denman *et al., In Press*; Sapp *et al.*, 2016), a shift which would be rendered meaningless through the original interpretation of Koch's postulates, but could be described as Decline induced dysbiosis. Furthermore, within wider studies describing the complexity of pathogen invasion, one theme emerges; pathogens use commensal microbiota derived molecules as signalling factors and nutrients for enhanced pathogenicity (Winter and Bäumler, 2014). One such example includes a study on the recovery of individuals affected by *Vibrio cholerae* in an outbreak in Bangladesh (Hsiao *et al.*, 2014). Recovery was mediated by greater diversity of the gut microbiota, as those individuals with a superior recovery outcome were colonised by a more diverse range of bacteria, including

Ruminococcus obeum, which produces a furanone autoinducer and represses *V. cholerae* colonisation factors. Individuals who contained a more diverse community of bacteria, had a microbiota recovery signature. Within the field of plant pathogenicity, this thesis aligns with these wider studies of bacterial pathogenicity, presenting evidence of interactions between putative bacterial pathogens and their plant host in a simulated lesion environment, and their active involvement within a necrotic AOD lesion, thereby furthering the research of plant polymicrobial infection, which challenges prevailing orthodoxies, and furthers the understanding of a threat to a keystone species within the United Kingdom. Therefore, using a combination of metagenomics, proteomics and metabolomics, the microbiota signature and predominant microbiota derived molecules in healthy and AOD affected trees could be studied to understand the complex interactions governing the pathogen phenotype.

7.8 Conclusions

The overarching aim of this thesis was to characterise the bacterial component of AOD, and through molecular analysis, understand the role of *G. quercinecans* and *B. goodwinii* as putative contributing factors to the Decline-disease. Molecular analysis of *G. quercinecans* FRB97 and *B. goodwinii* FRB141 has provided the first genomic and transcriptomic evidence of putative pathogenicity and polymicrobial infection in AOD. Novel insights were gained using contemporary molecular methods to unravel firstly, their genome encoded virulence potential and secondly, their idiosyncratic role within the lesion environment, which indicated a putative polymicrobial infection that does not conform with traditional views of plant pathogenesis. Typically, disease is defined through the application of Koch's postulates, however, the interchangeable contributory causes of Decline-disease cannot be bound within these narrow confines. Recently, polymicrobial studies have revealed that phytopathogen co-operation can amplify virulence (da Silva *et al.*, 2014b). Thus, polymicrobial infections are increasingly accepted as a fundamental component of plant pathology (Lamichhane and Venturi, 2015). Subtle interactions, such as the release of quorum sensing molecules, are the drivers of polymicrobial disease and require molecular approaches to illuminate their complexity. Future microbial AOD research should build upon previously employed techniques to describe polymicrobial infections such as within olive knot disease, that is, applying contemporary molecular tools to identify bacterial regulatory alterations which have substantial, disease causing effects. This thesis uses cutting edge research to provide detailed evidence of pathogenicity within bacteria consistently isolated from necrotic lesions on AOD affected trees, thus advancing the knowledge of polymicrobial infection as a pivotal component of an ecologically significant Decline-disease.

7.9 References

Bäumler, A.J. and Sperandio, V. (2016) Interactions between the microbiota and pathogenic bacteria in the gut. *Nature* **535**: 85–93.

Brady, C., Denman, S., Kirk, S., Venter, S., Rodríguez-Palenzuela, P., and Coutinho, T. (2010) Description of *Gibbsiella quercinecans* gen. nov., sp. nov., associated with Acute Oak Decline. *Syst. Appl. Microbiol.* **33**: 444–50.

Brown, N., Jeger, M., Kirk, S., Xu, X., and Denman, S. (2016) Spatial and temporal patterns in symptom expression within eight woodlands affected by Acute Oak Decline. *For. Ecol. Manage.* **360**: 97–109.

Buonaurio, R., Moretti, C., da Silva, D.P., Cortese, C., Ramos, C., and Venturi, V. (2015) The olive knot disease as a model to study the role of interspecies bacterial communities in plant disease. *Front. Plant Sci.* **6**: 1–12.

Datsenko, K.A. and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97**: 6640–5.

Day, W.R. (1927) The oak mildew *Microsphaera quercina* (schw.) burrill and *Armillaria mellea* (vahl) quel. in relation to the dying back of oak. *Forestry* **1**: 108–112.

Denman, S., Brady, C., Kirk, S., Cleenwerck, I., Venter, S., Coutinho, T., and De Vos, P. (2012) *Brenneria goodwinii* sp. nov., associated with acute oak decline in the UK. *Int. J. Syst. Evol. Microbiol.* **62**: 2451–2456.

Denman, S., Plummer, S., Kirk, S., Peace, A., and McDonald, J.E. Isolation studies reveal a shift in the cultivable microbiome of oak affected with Acute Oak Decline. *Syst. Appl. Microbiol.*

van Dijk, K., Tam, V.C., Records, A.R., Petnicki-Ocwieja, T., and Alfano, J.R. (2002) The SHcA protein is a molecular chaperone that assists in the secretion of the HopPsyA effector from the type III (Hrp) protein secretion system of *Pseudomonas syringae. Mol. Microbiol.* **44**: 1469–1481.

Garalde, D.R., Snell, E.A., Jachimowicz, D., Heron, A.J., Bruce, M., Warland, A., *et al.* (2016) Highly parallel direct RNA sequencing on an array of nanopores. *bioRxiv*.

Goldman, A.D. and Landweber, L.F. (2016) What is a genome? *PLOS Genet.* **12**: e1006181.

Hsiao, A., Ahmed, A.M.S., Subramanian, S., Griffin, N.W., Drewry, L.L., Petri, W.A., *et al.* (2014) Members of the human gut microbiota involved in recovery from *Vibrio cholerae* infection. *Nature* **515**: 423–6.

Jensen, M.A., Webster, J.A., and Straus, N. (1993) Rapid identification of bacteria on the basis of polymerase chain reaction-amplified ribosomal DNA spacer polymorphisms. *Appl. Environ. Microbiol.* **59**: 945–952.

Klappenbach, J. A., Dunbar, J.M., and Schmidt, T.M. (2000) rRNA operon copy number reflects ecological strategies of bacteria. *Appl. Environ. Microbiol.* **66**: 1328–1333.

Lamichhane, J.R. and Venturi, V. (2015) Synergisms between microbial pathogens in plant disease complexes: a growing trend. *Front Plant Sci* **6**: 1–12.

Loman, N.J. and Pallen, M.J. (2015) Twenty years of bacterial genome sequencing. *Nat. Rev. Microbiol.* **13**: 787-794.

Mansfield, J., Genin, S., Magori, S., Citovsky, V., Sriariyanum, M., Ronald, P., *et al.* (2012) Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol. Plant Pathol.* **13**: 614–29.

Marchi, G., Sisto, A., Cimmino, A., Andolfi, A., Cipriani, M.G., Evidente, A., and Surico, G. (2006) Interaction between *Pseudomonas savastanoi* pv. *savastanoi* and *Pantoea agglomerans* in olive knots. *Plant Pathol.* **55**: 614–624.

Mason, O.U., Hazen, T.C., Borglin, S., Chain, P.S., Dubinsky, E. A, Fortney, J.L., *et al.* (2012) Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill. *ISME.* **6**: 1715–1727.

Pallen, M.J., Beatson, S. a., and Bailey, C.M. (2005) Bioinformatics, genomics and evolution of non-flagellar type-III secretion systems: a Darwinian perpective. *FEMS Microbiol. Rev.* **29**: 201–229.

Sapp, M., Lewis, E., Moss, S., Barrett, B., Kirk, S., Elphinstone, J., and Denman, S. (2016) Metabarcoding of bacteria associated with the Acute Oak Decline syndrome in England. *Forests* **7**: 95.

da Silva, D.P., Castaneda-Ojeda, M.P., Moretti, C., Buonaurio, R., Ramos, C., and Venturi, V. (2014) Bacterial multispecies studies and microbiome analysis of a plant disease. *Microbiol. (United Kingdom)* **160**: 556–566.

Vieites, J.M., Guazzaroni, M.-E., Beloqui, A., Golyshin, P.N., and Ferrer, M. (2009) Metagenomics approaches in systems microbiology. *FEMS Microbiol. Rev.* **33**: 236–55.

Winter, S.E. and Bäumler, A.J. (2014) Why related bacterial species bloom simultaneously in the gut: Principles underlying the "like will to like" concept. *Cell. Microbiol.* **16**: 179–184.

APPENDIX I

Doonan J, Denman S, Gertler C, Pachebat JA, Golyshin PN, McDonald JE (2015) The intergenic transcribed spacer region 1 as a molecular marker for identification and discrimination of *Enterobacteriaceae* associated with acute oak decline. *J. Appl. Microbiol*, 118:193-201.

Journal of
Applied Microbiology

Journal of Applied Microbiology ISSN 1364-5072

ORIGINAL ARTICLE

The intergenic transcribed spacer region 1 as a molecular marker for identification and discrimination of *Enterobacteriaceae* associated with acute oak decline

J. Doonan[1], S. Denman[2], C. Gertler[1], J.A. Pachebat[3], P.N. Golyshin[1] and J.E. McDonald[1]

1 School of Biological Sciences, Bangor University, Bangor, UK
2 Centre for Ecosystems Society and Biosecurity, Forest Research, Surrey, UK
3 Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Aberystwyth, UK

Keywords
acute oak decline, *Brenneria goodwinii*, DNA gyrase B, *Enterobacteriaceae*, *Gibbsiella quercinecans*, ITS1.

Correspondence
James E. McDonald, School of Biological Sciences, Bangor University, Deiniol Road, Bangor, Gwynedd LL57 2UW, UK.
E-mail: j.mcdonald@bangor.ac.uk

2014/1743: received 22 August 2014, revised 21 October 2014 and accepted 21 October 2014

doi:10.1111/jam.12677

Abstract

Aims: We assessed the veracity of intergenic spacer region 1 (ITS1) ribotyping for the rapid, inexpensive and accurate identification of *Brenneria goodwinii* and *Gibbsiella quercinecans* that are associated with acute oak decline (AOD) in the UK.

Methods and Results: Agarose gel electrophoresis and polyacrylamide gel electrophoresis (PAGE) were applied for the typing of ITS1 PCR amplicons from strains of *B. goodwinii*, *G. quercinecans* and related species ($n = 34$). The number and length of ITS1 amplicons varied significantly between strains. ITS1 profiles generated via PAGE were used to differentiate species using a neighbour-joining phylogram. The ITS1 phylogram was compared against DNA gyrase B (*gyrB*) gene sequences from the same strains, demonstrating that ITS1 ribotyping is as effective as *gyrB* at resolving *G. quercinecans* and *B. goodwinii* to the species level.

Conclusions: The ITS1 gene has been successfully employed as a novel marker to resolve newly described AOD-associated *Enterobacteriaceae*, *B. goodwinii* and *G. quercinecans*, to species level.

Significance and Impact of the Study: ITS1 ribotyping of *B. goodwinii* and *G. quercinecans* provides equivalent sensitivity to the current standard method for strain identification (sequence analysis of the *gyrB* gene), but with reduced processing time and cost. Furthermore, the ITS1 gene is widely applicable as a rapid and inexpensive typing system for *Enterobacteriaceae*.

