Cooperative Research Centre
for National Plant Biosecurity

Final Report

CRC20054

Platforms to differentiate exotic pathovars of plant bacteria

Authors

Deborah Hailstones, Jo Luck, Simone Rochford, Dave Berryman,

Toni Chapman, Celia Smuts, Michelle Berg, Anna Englezou,

Vilnis Ezernieks, Nuttima Kositcharoenkul, Mike Jones, Mehmet Cakir

April 2012

Project Leader contact details:

Name: Dr Deborah Hailstones
Address: EMAI
Phone: 02 4640 6442
Fax: 02 4640 6300
Email: deborah.hailstones@dpi.nsw.gov.au

CRCNPB contact details:

Cooperative Research Centre for National Plant Biosecurity
LPO Box 5012
Bruce ACT 5012

Phone: +61 (0)2 6201 2882
Fax: +61 (0)2 6201 5067
Email: info@crcplantbiosecurity.com.au
Web: www.crcplantbiosecurity.com.au

Table of contents

1. Executive Summary

Many of the EPPs that pose the biggest threat to the biosecurity of Australia's plant industries are bacterial, but difficulties in identification to the subspecific or 'pathovar' level can seriously delay incursion management and affect market access. Pathovars are defined by host specificity so bioassays remain the definitive means of identification, but these require high level physical containment and can be slow and subjective, delaying diagnosis. Some pathovar-specific serological and molecular tests are available but better diagnostic methods are often required. This project used proteomics and metabolomics, platforms that identify functional molecules potentially associated with plant-pathogen interactions, to identify biomarkers that differentiate pathovars in species of *Xanthomonas*.

Membrane-associated proteins from a collection of bacterial isolates were compared on 2Dimensional gels. Proteins that were found to be differentially expressed between distinct pathovars may be important modulators of host specificity so they were identified and the genes that encode them located by reference to genomic sequences. DNA-based assays targeting these genes were designed and validated for their specificity to the pathovar level. We have developed two new assays that provide levels of specificity not reported elsewhere in the literature. These assays specifically target the bacteria causing the different forms of citrus canker, but without cross-reaction to the closely-related organisms causing bacterial blight on cotton and Citrus Bacterial Spot. The molecular assays will be incorporated into the National Diagnostic Protocol for citrus canker through the SPHDS process.

The metabolomics component has analysed metabolite expression in selected bacterial pathovars. Results showed separation between the different pathovars based on differential levels of expression of particular metabolites. These metabolites may be important determinants of pathogenicity.

Neither proteomics nor metabolomics had been implemented before in the study of phytopathogenic bacteria and whilst both proved to be technically demanding, each delivered new biomarkers that differentiate phytopathogenic bacteria to a subspecific level. This confirmed the viability of these approaches as platforms to discover novel diagnostic targets. The new methods developed will be implemented into the national incursion response capability, improving the specificity of diagnostic testing available and reducing the possibility of false positive diagnosis.

The project has fostered new collaborative partnerships both nationally (NSW, Victoria, WA) and internationally (to Thailand and the USA). The next phase of this work will provide a strong start-up project to the Plant Biosecurity Cooperative Research Centre (PBCRC). This project has directly enhanced the plant bacteriology capacity of NSW and Australia trough the recruitment and training of science professionals and an undergraduate student, and supported the specialist training of a Thai scientist through allied project CRC20093.

2. Aims and objectives

Many of the biggest threats to the biosecurity of Australia's plant industries are bacterial, for example 56 of the pathogens identified under the Emergency Plant Pest Response Deed are bacterial. Difficulties in the identification of phytopathogenic bacteria to the subspecific, or 'pathovar', level could seriously delay the management of major disease incursions and interfere with market access. By definition, bacterial pathovars are distinguished by their pathogenicity on different plant hosts so bioassays on intact plants remains the definitive means of identification. However, inoculation to a set of differential hosts to identify to pathovar level is a difficult process and not routinely incorporated into biosecurity protocols. In a few cases, rapid tests are available that have sufficient specificity and sensitivity to differentiate to the pathovar level but better diagnostic targets are required to rapidly and specifically differentiate many exotic pathovars from closely related endemic organisms.

An accurate, rapid method to identify biomarkers linked to pathovar specific traits is required. Markers linked to host specificity and/or pathogenicity are ideal and overcome many of the limitations associated with assays based on anonymous, plasmid-borne or incidental targets. This project aimed to evaluate two of the newer platforms that offer a more strategic approach to identifying functional determinants of plant/pathogen interactions. Proteomics and metabolomics were evaluated for their effectiveness as tools to streamline the discovery of biomarkers to differentiate specific sets of pathovars selected for use as the model to test these platforms. Both approaches have been shown to identify functional molecules associated with and/or determining infection, offering advantages over many other approaches.

The objectives of this project were:

- To evaluate new functional platforms for use as tools to discover biomarkers that effectively differentiate closely related pathovars within specific organisms used as models.
- To identify novel diagnostic targets for national and international validation, which can be fed into projects evaluating different delivery platforms.
- To increase national incursion response capability through improved specificity and turnaround time of diagnostic tests.
- To reduce the possibility of misdiagnosis and false positive diagnosis (Type 1 and Type 2 errors).
- To provide methods to improve tracking of bacterial pathovars in incursion.
- To foster new national and international scientific partnerships.
- To link to other projects enhancing plant bacteriology capacity through the training of specialist scientists and postgraduate students.

3. Key findings

3.1 Introduction

Many of the biggest threats to the biosecurity of Australia's plant industries are bacterial, with 56 of the pathogens identified under the Emergency Plant Pest Response Deed being bacteria. Difficulties in the identification of phytopathogenic bacteria to the subspecific, or 'pathovar', level could seriously delay the management of major disease incursions and interfere with market access. Financial losses stemming from wrongful diagnosis can be significant, as shown by the estimated loss of more than $15 million in trade following the inaccurate diagnosis of the fire blight pathogen, *Erwinia amylovora*, in 1997 (Rodoni *et al*., 2006).

By definition, bacterial pathovars are distinguished by their pathogenicity on different plant hosts so bioassays on intact plants remains the definitive means of identification. However, inoculation to a set of differential hosts to identify to pathovar level is not routinely incorporated into biosecurity protocols. Specific reasons for this include the requirement for QC3 level containment laboratories at minimum, which are rare in Australia (particularly at the time this project was initiated), that symptoms in such bioassays are often atypical or poorly expressed and that periods of weeks can be required for symptom expression. Rapid genomic, serological or lipid-based tests are available in some cases but few biomarkers with sufficient specificity and sensitivity to differentiate to the pathovar level are available.

Methods purported to differentiate pathovars are common in the literature but these are often found to be inadequate in practise. Many published molecular assays are based on anonymous or plasmid-borne targets (e.g. Bereswill *et al*., 1992) but such targets may mutate or be lost during replication if they do not directly contribute to an organism's fitness (pathogenicity). For some time this project team has strongly believed that most of these limitations can be overcome if determinants of pathogenicity or host specificity are used as diagnostic targets (Berg *et al*., 2005). Full validation, including against the background of Australian native microflora, is also essential prior to implementation.

Previous work by this team revealed a major weakness in the current diagnostic methods to identify the organism causing citrus canker. There is currently no simple method to separate this organism, *Xanthomonas citri subspecies citri* (*X. citri* subsp. *citri*), from a closely related organism causing bacterial blight of cotton (*X. citri* subsp*. malvacearum).* This is a significant issue, particularly as citrus and cotton are increasingly being grown alongside each other, for example cotton is now being grown in the Riverina in NSW, a major area of citrus production. The lack of specificity in the current 'best available' diagnostic means that the National Diagnostic Protocol must include a range of methods and that definitive identification of any Xanthomonad isolated from cotton in such areas could require bioassays to exclude the possibility of the organism causing citrus canker. During the incursion of the olive knot pathogen, *Pseudomonas syringae* pv *savastanoi* into Victoria in 2006, host testing in AQIS facilities returned ambiguous results.

Better diagnostic targets are required to rapidly and specifically differentiate many exotic pathovars from closely related endemic organisms. An accurate, rapid method to identify biomarkers linked to, or determining, pathovar specific traits is required. This project evaluated two of the newer platforms that offer a more strategic approach to identifying functional determinants of plant/pathogen interactions. Both proteomics and metabolomics identify functional molecules that could be associated with and/or direct pathogenicity, potentially fast tracking the identification of robust, stable, deterministic markers suitable as diagnostic targets. These platforms were evaluated in the same model organisms for their effectiveness as tools to streamline the discovery of biomarkers that differentiate specific sets of pathovars, biomarkers that could later be adapted to reliable, high throughput, rapid turn-around laboratory diagnostics.

Proteins modulate many of the interactions between pathogens and their hosts, particularly through the processes of adhesion and secretion (Gerlach and Hensel, 2007; Buttner *et al*., 2003). Proteomics is the analysis of proteins to identify and characterise proteins that are differentially expressed between similar organisms. Where these function as determinants of pathogenicity, or the means to regulate the initial interaction between the host and the pathogen, they can serve as the ideal diagnostic targets to discriminate the organisms. Proteins that are differentially-expressed between closely related pathovars can be identified using 2 dimensional (2D) gel electrophoresis and characterised by mass spectroscopy and/or peptide sequencing. Nucleotide sequences can be inferred from those peptides and genomic sequences in the international database screened to identify the genes encoding the protein. Conventional or real time PCR assays can be developed, validated and adapted based on these improved targets. Proteomics has been used in fast-tracking the development of human and veterinary diagnostics (Ansong *et al*., 2008; Cordwell *et al*., 2008; Ratcliffe *et al*., 2009), and by members of this project team to identify strains of nematode (Perera *et al*., 2009) but not previously in plant bacteriology.

Metabolomics is the study of metabolites and has potential to distinguish and identify different functional/physiological ecotypes of bacterial strains or species. For example, where genomic analysis failed to unambiguously distinguish between avirulent and virulent *Bacillus cereus* strains, multivariate pattern-recognition methods of metabolite profiles using 1H nuclear magnetic resonance (NMR) spectroscopy clearly separated the metabolite profiles of the two different ecotypes (Bundy *et al*., 2005). Spectral differences between different pathovars could lead to the identification of specific biomarkers and diagnostic NMR tests for high throughput analysis. This project is the first use of metabolomics in the study of plant pathogenic bacteria.

Bacteria of the genus *Xanthomonas* affect a wide variety of host plants. They produce a characteristic yellow pigment (xanthan gum) that may play a role in pathogenicity. Not all are pathogenic and individual strains may be infective to only one or a few host species, giving rise to the subspecific or "pathovar" terminology to indicate host range or the host from which they were originally isolated. This classification is problematic because typically extensive host range studies are not conducted, there can significant genetic heterogeneity and non-pathogenic species escape classification (Vauterin *et al*., 2000).

We had earlier conducted genomic analyses that developed an improved molecular diagnostic for the pathovars of *X. campestris* causing disease on brassicas (Berg *et al*., 2005), although this was a lengthy process. As discussed above, we had also identified acute problems at the pathovar level in definitively identifying the Xanthomonad causing citrus canker using rapid molecular methods. We therefore identified the need for improved diagnostics within the genus *Xanthomonas* and selected them as appropriate model organisms for this study.

The taxonomy of the genus *Xanthomonas* has been fluid across recent years and changes to taxonomy are often adopted slowly. A significant, multidisciplinary reclassification of all the most significant members of the genus was conducted in 1995 (Vauterin *et al*., 1995; also summarised in Vauterin *et al*., 2000). This redefined the species *X. campestris* as containing only the six pathovars pathogenic to crucifers (pvs *campestris, aberrans*, *armoraciae*, *raphani*, *incanae* and *barbarae*) and moved all the other Xanthomonads of unknown species affiliation (including the organism causing citrus canker) from *X. campestris* to *X. axonopodis,* acknowledging that the species *axonopodis* was a polyphyletic group (Vauterin *et al*., 2000). Within the species *X. axonopodis*, the citrus pathogens were named as pathovar *citri*, responsible for the classic, wide-host range Asiatic citrus canker (also known as canker pathotype A), and pathovar *aurantifolii*, causing classic symptoms but with a reduced host range and in practise rarely recorded (referred to as pathotypes B, C and D). Other pathovars within *X. axonopodis* included pathovar *citrumelo*, cause of the largely benign 'citrus bacterial spot' (also known as the E pathotypes), and pathovar *malvacearum* (cause of bacterial blight of cotton). At the time this project was being developed, *citri* and *malvacearum* were pathovars within the species *X. axonopodis*, but like many other taxa within the genus, they are still sometimes referred to under the old species name of *X. campestris*.

A second major revision in 2005 focussed on the organisms pathogenic to citrus and those organisms closely related to them (Schaad *et al*., 2005). This moved pathovars *citri* and *malvacearum* to be subspecies within their own species, *X. smithii*, subsequently emended to *X. citri* (Schaad *et al*., 2006), renaming them as *X. citri* subspecies *citri* and subsp. *malvacearum*. This reclassification also renamed the pathovar *aurantifolii* (B/C/D pathotypes) as *X. fuscans* subsp. *aurantifolii* and pathovar *citrumelo* (E pathotypes) as *X. alfalfae* subsp. *citrumelonis*.

Further subtypes or strains have now been described that are genetically similar to *X. citri* subsp. *citri* but have a limited host range, similar to *X. fuscans* subsp. *aurantifolii*, (Key/Mexican lime and alemow). These are referred to as strains of *X. citri* subsp. *citri*. The Aw types have been detected only in one area in Florida (Sun *et al*., 2004), the A* types are more heterogenous and originate from several areas of Asia (Verniere *et al*., 1998). These further complicate the issue because the detection of citrus canker with a limited pathogenicity would now require the identification of the causal organism to the sub-pathovar level, in order to determine the risk they might pose. Given the limitations of bioassays, this would be fraught with difficulty.

In this study, we have selected to use as model organisms a range of bacteria from the genus *Xanthomonas*, specifically those infecting citrus (*X. citri* subsp*. citri* pathotypes A, A* and Aw, B and C, and *X. alfalfae* subsp. *citrumelonis*), cotton (*X. citri* subsp*. malvacearum)*, brassicas (*X. campetris*, all pathovars) and soybeans (*X. axonopodis* pv *glycines*).

We did not seek to change the name of this project to reflect the most recent taxonomy but we have tried to use the updated nomenclature throughout this report.

3.2 Materials and Methods

3.2.1 Bacterial isolates and their handling

The bacterial isolates used during this project are shown in Tables 1, 2 and 3.

Pure bacterial isolates were obtained from various sources including reference collections, collaborators, and from the in-house collections of field isolates held by these authors. Bacteria were routinely grown on Nutrient Agar (NA) media at 25°C and subcultured every 3-4 days. Cultures were preserved onto Protect beads (Oxoid) by obtaining a loopful of culture and resuspending it into the Protect Bead solution. Cultures preserved in this way were kept at an indefinite length of time at -80°C.

3.2.2 Extraction of proteins from bacteria

For protein extraction, single colonies were streaked to NA plates to achieve a lawn of growth after incubation at 25°C for two nights. Bacteria were harvested using a sterile loop, washed in 100 µl ice-cold PBS and collected by centrifugation (16000 ×*g* for five minutes at 4°C) to produce a wet cell pellet equivalent to 50 µl. PBS was aspirated and the cell pellet stored overnight at -80 °C or processed immediately.

Hydrophobic membrane-bound protein fractions were extracted from bacteria using the ReadyPrep Protein Extraction Kit (Membrane 1) kit (BioRad), essentially according to manufacturer's instructions. Resulting extracts were frozen on dry ice or liquid Nitrogen and stored at -80°C if required before being freeze dried overnight. Samples were stored and shipped at ambient temperature.

3.2.3 Molecular analysis of DNA from bacterial cultures

DNA was extracted from pure cultures of bacteria on media using the Qiagen DNeasy™ kit, according to the manufacturer's protocol. Pure DNA from a range of other non-Xanthomonad bacteria, and from a range of other microorganisms pathogenic to citrus, was also available through our in-house collection (Table 3). These extracts were used in the final stages of validation of new diagnostic assays.

PCR was performed in an Eppendorf Master Cycler using 1 unit of Taq polymerase (Life Technologies) and the buffer supplied and using 200 nM dNTPs. Cycles used an initial denaturation of 3-5 mins at 95°C and a final extension of 72°C for 5-10 mins (except REP where the final extension was a 68°C). Other parameters of PCR that differed depending on the target are shown in Table 4 and primer sequences are shown in Table 5.