Introduction

Acute oak decline (AOD) is a syndrome partly derived from tissue necrosis of the inner bark of oak trees (Denman *et al.* 2014). The decline is relatively new in Britain (reported cases of AOD have been observed for 20–30 years) and has been identified in both species of native oak, with the number of reported cases on the increase. The precise biotic and abiotic causes of AOD are currently unknown, but there are likely to be multiple factors involved. Networks of galleries produced by the larvae of the buprestid beetle *Agrilus biguttatus* have been discovered in trees with symptoms of AOD and are typically associated in and around areas of lesion formation (Denman *et al.* 2014). The beetles themselves are rarely seen, but evidence of their occurrence is presented via larval galleries in the cambial zone of the oak tree. In addition, previous studies have described the consistent isolation of *Enterobacteriaceae* from necrotic lesions of affected oak trees in sampling sites across England (Brady *et al.* 2010; Denman *et al.* 2012). However, two newly described bacterial species belonging to the *Enterobacteriaceae*, *Gibbsiella quercinecans* and *Brenneria goodwinii*, are consistently isolated from necrotic lesions on oak trees, with other species such as *Rahnella* spp. isolated from both healthy and necrotic tissues. Consequently, the almost

exclusive association of *G. quercinecans* and *B. goodwinii* with the lesions of AOD-affected trees has led to the hypothesis that these species play a central role in tissue necrosis (Denman and Webber 2009), possibly via necrogenic enzymes or secretion systems and their associated effector proteins.

The ribosomal RNA (rrn) operon and in particular 16S rRNA are conserved genomic regions commonly used for bacterial genotyping and taxonomy (Woese and Fox 1977; Cedergren *et al.* 1988). However, due to the highly conserved nature of the 16S rRNA gene, resolution of closely related species within certain taxa such as the *Enterobacteriaceae* can be problematic (Janda and Abbott 2007; Naum *et al.* 2008). The dearth of sequence differences among closely related species contributes to a lack of phylogenetic power and limits its applicability for fine-scale taxonomic resolution. The intergenic or internal transcribed spacer region 1 (ITS1) is part of the *rrn* operon, found between the genes for small (16/18S) and large (23/28S) rRNA subunits. The ITS1 region is more susceptible to synonymous mutations than other parts of the operon due to its functional role being confined to the coding of tRNAs and processing rRNA (Scheinert *et al.* 1996). Consequently, it is an excellent candidate for molecular differentiation between both diverse and closely related species. ITS analysis, sometimes called RISA (ribosomal intergenic spacer analysis) is a method most commonly applied to examine the extent of microbial diversity in environmental samples from soil and marine communities, for example Brown and Fuhrman (2005) and Rappe *et al.* (2002), and it has been well described in medical microbiology for identifying species in clinical samples, for example Gurtler and Stanisich (1999). However, its use in arboreal microbial analysis has not previously been tested. Molecular studies typically use the 16S rRNA gene for taxonomic identification at the familial level, and although this is the gold standard of PCR-based identification of taxonomic identity, it often lacks accuracy in discriminating at the species level (Gurtler and Stanisich 1999). Conversely, the ITS1 region is highly variable in length and sequence in all bacteria and may thus be effective at distinguishing bacterial species and strains. It should be noted that many bacteria contain more than one copy of the *rrn* operon, which usually show strong sequence homogeneity (>98%) but may differ through insertions or deletions by a range of 2–301 bp, with an average difference of 166 bp (Klappenbach *et al.* 2001; Larkin *et al.* 2007). ITS1 amplicons will therefore exhibit substantial variation in size as amplicon range can vary from 150 to 1500 bp with 85–90% of amplicons being between 150 and 600 bp (Fisher and Triplett 1999). Often the ITS1 region will code for tRNA genes, these are most frequently found in Gram-negative bacteria and give extra length to the sequence. However, if present, they are usually more conserved than other parts of ITS1 (Daffonchio *et al.* 1998). Typically, the ITS1 amplicon pattern resolved using agarose gels can be used to identify species, whereas the hypervariable indels (mutations) can be used to differentiate between strains (Garcia-Martinez *et al.* 1999).

Due to the conserved nature of 16S rRNA genes across the *Enterobacteriaceae*, *gyrB* is used as the phylogenetic marker gene of choice for the identification and taxonomic resolution of strains isolated from the necrotic lesions of oaks affected by AOD. This approach has provided important insights into the ecology and epidemiology of *G. quercinecans* and *B. goodwinii* and their almost exclusive association with necrotic tissue in AOD-affected trees. However, *gyrB* gene PCR amplification, sequencing and phylogenetic analysis of hundreds/thousands of isolated strains from affected trees is a costly and laborious process, representing a barrier to elucidating the ecology and epidemiology of *B. goodwinii* and *G. quercinecans*. This is a pertinent issue, as there are numerous ongoing investigations into the role of bacteria in AOD and the need to confirm the identity and purity of strains is of paramount importance.

The aim of this study was therefore to develop a rapid molecular diagnostic test for *B. goodwinii* and *G. quercinecans*, based on ITS1 profiling, to resolve species identity and purity of strains. We hypothesized that the ITS1 region represents a suitable genotypic marker for identification and differentiation of *B. goodwinii* and *G. quercinecans* strains associated with AOD and negates the requirement for gene sequencing, as multiple copies of the ITS1 region are present in the bacterial genome, with significant size variation in the length of the sequence.

Here, the ITS1 profiles of bacterial isolates from the necrotic lesions of oak trees were characterized and validated using a polyphasic analysis. Two DNA fingerprinting methods were tested and compared to verify and validate the use of ITS1 profiles in species identification and resolution. Amplified ITS1 PCR products were resolved via (i) 3% agarose gel electrophoresis and (ii) polyacrylamide gel electrophoresis. Finally, the ability to resolve isolated strains to species level was determined by a comparative phylogram of ITS1 profiles and the DNA gyrase B gene sequence of each strain.

Methods

Maintenance of bacterial strains

Enterobacteriaceae strains were isolated by Forest Research, from oak trees affected by AOD. Most strains had been previously identified to species level through

DNA gyrase B (*gyrB*) sequencing and DNA–DNA hybridization (Brady *et al.* 2010; Denman *et al.* 2012). The strains were stored in glycerol stocks at −80°C and maintained on nutrient agar (Oxoid) at 20°C.

ITS1 PCR amplification, agarose gel electrophoresis and sequencing

Each strain was subcultured from an individual colony five times, creating five subisolates per strain. Genomic DNA was prepared using the colony extraction method for ITS1 PCRs: one colony of bacterial cells was picked from an agar plate and added directly to the PCR assay tube (colony PCR). The ITS1 region of each subisolate was amplified using ITS1-specific oligonucleotide primers, designed for environmental bacterial communities ITSF (5′-GTCGTAACAAGGTAGCCGTA-3′) and ITSReub (5′-GCCAAGGCATCCACC-3′) (Cardinale *et al.* 2004). PCRs were performed in 50 μl reaction volumes containing a colony of each isolate, 1 × MyTaq Red Mix (Bioline), 50 μmol l^{-1} each forward and reverse primer. The solution was made up to its final volume of 50 μl with molecular grade double-distilled water. The thermal cycle consisted of initial denaturation at 95°C for 60 s followed by 25 cycles of denaturation at 95°C for 15 s, primer annealing at 55°C for 15 s and elongation at 72°C for 10 s. PCR amplification products were visualized using 3% agarose gel at 120 V for 135 min, and amplicon size was calculated using Hyperladder I (Bioline). ITS1 amplicons were excised from agarose gels and extracted directly using the QIAEX II get extraction kit (Qiagen, Manchester, UK) for all sequencing reactions. Purified PCR amplicons were sequenced by Macrogen Inc. These sequence data have been deposited to GenBank under Accession Numbers KJ418748 to KJ418834.

gyrB PCR amplification, gel electrophoresis and sequencing

The type II topoisomerase *gyrB* gene was selected as a quality control marker to positively identify bacterial species and verify the utility of ITS1 as a molecular marker for identification and discrimination of bacterial species. Oligonucleotide primers were *Enterobacteriaceae* specific as described by Brady *et al.* (2010), gyrB01F (5′-TAART TY GAYGAYAACTCYTAYAAAGT-3′) and gyrB02R (5′-C M CCYTCCACCARGTAMAGT-3′). A separate forward primer was used for sequencing of *gyrB* PCR amplicons, gyrB07F (5′-GTVCGTTTCTGGCCVAG-3′). Genomic DNA was prepared using the boil prep method, where a colony of the isolate was picked from nutrient agar plates and suspended in 20 μl molecular grade water and incubated at 100°C for 5 min before adding 1 μl to the PCR mixture. PCRs were performed as described for ITS1 with an annealing temperature of 50°C. PCR amplification products were visualized via 1% agarose gel electrophoresis at 140 V for 35 min. *gyrB* amplicons were excised from the gel and purified using the QIAEX II kit (Qiagen). Purified amplicons were sequenced at Macrogen Inc, Amsterdam, the Netherlands.

Polyacrylamide Gel Electrophoresis (PAGE)

The ITS1 region of 34 bacterial isolates from AOD-affected trees was amplified as described for the agarose gel analysis. Electrophoresis was conducted using an Ingeny PhorU electrophoresis unit (Ingeny, Leiden, the Netherlands). ITS1 PCR products (50 μl, prepared as described in section 'ITS1 PCR amplification, agarose gel electrophoresis and sequencing') were loaded into a 15% polyacrylamide gel and allowed to migrate through the gel for 16 h at 100 V. Amplicons were then visualized using a SYBR Gold stain (Invitrogen, Paisley, UK).

Phylogram analyses of ITS1 amplicon patterns and *gyrB* DNA sequences

Phylogenetic analyses of all 34 AOD isolates based on the separation of their ITS1 amplicons from the PAGE analysis were conducted in Dendroscope (Huson and Scornavacca 2012). The size (in base pairs) of ITS1 amplicons was scored individually using Bio-Rad Chemidoc software, which aligned matching amplicons in the presence–absence test (with a sensitivity setting of 2) giving a binary output. The binary output was then altered, changing '0' to 'T' and '1' to 'C'. This altered binary data were entered into ClustalW (Larkin *et al.* 2007), which produced a neighbour-joining phylogenetic tree based on the presence and absence of amplicons. The resulting tree was entered into Dendroscope and modified into both a rectangular phylogram and a tanglegram. The phylogram was used to display ITS1 fingerprints from the polyacrylamide gel. The tanglegram directly compared the PAGE analysis against DNA gyrase B gene sequences for 33 AOD strains.

Results

Validation of ITS1 typing as a tool for species resolution

Six bacterial strains (two × *Gibbsiella quercinecans*, two × *Brenneria goodwinii* and two × *Rahnella* spp.) were used to validate the veracity of the ITS1 amplicon typing approach. To confirm the purity of these strains, each strain was consecutively plated from a single colony five times, and this subset (n = 30 subisolates) was characterized by ITS1 PCR amplicon typing and *gyrB* gene

sequencing to confirm that each isolate was correctly identified (Table S1).

ITS1-specific PCR amplicons were generated for each of the thirty subisolates and resolved via 3% agarose gel electrophoresis (e.g. Fig. 1). The number of visible ITS1 amplicons for each strain tested varied from three to seven, with both strains of *G. quercinecans* possessing six ITS1 amplicon bands. The two *B. goodwinii* strains possessed different numbers of visible ITS1 amplicons (four and six, respectively), and *Rahnella* sp. strains also possessed a different number of amplicons (four and seven) (Table S1). These data demonstrate that ITS1 typing provides an effective method for the resolution of the six different *Enterobacteriaceae* strains tested at the species level. Furthermore, despite the presence of six amplicons for both *G. quercinecans* strains tested, differences in the length of these ITS1 amplicons provided strain level differentiation, and this was also the case for *B. goodwinii* and *Rahnella* sp. strains, which demonstrated the potential for strain-specific resolution, where both the number and length of ITS1 amplicon varied between the two strains of each species.

Visible ITS1 amplicons for all subisolates were subsequently excised and purified from agarose gels, and where sufficient DNA was retrieved, each amplicon was sequenced to confirm its identity. ITS1 amplicon sequencing ($n = 87$) and BLASTn searches (Altschul *et al.* 1990) confirmed that all observed ITS1 amplification products were *bona fide* ITS1 sequences. In addition, each strain was identified via a confirmatory *gyrB* sequencing reaction (Table S1). However, for one of the strains (*G. quercinecans* FRB98), agarose gel electrophoresis of five subisolates obtained from the same agar culture plate revealed two distinct ITS1 amplicon types (Fig. 1), where the presence of a contaminant is clearly visible, as three subisolates have three visible amplicons and two isolates have six amplicons. These data suggested the presence of a second bacterial strain in the culture of *G. quercinecans* FRB98, and *gyrB* sequencing confirmed the presence of two strains within the culture, *G. quercinecans* FRB98 (lanes 5 and 6, Fig. 1) and *Enterobacter cloacae* (lanes 2–4, Fig. 1), demonstrating an additional application of ITS1 ribotyping in strain maintenance and the identification of culture contaminants in laboratory strains. The pilot study therefore validated the use of ITS1 typing to separate members of the *Enterobacteriaceae* at both species and strain level, and this approach was subsequently tested and validated on three commonly used and independent laboratory methods for DNA fingerprinting.

Comparison of three independent methods for the resolution of ITS1 ribotypes in *Enterobacteriaceae* strains associated with AOD

3% agarose gel electrophoresis

Based on the results of the pilot study, the ITS1 genomic region was proposed as an appropriate molecular marker for the resolution of members of the *Enterobacteriaceae*. The agarose verification method was subsequently extended to a wider selection of *Enterobacteriaceae* strains ($n = 34$) (Table S2). However, variations in ITS1 amplicon numbers for some strains were observed when the same strain was run on replicate agarose gels despite identical quantities of PCR-amplified ITS1 fragments being loaded onto the gel. It is possible that in independent PCRs, certain amplicons are preferentially amplified due to PCR bias, and agarose is not a sensitive tool to visualize DNA fragments. Nevertheless, the sensitivity of detection using 3% agarose gel separation of ITS1 fragments did appear to provide a rudimentary resolution of strains and species, but the ITS1 amplicon differences observed between the same strain run on replicate agarose gels suggest that alternative and more sensitive method for ITS1 typing would be more appropriate. Consequently, polyacrylamide gel electrophoresis analysis of ITS1 PCR products was tested as an alternative

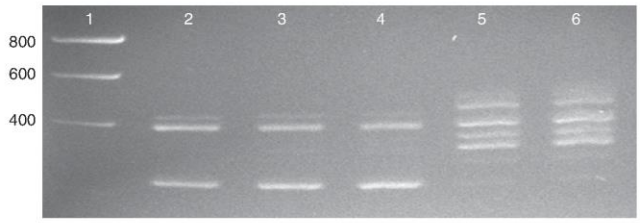


Figure 1 Five subisolates (lanes 2–6) from a culture of *Gibbsiella quercinecans* strain FRB98 revealed two distinct ITS1 amplicon patterns on a high-resolution 3% agarose gel, suggesting that the culture was impure. *gyrB* sequencing of subisolates in lanes 5 and 6 confirmed that these subisolates were *G. quercinecans*, whereas *gyrB* sequencing of subisolates in lanes 2–4 established that the subisolates were *Enterobacter cloacae*. Lane 1 contains Hyperladder I (Bioline). Ladder size is given in base pairs.

method for the resolution of ITS1 ribotypes and compared with phylogenetic resolution against the *gyrB* gene phylogeny for each strain.

PAGE resolution of Enterobacteriaceae ITS1 ribotypes

The ITS1 ribotype profiles of the same 34 *Enterobacteriaceae* strains tested via agarose gel electrophoresis were analysed using PAGE to further test the utility of the ITS1 region as a marker for strain discrimination and provide greater resolution of the ITS1 amplicons (Fig. 2).

Visual inspection of ITS1 profiles for *G. quercinecans* and *B. goodwinii* (Fig. 2) suggests similarities between the ITS1 ribotypes of isolates of the same species. This discrimination was further demonstrated using a neighbour-joining phylogenetic tree to infer relationships between species (Mount 2008).

Phylogenetic resolution of the 34 AOD isolates based on their ITS1 profile using PAGE successfully separated *G. quercinecans* and *B. goodwinii* strains into two lineages (Fig. 3). Twenty of the 34 strains were *G. quercinecans* and *B. goodwinii* isolates, and to further validate the use of ITS1 as a molecular marker for the discrimination of *G. quercinecans* and *B. goodwinii*, a broad spectrum of related AOD strains ($n = 14$) from the *Enterobacteriaceae* was also separated and compared using ITS1 ribotyping.

An amplicon presence/absence matching alignment was used to generate the phylogenetic tree in Fig. 3, to demonstrate the relationship between isolates. One potential caveat to this approach is that similarly sized ITS1 amplicons can erroneously be grouped together, as the QUANTITYONE software (Bio-Rad, Hemel Hempstead, UK) used in this study (or the similar amplicon matching open source software PYELPH 1.4 (Pavel and Vasile 2012)) clusters the amplicons of a similar but not identical size. However, our data suggest that this does not result in the misidentification of strains at the species level at least. These programs do have an adjustable sensitivity gauge, but ultimately they must be manually validated. An appropriately spaced molecular ladder (i.e. a long gap between each rung) can reduce bias by creating a larger area on the gel image.

The PAGE-derived ITS1 phylograms were compared against *gyrB* sequences using a tanglegram (Fig. 4). This contrasting phylogram reveals the differences in resolution between *gyrB* sequencing and ITS1 ribotyping methods, but demonstrates the ability of both methods to resolve *G. quercinecans* and *B. goodwinii* at the species level. Furthermore, some strains could be resolved through their ITS1 amplicon pattern; for example, *G. quercinecans* FRB92 has five amplicons, whereas *G. quercinecans* FRB93 has four. PAGE successfully resolved ITS1 ribotypes to the species level for all *G. quercinecans* and *B. goodwinii* isolates, barring one exception; *Gibbsiella gregii* USA42 clustered with *G. quercinecans* FRB185, and the latter strain has therefore not been successfully resolved according to species. However, it is clear from the PAGE profiles that *G. quercinecans* strain FRB185 has a different ITS1 PAGE profile to both *G. gregii* USA42 and also the other members of that species, but as it is currently a singleton strain for this ITS1 ribotype, it appears to be an outlier. Furthermore, *G. quercinecans* strain FRB185 also failed to cluster with the other members of this species via *gyrB* sequence analysis (Fig. 4) and clustered with *Gibbsiella dentisursi* and *G. papillionis*, suggesting that the inability of both ITS1 ribotyping and *gyrB* sequencing to resolve FRB185 to species level is an issue specific to that strain, rather than a methodological issue. Nevertheless, *G. quercinecans*

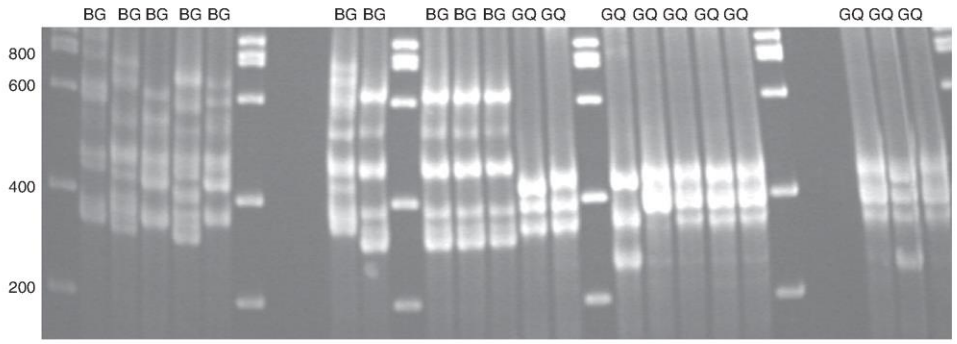


Figure 2 Polyacrylamide gel with ITS1 profiles for: *Brenneria goodwinii*, lanes 2–6, 8–9 and 11–13; *Gibbsiella quercinecans*, lanes 14–15, 17–21 and 23–25. Lanes 1,7,10,16 and 22 contain Hyperladder I (Bioline). Blank lanes are not included in the lane numbering. Ladder size is given in base pairs.

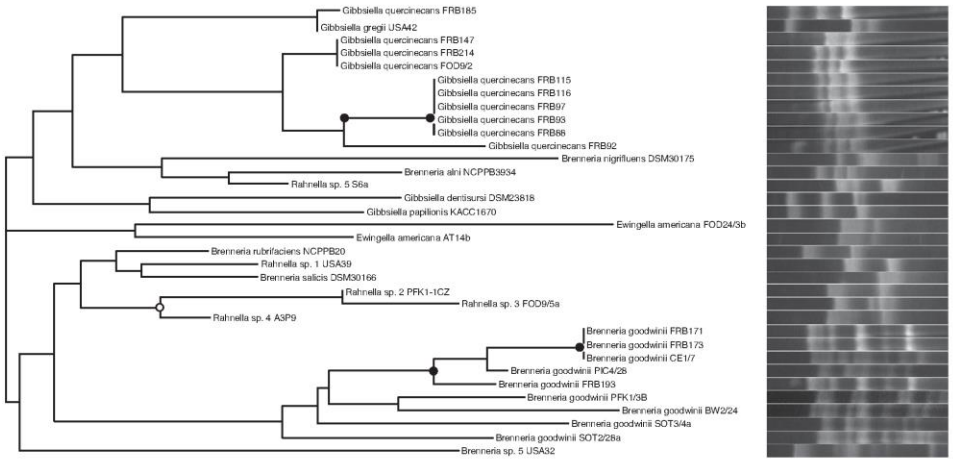


Figure 3 Neighbour-joining phylogram of acute oak decline isolates and their ITS1 fingerprints resolved on a polyacrylamide gel. Nodes in which bootstrap value >95% are denoted as filled circles, and those between 75 and 95% as unfilled circles.