Table 1: Isolates of Xanthomonas that are pathogenic to citrus, that were used in the different components of this study (a tick (✓) in a column indicates that the isolate was used in the corresponding analysis)

Isolate Number	Synonym	Origin	Host/Name substrate	Proteomic Analysis	Metabolomic Analysis	Genomic Analysis	Leaf Assay
***X. citri* subsp. *citri* pathotype A**							
DAR77358	MN04/5578 60.1	Emerald, QLD (2004)	Unknown		✓	✓	
MN04/8516 699A		Emerald, QLD (2004)	unknown	✓	✓	✓	
DAR77356	MN04/5370 3.2A	Emerald, QLD (2004)	*Citrus sinensis* (L.) Osbeck			✓	
MN04/5703 389.7A		Emerald, QLD (2004)	mixed			✓	
DAR77357	MN04/5370 8.2	Emerald, QLD (2004)	*Citrus sinensis* (L.) Osbeck			✓	
MN04/5703 393.1		Emerald, QLD (2004)	*Citrus reticulata* Blanco	✓		✓	✓
DAR77359	MN04/5703 394.2	Emerald, QLD (2004)	*Citrus reticulata* Blanco			✓	
MN04/5703 388.4		Emerald, QLD (2004)	*Citrus reticulata* Blanco			✓	
DAR65863		Humpty Doo, NT (1991)	*Citrus grandis* Hassk. Cv. Termat (Rutaceae)		✓	✓	
DAR65869		Torres Strait (1991)	*Citrus sinensis* (L.) Osbeck (Rutaceae)		✓	✓	
DAR69978a		Thailand (1994)	*Citrus hystrix* DC. (Rutaceae)			✓	
DAR69978b		Thailand (1994)	*Citrus hystrix* DC. (Rutaceae)			✓	
DAR73871		Asia (1999)	*Citrus* sp. (Rutaceae)			✓	
DAR73872	Q99/470	Thailand (1999)	*Citrus hystrix* DC. (Rutaceae)	✓	✓	✓	
DAR73889		Thailand	*Citrus* sp. (Rutaceae)			✓	
DAR75280		East Timor (2005)	*Citrus* sp. (Rutaceae)			✓	
XS97-3-1		Florida, USA (1997)	*Citrus x paradisi* (Rutaceae)			✓	
XS97-5-1		Florida, USA (1997)	*Citrus x paradisi* (Rutaceae)			✓	
ATCC49118		Florida, USA	*Citrus aurantifolia*		✓	✓	
X03-2914		Florida, USA (2003)	*Citrus sinensis*		✓	✓	
X02-10		Florida, USA (2002)	*Citrus x paradisi* (Rutaceae)			✓	
X04-1469		Florida, USA (2004)	*Citrus* sp.		✓	✓	
207		Oman (1989)	*Citrus* sp. (lime)			✓	
209		Oman (1989)	*Citrus* sp. (lime)			✓	

X.* subsp. *citri* pathotype A							
Xcc290		Southwest Asia	*Citrus aurantifolia*		✓	✓	
169		India (1988)	*Citrus* sp. (lime)		✓	✓	
206		Oman (1989)	*Citrus* sp. (lime)		✓	✓	
269		Saudi Arabia (1988)	*Citrus* sp. (lime)		✓	✓	
270		Saudi Arabia (1988)	*Citrus* sp. (lime)			✓	
271		Saudi Arabia (1988)	*Citrus* sp. (lime)			✓	
276		Saudi Arabia (1988)	*Citrus* sp. (lime)		✓	✓	
289		Saudi Arabia (1988)	*Citrus* sp. (lime)			✓	
323		Saudi Arabia (1990)	*Citrus* sp. (lime)			✓	
328		Saudi Arabia (1991)	*Citrus* sp. (lime)			✓	
329		Saudi Arabia (1991)	*Citrus* sp. (lime)			✓	
Xcc406		Southwest Asia	*Citrus aurantifolia*	✓	✓	✓	
DAR72029		Singapore (1997)	*Citrus* sp. (Rutaceae)		✓	✓	
DAR73909		Thailand (2000)	*Citrus* sp. (Rutaceae)			✓	
***X. citri* subsp. *citri* - Not identified to the level of pathotype**							
284		Saudi Arabia (1988)	*Citrus* sp. (lime)			✓	
324		Saudi Arabia (1990)	*Citrus* sp. (lime)			✓	
327		Saudi Arabia (1991)	*Citrus* sp. (lime)			✓	
***X. citri* subsp. *citri* pathotype Aw**							
X2002-12884	DAR76571	Florida, USA (2000)	*Citrus aurantifolia* (Christm.) Swingle		✓	✓	
X2002-00005	DAR76572	Florida, USA (2001)	*Citrus aurantifolia* (Christm.) Swingle		✓	✓	
X2001-00032	DAR76573	Florida, USA (2001)	*Citrus aurantifolia* (Christm.) Swingle		✓	✓	
X2003-01008	DAR76575	Florida, USA (2003)	*Citrus aurantifolia* (Christm.) Swingle		✓	✓	
X2003-01012	DAR76576	Florida, USA (2003)	*Citrus aurantium* L.			✓	
X2003-01029	DAR76577	Florida, USA (2003)	*Citrus aurantifolia* (Christm.) Swingle		✓	✓	
X2003-01036	DAR76578	Lake Worth, USA (2003)	*Citrus aurantifolia* (Christm.) Swingle (Rutaceae)		✓	✓	
X03-1003		Florida, USA (2003)	*Citrus aurantifolia*	✓	✓	✓	
X02-1035		Florida, USA (2002)	*Citrus aurantifolia*		✓	✓	
***X. fuscans* subsp. *aurantifolii* pathotype B**							
1622		Argentina (2003)	*Citrus limon*			✓	
***X. fuscans* subsp. *aurantifolii* pathotype C**							
ATCC51302		Brazil	*Citrus aurantifolia*			✓	
X. alfalfae* pv. *citrumelonis							
XSol-4		Florida, USA (2001)	*Citrus aurantifolia*	✓	✓	✓	✓
P03-83		Florida, USA (2003)	*Citrus x paradisi*	✓	✓	✓	✓

NB. A tick (✓) in a column indicates that the isolate was used in the corresponding analysis as part of this study.

Table 2: Isolates of Xanthomonas that are pathogenic to non-Citrus hosts, that were used in the different components of this study (a tick (✓) in a column indicates that the isolate was used in the corresponding analysis)

Isolate Number	Synonym	Origin	Host/Name substrate	Proteomic Analysis	Metabolomic Analysis	Genomic Analysis	Leaf Assay	Plant Assay Pathogenic (+)/Non-pathogenic (-)
X. citri* subsp. *malvacearum								
DAR26840		Narrabri, NSW (1974)	*Gossypium hirsutum* L. cv. Deltapine 16		✓	✓		+
DAR26843		Lockyer Valley, QLD (1967)	*Gossypium hirsutum* L.		✓	✓		+
DAR26866		Kununurra, WA (1975)	*Gossypium hirsutum* L.	✓	✓	✓		+
DAR26889		Macquarie Valley, NSW (1977)	*Gossypium hirsutum* L. cv. Deltapine 16			✓		-
DAR26927		Texas, USA (1978)	*Gossypium hirsutum* L.	✓	✓	✓		+
DAR26928		Texas, USA (1978)	*Gossypium hirsutum* L.			✓		+
DAR26934		Texas, USA (1978)	*Gossypium hirsutum* L.	✓		✓		+
DAR26936		Texas, USA (1978)	*Gossypium hirsutum* L.			✓		+
DAR26941		Kununurra, WA (1975)	*Gossypium hirsutum* L.			✓		+
DAR29063		Rydalmere, NSW (1977)	*Gossypium hirsutum* L. cv Akala	✓	✓	✓		+
DAR30521		Rydalmere, NSW (1978)	*Gossypium hirsutum* L.	✓		✓	✓	+
DAR30522		Rydalmere, NSW (1978)	*Gossypium hirsutum* L.	✓		✓		+
DAR33386		Emerald, QLD (1979)	*Gossypium hirsutum* L.	✓	✓	✓		+
DAR33396		Emerald, QLD (1979)	*Gossypium hirsutum* L.			✓		+
DAR35718		Warren, NSW (1982)	*Gossypium hirsutum* L. cv. Deltapine 61			✓		+
DAR63291		Narrabri, NSW (1989)	*Gossypium barbadense* L.			✓		+
X. axonopodis* pv. *glycines								
DAR26926		Kingaroy, QLD (1977)	Glycine Max (L.) Merr.	✓				-
DAR33424		Woodburn, NSW (1980)	Glycines max (L.) Merr. Cv. Bossier	✓				-
DAR34859		Woodburn, NSW (1980)	Glycines max (L.) Merr. Cv. Bossier	✓				+
DAR34864		Casino, NSW (1978)	Glycine max (L.) Merr.	✓			✓	-
DAR34868		Rydalmere, NSW (1980)	Glycine max (L.) Merr. Cv. Bethel	✓				+
DAR49310		Trangie, NSW (1984)	Glycine max (L.) Merr. Cv. WIC 36	✓				-
DAR58707		Wardell, NSW (1987)	Glycine max (L.) Merr. Cv. WL33					-

X. campestris* pv. *campestris								
DAR76566		Leppington, NSW (2003)	*Brassica oleracea L.* (broccoli cv. Grevillea)			✓		
DAR76567		Leppington, NSW (2003)	*Brassica* oleracea L. (cauliflower cv. Omeo)			✓		
DAR75950	ATCC 33913, NCPPB 528, LMG568, ICMP 13	Wye, United Kingdom (1957)	*Brassica oleracea* L. var. *gemmifera* DC. (brussel sprouts)	✓	✓	✓		+
X. campestris* pv. *aberrans								
DAR75944		Redland Bay, QLD (1975)	*Brassica oleracea* L. var. *capitata* (L.) Alef. (cabbage)	✓	✓	✓		+
X. campestris* pv. *armoraciae								
DAR75942		Tanzania, Africa (1954)	*Iberis* sp. (candytuft)	✓	✓	✓		+
X. campestris* pv. *barbareae								
DAR75945		USA (1939)	*Barbarea vulgaris* R. Br.	✓	✓	✓		+
X. campestris* pv. *incanae								
DAR75959		USA (1950)	*Matthiola incana* (L.) R. Br.	✓	✓	✓		+
X. vesicatoria								
DAR26932		Mullamuddy, NSW (1977)	*Datura ferox* L.			✓		
DAR26933		Redland Bay, QLD (1977)	*Capsicum annuum* L.			✓		
DAR33341		Peats Ridge, NSW (1979)	*Capsicum annuum* L.			✓		
DAR73877		Howard Springs, NT (2000)	*Capsicum annuum* L.			✓		
DAR80973		QLD, Australia (2010)	*Lycopersicon esculentum* Mill.			✓		
X. campestris* pv. *raphani								
DAR75983		USA	*Raphanus sativus* L.			✓		
DAR75960		USA (1940)	*Raphanus sativus* L.	✓	✓	✓		+
X. axonopodis* pv. *phaseoli								
DAR75943		USA (1954)	*Phaseolis vulgaris* L.			✓		
DAR58726		Bileola, QLD (1987)	*Phaseolus vulgaris* L.			✓		
DAR65944		Harrington, NSW (1992)	*Phaseolus vulgaris* L.			✓		
X. cucurbitae								
DAR75961		Auckland, NZ (1968)	*Cucurbita maxima Duchesne*			✓		

X. axonopodis* pv. *sesame								
DAR75547		Katherine, NT (2002)	*Sesamum indicum*			✓		
DAR75558		Sudan (1958)	*Sesamum indicum*			✓		
DAR75559		Sudan (1971)	*Sesamum indicum*			✓		
X. axonopodis* pv. *fici								
DAR61743		Humpty Doo, NT (1989)	*Ficus microcarpa* L. f.			✓		
X. albilineans								
DAR34874		Tully, QLD (1980)	*Saccharum officinarum L.*			✓		
X. hortorum* pv. *carotae								
DAR49849		Gosford, NSW (1985)	*Daucus carota L.*			✓		
X. axonopodis* pv. *vitians								
PHDS 02/559		NSW (2002)	*Lactuca sativa* L. var *longifolia*			✓		
X. oryzae* pv. *oryzae								
DAR61716		Mareeba, QLD (1987)	*Oryza sativa* L.			✓		
***Xanthomonas* sp.**								
DAR72016		Kununurra, WA (1997)	*Citrus paradisi* Macfad	✓		✓	✓	
PHDS 08/318		Broome, NT (2008)	*Citrus aurantifolia* (christm.) Swingle	✓		✓		

Table 3: Other microorganisms from which purified DNA (only) was used in this study.

Isolate Number	Synonym	Isolates	Origin	Host/Name substrate
DAR76591		*Pseudomonas syringae* pv. *syringae*	United Kingdom (1950)	*Syringa vulgaris* L.
DAR54698		*Pseudomonas syringae* pv. *syringae*	Invergordon, Vic (1985)	*Prunus armeniaca*
PHDS 01/523		*Pseudomonas syringae* pv. *lachrymans*	Rossmore, NSW (2001)	*Cucumis sativus*
DAR73894		*Pseudomonas viridiflava*	Wagga Wagga, NSW (1999)	*Sagittaria montevidensis* Cham & Schltdl
DAR58720		*Pseudomonas cichorii*	Duffys Forest, NSW (1987)	*Lactuca sativa*
DAR30556		*Pseudomonas syringae* pv. *maculicola*	Wagga Wagga, NSW (1978)	*Brassica rapa* L.
PHDS 01/275		*Burkholderia andropogonis*	Arcadia (2001)	Sweet Williams
DAR75243		*Burkholderia andropogonis*	USA (1957)	Carnation
DAR72022		*Pantoea agglomerans* prev. *Erwinia herbicola*	Melbourne, Vic (1997)	*Sorbus* sp.
DAR65858		*Pectobacterium carotovorum* prev. *Erwinia carotovora* pv. *carotovora*	Guyra, NSW (1991)	*Solanum tuberosum* L.
DAR73906		*Erwinia chrysanthemi*	North Arm, QLD (1998)	*Zingiber officinale* Rosc.
DAR72017	Ea322	*Erwinia amylovora*	France (1972)	*Pyrus communis* L.
PHDS 09/755B 2		*Erwinia billingiae*	Bilpin, NSW (2009)	Ornamental Cherry (*Prunus* sp.)
PHDS 09/755B 5		*Erwinia billingiae*	Bilpin, NSW (2009)	Ornamental Cherry (*Prunus* sp.)
4080	LMG5602	*Clavibacter michagenensis*	New Zealand (1967)	*Cyphomandra betacea*
4087	LMG5643	*Clavibacter michagenensis*	Canada (1982)	*Lycopersicon esculentum*
4071	LMG3685	*Clavibacter michagenensis*	USA (1939)	*Lycopersicon esculentum*
4088	LMG5644	*Clavibacter michagenensis*	Canada (1982)	*Lycopersicon esculentum*
DAR76574	2048	*Clavibacter michagenensis*	Murrumbateman, NSW (2004)	*Lycopersicon esculentum* Miller
M21		*Mycospharella* sp.	Australia (2001)	*Citrus reticulata* Blanco
M24		*Mycospharella* sp.	Australia (2002)	*Citrus gracilis*
DAR56256		*Guignardia citricarpa*	Cornwallis, NSW (1986)	*Citrus sinensis* (L.) Osbeck cv. Washington Navel
DAR33474		*Guignardia citricarpa*	Kulnura, NSW (1978)	*Citrus sinensis* (L.) Osbeck cv Valencia
1445		*Liberobacter asiaticus*	India (2005)	*Citrus jambhiri*
1318		*Liberobacter asiaticus*	Papua New Guinea	

Table 4: Summary of PCR conditions as used in this project

Target in PCR	Mg $^{2+}$ mM	Primers uM	Denaturation	Annealing	Extension	Number of Cycles	Fragment length in bp
J-pth	3	0.1	94°C/30s	58°C/30s	72°C/30s	30	197bp
BOX	6	1.2	94°C/30s	50°C/30s	72°C/1min	40	-
REP	4	1.0	94°C/30s	54°C/30s	68°C/1min	40	-
IS50	4	2	94°C/1min	38°C/1min	72°C/3.5mins	40	-
Spots 123 selective	2	1	95°C/20s	58°C/20s	72°C/20s	25	486bp
Spots 123 generic	2	1	95°C/40s	58°C/40s	72°C/40s	40	1029bp
Spot 8 selective	2	1	95°C/20s	58°C/20S	72°C/20S	25	177bp
Spot 8 generic	2	1	95°C/40s	69°C/40s	72°C/40s	6	649bp
			95°C/40s	64°C/40s	72°C/40s	34	
Spots 123 selective	2	1	95°C/20s	58°C/20s	72°C/20s	25	486bp
Spots 123 generic	2	1	95°C/40s	58°C/40s	72°C/40s	40	1029bp

Table 5: Summary of the primers used and the expected size in bp of fragments generated

Primer	Target	Primer sequence -5' to 3'	Expected size (bp)	Reference
J-pth1 J-pth2	*pthA* gene	CTTCAACTCAAACGCCGGAC CATCGCGCTGTTCGGGAG	197	Cubero and Graham, 2002
BOX1AR		CTACGGCAAGGCGACGCTGACG	-	Koeuth *et al.*, 1995
REP1R REP21		IIIICGICGICATCIGGC ICGICTTATCIGGCCTAC	-	Versalovic *et al.*, 1991
IS50		CAGGACGCTACTTGTGT		Ullrich *et al.*, (1993)
s8s2F s8s2R	*ompP1* gene for outer membrane protein W (protein spot 7,8) – selective	CCTTGGCCGCGCATTCGCA CACCGGCGCGCCGTTGGCGTAGGCC	177	This study
s8ns2F s8ns2R	*ompP1* gene for outer membrane protein W (protein spot 7,8) – generic	GTTATTGTTCGGTTAATCGTC ACGCGACGTTGAAGTCGATCG	649	
s123sF s123sR	*mopB* gene for outer membrane protein (protein spots 1,2,3) – selective	CGACCAGAGCGTTGCTGCTCCGCAG GATCGGGCCGACCAGGCGCGAT	486	
s123nsF s123nsR	*mopB* gene for outer membrane protein (protein spots 1,2,3) – generic	GCCCAGGCTGCTTCCGCGCAGGAG CGGTACGACGGTTCTTCGCACG	1029	

PCR products were electrophoresed on 1% agarose gels (Promega) in 1X TBE buffer; gels were stained with Gel Red (Biotium) and images captured by UV transillumination using a Gel Doc 2000 (Bio-Rad, Gladesville, NSW, Australia). All fragments were co-electrophoresed with a 1Kb plus marker (Invitrogen).