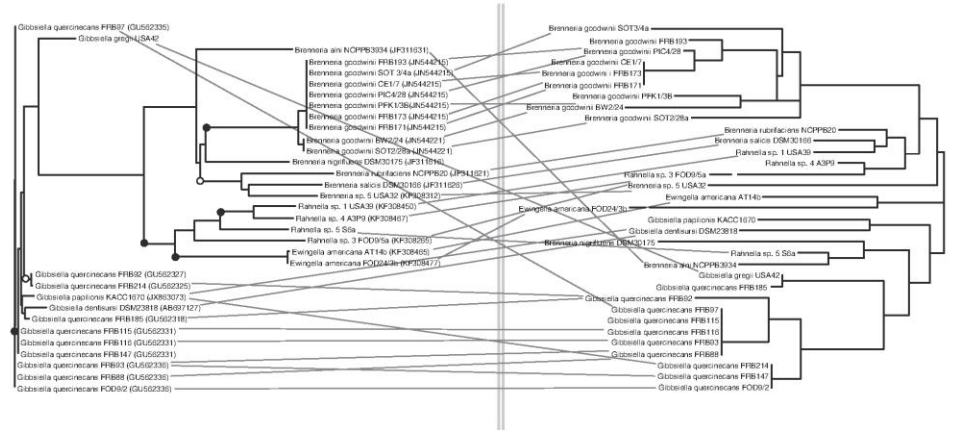


Figure 4 Tanglegram comparing *gyrB* phylogram (left) and polyacrylamide phylogram (right). GenBank Accession Numbers are given in parentheses. Bootstrap values >95% are denoted as filled circles; values between 75 and 95% are denoted as clear circles.

FRB185 occupies the same clade as the other *G. quercinecans* strains, and it is likely that future typing of additional *G. quercinecans* strains will further resolve the intraspecies and genus relationships between strains.

Discrimination of closely related bacterial strains using ITS1 amplicon length may sometimes be challenging as it is not possible to ensure each copy of the ITS1 region is amplified with equal concentration. Therefore, some ITS1 copies which may be present in the genome may not appear when visualized on a gel image or appear as a shadow. It should be noted that it is possible if not probable that not all ITS1 copies are resolved. However, resolved ITS1 fragments were visualized using PAGE consistently and reproducibly, and the method was capable of validation of the ITS1 maker and separation of the AOD isolates. On a cautionary note, there may be

similarly weighted ITS1 fragments shared between unrelated bacteria causing relatedness between bacteria to be inferred when no such genotypic relationship exits (Jensen *et al.* 1993). This is likely to be a rare event, but it should be noted as overlapping intergenic spacer size classes could lead to misinterpretation of results.

gyrB phylograms *vs* ITS1 typing

16S rRNA gene sequencing represents the most commonly used phylogenetic tool for estimating microbial diversity and would be improperly applied to the *Enterobacteriaceae* here due to the conserved nature of the phylum at this locus (Mollet *et al.* 1997). Therefore, to resolve interspecies taxonomic relatedness, the *gyrB* gene is the marker of choice for taxonomic identification of *Enterobacteriaceae* strains isolated from AOD-affected trees (Brady *et al.* 2010; Denman *et al.* 2012). Relationships between *Enterobacteriaceae* strains were revealed using the *gyrB* molecular marker to create a neighbour-joining topology. This topology was compared to a neighbour-joining phylogram from the polyacrylamide gel generated via the ITS1 molecular marker using a tanglegram (Fig. 4). The tanglegram reveals that species-level resolution of ITS1 is very similar to that of *gyrB*. Furthermore, strain separation may be possible using ITS1, especially in strains with a higher number of ITS1 copies, as this gives a greater prospect of variation. For example, *G. quercinecans* has fewer ITS1 copies than *B. goodwinii*, and this increases the phylogenetic resolution of ITS1 in *B. goodwinii* compared to *G. quercinecans*. Additionally, as Fig. 4 reveals, *B. goodwinii* FRB171 and *B. goodwinii* PFK1/3B are identical in their *gyrB* sequence, whereas the ITS1 copy number differs, and they are therefore separated into different branches of the tree. Therefore, as ITS1 copy number increases, the marker contains ever greater phylogenetic power to differentiate between closely related strains to a greater extent than that of *gyrB*. The tanglegram analysis conclusively demonstrates the suitability of ITS1 ribotyping for the resolution of *Enterobacteriaceae* strains.

Discussion

Gibbsiella quercinecans and *Brenneria goodwinii,* belonging to the family *Enterobacteriaceae,* are strongly suspected to play a leading role in AOD, but one of the difficulties faced by microbiologists working in the field is culture-based differentiation of AOD bacterial isolates, which are morphologically similar and require strain purification, DNA extraction, PCR amplification of the *gyrB* gene, sequencing and phylogenetic analysis for taxonomic identification. This study demonstrates that ITS1 profiling can discriminate between species among the family *Enterobacteriaceae* to the same resolution as the most commonly used differentiation marker, DNA gyrase B (*gyrB*). The advantage that ITS1 ribotyping has over *gyrB* is that it negates the requirement for a sequencing reaction and DNA sequence analysis, thereby reducing both cost and time. Polyacrylamide gel electrophoresis successfully separated ITS1 amplicons using intergenic heterogeneity in the ITS1 region using standard protocols. *gyrB* sequencing of individual AOD bacterial isolates validated the ITS1 amplicon separation and phylogenetic analyses providing evidence that it could be used as a relatively rapid and inexpensive environmental bacterial identification method (Fig. 4).

Beyond the successful resolution of the two species of interest (*G. quercinecans* and *B. goodwinii*), our data suggest that ITS1 ribotyping analysis is not suitable for the discrimination of genus-level relationships. As described above, low fidelity of the ITS1 region provides the basis for this study. However, variability of the marker is such that intergenus relationships are idiosyncratic, and meaningful comparisons are limited. This is demonstrated via Figs 3 and 4 where the *G. quercinecans* and *B. goodwinii* species are robustly separated into distinct clades. However, drawing any phylogenetic conclusions between genera for the AOD isolates used in this study is not supported via bootstrapping (Fig. 3). For example, on the PAGE phylogenetic tree, *Rahnella sp.*1 USA39 and *Brenneria salicis* DSM30166 are placed on the same phylogenetic node, yet based on the random sampling of the bootstrapping analysis, this node is not strongly supported; hence, it could not be inferred that these isolates are closer than other intergenus relationships on the tree. Therefore, our data have shown similar to previous studies (Gurtler and Stanisich 1999) that ITS1 copy numbers are strain specific with intergenus-level relationships unclear.

In this study, ITS1 ribotyping has been tested and validated on arboreal *Enterobacteriaceae,* using two independent electrophoresis methods. However, it could equally be applied to other members of the *Enterobacteriaceae* including bacterial groups in clinical studies. There are reports in the literature on the inefficiency of the 16S rRNA gene or other conventional identification methods for screening bacterial isolates due to the sequence similarity between closely related species (Fukushima *et al.* 2002; Paradis *et al.* 2005; Pavlovic *et al.* 2011). For example, *Escherichia coli* and *Shigella* spp. pathovars are closely related with few disparate biochemical characteristics. Consequently, separation using biochemical tests or the 16S rRNA gene will often fail to resolve strains. Therefore, the method we propose here could thus be applied to other situations. Equally, the food industry requires diagnostic mechanisms for *Enterobacteriaceae* species, as these through a variety of species- or strain-specific

mechanisms can spoil meat products (Doulgeraki *et al.* 2011; Mofokeng *et al.* 2011). Through building ITS1 profiles for individual *Enterobacteriaceae* species, the method offers a rapid fingerprinting mechanism without the need for expensive and time-consuming sequencing reactions, which is not only applicable in arboreal studies but across all disciplines of microbiology.

This is a simple, accurate and inexpensive method for the rapid screening of *Enterobacteriaceae* samples and provides robust species-specific profiles. The method allows environmental screening of a large number of samples providing species identification and differentiation data.

Acknowledgements

We thank Susan Kirk and Sarah Plummer for providing AOD isolates.

Funding

J.D. is funded by a scholarship from The Rufford Foundation, to whom we are grateful for their support.

Conflict of interest

No conflict of interest declared.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**, 403–410.

Brady, C., Denman, S., Kirk, S., Venter, S., Rodriguez-Palenzuela, P. and Coutinho, T. (2010) Description of *Gibbsiella quercinecans* gen. nov., sp. nov., associated with Acute Oak Decline. *Syst Appl Microbiol* **33**, 444–450.

Brown, M.V. and Fuhrman, J.A. (2005) Marine bacterial microdiversity as revealed by internal transcribed spacer analysis. *Aquat Microb Ecol* **41**, 15–23.

Cardinale, M., Brusetti, L., Quatrini, P., Borin, S., Puglia, A.M., Rizzi, A., Zanardini, E., Sorlini, C. *et al.* (2004) Comparison of different primer sets for use in automated ribosomal intergenic spacer analysis of complex bacterial communities. *Appl Environ Microbiol* **70**, 6147–6156.

Cedergren, R., Gray, M.W., Abel, Y. and Sankoff, D. (1988) The evolutionary relationships among known life forms. *J Mol Evol* **28**, 98–112.

Daffonchio, D., Borin, S., Frova, G., Manachini, P.L. and Sorlini, C. (1998) PCR fingerprinting of whole genomes: the spacers between the 16S and 23S rRNA genes of intergenic tRNA gene regions reveal a different intraspecific genomic variability of *Bacillus cereus* and *Bacillus licheniformis*. *Int J Syst Evol Microbiol* **48**, 107–116.

Denman, S. and Webber, J. (2009) Oak declines: new definitions and new episodes in Britain. *Quart J Forest* **103**, 285–290.

Denman, S., Brady, C., Kirk, S., Cleenwerck, I., Venter, S., Coutinho, T. and De Vos, P. (2012) *Brenneria goodwinii* sp. nov., a novel species associated with Acute Oak Decline in Britain. *Int J Syst Evol Microbiol* **62**, 2451–2456.

Denman, S., Brown, N., Kirk, S.A., Jeger, M. and Webber, J.F. (2014) A description of the symptoms of Acute Oak Decline in Britain and a comparative review of causes of similar disorders on Oak in Europe. *Forestry* **87**, 535–551.

Doulgeraki, A.I., Paramithiotis, S. and Nychas, G.J.E. (2011) Characterisation of the *Enterobacteriaceae* community that developed during storage of minced beef under aerobic or modified atmosphere packaging condition. *Int J Food Microbiol* **145**, 77–83.

Fisher, M.M. and Triplett, E.W. (1999) Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. *Appl Environ Microbiol* **65**, 4630–4636.

Fukushima, M., Kakinuma, K. and Kawaguchi, R. (2002) Phylogenetic analysis of *Salmonella*, *Shigella*, and *Escherichia coli* strains on the basis of the *gyrB* gene sequence. *J Clin Microbiol* **40**, 2779–2785.

Garcia-Martinez, J., Acinas, S.G., Anton, A.I. and Rodriguez-Valera, F. (1999) Use of the 16S-23S ribosomal genes spacer region in studies of prokaryotic diversity. *J Microbiol Methods* **36**, 55–64.

Gurtler, V. and Stanisich, V.A. (1999) New approaches to typing and identification of bacteria using the 16S-23S rDNA spacer region. *Microbiology* **142**, 3–16.

Huson, D.H. and Scornavacca, C. (2012) Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst Biol* **61**, 1061–1067.

Janda, J.M. and Abbott, S.L. (2007) 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J Clin Microbiol* **45**, 2761–2764.

Jensen, M.A., Webster, A. and Straus, N. (1993) Rapid identification of bacteria on the basis of polymerase chain reaction amplified ribosomal DNA spacer polymorphisms. *Appl Environ Microbiol* **59**, 945–952.

Klappenbach, J.A., Saxman, P.R., Cole, J.R. and Schmidt, T.M. (2001) rrndb: the ribosomal RNA operon copy number database. *Nucleic Acids Res* **29**, 181–184.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M. *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948.

Mofokeng, L., Cawthorn, D.M., Witthuhn, R.C., Anelich, L.E.C.M. and Jooste, P.J. (2011) Characterisation of

Cronobacter species (*Enterobacter sakazakii*) isolated from various South African food sources. *J Food Saf* **31**, 98–107.
Mollet, C., Drancourt, M. and Raoult, D. (1997) *rpoB* sequence analysis as a novel basis for bacterial identification. *Mol Microbiol* **26**, 1005–1011.
Mount, D.W. (2008) Choosing a method for phylogenetic prediction. *Cold Spring Harb Protoc* **3**, 1–3.
Naum, M., Brown, E.W. and Mason-Gamer, R.J. (2008) Is 16S rDNA a reliable phylogenetic marker to characterize relationships below the family level in the *Enterobacteriaceae*? *J Mol Evol* **66**, 630–642.
Paradis, S., Boissinot, M., Paquette, N., Belanger, S.D., Martel, E.A., Boudreau, D.K., Picard, F.J., Ouellette, M. *et al.* (2005) Phylogeny of the *Enterobacteriaceae* based on genes encoding elongation factor Tu and F-ATPase β-subunit. *Int J Syst Evol Microbiol* **55**, 2013–2025.
Pavel, A.B. and Vasile, C.I. (2012) PyElph – a software tool for gel images analysis and phylogenetics. *BMC Bioinformatics* **13**, 9.
Pavlovic, M., Luze, A., Konrad, R., Berger, A., Sing, A., Busch, U. and Huber, I. (2011) Development of a duplex real-time PCR for differentiation between *E. coli* and *Shigella* spp. *J Appl Microbiol* **110**, 1245–1251.
Rappe, M.S., Connon, S.A., Vergin, K.L. and Giovannoni, S.J. (2002) Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* **418**, 630–633.
Scheinert, P., Krausse, R., Ullmann, U., Soller, R. and Krupp, G. (1996) Molecular differentiation of bacteria by PCR amplification of the 16S-23S rRNA spacer. *J Microbiol Methods* **26**, 103–117.
Woese, C.R. and Fox, G.E. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci USA* **74**, 5088–5090.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1 Validation of the ITS1 method, with complete characterisation of six bacterial AOD isolates.

Table S2 Identification of AOD isolates using two molecular markers (*gyrB* and ITS1) and two methods; agarose gel electrophoresis (column 4), *gyrB* sequencing (column 2) and re-sequencing to test purity of selected laboratory cultures (column 5).

APPENDIX II

Supplementary Table for Chapter 2

Supplementary Table 2.1. Validation of the ITS1 method, with complete characterisation of six bacterial AOD isolates. Column 1 gives the original bacterial identification using *gyrB* sequencing of a single bacterial isolate, column 2 gives the results of subsequent re-sequencing of the *gyrB* gene for 5 sub-isolates, column 3 gives the number of amplicons produced from ITS1 amplification using an agarose gel for visualisation and column 4 gives the identity of these amplicons when extracted from an agarose gel and sequenced, GenBank accession numbers are in parentheses. N.B. All amplicon sizes are approximations.

Isolate (sub-isolate)	Sequence ID (*gyrB*)	Number of ITS1 amplicons and estimated amplicon size (bp)	Sequence ID (ITS1)
Gibbsiella quercinecans FRB98 (I8A)	*Enterobacter cloacae*	3 (230, 400, 420)	3 amplicons sequenced, all ITS1 (KJ418767-9)
Gibbsiella quercinecans FRB98 (I8B)	*Enterobacter* sp.	3 (230, 400, 420)	3 amplicons sequenced, all ITS1 (KJ418770-2)
Gibbsiella quercinecans FRB98 (I8C)	*Enterobacter cloacae*	3 (230, 400, 420)	3 amplicons sequenced, all ITS1 (KJ418773-5)
Gibbsiella quercinecans FRB98 (I8D)	*Gibbsiella quercinecans* FRB98	6 (340, 360, 380, 390, 400, 450)	4 amplicons sequenced, all ITS1 (KJ418748-51)
Gibbsiella quercinecans FRB98 (I8E)	*Gibbsiella quercinecans* FRB98	6 (340, 360, 380, 390, 400, 450)	5 amplicons sequenced, all ITS1 (KJ418752-6)
Ewingella sp.(AT14bA)	*Ewingella americana*	7 (370, 400, 450, 475, 500, 550,580)	3 amplicons sequenced, all ITS1 (KJ418757-9)
Ewingella sp.(AT14bA)	*Ewingella americana*	7 (370, 400, 450, 475, 500, 550,580)	1 amplicon sequenced, ITS1 (KJ418760)
Ewingella sp.(AT14bA)	*Ewingella americana*	7 (370, 400, 450, 475, 500, 550,580)	2 amplicons sequenced, all ITS1 (KJ418761-2)
Ewingella sp.(AT14bA)	*Ewingella americana*	7 (370, 400, 450, 475, 500, 550,580)	2 amplicons sequenced, all ITS1 (KJ418763-4)
Ewingella sp.(AT14bA)	*Ewingella americana*	7 (370, 400, 450, 475, 500, 550,580)	2 amplicons sequenced, all ITS1 (KJ418765-6)
Brenneria goodwinii FRB184 (I67A)	*Brenneria goodwinii* FRB184	6 (340, 360, 370, 400, 430, 520)	3 amplicons sequenced, all ITS1 (KJ418803-5)
Brenneria goodwinii FRB184 (I67B)	*Brenneria goodwinii* FRB184	6 (340, 360, 370, 400, 430, 520)	2 amplicons sequenced, all ITS1 (KJ418806-7)
Brenneria goodwinii FRB184 (I67C)	*Brenneria goodwinii* FRB184	6 (340, 360, 370, 400, 430, 520)	3 amplicons sequenced, all ITS1 (KJ418808-10)

Brenneria goodwinii FRB184 (I67D)	*Brenneria goodwinii* FRB184	6 (340, 360, 370, 400, 430, 520)	2 amplicons sequenced, all ITS1 (KJ418811-2)
Brenneria goodwinii FRB184 (I67E)	*Brenneria goodwinii* FRB184	6 (340, 360, 370, 400, 430, 520)	1 amplicon sequenced, ITS1 (KJ418813)
Brenneria goodwinii WW2 (I77A)	*Brenneria goodwinii* WW2	4 (370, 420, 570, 590)	1 amplicon sequenced, ITS1 (KJ418776)
Brenneria goodwinii WW2 (I77B)	*Brenneria goodwinii* WW2	4 (370, 420, 570, 590)	3 amplicons sequenced, all ITS1 (KJ418777-9)
Brenneria goodwinii WW2 (I77C)	*Brenneria goodwinii* WW2	4 (370, 420, 570, 590)	4 amplicons sequenced, all ITS1 (KJ418780-3)
Brenneria goodwinii WW2 (I77D)	*Brenneria goodwinii* WW2	4 (370, 420, 570, 590)	3 amplicons sequenced, all ITS1 (KJ418784-6)
Brenneria goodwinii WW2 (I77E)	*Brenneria goodwinii* WW2	4 (370, 420, 570, 590)	4 amplicons sequenced, all ITS1 (KJ418787-90)
Rahnella sp. (USA39A)	*Rahnella* sp.	4 (350, 500, 620, 850)	3 amplicons sequenced, all ITS1 (KJ418791-3)
Rahnella sp. (USA39B)	*Rahnella* sp.	4 (350, 500, 620, 850)	3 amplicons sequenced, all ITS1 (KJ418794-6)
Rahnella sp. (USA39C)	*Rahnella* sp.	4 (350, 500, 620, 850)	2 amplicons sequenced, all ITS1 (KJ418797-8)
Rahnella sp. (USA39D)	*Rahnella* sp.	4 (350, 500, 620, 850)	3 amplicons sequenced, all ITS1 (KJ418799-801)
Rahnella sp. (USA39E)	*Rahnella* sp.	4 (350, 500, 620, 850)	1 amplicon sequenced, ITS1 (KJ418802)
Gibbsiella quercinecans FRB147 (17A)	*Gibbsiella quercinecans FRB147*	6 (330, 370, 380, 440, 560, 580)	5 amplicons sequenced, all ITS1 (KJ418814-18)
Gibbsiella quercinecans FRB147 (17B)	*Gibbsiella quercinecans FRB147*	6 (330, 370, 380, 440, 560, 580)	3 amplicons sequenced, all ITS1 (KJ418819-21)
Gibbsiella quercinecans FRB147 (17C)	*Gibbsiella quercinecans FRB147*	6 (330, 370, 380, 440, 560, 580)	4 amplicons sequenced, all ITS1 (KJ418822-25)
Gibbsiella quercinecans FRB147 (17D)	*Gibbsiella quercinecans FRB147*	6 (330, 370, 380, 440, 560, 580)	4 amplicons sequenced, all ITS1 (KJ418826-29)
Gibbsiella quercinecans FRB147 (17E)	*Gibbsiella quercinecans FRB147*	6 (330, 370, 380, 440, 560, 580)	5 amplicons sequenced, all ITS1 (KJ418830-4)

APPENDIX III

Supplementary Table for Chapter 2

Supplementary Table 2.2. Identification of AOD isolates using two molecular markers (*gyrB* and ITS1) and two methods; agarose gel electrophoresis (column 4), *gyrB* sequencing (column 2) and re-sequencing to test purity of selected laboratory cultures (column 5). Each isolate was sub-isolated five times and designated A-E. N.B. ITS1 amplicon sizes are approximations.