PCR products prepared for nucleotide sequencing analysis were purified using JETquick columns (Astral Scientific) and sent to the Australian Genome Research Facility (Sydney) for sequencing. Nucleotide sequences were analysed and aligned with the Bionumerics (version 6.6.4) software using the multiple alignment settings of the Needleman-Wunsch algorithm, pairwise alignment (open gap penalty=15; Extend gap penalty=6.7), multiple alignment (Open gap penalty=15; Extend gap penalty=6.7; Delay divergent sequences=30; DNA transition weight =0.50) Similarity calculation (DNA weight matrix=IUB).

Clustering settings: Needleman-Wunsch algorithm, Pairwise alignment (open gap penalty=15; extend gap penalty=6.7), Clustering: Complete linkage, Similarity calculation: DNA weight matrix.

3.2.4 Whole genome sequencing using 454 GS Junior

Sample preparation – Libraries were prepared using the Rapid Library Preparation kit (05608228001) as per the manufacturers instructions (Rapid Library Preparation Method Manual, June 2010, Roche). Emulsion breaking and enrichment of DNA was conducted using the GS Junior Titianium emPCR kit (05996481001) as per manufactures instructions (emPCR Amplification Method Manual-Lib-L, June 2010, Roche). Sequencing was conducted using the GS Junior Titanium Sequencing kit (0599655401) as per the manufactures instructions (Sequencing Method Manual GS Junior Titanium Series (June 2010, Roche).

Genomes were assembled and annotated using Reference mapper (Roche), a reference guided and variation detection software system provided for the resequencing of 454 GS Junior data. The reference strain *X. citri* subsp. *citri* strain 306 (Original name *X. axonopodis citri* Genbank Accession number AE008923) was used as the scaffold for the assembly of the 454 genome data. The reference strain data was uploaded in fasta format, while the 454 data was in .sff form. The analysis is a point and click with the new data being assembled against the reference scalfold, the output files being .ace and .fna files which were either visualized in Tablet (bioinf.scri.ac.uk/tablet) using the .ace files or in artemis.jarI http://www.sanger.ac.uk/) for the .fna files. The .fna files were then uploaded into xBASE (xbase.ac.uk) for annotation. The annotation files (.gbk) were then analysed in Artemis. The sequence comparison tool ACT (www.sanger.ac.uk/) was used to compare the bacterial genomes following assembly in Artemis.

3.2.5 Two-dimensional gel electrophresis of proteins

Following extractions, the proteins in the insoluble (membrane bound) fraction were reconstituted in 200µl of ddH_20 at room temperature for 5-10 minutes. Protein content was quantified using the RC DC protein assay kit (Bio-Rad, USA).

For each sample, 80µg of protein was made up to 100µl using M1 extraction buffer. Salts were removed from the sample using the ReadyPrep 2D cleanup kit (Bio-Rad) according to the manufacturer's instructions, and the sample solubilised in 220µl of protein solubilisation buffer (PSB) containing ampholytes (Bio-Lyte 3/10 ampholytes, Bio-Rad). Immobilised pH gradient (IPG) strips (pH 3-10 and pH4-7, Bio-Rad) were rehydrated with 200µl of this sample in Protean IEF disposable trays at room temperature for 12-18 hours. First dimension separation of the proteins was achieved using the following protocol at 20ºC: 240V for 20 minutes, rapid ramp to 8,000V for 2h30min, rapid ramp to 8000V for 45,000Vh and then rapid ramp to holding step at 400V.

Electrophoresed IPG strips were equilibrated using 0.1% DL-dithiothreitol (Sigma, USA) in urea, SDS and glycerol for 20 minutes, then 0.25% iodoacetamide (Bio-Rad, USA) for 20 minutes at room temperature. Strips were rinsed in XT MOPS running buffer (Bio-Rad) and run on Criterion XT 12% acrylamide gels (Bio-Rad) at 200V, 400mA for 55 minutes at room temperature. Completed gels were silver stained and imaged using ProXPRESS 2D Proteomic Imaging System (PerkinElmer, USA).

Duplicate gels were initially run on two biological preparations from two or three isolates to determine reproducibility, then individual gels were run on each isolate sample. Silver staining was done using the protocol by Shevchenko *et al* (1996). Gels were stained using SyproRuby fluorescent stain overnight and imaged with PerkinElmer's ProXPRESS 2D Proteomic Imaging System, excitation wavelength

480/30nm and emission 620/30nm. Gels stained with Coomassie were stained for 90 minutes, then destained in distilled water overnight.

3.2.6 Analysis of differentially expressed proteins

Analysis of the scanned gels was performed using Progenesis SameSpots (Nonlinear, UK) and evaluated using the attached statistical software. Spots were also manually checked for differences between the groups being evaluated. All the gels from each isolate of *X. citri* subsp. *malvacearum* and *X. a.* pv. *glycines* were compared to all the gels from *X. citri* subsp. *citri*. Different pathotypes of *X. citri* subsp. *citri* were compared to each other.

Spots present in gels from *X. citri* subsp. *citri* but not *X. citri* subsp. *malvacearum* were manually excised immediately after imaging and in-gel digestion using trypsin (Roche) performed according to Shevchenko et al. 1996 and Shevchenko *et al*. 2006. Supernatants were dried under vacuum, reconstituted in 10µl 0.1% TFA and incubated at 4°C overnight and 37°C for a further hour. Salts were removed using C18 tips (Varian, Omix, USA) according to their protocol.

Samples were mixed 1 in 10 in alpha hydroxyl cinnamic acid (Sigma Aldrich, USA) matrix and 1µl aliquots placed on mass spectrometry plates, allowed to dry and 1µl of the mixture added to each spot. Spectra were obtained using a Voyager mass spectrometer with settings 23000V acceleration, 2250 laser output and an output of mass-to-charge ratio of 500-4000 m/z. Mascot (UK) was used for peptide identification.

Peptide sequencing was done by Proteomics International (Australia). Protein spots stained with Coomassie or SyproRuby were trypsin digested and peptides extracted according to standard techniques. Peptides were analysed by electrospray ionisation mass spectrometry using the Ultimate 3000 nano HPLC system (Dionex) coupled to a 4000 Q TRAP mass spectrometer (Applied Biosystems). Tryptic peptides were loaded onto a C18 PepMap100, 3 µm (LC Packings) and separated with a linear gradient of water/acetonitrile/0.1% formic acid (v/v). Spectra were analysed to identify proteins of interest using Mascot sequence matching software (Matrix Science) with Ludwig NR Database and taxonomy set to Eubacteria.

Proteins identified from the sequenced peptides were compared to genomic DNA from *X. citri* subsp. *citri* using BLAST (NCBI).

3.2.7 Extraction of metabolites from bacteria

Seven replicates, each inoculated from a single colony, were prepared for each bacterial isolate to be studied. Cultures were grown from a single colony inoculated into 10ml KB broth at 25°C and shaken horizontally at 200 rpm to achieve an OD of 1.0 at 600nm (obtained by comparing to McFarland Standards). The culture was chilled in 80% methanol/water overnight at -20°C, the cell pellet collected by centrifugation at 4000rpm for 10 mins . The pellet was resuspended in 5mls of pre-chilled 80% methanol/water, vortexed for 20 seconds, and sonicated for 10 minutes. The pellet was collected again by centrifugation at 5500rpm at 4°C for 10 minutes, and supernatant (containing the metabolites) transferred into labelled tubes. Samples were stored at -80C and despatched on dry ice via overnight dangerous goods courier to DPI Victoria.

3.2.8 Analysis of extracted metabolites

Proton NMR spectra were obtained on a Bruker AvanceIII spectrometer operating at 800.13 MHz and equipped with a CryoProbe (Bruker, Rheinstetten, Germany). Extracts were reconstituted in D_2O (650 µL, 1% TSP, NaN_3) and 600 µL transferred to a 5 mm diameter NMR tube for analysis. Data were acquired using 1D pulse (zgesgp) sequence with 240 transients over a sweep width of 16025 Hz and a total acquisition time of 2.05 s. A line broadening of 0.3 Hz was applied to all free induction decays (FIDs) prior to Fourier transformation. Spectra were manually phased and baseline corrected in Topspin 3.0 (Bruker). Aqueous samples were referenced to TSP at δ0.0.

Spectra were imported into MatLab R2011b (Version 7.13.0.564) using ProMetab_v1_1 script. Data were imported over 0.2 to 10.5 ppm range. The data was aligned using a correlated optimised warping protocol. The data points corresponding to the residual water peak, methanol and TSP were removed. This modified data was used for all subsequent analysis. Data were statistically analysed in MatLab using PLS-Toolbox (Eigenvector Research, Inc., Version: 6.2.1). Principal component analysis (PCA) was carried out using normalized (to 1) mean centred data. PLSDA models were optimised by OSC (orthogonal signal correction) with autoscaling. Cross validation and permutation analysis was also carried out.

3.3 Results

Whilst conducting research to develop Australia's National Diagnostic Protocol for X. *citri* subsp. *citri,* the organism causing citrus canker, our team had identified the need for a more robust diagnostic for this organism. Specifically, the primer pair selected for use in PCR, the Jpth primers described by Cubero and Graham in 2002, are not completely specific to this organism. This is demonstrated in Figure 1, which shows that the Jpth primers amplify a fragment of the same size and sequence from isolates of both X. *citri* subsp. *citri* and *X. citri* subsp. *malvacearum,* the causal agent of bacterial blight of cotton. Other available primer sets for X. *citri* subsp. *citri* have the same limitations (results not shown).

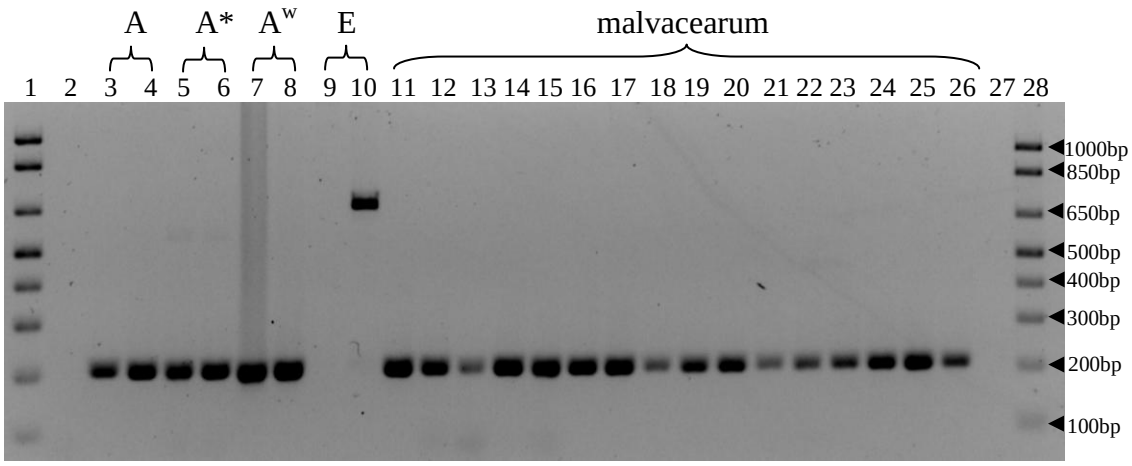


Figure 1: Jpth PCR 197bp. Primers used were J-pth1 and J-pth2. Lanes 1 & 28: 1kb + Marker; Lanes 2 & 27: negative control; Lane 3: DAR77358; Lane 4: MN04/8516 699A; Lane 5: 169; Lane 6: 206; Lane 7: X02-1035; Lane 8: X2002-12884; Lane 9: XSol-4; Lane 10: P03-83; Lane 11: DAR26840; Lane 12: DAR26843; Lane 13: DAR26866; Lane 14: DAR26889; Lane 15: DAR26927; Lane 16: DAR26928; Lane 17: DAR26934; Lane 18: DAR26936; Lane 19: DAR26941; Lane 20: DAR29063; Lane 21: DAR30521; Lane 22: DAR30522; Lane 23: DAR33386; Lane 24: DAR33396; Lane 25: DAR35718; Lane 26: DAR63291. Refer to Table 1 & 2 for further detail regarding isolates.

This is a significant issue given that citrus and cotton are more frequently being grown in the same regions, for example the Riverina of NSW. The close relationship between X. *c.* subsp. *citri* and other important pathogens within the Xanthomonas species had also been supported by earlier work by our team using fatty acid methyl ester (FAME) profiling to analyse the lipid composition of isolates of X. *c.* subsp. *citri* and other related organisms (Cother, Hailstones and Noble, unpublished). Whilst FAME often identified isolates of X. *c.* subsp. *citri* correctly, in many cases their closest match was reported to be to library specimens taxonomically identified as X. *c.* subsp. *malvacearum* or *X. axonopodis pv glycines.* On this basis, we selected isolates of these species as key organisms for use in this study. Given that specific diagnostic methods were also not available to differentiate between the pathotypes within X. *c.* subsp. *citri* (pathotypes A, A* and Aw), these were also included.

A collection of pure cultures of *Xanthomonas* isolates was assembled for use in this study, covering isolates infecting citrus (Table 1), Xanthomonads causing disease on other hosts (Table 2), and other hosts (Table 3). DNA was extracted from each and analysed using a number of genomic fingerprinting techniques such as REP-, BOX- and IS50-PCRs. A representative example, shown in Figure 2, indicates that multiple profiles are generated by the IS50 primer from within the 17 isolates of X. *c.* subsp. *malvacearum* tested. Isolates were grouped based on these patterns, and this informed the selection of isolates from within one species that best represent the molecular diversity within the species, for use in subsequent analyses (bioassays, proteomics, metabolomics).