Code	*gyrB* identification	Isolate number	No. of ITS1 amplicons and estimated size (agarose gel)	*gyrB* re-sequencing
1	*G. quercinecans*	FRB24		
6	*G. quercinecans*	FRB88	6 (300-700)	
7	*G. quercinecans*	FRB89	6 (350-600)	
8	*G. quercinecans*	FRB91	3 A-C (300-400), 6 D-E (350-600)	*Enterobacter cloacae* A-C, *Gibbsiella quercinecans* D-E
10	*G. quercinecans*	FRB92		
11	*G. quercinecans*	FRB93		
13	*G. quercinecans*	FRB97(T)	6 (350-600)	
14	*G. quercinecans*	FRB98	6 (450-750)	
16	*G. quercinecans*	FRB115		
17	*G. quercinecans*	FRB116	6 (350-600)	*G. quercinecans* A-E
18	*G. quercinecans*	FRB124		
19	*G. quercinecans*	FRB147		

20	*G. quercinecans*	FRB169	7 (350-600)	*G. quercinecans* A-E
21	*G. quercinecans*	FRB185	6 (220-600)	
22	*G. quercinecans*	FRB214		
24	*G. quercinecans*	N78		
26	*G. quercinecans*	N79	3 (350-550)	
31	*G. quercinecans*	FOD 9/25		
32	*G. quercinecans*	FOD 9/25	5 (350-550)	
34	*G. quercinecans*	FOD 23/4	5 (350-600)	
37	*G. quercinecans*	BH 1/44b	5 (350-550)	
39	*G. quercinecans*	BH 1/65b	4 (350-550)	
40	*G. quercinecans*	USA 42	5 (300-600)	
41	*G. quercinecans*	W10	4 (350-550)	*G. quercinecans* A-E
43	*G. quercinecans*	BW 2/28		
46	*G. quercinecans*	BER 12		
49	*G. quercinecans*	BER 14	7 (350-550)	*G. quercinecans* A-E
52	*G. quercinecans*	SOT 3/3a	6 (350-550)	
57	*B. goodwinii*	FRB135	4 (450-650)	
59	*B. goodwinii*	FRB141	4 (450-650)	
60	*B. goodwinii*	FRB171	6 (350-600)	*B. goodwinii* A-E
61	*B. goodwinii*	FRB173	6 (450-600)	
62	*B. goodwinii*	FRB177		
64	*B. goodwinii*	FRB182		
67	*B. goodwinii*	FRB184	6 (350-550)	*B. goodwinii* A-E
68	*B. goodwinii*	FRB186	5 A-D (350-600), E did not amplify	
70	*B. goodwinii*	FRB193		

71	*B. goodwinii*	SOT 2/28a		
73	*B. goodwinii*	SOT 3/4a		
75	*B. goodwinii*	CE 1/7		
77	*B. goodwinii*	WW2	5 (350-600)	*B. goodwinii* A-E
79	*B. goodwinii*	PIC 4/28	5 (350-600)	
82	*B. goodwinii*	BH 4/25		
84	*B. goodwinii*	BH 4/41	6 (350-600)	*B. goodwinii* A-E
87	*B. goodwinii*	BW 2/24		
88	*B. goodwinii*	PFK 1/3B	5 (350-550)	

APPENDIX IV

Supplementary Table for Chapter 4

Supplementary Table 4.1. Annotated genes downstream, upstream, and spanning contig breaks of the *G. quercinecans* FRB97 Illumina sequenced genome assembly. Contigs ends were analysed beginning from the 3' end of contig 1. Spanning genes covered a portion of the 3' contig and the subsequent 5' contig. NA – non-applicable, as no gene was aligned against contig segment.

Contig	Downstream gene	Spanning gene	Upstream gene
1	Hypothetical	NA	Ureidoglycolate lyase
2	Hypothetical	NA	hypothetical
3	Glucoronate transporter	NA	Prophage regulatory protein
4	Transposase	NA	Transposase
5	Exodeoxyribonuclease	Integrase	Putrescine importer
6	Colicin I receptor	NA	Iron hydroxamate import ATP-binding protein

7	Dehydrorhamnose reductase	Dehydrorhamnose epimerase	Glucose-1-phosphate thymidylyltransferase
8	Dehydrorhamnose reductase	Dehydrorhamnose epimerase	Glucose-1-phosphate thymidylyltransferase
9	Hypothetical	NA	hypothetical
10	Hypothetical	NA	hypothetical
11	Glycerate 2-kinase	Transposase	Hfq binding sRNA
12	Clp protease	Transposase	Hfq binding sRNA
13	Hypothetical	NA	Integrase
15	Integrase	NA	Ribose-5-phosphate isomerase B
16	Nitrate reductase	Integrase	Transcriptional repressor
17	Hypothetical	NA	16S rRNA
18	Glutamylputrescine oxidoreductase	NA	Spermidine/putrescine-binding periplasmic protein
19	tRNA	Integrase	Ferrous iron transport protein
20	5S rRNA	NA	23S rRNA
21	16S rRNA	NA	Bifunctional purine biosynthesis protein
22	Translocase subunit secE	Elongation factor Tu	tRNA
23	23S rRNA	NA	16S rRNA

24	Transcriptional activator protein	NA	DNA-3-methyladenine glycosylase
25	16S rRNA	NA	23S rRNA
26	Hypothetical	NA	Hypothetical
27	16S rRNA	NA	23S rRNA
28	Ketol-acid reductoisomerase	Integrase	2-dehydro-3-deoxygluconokinase (KdgK)
29	CDP-diacylglycerol pyrophosphatase	Inner membrane metabolite transport protein	Inner membrane protein
30	Elongation factor G	Elongation factor Tu	Bacterioferritin-associated ferredoxin
31	16S rRNA	NA	23S rRNA
32	Hypothetical	Hypothetical	hypothetical
33	Osmolarity sensor protein	NA	Integrase
34	Fumarate reductase flavoprotein	NA	tartrate/succinate antiporter
35	Cation-transporting ATPase	NA	Multidrug resistance protein
36	Transposase	NA	Acetyltransferase
37	Transketolase	NA	Malate dehydrogenase
38	16S rRNA	NA	23S rRNA
39	Transcription regulator GntR family	Transposase	Phospholipase

40	Carboxyvinyl-carboxyphosphonate	Integrase	Glutathione-regulated potassium-efflux system ancillary protein
41	Hfq binding sRNA	Transposase	Hfq binding sRNA
42	Integrase	NA	Transketolase
43	Sorbitol operon regulator	Transketolase	Transaldolase
44	Transketolase	Transketolase	Transaldolase
45	Deoxynucleotidase	Integrase	Glutamate-pyruvate aminotransferase
46	Hfq binding sRNA	Transposase	Aliphatic sulfonates import ATP-binding protein SsuB
47	Cryptic outer membrane porin	Integrase	Carboxymethylenebutenolidase
48	Transcriptional regulator LysR family	NA	Decarboxylase
49	Methanesulfonate monooxygenase	NA	Transcriptional regulator LysR family
50	Bacilysin exporter	Alanine-anticapsin ligase	Purine efflux pump
51	Deoxyribonuclease	Transposase	Transposase
52	RyeB ncRNA	NA	Abortive infection phage resistance (AIPR) protein
53	Hfq binding sRNA	Transposase	Phospholipase
54	Hfq binding sRNA	Transposase	Superoxide dismutase [Fe]
55	Beta-lactamase	NA	Melibiase

56	Transcriptional regulator LysR family	NA	Transposase

APPENDIX V

Supplementary Table for Chapter 4

Supplementary Table 4.2. Annotated genes downstream, upstream, and spanning contig breaks of the *B. goodwinii* FRB141 Illumina sequenced genome assembly. Contigs ends were analysed beginning from the 3' end of contig 1. Spanning genes covered a portion of the 3' contig and the subsequent 5' contig. NA – non-applicable, as no gene was aligned against contig segment.

Contig	Downstream	Spanning	Upstream
1	Hypothetical	NA	Acyl-CoA dehydrogenase
2	Aliphatic amidase expression-regulating protein	NA	Hypothetical
3	Nucleoside triphosphate pyrophosphohydrolase	NA	Glutamylputrescine oxidoreductase
4	Enolase	Transposase	Glucuronide carrier protein
5	16S rRNA	NA	23S rRNA
6	Shikimate kinase	NA	Inner membrane transport protein
7	Transcriptional regulator YgaV	NA	Hypothetical

8	Hypothetical	NA	NA
9	NA	NA	NA
10	NA	NA	Hypothetical
11	Hypothetical	NA	DNA repair protein
12	DNA repair protein	DNA repair protein	Hypothetical
13	Hypothetical	Hypothetical	Hypothetical
14	Glucans biosynthesis glucosyltransferase	NA	Hypothetical
15	Mobilization protein A	NA	Hypothetical
16	Hypothetical	DNA repair protein	Hypothetical
17	Conjugal transfer protein TraD	NA	Hypothetical
18	Hypothetical	NA	DNA binding protein
19	Hypothetical	NA	Pilus protein
20	Shufflon protein	NA	Hypothetical
21	Hypothetical	NA	Hypothetical
22	mRNA interferase	Transposase	Integrase
23	16S rRNA	NA	Aspartate 1-decarboxylase

24	Cyclic di-GMP phosphodiesterase	NA	Pilus protein
25	Type I SS agglutinin Rtx	NA	Type I SS agglutinin Rtx
26	Type I SS agglutinin Rtx	NA	Type I SS agglutinin Rtx
27	Type I SS agglutinin Rtx	NA	Type I SS agglutinin Rtx
28	Type I SS agglutinin Rtx	NA	Type I SS agglutinin Rtx
29	Type I SS agglutinin Rtx	NA	Type I SS agglutinin Rtx
30	Type I SS agglutinin Rtx	NA	Type I SS agglutinin Rtx
31	Type I SS agglutinin Rtx	NA	Type I SS agglutinin Rtx
32	Type I SS agglutinin Rtx	NA	Type I SS agglutinin Rtx
33	Type I SS agglutinin Rtx	NA	Type I SS agglutinin Rtx
34	Type I SS agglutinin Rtx	Type I SS agglutinin Rtx	Type I SS agglutinin Rtx
35	Hypothetical	DNA-binding transcriptional regulator	Hypothetical
36	Dicarboxylate transporter	NA	Outer membrane pore protein
37	Inositol-1-monophosphatase	NA	Hypothetical
38	Galactokinase	Phage-related baseplate assembly protein	Hypothetical
39	Methyl-accepting chemotaxis protein	NA	O-methyltransferase

40	DNA binding protein	NA	3-octaprenyl-4-hydroxybenzoate carboxyl-lyase
41	ABC transporter ATP-binding protein	NA	Hydrogenase
42	Nodulation protein	NA	Hypothetical
43	Major myo-inositol transporter	NA	Glucose--fructose oxidoreductase
44	Inositol 2-dehydrogenase	NA	Hydrolase
45	5-dehydro-2-deoxygluconokinase	NA	Inositol 2-dehydrogenase
46	Hypothetical	Hypothetical	Hydrolase
47	Chaperone protein ClpB	Phage-related baseplate assembly protein	Lipase
48	Hypothetical	Hypothetical	Hypothetical
49	Type VI secretion protein VasK	Phage-related baseplate assembly protein	PAAR motif protein
50	Hypothetical	Hypothetical	Hypothetical
51	Hypothetical	Hypothetical	Hypothetical
52	Hypothetical	NA	Hypothetical
53	Hypothetical	NA	Hypothetical
54	Hypothetical	Hypothetical	Hypothetical
55	Hypothetical	NA	Hypothetical

56	Hypothetical	NA	Type VI secretion protein VasK
57	Phage-related baseplate assembly protein	NA	Hypothetical
58	Hypothetical	NA	Hypothetical
59	Hypothetical	NA	Nitroreductase
60	UvrABC system protein	NA	Fumarate reductase flavoprotein
61	Hypothetical	Hypothetical	Hypothetical
62	Glutamine-binding periplasmic protein	NA	DNA protection during starvation protein
63	Hypothetical	NA	Hypothetical
64	Hypothetical	NA	Hypothetical
65	Hypothetical	Nucleoid-associated protein	tRNA
66	Hypothetical	NA	Homoserine/homoserine lactone efflux protein
67	Lipid A biosynthesis lauroyl acyltransferase	NA	NADPH dehydrogenase
68	Lipoate-protein	Diguanylate phosphodiesterase	Hypothetical
69	Phosphoenolpyruvate synthase;	NA	Thiamine biosynthesis lipoprotein
70	Hypothetical	NA	Iron-binding protein
71	Aliphatic sulfonates-binding protein	NA	Aliphatic sulfonates-binding protein