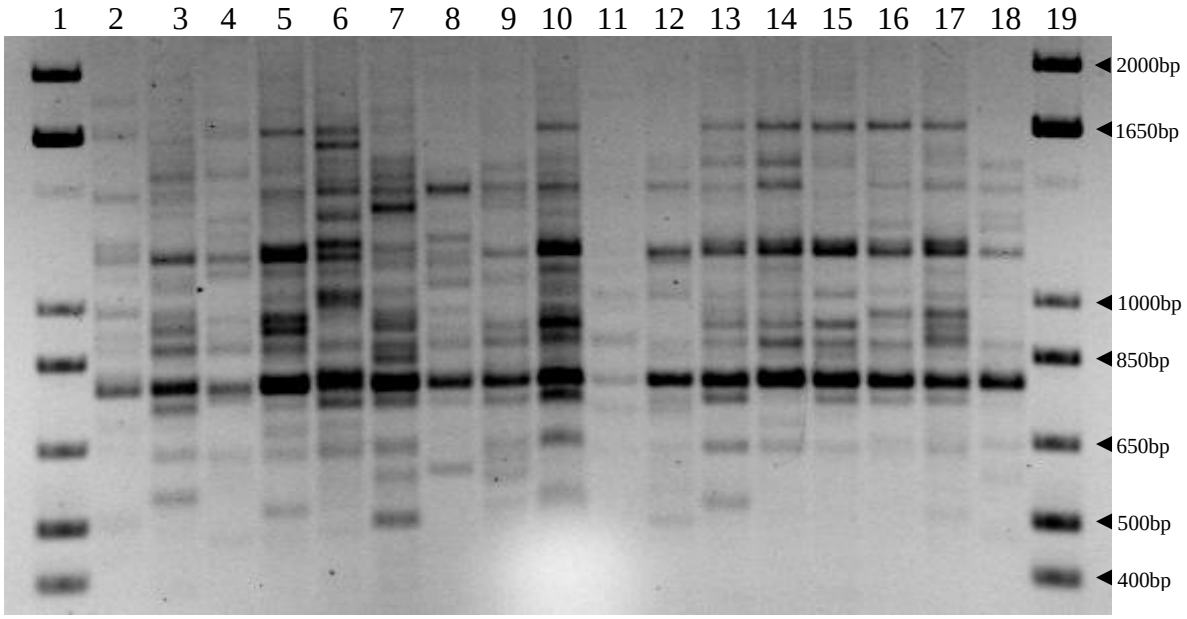


Figure 2: Comparing the IS50 profile of all the *X. c*. subsp. *malvacearum* isolates at EMAI grown on NA. Lane 1 = 1kb+ marker; Lane 2 = DAR26840; Lane 3 = DAR26843; Lane 4 = DAR26866; Lane 5 = DAR26889; Lane 6 = DAR26901; Lane 7 = DAR26927; Lane 8 = DAR26928; Lane 9 = DAR26934; Lane 10 = DAR26936; Lane 11 = DAR26941; Lane 12 = DAR29063; Lane 13 = DAR30521; Lane 14 = DAR30522; Lane15 = DAR33386; Lane 16 = DAR33396; Lane 17 = DAR35718; Lane 18 = DAR63291; Lane 19 = 1kb+ marker.

3.3A – Proteomics

The proteomics component focussed on identifying differentially expressed proteins to develop DNA-based tests that could distinguish between *X. citri* subsp*. citri,* and the closely related *X. citri* subsp*. malvacearum*. A secondary aim was to develop DNA-based tests to differentiate between the xanthomonads that infect citrus, specifically the pathogenic *X. citri* subsp*. citri* from the less aggressive *X. alfalfae* pv *citrumelonis*.

3.3A.1 Phase 1 - Preliminary experiments

The initial stages of the project focussed on project planning and skills expansion, with the team from NSW travelling to Perth to for training in preparing, handling and analysing protein extracts and the team from WA travelling to EMAI for training in microbiology and bacterial methods. A proprietary (kit-based) method to extract membrane bound proteins from bacteria was tested for ease of use and reproducibility, with extractions performed on the same organism in both states. This also confirmed that extracts prepared in this way were amenable to transport interstate without the risk of deterioration.

Preliminary work to optimise protein profiling using 2-Dimensional (2D) gels analysis determined the optimum sample loading, IEF range and resolution conditions and included extensive work building familiarity with new, improved software and reporting applications. Subsequent analyses confirmed the quality and integrity of extracts prepared in NSW and shipped to WA, the fidelity of "replicate" extracts, and that different profiles were observed for the membrane-bound proteins extracted from different bacterial species.

Selected isolates were passaged through sensitive and tolerant hosts in a series of detached leaf bioassays to characterise and confirm their host specificity. Once symptoms were scored and the bioassay concluded, bacteria were re-isolated from the resulting lesions or the site of inoculation. The

membrane-associated proteins were extracted from these re-isolated bacterial strains and from isolates of the same bacteria maintained on media plates alone. These extracts were analysed for any effect of passage through a plant host on the expression of surface mounted proteins. Profiles were not remarkably different for the same isolate when cultured on an agar plate compared to when the same isolate had been passaged through a sensitive host. We concluded that cultures on media plates were an appropriate source of useful protein extracts, simplifying the process of generating suitable protein extracts for this study.

3.3A.2 Phase 2 – data collection and analysis

More detailed 2D gel analysis of proteins from specific pairs of pathovars was performed to identify differentially-expressed proteins. A moderate number of spots were obtained from hydrophobic cell membrane extracts with many protein spots present at around pH 4 and many with mass of 60-100kD. Horizontal chains of spots were sometimes present, potentially related to post translational modification, sample degradation or differences in charge. The proteins profiled within replicated gels were compared using principle component analysis (NonLinear Dynamics Samespots software).

Matching of gels and evaluation by Principal Component Analysis (PCA) showed clustering of isolates of the same pathovar and separate from those of a different pathovar. For example, Figure 3 shows differentiation between the profiles from strains of Xanthomonads from citrus (pink dots) and *X. citri* subsp*. malvacearum* (blue dots). In contrast, there was a large degree of similarity between *X. citri* subsp*. citri* and many *X. a.* pv. *glycines* strains. Far fewer differences were observed between the pathotypes A, A* and Aw within *X. citri* subsp*. citri*. The results confirm separate clustering of most of the samples from citrus-pathogenic isolates compared to the *X. citri* subsp*. malvacearum* samples from cotton.

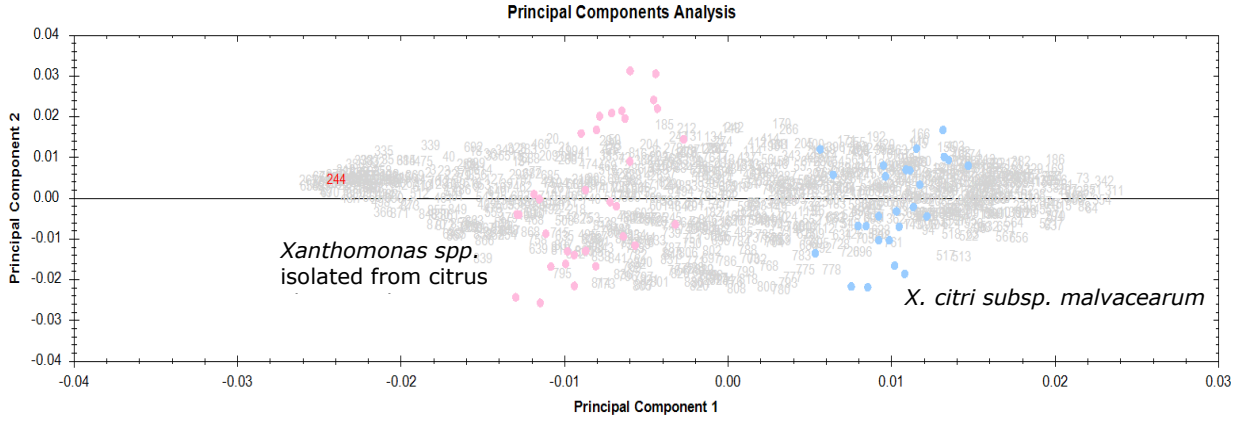


Figure 3: PCA comparing 2D gel electrophoresis of proteins from the hydrophobic cell wall fraction of xanthomonads originally isolated from cotton or citrus plants and grown on nutrient agar.

Analysis of replicated extracts of total membrane-associate proteins from biological replicates of bacteria indicated particular 2D protein profiles are maintained within isolates of a pathovar. The results suggest that 2D protein profiling could potentially be used as a confirmatory test to identify these pathovars, by reference to the local data 'library' that we have created. This is the first demonstration of its kind in plant pathogenic bacteria.

More detailed analysis of individual protein profiles from a range of *Xanthomonas* species and pathovars that are pathogenic to citrus and different isolates from cotton (Tables 1 and 2) identified protein spots that were differentially expressed between bacterial pathovars attacking different hosts (e.g. between *X. citri* subsp*. citri and X. citri* subsp*. malvacearum*). It also showed higher capacity to differentiate between more closely related organisms, for example Figure 4 identifies proteins that are differentially expressed between *X. citri* subsp*. citri* pathotype A and *X. alfalfae* pv*. citrumelonis,* both of which attack citrus.

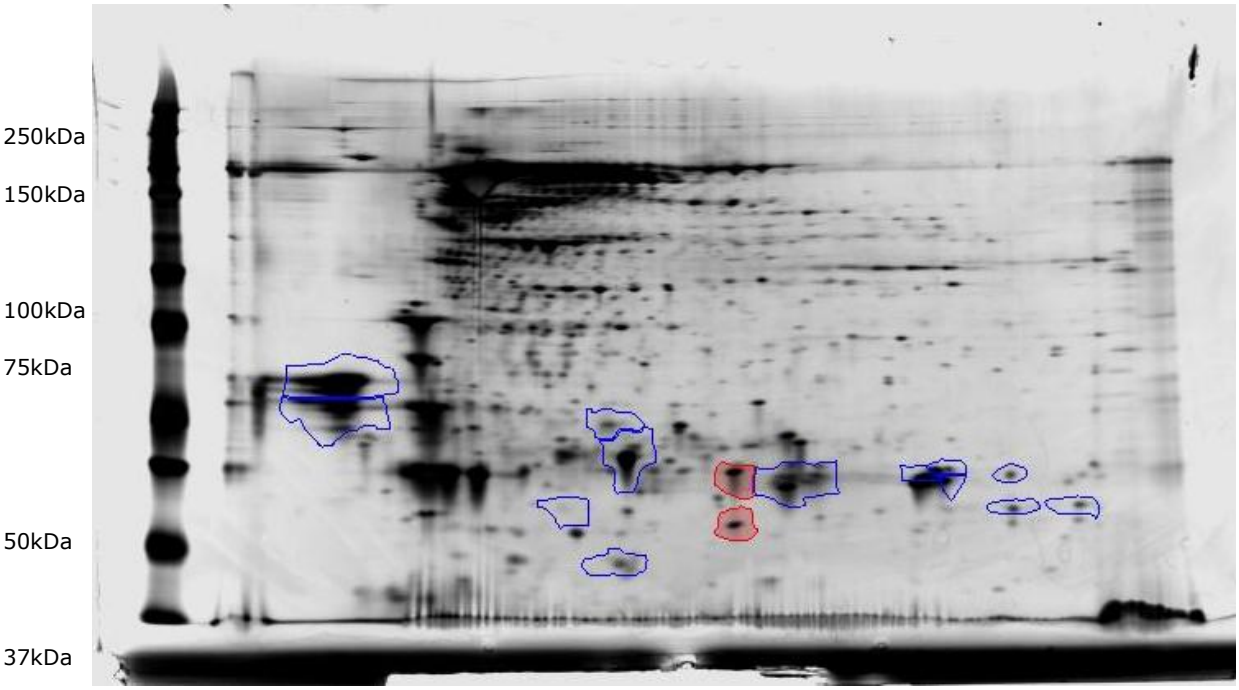


Figure 4: 2D gel showing the differential spot expression between *X. citri* subsp*. citri* type A and *X. alfalfae* pv *citrumelonis*. Spots not present in pv *citrumelonis* are outlined in red and those present only in *X. citri* subsp*. citri* type A are outlined in blue.

These differentially expressed proteins are candidate biomarkers on which pathovar-specific diagnostic tests can be developed. Their potential was validated by confirming that the differences in protein expression are seen in multiple, completely independent bacterial isolates of these pathovars (collected from different countries and host species and at different times).

Proteins that were found to be differentially expressed between *X. citri* subsp*. citri* and *X. citri* subsp. *malvacearum*, and between strains of E and A (A, Aw and A*) pathotypes were chosen for further analysis. Specifically, this involved peptide sequencing and the screening of protein databases, inferred from whole genome reference sequences, to identify first the differentially expressed proteins and then the genes that encode them. Amongst the Xanthomonads attacking citrus, the international nucleotide sequence databases only contain the whole genome sequence for one isolate of *X. citri* subsp*. citri* (da Silva *et al*., 2002). There are no reference whole genome sequences for any isolates of *X. citri* subsp*. malvacearum* or *X. alfalfae pv. citrumelonis.* It was therefore necessary to confine our subsequent analyses to only those spots that were identified as being present in the *citri* pathovar but absent in other isolates. Whole genome sequences for isolates of *X. campestris pv campestris* (Genbank accession number NC_007086, Qian *et al.,* 2005) and *X. vesicatoria* (Genbank accession number NC_007508 Thieme *et al.,* 2005) were also available and were used for comparative purposes.

Differentially expressed protein spots were excised, digested with trypsin and subjected to mass spectrometry to provide peptide fingerprints. Peptide fingerprints were compared to peak profiles available in the 'Mascot' and 'ProTOF' databases but no direct matches were observed.

Candidate peptides were then subjected to amino acid sequencing and these peptide sequences were compared to protein databases. This provided a list of possible protein matches in various species, including to *X. citri* subsp*. citri* , *X. campestris pv. campestris* and *X. vesicatoria.* This process putatively identified some of the proteins as being consistent with plant pathogen interactions, such as outer membrane proteins (confirming the validity of the extraction method), selective porins and enzyme subunits. Some proteins were more novel, not yet having had a function assigned to them.

Table 6 summarises the proteins identified in this way. Most differentially expressed protein spots occurred at only one unique location on the 2D gel, however the spots numbered 1, 2 and 3 all corresponded to the same protein when sequenced, and similarly those numbered 7 and 8 were derived

from the same protein. Presumably the different spots on the gel were different modifications of the same protein.

Table 6 : Differentially expressed proteins and their characteristics

Protein spot number	Putative ID	Function	No. of Primers	Names of Primers
1,2,3	Outer membrane protein	Unknown	4	s123sF, s123sR, s123nsF, s123nsR
4	Putative uncharacterised protein	Unknown	4	s4sF, s4sR, s4nsF, s4nsR
5	Outer membrane protein	Unknown	4	s5sF, s5sR, s5nsF, s5nsR
6	Outer membrane protein P6/putative uncharacterised protein	Unknown	8	s6sF, s6sR, s6nsF, s6nsR, ompP6 sF, ompP6 sR, ompP6 nsF, ompP6 nsR
7,8	Outer membrane protein W	Higher levels of expression in oxidative stress (Asakura *et al*., 2008); modulation of expression in cells grown under stress such as elevated temps., high salt and low aeration (Nandi *et al*., 2005)	4	ompW sF, ompW sR, ompW nsF, ompW nsR
			11	s8sF, s8sR, s8s2F, s8s2R, s8s3F, s8s3R, s8s4F, s8nsF, s8nsR, s8ns2F, s8ns2R
9	Outer membrane protein	unknown	4	s9sF, s9sR, s9nsF, s9nsR
14	Alkyl hydroperoxide reductase subunit C	Catalytic subunit responsible for alkyl peroxide metabolism (Mongkolsuk *et al*., 2000)	6	ahrsc s1, ahrsc s1R, ahrsc s2F, ahrsc s2R, ahrsc nsF, ahrsc nsR
15	Polyphosphate-selective porin O	Small ion channel, including under conditions of overnight phosphate starvation (Hancock *et al*., 1992)	8	ppO F 2, ppO R 3, ppO R 4, , ppO R 5, ppO R 6, ppO F 7, ppO F 8, ppO F 9
16	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MECDP)	An enzyme in the mevalonate-independent isoprenoid biosynthetic pathway (Richard *et al*., 2002). Catalyses the conversion of 4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate to MECDP.	4	s16sF, s15sR, s16nsF, s16nsR

3.3A.3 Development and validation of novel molecular diagnostics

Beginning with proteins that are expressed in *X. citri* subsp*. citri* but not other related pathovars, the available whole genome sequences for Xanthomonad species were screened to identify the genes encoding those proteins. Comparison of the sequence of these genes in *X. citri* subsp*. citri* to the corresponding genes from other available Xanthomonads informed the design of multiple primer sets for each of the targeted genes. A minimum of two sets of primers were identified for each gene – one generic, designed to amplify the entire gene, and at least one that was intended to provide selectivity to a particular pathovar (for potential use as the diagnostic tool). Table 6 indicates the number of primers designed for each protein spot. Sequences of primers tested but not detailed in Figure 5 are available from these authors on request.