72	Hypothetical	NA	5-keto-4-deoxy-D-glucarate aldolase
73	Glucarate transporter;	NA	Transcriptional regulator
74	Ferrienterobactin receptor	NA	Isochorismate synthase
75	L-lysine N6-monooxygenase	NA	Lysine-arginine-ornithine-binding periplasmic protein
76	Methyl-accepting chemotaxis protein	NA	Signal transduction histidine-protein
77	Hypothetical	Hypothetical	Hypothetical
78	Hypothetical	NA	Hypothetical
79	Hypothetical	NA	FMN-dependent NADH-azoreductase
80	Toxin-antitoxin protein SymE	NA	O-succinylbenzoate synthase
81	Hypothetical	NA	Hypothetical
82	Outer membrane protein W	NA	Flagellin
83	Serine dehydratase	NA	Magnesium and cobalt efflux protein
84	Ferritin	NA	DNA polymerase III
85	Phosphatase	tRNA	Ribosome modulation factor
86	Dihydroorotate dehydrogenase	NA	Lipoprotein
87	Flagellar transcriptional regulator FlhD	NA	Methyl-accepting chemotaxis protein

88	tRNA	NA	D-serine/D-alanine/glycine transporter
89	Amidohydrolase (EC 3.5.1.47)	Phage-related baseplate assembly protein	Hypothetical
90	Fructose-6-phosphate aldolase	NA	Hypothetical
91	6-oxocamphor hydrolase	NA	Transcriptional regulator LysR
92	Taurine-binding periplasmic protein	NA	Inner membrane ABC transporter
93	Hypothetical	NA	Anaerobic C4-dicarboxylate transporter
94	Oxidoreductase	NA	Sulfonate ABC transporter
95	Formate hydrogenlyase	NA	Proton glutamate symport protein
96	Hypothetical	NA	Iron utilisation protein
97	FhuE receptor	NA	Iron-hydroxamate import protein
98	Iron-hydroxyapatite import ATP-binding protein	NA	Aminolevulinic acid dehydratase
99	Pyrogallol hydroxytransferase	NA	Transport protein HsrA
100	LexA repressor	NA	Cyclic-di-GMP phosphodiesterase
101	Type-1 fimbrial protein	Transcriptional regulator	Transcriptional regulator
102	Transcriptional regulator	Transcriptional regulator	PTS system N-acetylmuramic acid
103	Hypothetical	Hypothetical	D-inositol 3-phosphate glycosyltransferase

104	Hypothetical	NA	Hypothetical
105	Polysaccharide pyruvyl transferase	Polysaccharide pyruvyl transferase	Polysaccharide pyruvyl transferase
106	Polysaccharide pyruvyl transferase	Polysaccharide pyruvyl transferase	Polysaccharide biosynthesis protein
107	FMN reductase	NA	Glucosyltransferase
108	Glucosyltransferase	NA	Dehydrorhamnose reductase
109	UDP-glucose	Undecaprenyl-phosphate	Tyrosine-protein kinase
110	Formate dehydrogenase	NA	ABC transporter
111	Hypothetical	NA	Potassium transport regulator
112	Hypothetical	NA	Transcriptional regulator CadC
113	Hypothetical	NA	Hypothetical
114	Octanoyltransferase	NA	Lipoyl synthase
115	Inner membrane transport protein	NA	Hypothetical
116	Hypothetical	DNA replication and repair protein RecF	HNH endonuclease
117	HNH endonuclease	NA	Hypothetical
118	tRNA threonylcarbamoyladenosine dehydratase	Hypothetical	Dimethyl sulfoxide
119	Plipastatin synthase	NA	Tyrocidine synthase

120	Disulfide-bond oxidoreductase	NA	Outer membrane protein OprM
121	Gluconate 5-dehydrogenase	NA	Starvation-sensing protein RspA
122	Chromosome partition protein Smc	NA	Phage/plasmid primase P4
123	Homoserine/homoserine lactone efflux protein	NA	Negative regulatory protein PAI
124	Glutamine transport ATP-binding protein	NA	Homoserine kinase
125	Hypothetical	NA	Hypothetical
126	Fec operon regulator	NA	Fe-pyochelin receptor
127	Limonene 2-monooxygenase	Transcriptional regulator LacI	PTS system mannose
128	PTS system fructose	NA	Oligo-6-glucosidase (EC 3.2.1.10)
129	Restriction enzyme	NA	Hypothetical
130	Hypothetical	NA	Transcriptional regulator OmpR
131	Sugar-binding periplasmic protein	NA	β-glucosidase (EC 3.2.1.21)
132	2-octaprenylphenol hydroxylase	NA	Salicylate biosynthesis protein
133	Antitoxin igA-2	ParB-like nuclease domain protein	Modification methylase DpnIIA
134	Modification methylase DpnIIA	Modification methylase DpnIIA	Hypothetical
135	Hypothetical	NA	Diacetyl reductase

136	Methyl-accepting chemotaxis protein	NA	Peptidase
137	DNA-binding protein Fis	Hypothetical	Hypothetical
138	23S rRNA	NA	Hypothetical
139	23S rRNA	NA	16S rRNA
140	Acetyltransferase	NA	Amino acid binding protein
141	Transcriptional regulator	NA	2-diketo-L-gulonate reductase
142	Cellulose synthase operon protein C	NA	Haemolysin transporter protein
143	5S rRNA	NA	16S rRNA
144	tRNA	NA	Hypothetical
145	Hypothetical	NA	Adenine permease
146	Acyl carrier protein	NA	Hypothetical
147	Hypothetical	NA	Hypothetical
148	L-tartrate/succinate antiporter	NA	NA
149	NA	NA	Fumarate reductase
150	Glyoxylate/hydroxypyruvate reductase	NA	Aldolase
151	Multiple stress resistance protein BhsA	NA	TolB

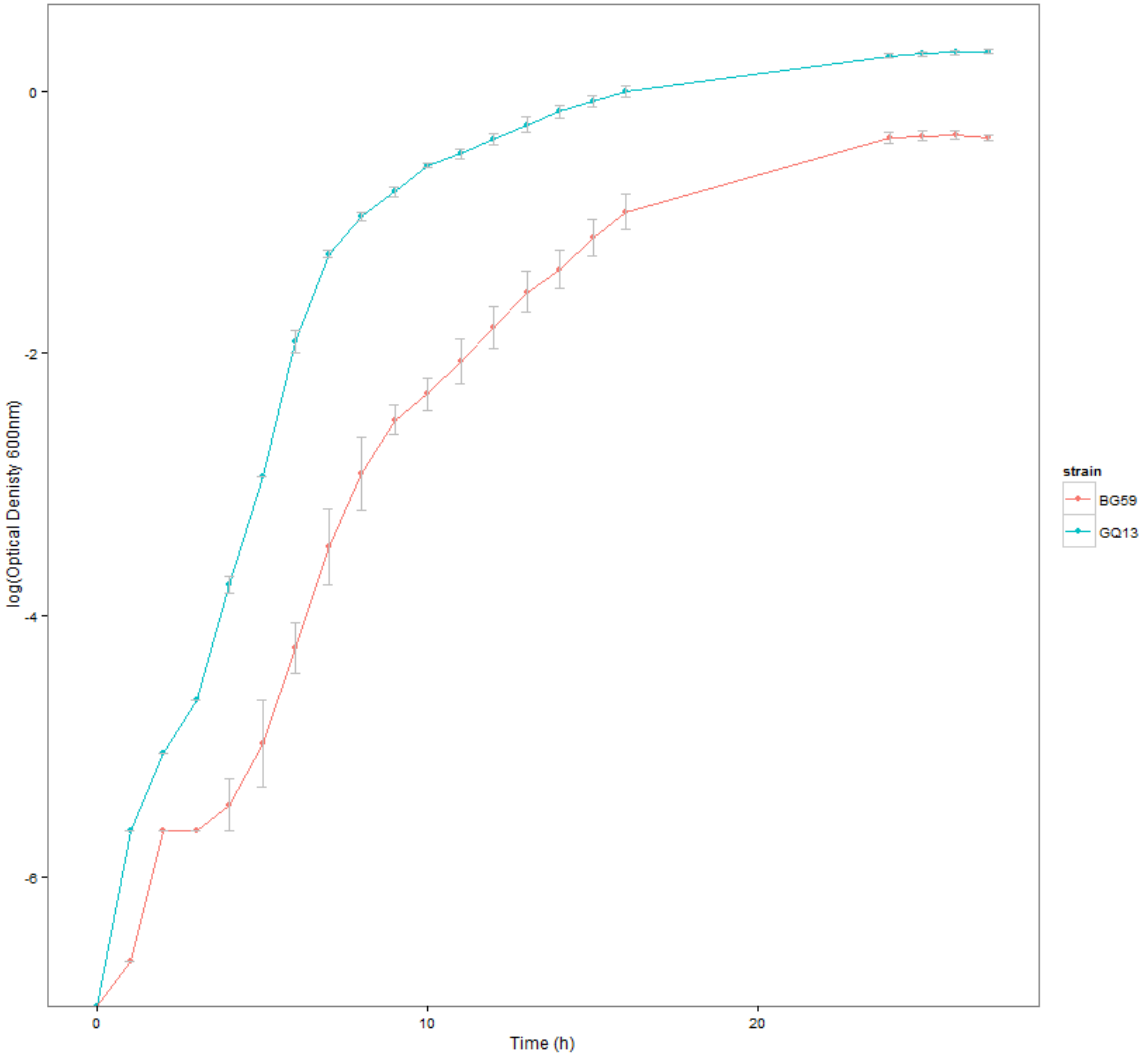
152	TolB	TolB	Hypothetical
153	Aspartate/alanine antiporter	NA	Malate transporter
154	Citrate/malate transporter	NA	Periplasmic dipeptide transport protein
155	NAD-dependent dihydropyrimidine dehydrogenase	NA	D-hydantoinase/dihydropyrimidinase
156	Hypothetical	NA	Hypothetical
157	Glycine--tRNA ligase	NA	Aldehyde-alcohol dehydrogenase
158	Hypothetical	KAP family P-loop domain protein	KAP family P-loop domain protein
159	KAP family P-loop domain protein		radC RNA motif
160	Formamidopyrimidine-DNA glycosylase	Lipopolysaccharide heptosyltransferase	RES domain protein
161	3-deoxy-D-manno-octulosonic acid transferase	NA	Glucosyltransferase
162	Lipopolysaccharide core heptosyltransferase	O-antigen ligase	JumpStart RNA motif
163	16S rRNA	16S rRNA	23S rRNA
164	Sensor histidine kinase	NA	Lysine-arginine-ornithine-binding periplasmic protein
165	Elongation factor G	Elongation factor Tu	Bacterioferritin-associated ferredoxin
166	16S rRNA	23S rRNA	5S rRNA
167	Tyrosine recombinase	NA	Hypothetical

168	Hypothetical	NA	Hypothetical
169	Hypothetical	NA	Hypothetical
170	Gluconate 5-dehydrogenase	NA	2-isopropylmalate synthase
171	Glycine cleavage system	NA	Transposase
172	Hypothetical	Exoglucanase	Exoglucanase
173	Exoglucanase	Exoglucanase	Exoglucanase
174	Exoglucanase	Exoglucanase	Exoglucanase
175	Cytosine kinase	Transcriptional regulator	Acetyltransferase
176	Hypothetical	NA	Hypothetical
177	Thiamine pyrophosphate-containing protein	NA	Transcriptional regulator
178	6-phosphofructokinase	NA	α-glucosidase
179	Transcriptional regulator	NA	Metalloprotease
180	Diguanylate cyclase	NA	Molybdopterin adenylyltransferase
181	Fumarate reductase	NA	Alkanesulfonate transporter
182	L-alanine exporter	NA	MraZ
183	Regulatory protein AmpE	Phage-related baseplate assembly protein	VRR-NUC domain protein