The diagnostic primer sets were initially evaluated under high-stringency conditions against a small panel of isolates as a 'first pass'. In most cases the fragment was amplified in both *X. citri* subsp*. citri* and *X. citri* subsp*. malvacearum,* indicating that the gene was actually present in both subspecies and that particular primer pair was therefore unlikely to be suitable for further evaluation. In these cases, the generic primer set was subsequently employed to amplify the entire gene, which was then subjected to nucleotide sequencing to investigate differences that might be exploited in designing other primer pair(s) that may better facilitate a diagnostic test.

Primer sets that showed potential to discriminate between operative taxa were evaluated against DNAs from a comprehensive panel of bacterial isolate.

New molecular diagnostic differentiating *X. citri* subsp*. citri* from *X. citri* subsp*. malvacearum*

One primer set targeting a fragment within the gene for differentially expressed protein number 7,8 (Table 6), strongly amplified the diagnostic fragment (177 bp) from all *X. citri* subsp*. citri* pathotypes but did not amplify isolates of *X. citri* subsp*. malvacearum* (Figure 5).

This pair was validated against a total of 69 different *Xanthomonas* isolates and other pathogens of citrus as indicated in Tables 1, 2 and 3.

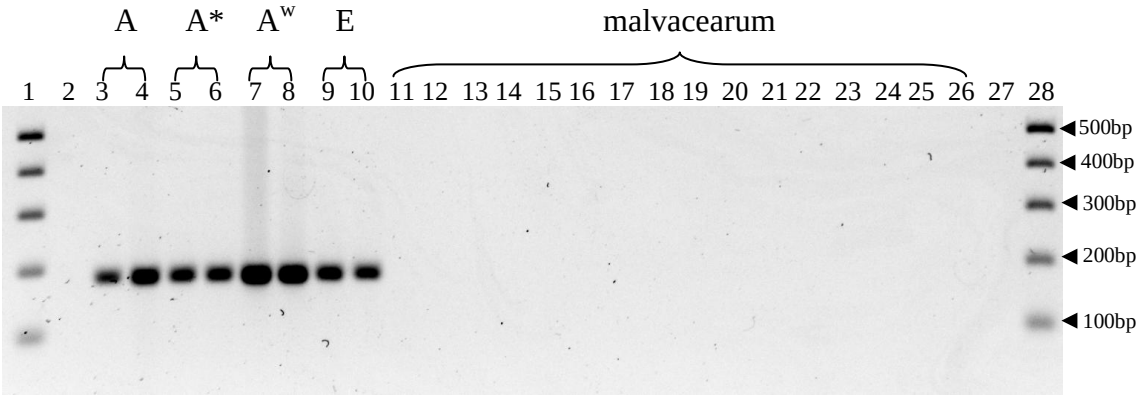


Figure 5: New diagnostic to discriminate *citri* from *malvacearum,* targeting a region of the gene encoding differentially expressed protein Spot 7,8. Lanes 1 & 28: 1kb + Marker; Lanes 2 & 27: negative control; Lane 3: DAR77358; Lane 4: MN04/8516 699A; Lane 5: 169; Lane 6: 206; Lane 7: X02-1035; Lane 8: X2002-12884; Lane 9: XSol-4; Lane 10: P03-83; Lane 11: DAR26840; Lane 12: DAR26843; Lane 13: DAR26866; Lane 14: DAR26889; Lane 15: DAR26927; Lane 16: DAR26928; Lane 17: DAR26934; Lane 18: DAR26936; Lane 19: DAR26941; Lane 20: DAR29063; Lane 21: DAR30521; Lane 22: DAR30522; Lane 23: DAR33386; Lane 24: DAR33396; Lane 25: DAR35718; Lane 26: DAR63291. Refer to Table 1 & 2 for further detail regarding isolates.

False positives were occasionally observed with this set of primers, that is the diagnostic fragment was amplified from isolates other than *X. citri* subsp *citri* (Table 6). For example, three isolates of the *X. axonopodis* pv. *sesame* and one of two *X. fuscans* subsp. *aurantifolii* were positive for the spot 7,8 fragment (information shown in Table 7). An *in silico* analysis of the primer sequences also indicated that this pair would also amplify a fragment from *X. vesicatoria*, and this was confirmed by analysis of four isolates. This was not considered to compromise the integrity of the test because, unlike cotton, these hosts are not grown in the same areas as citrus.

Table 7: Isolates used to validate the new diagnostic assays, designed to selectively amplify fragments of the genes encoding the differentially expressed proteins Spots 7,8 (ompP1 – ~ 177bp) and Spots 1,2,3 (mopB gene – ~ 486bp). Unk. = Unknown.

Isolate	**Pathotype / subsp./ pathovar**	**Spots 7,8 (ompP1)**			**Spots 1,2,3 (mopB)**		
		Number of isolates	Number positive	Number negative	Number of isolates	Number positive	Number negative
X. citri	*citri,* A	24	24		24	24	
	citri, A*	15	14	1	15	15	
	citri, Aw	9	9		9	9	
	citri, Unk.	4	3	1	4	4	
X. fuscans	*aurantifolii*, B	1	1		1	1	
	aurantifolii, C	1		1	1		1
X. axonopodis	*malvacearum*	16		16	16	16	
	phaseoli	3	3		3	2	1
	vitians	1		1	1		1
	sesame	3	3		3		3
	fici	1	1		1		1
X. alfalfae	*citrumelonis,* E	2	2		2		2
X. campestris	*campestris*	3		3	3		3
	aberrans	1		1	1		1
	armoraciae	1		1	1		1
	barbareae	1		1	1		1
	incanae	1		1	1		1
	raphani	2		2	2		2
X. albilineans		1		1	1		1
X. cucurbitae		1		1	1		1
X. vesicatoria		5	4	1	5	1	4
X. hortorum	*carotae*	1		1	1		1
X. ssp.	Unknown	1	1		1		1
X. oryzae	*oryzae*	1		1	1		1
Pseudomonas syringae	*syringae*	1		1	1		1
	lacrymans	1		1	1		1
	maculicola	1		1	1		1
Pseudomonas viridiflava		1		1	1		1
Pseudomonas cichorri		1		1	1		1
Pectobacterium carotovorum		1		1	1		1
Pantoea agglomerans		1		1	1		1
Erwinia chrysanthemi		1		1	1		1
Erwinia amylovora		1		1	1		1
Erwinia billingiae		2		2	2		2
Clavibacter michagenensis		5		5	5		5
Mycospharella sp.		2		2	2		2
Guignardia citricarpa		2		2	2		2
Candidatus Liberibacter	*asiaticus*	2		2	2		2
Burkholderia andropogonis		2		2	2		2

New molecular diagnostic differentiating *X. citri* from *X. alfalfae pv citrumelonis*

A second pair of primers targeted the gene for differentially expressed protein corresponding to Spot 1/2/3 and amplified a fragment of 486 bp. This fragment was reliably present in strains of *X. citri* subsp. *citri* but absent from *X. alfalfae pv citrumelonis* (Figure 6).

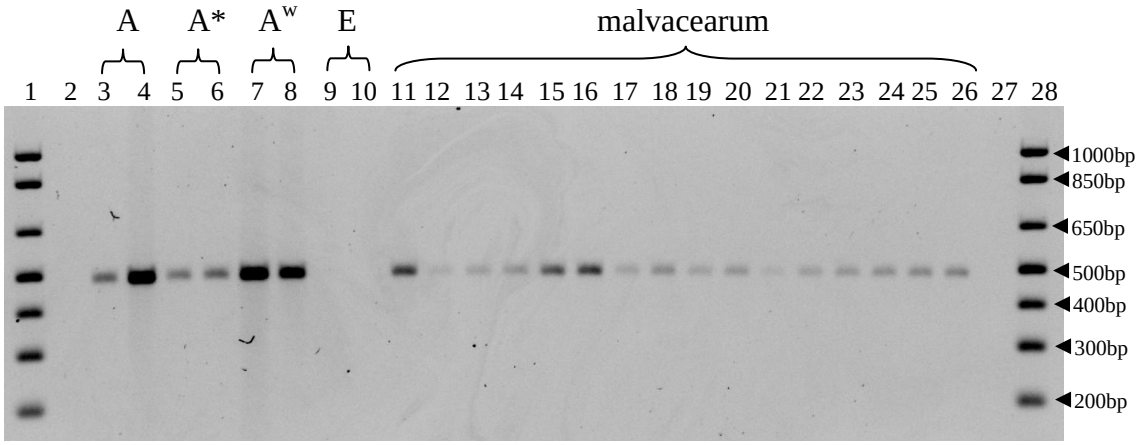


Figure 6: New diagnostic to discriminate *citri* from *citrumelonis (pathotype 'E'),* targeting a region of the gene encoding the differentially expressed protein Spot 1,2,3.. Lanes 1 & 28: 1kb + Marker; Lanes 2 & 27: negative control; Lane 3: DAR77358; Lane 4: MN04/8516 699A; Lane 5: 169; Lane 6: 206; Lane 7: X02-1035; Lane 8: X2002-12884; Lane 9: XSol-4; Lane 10: P03-83; Lane 11: DAR26840; Lane 12: DAR26843; Lane 13: DAR26866; Lane 14: DAR26889; Lane 15: DAR26927; Lane 16: DAR26928; Lane 17: DAR26934; Lane 18: DAR26936; Lane 19: DAR26941; Lane 20: DAR29063; Lane 21: DAR30521; Lane 22: DAR30522; Lane 23: DAR33386; Lane 24: DAR33396; Lane 25: DAR35718; Lane 26: DAR63291. Refer to Table 1 and 2 for further detail regarding isolates.

This primer pair was similarly validated against our collection of isolates and our other non-Xanthomonas bacteria and other microorganisms commonly found on citrus (Table 7). These included other non-Xanthomonas bacteria including species of *Pseudomonas, Erwinia and Clavibacter*, and from other exotic and endemic bacteria occurring on citrus, such *Guignardia, Alternaria, Mycosphaerella and Liberibacter*. The range of strains against which these two primer sets were validated is presented in Table 8.

Table 8: Summary of isolates used to validate selectivity of PCRs for Spots 7, 8 and Spots 1,2,3

Isolate	Spots 1,2,3 - 486bp	Spot 7,8 -177bp
***X. citri* subsp. *citri*, pathotype A**		
DAR77358 [MN04/5578 60.1]	+	+
MN04/8516 699A	+	+
DAR77356 [MN04/5370 3.2A]	+	+
MN04/5703 389.7A	+	+
DAR77357 [MN04/5370 8.2]	+	+
MN04/5703 393.1	+	+
DAR77359 [MN04/5703 394.2]	+	+
MN04/5703 388.4	+	+
DAR65863	+	+
DAR65869	+	+
DAR69978a	+	+
DAR69978b	+	+
DAR73871	+	+
DAR73872	+	+
DAR73889	+	+
DAR75280	+	+
XS97-3-1	+	+
XS97-5-1	+	+
ATCC49118	+	+
X03-2914	+	+
X02-10	+	+
X04-1469	+	+
207	+	+
209	+	+
X. citri* subsp. *citri*, pathotype A		
Xcc290	+	+
169	+	+
206	+	+
269	+	+
270	+	+
271	+	+
276	+	+
289	+	+
323	+	+
328	+	+
329	+	+
Xcc406	+	+
DAR72029	+	+
DAR73909	+	+
***X. citri* subsp. *citri*, pathotype Aw**		
X02-1035	+	+
X2002-12884 [DAR76571]	+	+
X2001-00005 [DAR76572]	+	+
X2001-00032 [DAR76573]	+	+
X2003-01008 [DAR76575]	+	+
X2003-01012 [DAR76576]	+	+
X2003-01029 [DAR76577]	+	+
X2003-01036 [DAR76578]	+	+
X03-1003	+	+
***X. fuscans* subsp. *aurantifolii*, B strain**		
1622	+	+

***X. fuscans* subsp. *aurantifolii*, C strain**		
ATCC51302	-	-
***X. alfalfae* subsp. *citrumelonis*, E type**		
P03-83	-	+
XSol-4	-	+
***X. citri* subsp. *citri*, Not Typed**		
284	+	+
324	+	+
327	+	+
***Xanthomonas* sp., Not Typed**		
DAR72016	-	+
X. citri* subsp. *malvacearum		
DAR26840	+	-
DAR26843	+	-
DAR26866	+	-
DAR26889	+	-
DAR26927	+	-
DAR26928	+	-
DAR26934	+	-
DAR26936	+	-
DAR26941	+	-
DAR29063	+	-
DAR30521	+	-
DAR30522	+	-
DAR33386	+	-
DAR33396	+	-
DAR35718	+	-
DAR63291	+	-
X. albilineans		
DAR34874	-	-
X. axonopodis* pv. *phaseoli		
DAR75943	-	+
DAR58726	+	+
DAR65944	+	+
X. axonopodis* pv. *vitians		
PHDS 02/559	-	-
X. campestris* pv. *aberrans		
DAR75944	-	-
X. campestris* pv. *armoraciae		
DAR75942	-	-
X. campestris* pv. *barbareae		
DAR75945	-	-
X. campestris* pv. *campestris		
DAR75950	-	-
DAR76566	-	-
DAR76567	-	-
X. axonopodis* pv. *fici		
DAR61743	-	+
X. campestris* pv. *incanae		
DAR75959	-	-
Xanthomonas campestris* pv. *raphani		
DAR75960	-	-
DAR75983	-	-

Xanthomonas axonopodis* pv. *sesame		
DAR75558	-	+
DAR75547	-	+
DAR75559	-	+
Xanthomonas cucurbitae		
DAR75961	-	-
Xanthomonas hortorum* pv. *carotae		
DAR49849	-	-
Xanthomonas oryzae* pv. *oryzae		
DAR26932	-	-
Xanthomonas vesicatoria		
DAR26932	-	-
DAR26933	-	+
DAR33341	+	+
DAR73877	-	+
DAR80973	-	+
Pseudomonas syringae* pv. *syringae		
DAR54698	-	-
Pseudomonas syringae* pv. *lachcrymans		
PHDS 01/523	-	-
Pseudomonas viridiflava		
DAR73894	-	-
Pseudomonas cichorri		
DAR58720	-	-
Pseudomonas maculicola		
DAR30556	-	-
Burkholderia andropogonis		
PHDS 01/275	-	-
DAR75243	-	-
Pectobacterium carotovorum* prev. *Erwinia carotovora* pv. *carotovora		
DAR65858	-	-
Pantoea agglomerans* prev. *Erwinia herbicola		
DAR72022	-	-
Erwinia chrysanthemi		
DAR73906	-	-
Erwinia amylovora		
DAR72017	-	-
Erwinia billingiae		
PHDS 09/755B 2	-	-
PHDS 09/755B 5	-	-
Clavibacter michagensis		
4080	-	-
4087	-	-
4071	-	-
2048	-	-
4088	-	-
Mycosphaerella sp.		
M21	-	-
M24 Wendy G	-	-

Guignardia citricarpa		
DAR56256	-	-
DAR33474	-	-
Liberobacter asiaticus		
HLB 1445	-	-
HLB 1318	-	-

Nucleotide sequences for the diagnostic fragments for significant isolates have been uploaded into GenBank, an international nucleotide sequence database. This allows the wider scientific community to access these data. Details on Accession Numbers etc are presented in Appendix 1 of this report.

Full nucleotide sequence analysis of the entire genes encoding the differentially expressed proteins 7,8 and 1/2/3 was performed using the generic primer sets. Sequences were aligned and the relationship between isolates for the genes plotted as dendograms.

The dendogram for spot 7,8 (Figure 7) shows that the majority of the *X. citri* subsp *citri* isolates were identical in sequence and group together, with more than 98% identity overall. Similarly, the *X. citri* subsp. *malvacearum* isolates cluster strongly together and share more than 98% identity to each other. Importantly, the isolates of *X. citri* subsp *citri* show only 92% identity to the isolates of *X. citri* subsp *malvacearum* in this region of the genome. The location and distribution of these nucleotide sequence disparities forms the basis for the diagnostic assay.

The dendogram for spots 1,2,3 (Figure 8) also shows strong clustering within the isolates *X. citri* subsp *citri* and *X. citri* subsp *malvacearum* (100% identical to each other). In this region of the genome, however, but two clusters show 98.5% similarity between the two subspecies, confirming that the molecular assay to this gene detects both *X. citri* subsp *citri* and subsp. *malvacearum.* Significantly, in this case the two available isolates of *X. alfalfae pv citrumelonis* show differences between them, and more than 3% difference from sufficient difference from *X. citri* subsp *citri* and subsp. *malvacearum.* The location of these differences underpin the development of the second novel diagnostic assay.

Both dendograms (Figure 7 & 8) indicate that one of the vesicatoria isolates (DAR33341) is very close in similarity to the citri isolates, in these regions of the genome.

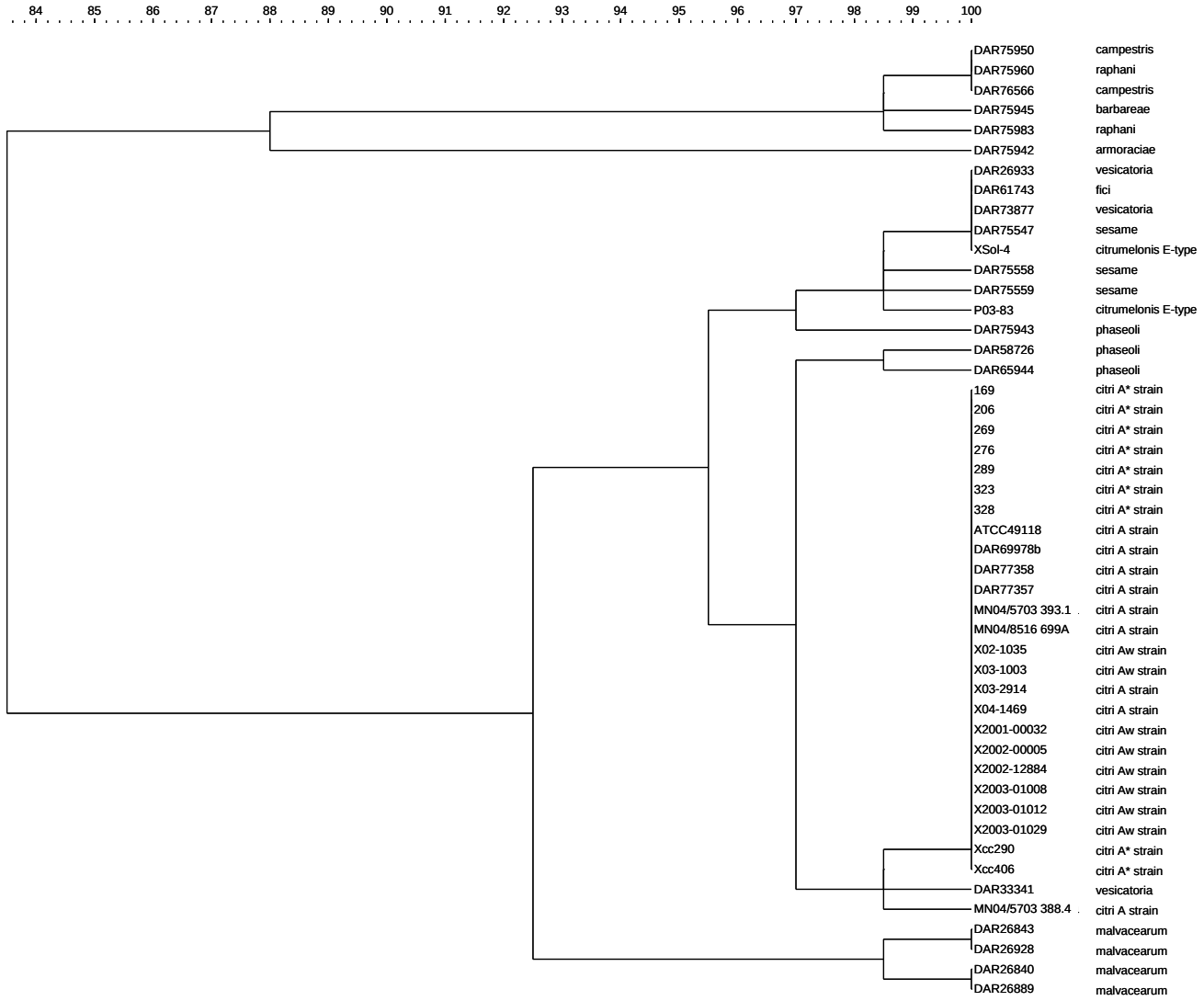


Figure 7: Dendogram of the alignment of gene sequences for outer membrane protein P1 (spot 7,8) ~ 540bp fragment. This was analysed using the Needleman-Weishmer UPGAMA complete linkage analysis This was analysed using the multiple alignment settings of the Needleman-Wunsch algorithm, pairwise alignment (open gap penalty=15; Extend gap penalty=6.7), multiple alignment (Open gap penalty=15; Extend gap penalty=6.7; Delay divergent sequences=30; DNA transition weight =0.50) Similarity calculation (DNA weight matrix=IUB). Clustering settings: Needleman-Wunsch algorithm, Pairwise alignment (open gap penalty=15; extend gap penalty=6.7), Clustering: Complete linkage, Similarity calculation: DNA weight matrix IUB.

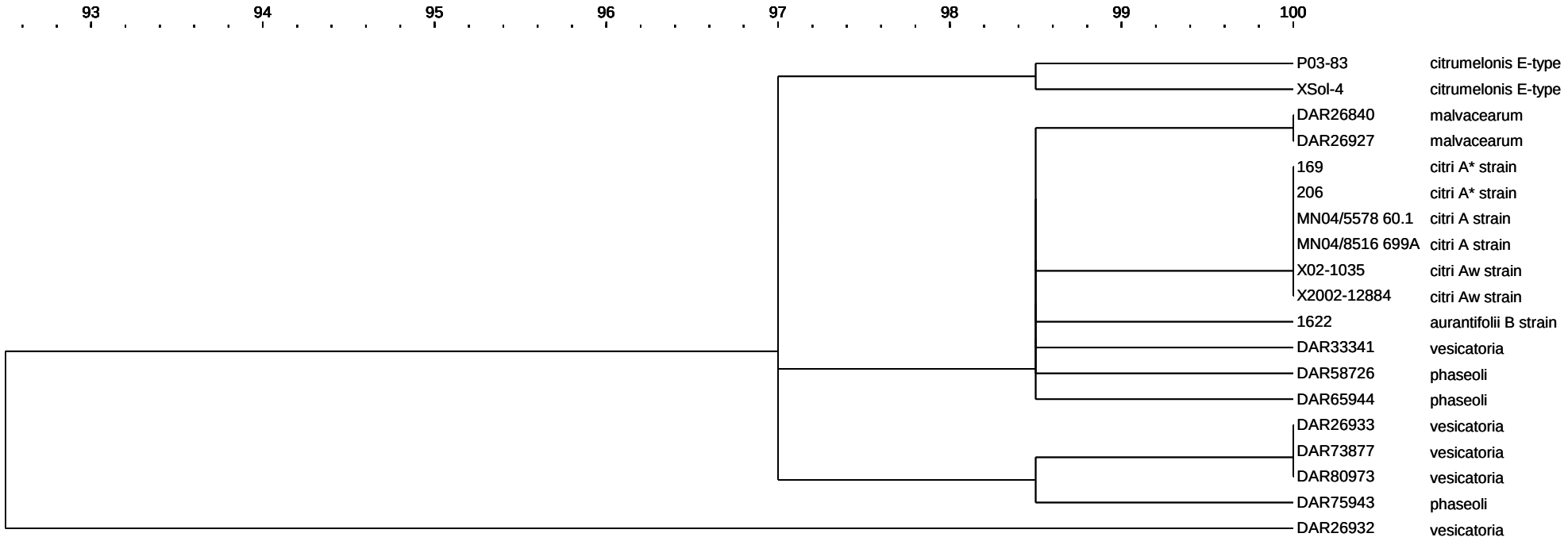


Figure 8: Dendogram of the alignment of gene sequences for the *mopB* gene (spots 1,2,3) ~ 971bp fragment. This was analysed using the multiple alignment settings of the Needleman-Wunsch algorithm, pairwise alignment (open gap penalty=15; Extend gap penalty=6.7), multiple alignment (Open gap penalty=15; Extend gap penalty=6.7; Delay divergent sequences=30; DNA transition weight =0.50) Similarity calculation (DNA weight matrix=IUB). Clustering settings: Needleman-Wunsch algorithm, Pairwise alignment (open gap penalty=15; extend gap penalty=6.7), Clustering: Complete linkage, Similarity calculation: DNA weight matrix IUB.

3.3A.4 Whole genome sequencing

Extracts of pure DNA from selected isolates were subjected to whole genome sequence analysis. This generated a greater than 95% coverage of the approximate 5 Mbp genome of 1 x *X. citri* subsp. *citri* pathotype A (56 Mbp amplified), 2 x *X. citri* subsp. *citri* pathotype A* (66 Mbp and 51 Mbp amplified) and 2 x *X. citri* subsp. *malvacearum* (48 MBp and 80 Mbp amplified).

Primary annotation of each genome has been completed and the output of the gene profiles will follow in supplemental files uploaded to IMAP. The overview of the composition of each genome is described and compared to the reference genome (*X. citri* subsp. *citri* isolate 306, da Silva *et al*., 2006) in Table 9.

All genomes show strong similarity at the global scale, and Figure 9a depicts the circular view of the genome of one isolate as a representative, this being *X. citri* subsp. *citri* pathotype A*. The two genes that serve as the novel diagnostic targets in the assays developed during this study (mopB and ompP1) are annotated. Also annotated are the location of the genes involved in betaine glycine production, a pathway highlighted in the metabolomics results.

Strong relationships were observed between the genome sequence of isolates we had selected for this part of our study and genome of the reference isolate, *X. citri* subsp. *citri* isolate 306 (da Silva *et al*., 2006). A visual representation of each of the pair wise comparisons (306 to our isolate) is presented in Figure 9b to 9f, where for each of the five genomes sequenced the red region represents areas of sequence matching that in the reference isolate 306, the white regions represent regions present in the reference strain but absent in our sequenced strains, and the blue lines represent areas of sequence that appear in both strains but are translocated between the two. Not surprisingly, the greatest differences were observed between the reference isolate *X. citri* subsp. *citri* 306 and the sequences for our two isolates of *X. citri* subsp*. malvacearum.* Reasons why regions are identified as 'missing' from our sequence data could be attributed, amongst other things, to gaps in our data or its assembly (incomplete sequence) or these could be regions of DNA that are genuinely not present in our isolates. These possibilities will be investigated further following the conclusion of this project.

Table 9: Overview of the whole genome sequence of five isolates as determined using the 454 GS Junior Sequencer, including comparison to reference genome for *X. citri* subsp. *citri* strain 306 (da Silva *et al*., 2006)

Name	Number of bases	Number of genes	Gene sequence composition	rRNA bases	rRNA Base composition	tRNA bases	tRNA sequence composition	Overall sequence composition
X. citri subsp. *citri* reference strain 306	5175554	4312	A 17.53% C 32.22% G 32.89% T 17.35%	9094	A 25.55% C 22.8% G 31.33% T 20.29%	4243	A 18.83% C 29.15% G 31.81% T 20.19%	A 17.62% C 32.42% G 32.34% T 17.6%
X. citri subsp. *citri* pathotype A*	5029423	4585	A 17.37% C 32.37% G 33.05% T 17.19%	230	A 23.47% C 26.95% G 32.17% T 17.39%	4075	A 18.94% C 29.2% G 31.63% T 20.22%	A 17.57% C 32.46% G 32.41% T 17.54%
X. citri subsp. *citri* pathotype A*	4916308	4485	A 17.33% C 32.38% G 33.07% T 17.21%	230	A 23.47% C 26.95% G 32.17% T 17.39%	4004	A 18.93% C 29.14% G 31.66% T 20.25%	A 17.52% C 32.49% G 32.48% T 17.49%
X. citri subsp. *malvacearum*	4382336	4504	A 17.2% C 32.29% G 33.22% T 17.27%	114	A 23.68% C 26.31% G 31.57% T 18.42%	3361	A 18.89% C 29.03% G 31.59% T 20.47%	A 17.41% C 32.58% G 32.61% T 17.39%
X. citri subsp. *malvacearum*	4568385	4167	A 17.02% C 32.43% G 33.14% T 17.21%	230	A 23.47% C 26.95% G 32.17% T 17.39%	4004	A 18.9% C 29.17% G 31.66% T 20.25%	A 17.38% C 32.6% G 32.63% T 17.37%
X. citri subsp. *citri* pathotype A	4568385	4167	A 17.2% C 32.43% G 33.14% T 17.21%	230	A 23.47% C 26.95% G 32.17% T 17.39%	4004	A 18.9% C 29.17% G 31.66% T 20.25%	A 17.38% C 32.6% G 32.63% T 17.37%

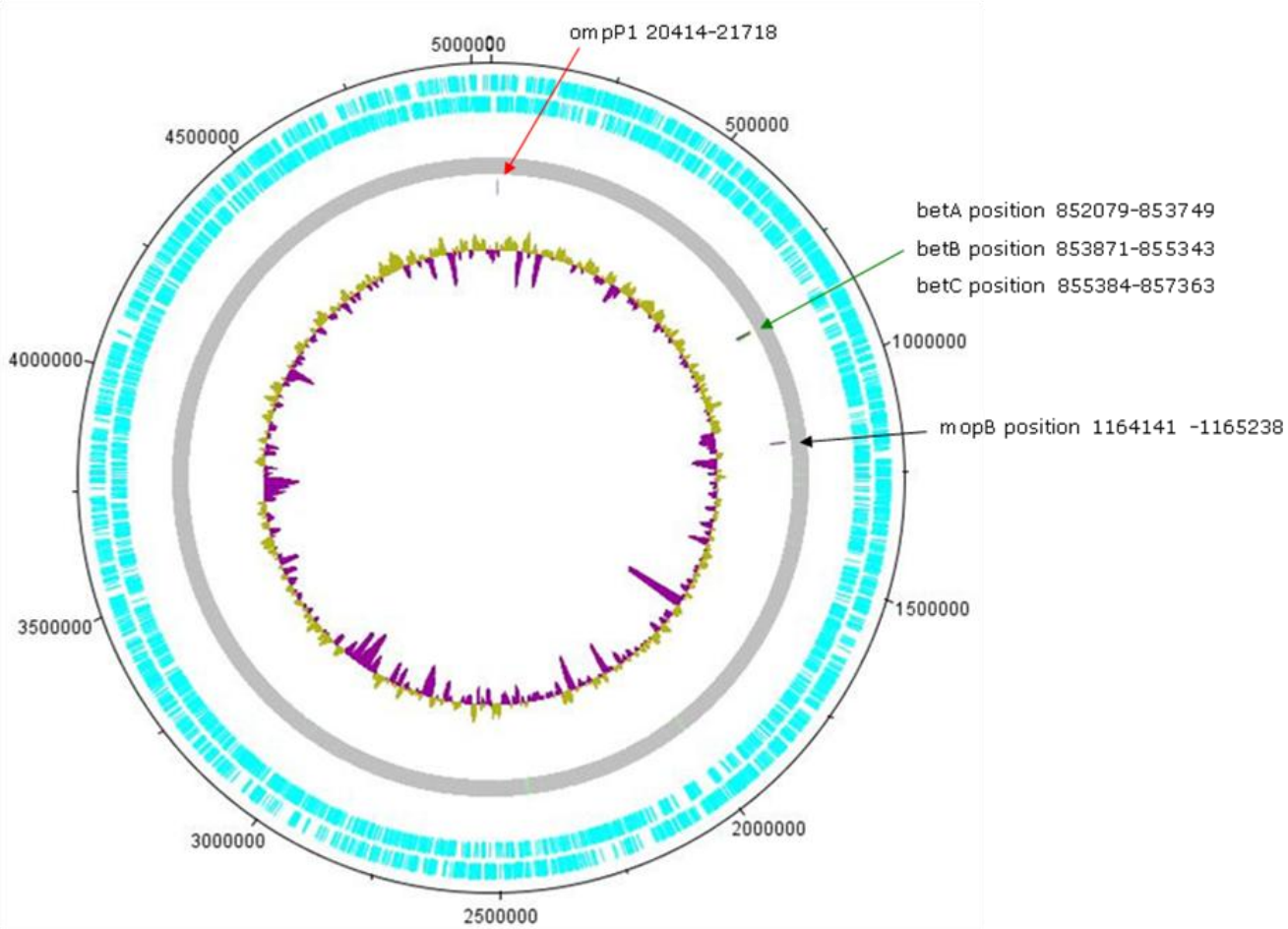


Figure 9a: Circular view of the whole genome of *X. citri* subsp. *citri* pathotype A*, including annotation to indicate the position of genes used in this study. The purple and green are indicative of the GC skew.