184	Hypothetical	NA	Reverse gyrase
185	Phage tail protein	NA	Hypothetical

APPENDIX VI

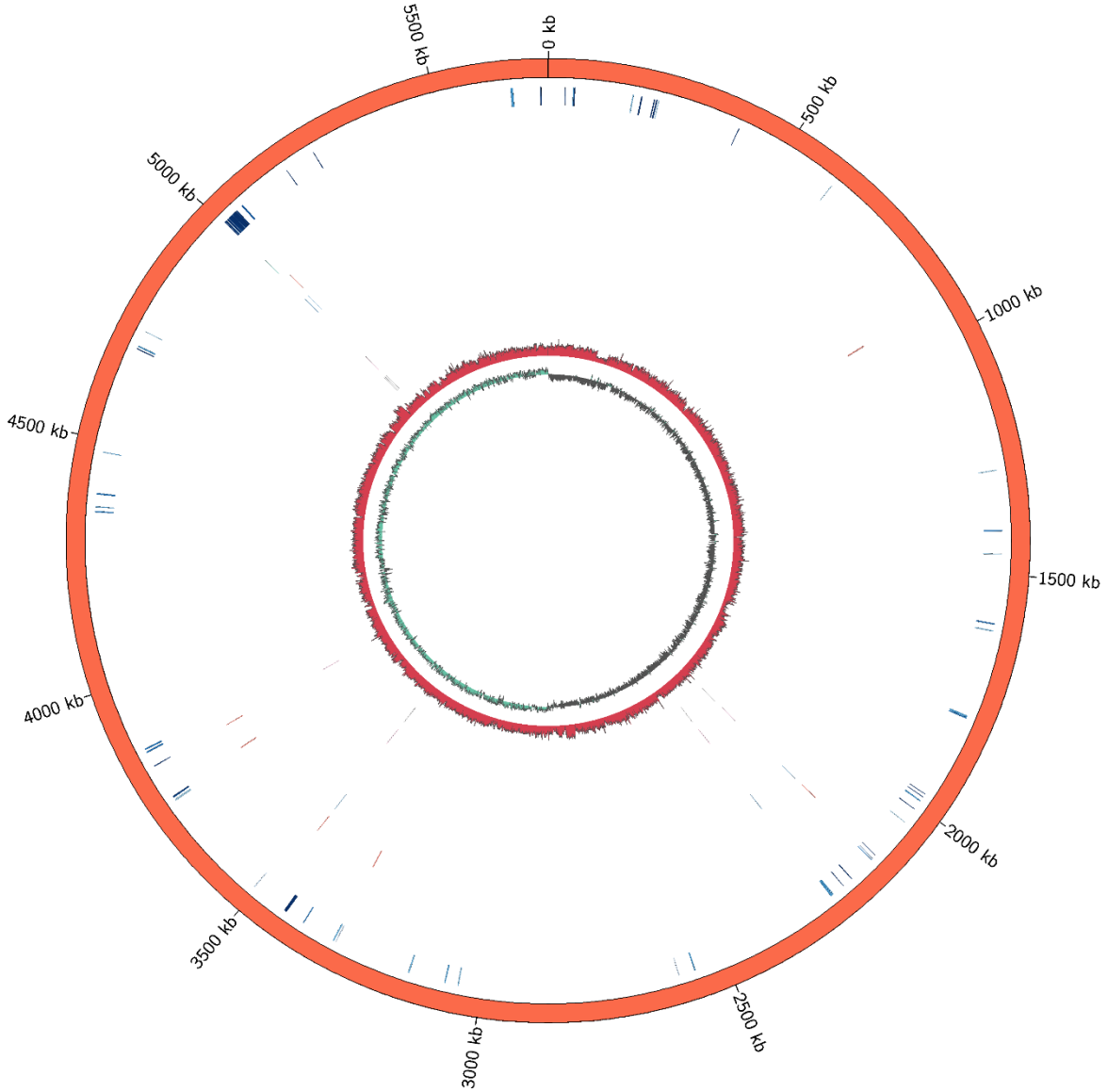
Supplementary Figure for Chapter *5*


0
-2
-4
-6
log(Optical Denisty 600nm)
0
10
20
Time (h)
strain
BG59
GQ13

Supplementary Figure 5.1. Growth curve of *Gibbsiella quercinecans* FRB97 and *Brenneria goodwinii* FRB141 over 24 hours. Samples were incubated in 125 ml culture flasks and aerated by shaking at 100 rpm. Optical density measurements were taken every 2 hours at 600 nm. Blue = *G. quercinecans* FRB97, Red = *B. goodwinii* FRB141.

APPENDIX VII

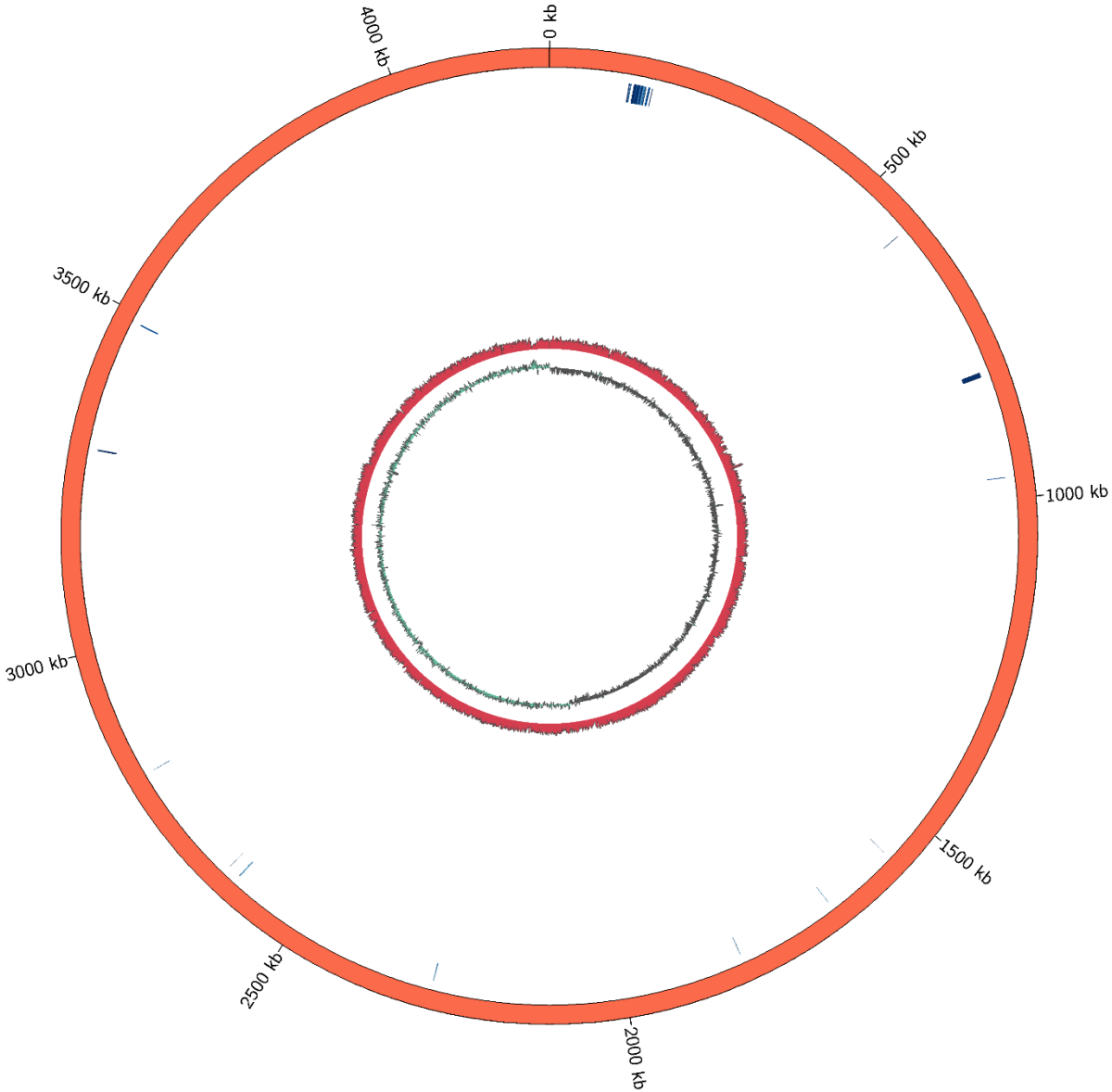
Supplementary Figure for Chapter 6


0 kb
500 kb
1000 kb
1500 kb
2000 kb
2500 kb
3000 kb
3500 kb
4000 kb
4500 kb
5000 kb
5500 kb

Supplementary Figure 6.1. AOD metagenome coding domains and metatranscripts aligned against the *Paenibacillus polymyxa* SC2 genome. The outermost circle represents a circular reconstruction of *P. polymyxa* SC2 genome. The following circles represent, from outside to inside: 1) metatranscriptome heatmap, with increased blue saturation representing increasing read alignment, 2) Attingham healthy (AH) aligned metagenome coding domains, 3) Attingham mid asymptomatic (AMA) aligned metagenome coding domains, 4) Attingham early symptomatic (AES) aligned metagenome coding domains, 5) Attingham midstage symptomatic (AMS) aligned metagenome coding domains, 6) Attingham late symptomatic (ALS) aligned metagenome coding domains, 7) Runs Wood early symptomatic (RES) aligned metagenome coding domains, 8) Runs Wood mid symptomatic (RMS) aligned metagenome coding domains, 9) Runs Wood late symptomatic (RLS) aligned metagenome coding domains, 10) G+C content across the *P. polymyxa* SC2 genome, 11) G+C skew across the *P. polymyxa* SC2 genome.

APPENDIX VIII

Supplementary Figure for Chapter 6


0 kb
500 kb
1000 kb
1500 kb
2000 kb
2500 kb
3000 kb
3500 kb
4000 kb

Supplementary Figure 6.2. AOD metagenome coding domains and metatranscripts aligned against the *Bacillus licheniformis* ATCC14580 genome. The outermost circle represents a circular reconstruction of *B. licheniformis* ATCC14580 genome. The following circles represent, from outside to inside: 1) metatranscriptome heatmap, with increased blue saturation representing increasing read alignment, 2) Attingham healthy (AH) aligned metagenome coding domains, 3) Attingham mid asymptomatic (AMA) aligned metagenome coding domains, 4) Attingham early symptomatic (AES) aligned metagenome coding domains, 5) Attingham midstage symptomatic (AMS) aligned metagenome coding domains, 6) Attingham late symptomatic (ALS) aligned metagenome coding domains, 7) Runs Wood early symptomatic (RES) aligned metagenome coding domains, 8) Runs Wood mid symptomatic (RMS) aligned metagenome coding domains, 9) Runs Wood late symptomatic (RLS) aligned metagenome coding domains, 10) G+C content across the *B. licheniformis* ATCC14580 genome, 11) G+C skew across the *B. licheniformis* ATCC14580 genome.