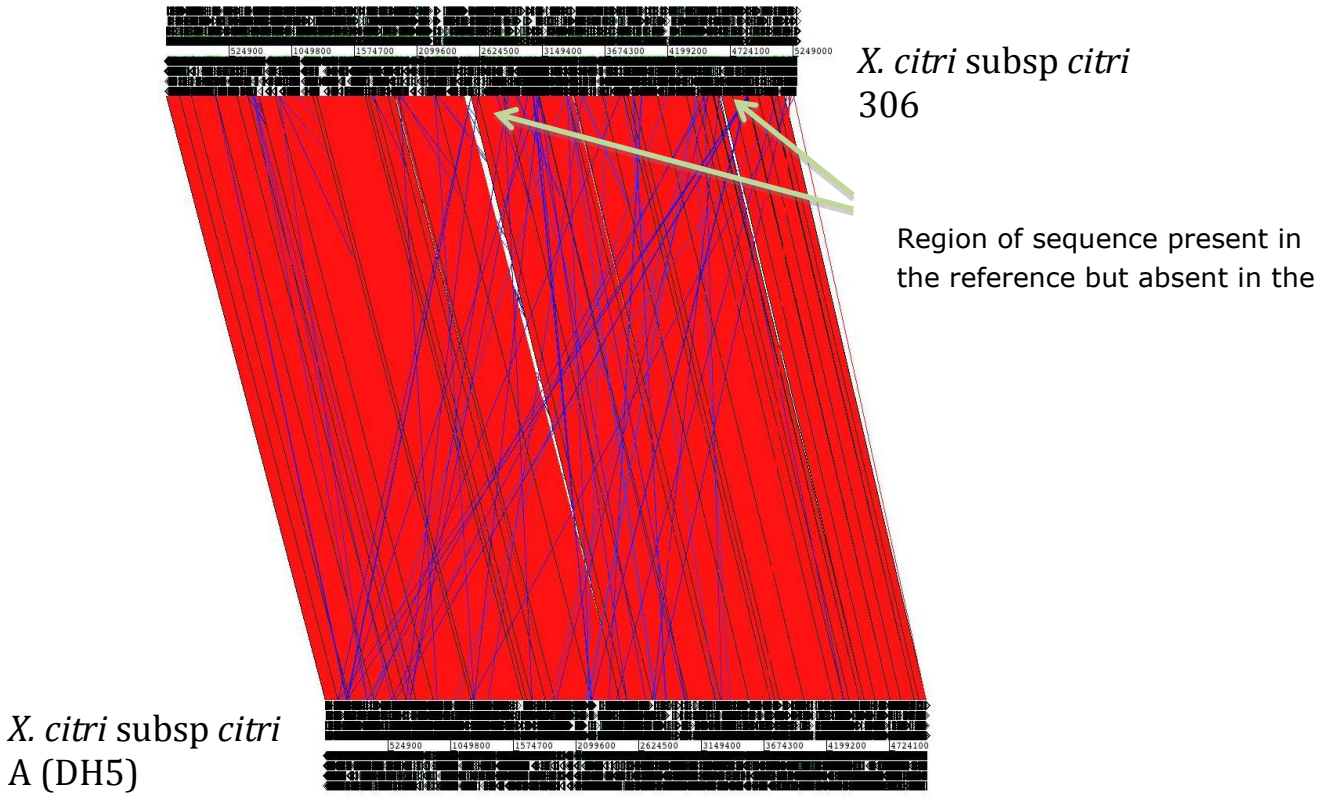


Figure 9b: Comparison showing the alignment of the reference *X. citri* subsp *citri* 306 strain with *X. citri* subsp. *citri* A (isolate DH5). Regions shown in white represent regions missing in the DH5 strain compared to the reference and blue lines represent areas translocated compared to the reference.

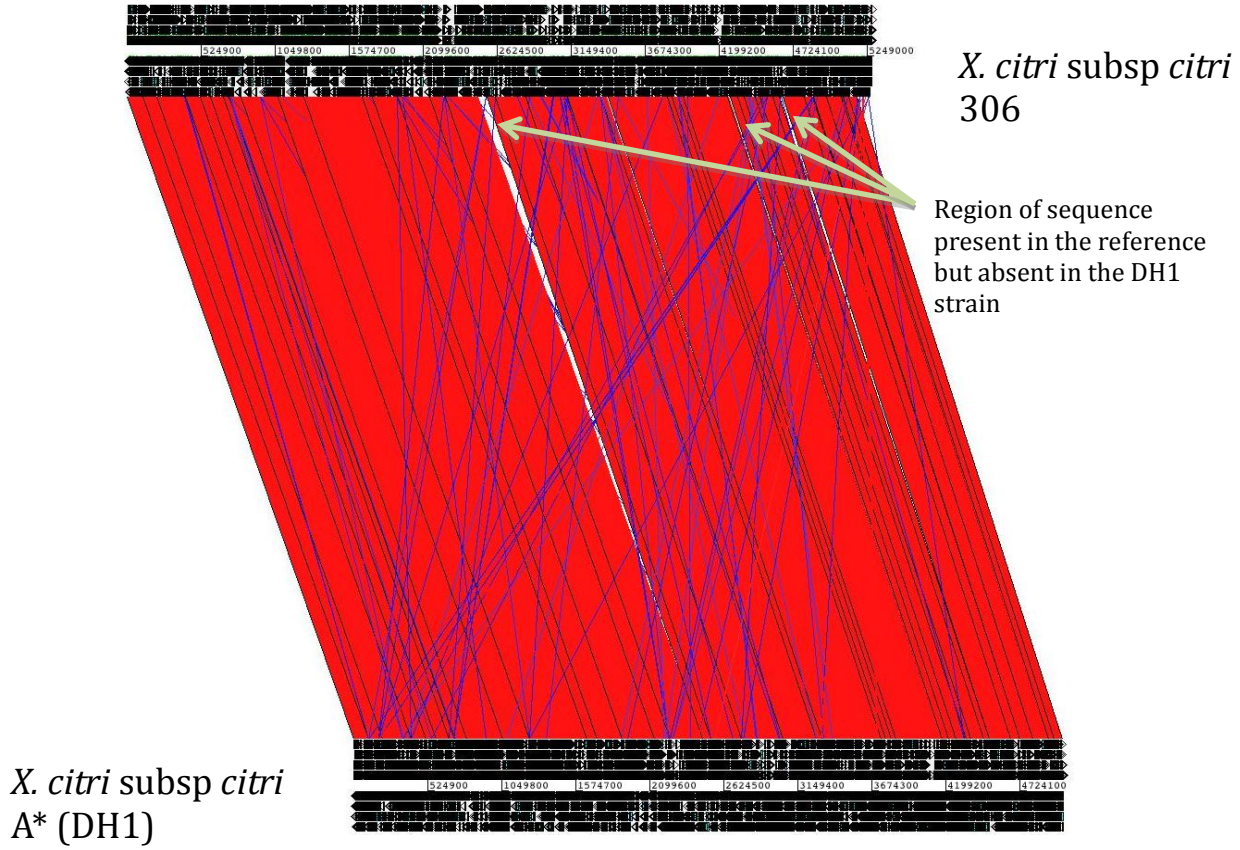


Figure 9c: Comparison showing the alignment of the reference *X. citri* subsp *citri* 306 strain with *X. citri* subsp. *citri* A* (isolate DH1). Regions shown in white represent regions missing in the DH1 strain compared to the reference and blue lines represent areas translocated compared to the reference.

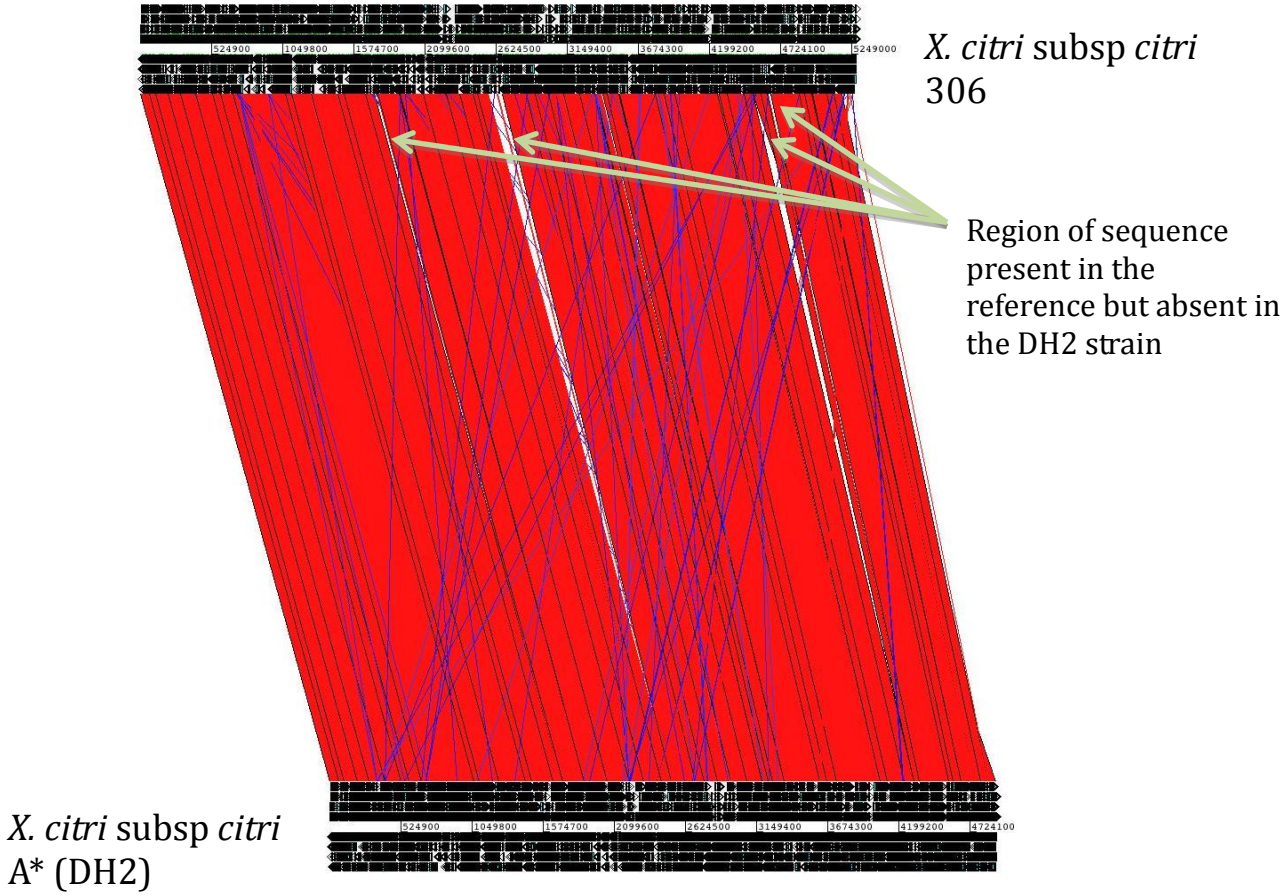


Figure 9d: Comparison showing the alignment of the reference *X. citri* subsp *citri* strain 306 with the *X. citri* subsp. *citri* A* (isolate DH2). Regions shown in white represent regions missing in the DH2 strain compared to the reference and blue lines represent areas translocated compared to the reference.

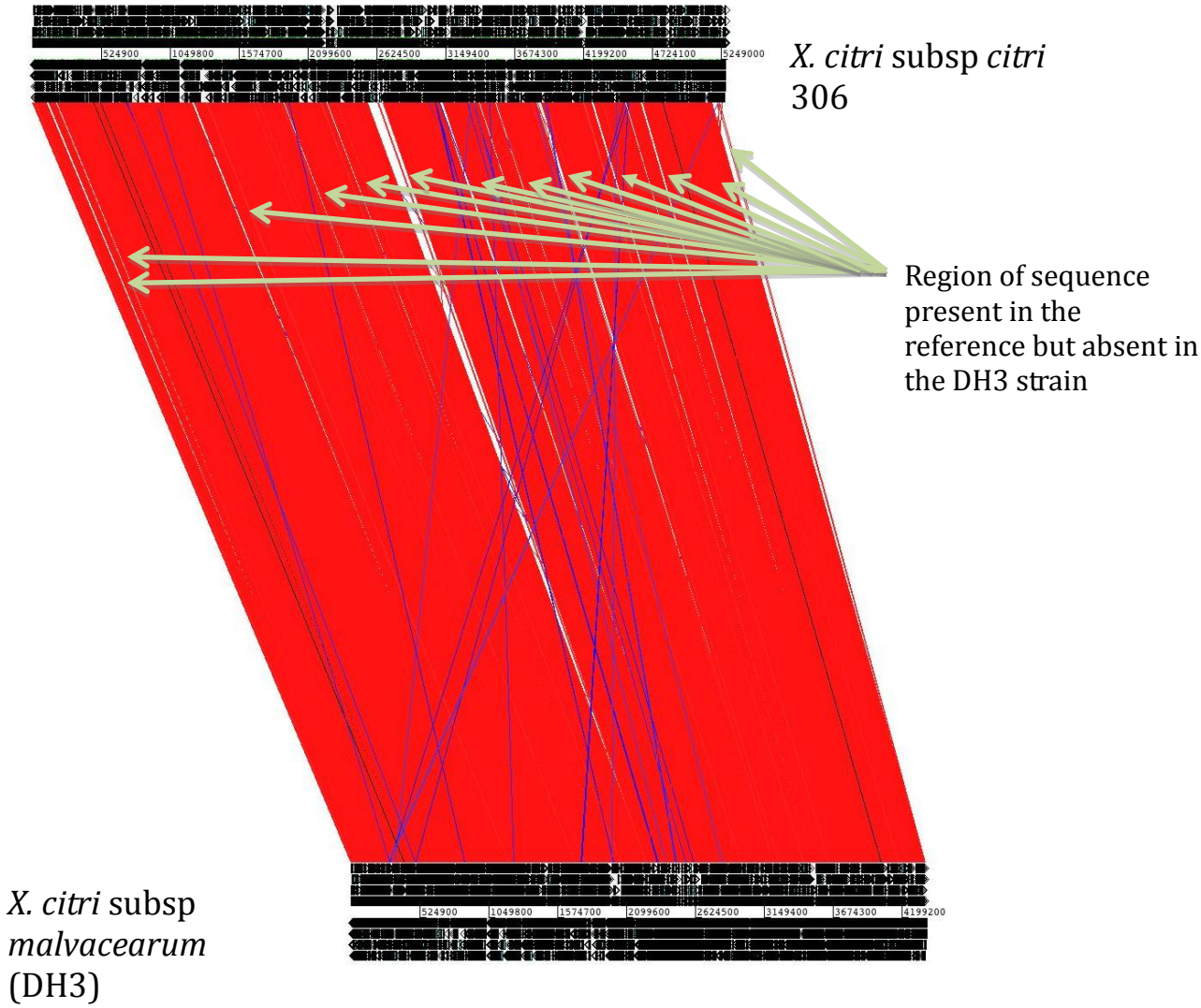


Figure 9e: Comparison showing the alignment of the reference *X. citri* subsp *citri* strain 306 with *X. citri* subsp. *malvacearum* (isolate DH3). Regions shown in white represent regions missing in the DH3 strain compared to the reference and blue lines represent areas translocated compared to the reference.

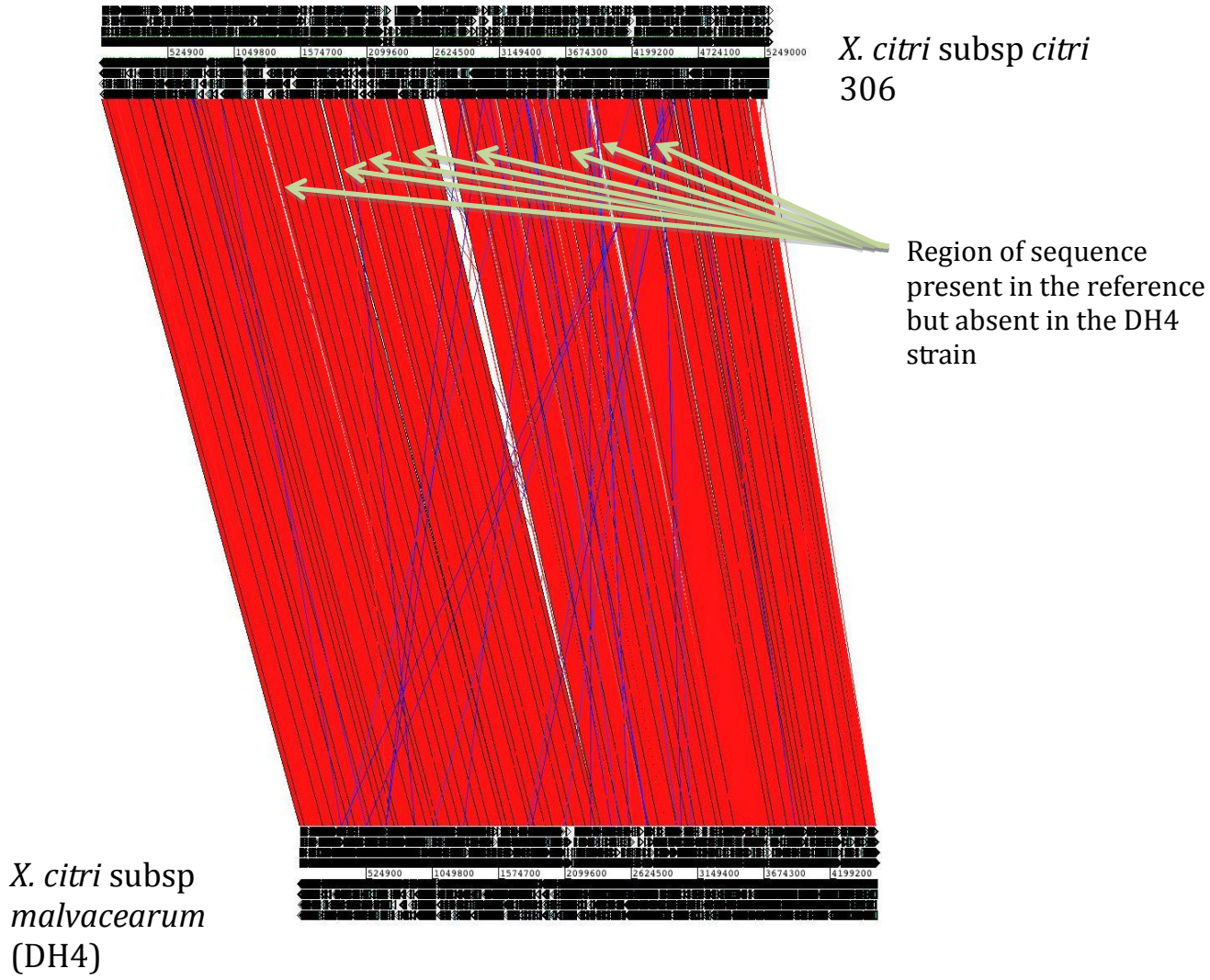


Figure 9f: Comparison showing the alignment of the reference *X. citri* subsp *citri* strain 306 with *X. citri* subsp. *malvacearum* (isolate DH4). Regions shown in white represent regions missing in the DH4 strain compared to the reference and blue lines represent areas translocated compared to the reference.

3.3B – Metabolomics

The initial study within this component developed NMR spectral data from metabolite extracts from six replicates of the six pathovars within the species *Xanthomonas campestris,* these being *X. campestris* pv *armoraciae*, *aberrans, barbarae, campestris, incanae* and *raphani*. In this proof of concept study, NMR and LCMS metabolomics showed good discrimination was achieved for four pathovars, suggesting that the approach has promise.

To develop the concept further, glasshouse grown seedlings of broccoli, cauliflower and cabbage were inoculated in two independent experiments with different isolates of *Xanthomonas campestris* pv *campestris* that had been derived originally different species of host brassica plants, and also of *X. campestris pvs incanae* and *raphani*. Symptom expression conformed to expectations, with severe symptoms on plants inoculated with pathovars derived from cabbage (which showed the most severe response), broccoli and cauliflower and none on plants inoculated with pathovars derived from non-hosts, or negative controls. The metabolite profiles of the whole plants were analysed by NMR and LCMS after seven and 14 days. Subsequent analysis of LCMS data showed separation between the different samples, the first demonstration of this kind in a plant system and we consider that we have achieved the proof of concept for this potential application of metabolomics. In practise, time constraints and quarantine concerns prompted the team to discontinue plant responses to bacterial infection, and to concentrate in stead on the ability of metabolomics to differentiate between bacterial taxa.

A rigorous analysis was conducted to assess the metabolites expressed in replicated cultures of selected bacterial isolates, focussing on the same suite of isolates as used in the proteomics section. This study concentrated on *X. citri* subsp. *citri* (eight each of pathotypes *A, A** and *Aw*), six *X. citri* subsp. *malvacearum* and two *X. alfalfae pv citrumelo* (Table 1 and 2) and their analysis by NMR. A typical spectrum of a cell pellet extract is shown in Figure 10.

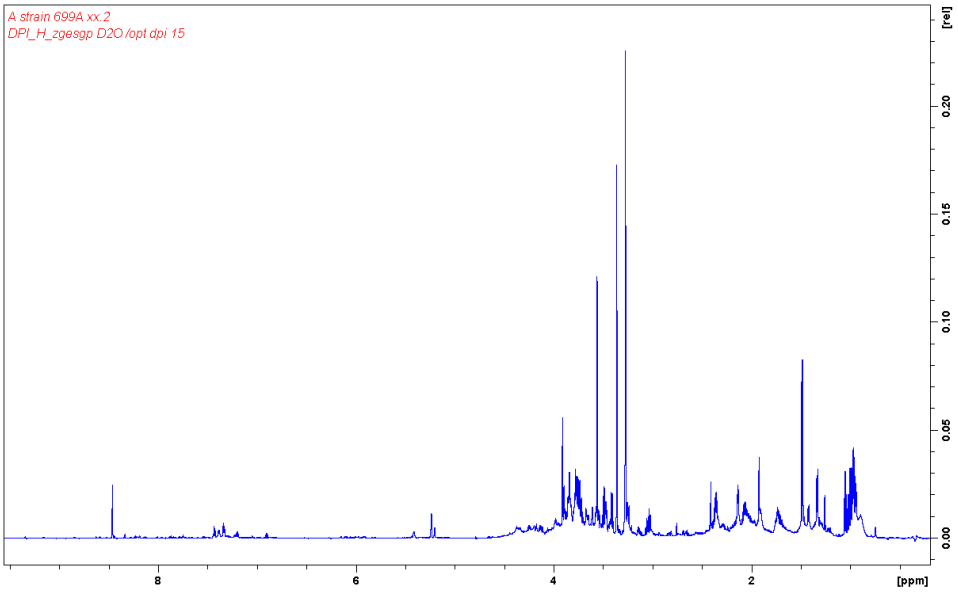


Figure 10: Typical NMR spectrum of the cell pellet extract

Visualisation of the data with each class represented as its mean (Figure 11) revealed that although there were no obvious resonances unique to a particular pathovar there were differences in concentration of various metabolites between the pathovars. The most obvious differences occurred for betaine, glucose and certain aromatics (Figure 12).

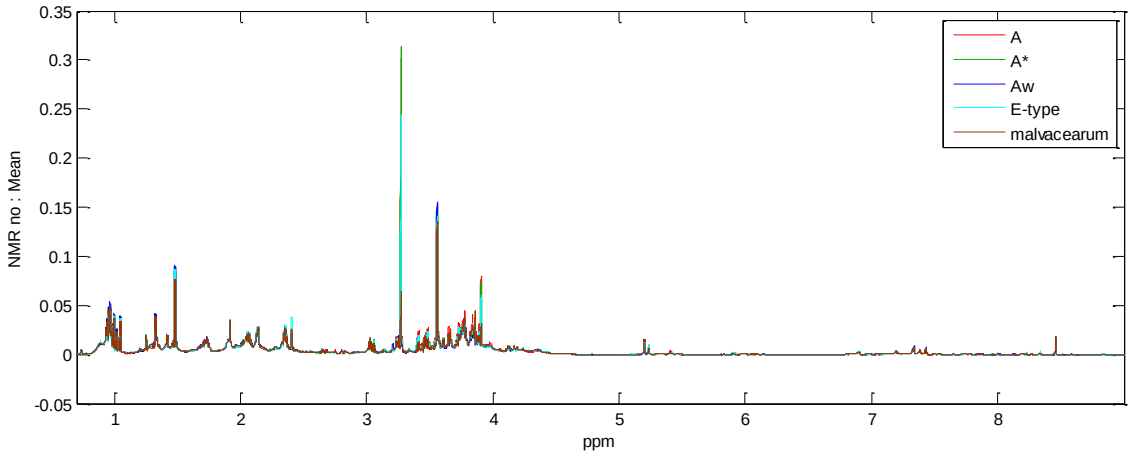


Figure 11: Mean NMR spectra of the different pathovars.

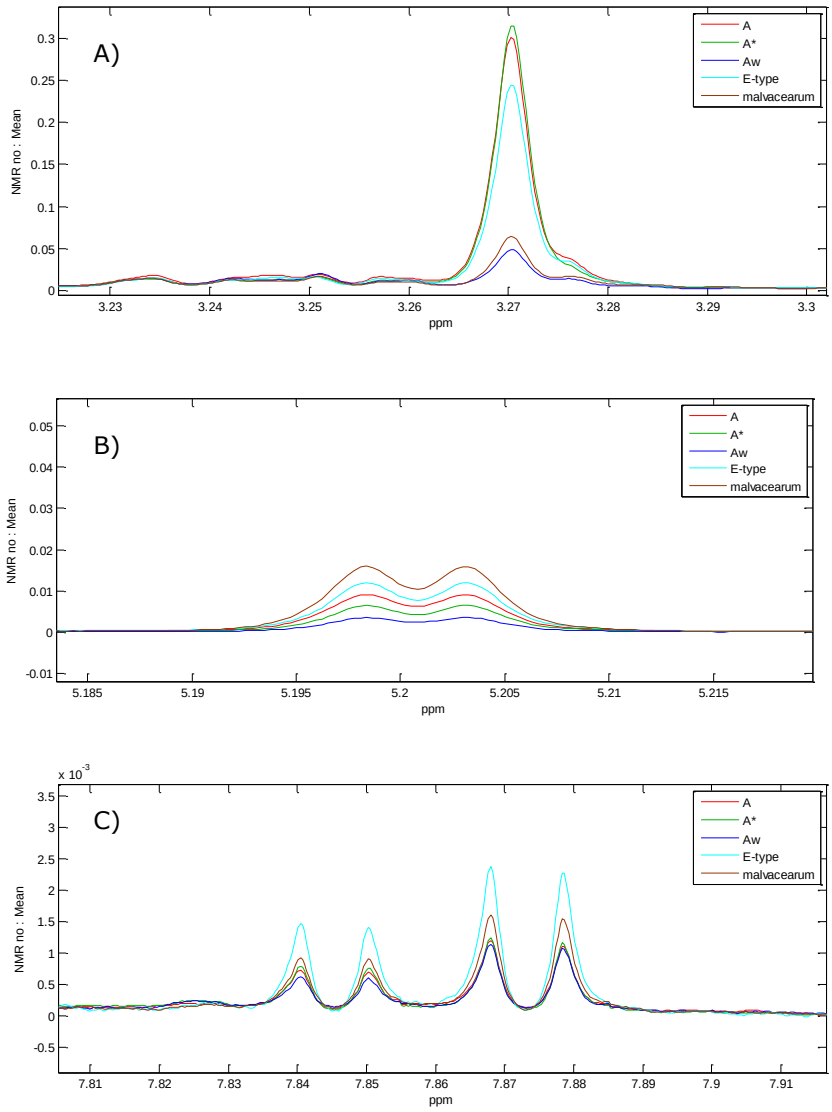


Figure 12: Certain metabolites are differentially expressed. A) betaine, B) glucose C) unknown aromatic metabolite.

Principal component analysis revealed that the *X. citri* subsp. *citri* Aw pathotypes and *X. citri* subsp. *malvacearum* clustered together and away from *X. citri* subsp. *citri* pathotypes A and A* and *X. alfalfae pv citrumelo* (E-type) (Figure 13). The data also shows that there is some clustering due to isolate as well.

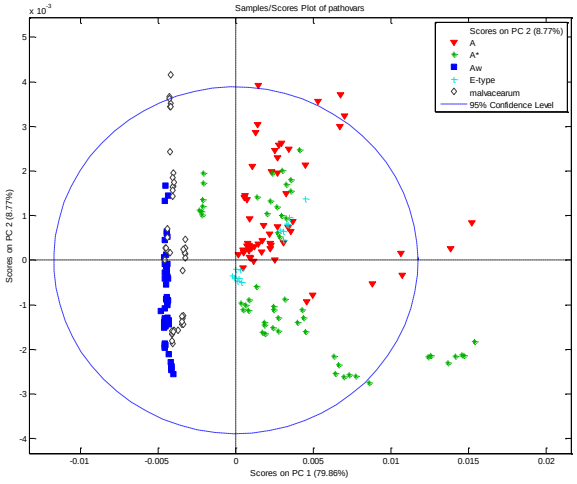


Figure 13: PCA of pathovar NMR data.

Principal Components Analysis Model
Developed 25-Nov-2011 09:49:43.95
Author: sr60@EDA101724
X-block: pathovars 222 by 27020
(sr60@EDA101724@20110609T132431.53189196
m:20110610121930.500)
Included: [1-222] [1-27020]
Included (in axis units): [n/a] [0.700593-8.99997]
Preprocessing: Normalize (1-Norm, Area = 1), Baseline
(Weighted Least Squares), Mean Center
Num. PCs: 6
Algorithm: SVD
Percent Variance Captured by PCA Model

Principal Component Number	Eigenvalue Cov(X) This PC	% Variance Captured This PC	% Variance Captured Total
1	2.25e-005	79.86	79.86
2	2.48e-006	8.77	88.63
3	7.88e-007	2.79	91.42
4	4.66e-007	1.65	93.08
5	3.77e-007	1.34	94.41
6	3.19e-007	1.13	95.54

Analysis of the loadings plots showed the separation on PC1 is due mainly to betaine (Figure 14). Betaine is higher in the *X. citri* subsp. *citri* pathotypes A, A* and *X. alfalfae pv citrumelo* isolates compared to the *X. citri* subsp. *citri* Aw and *X. citri* subsp. *malvacearum* isolates.

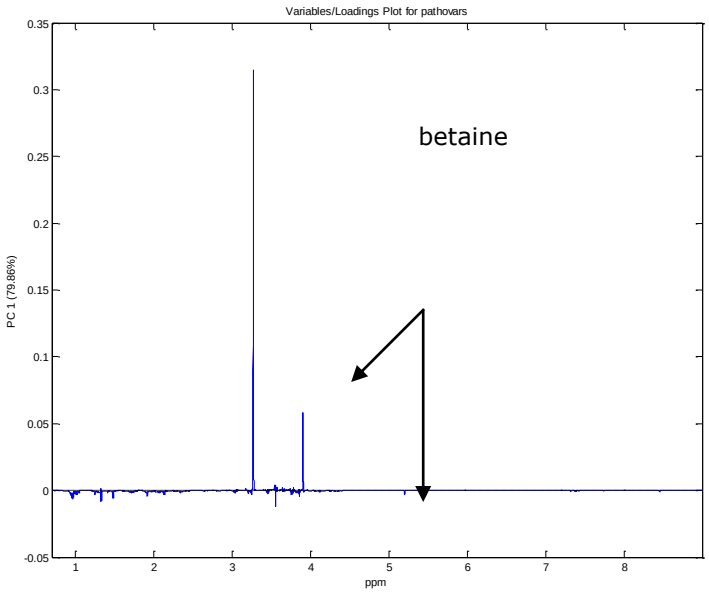
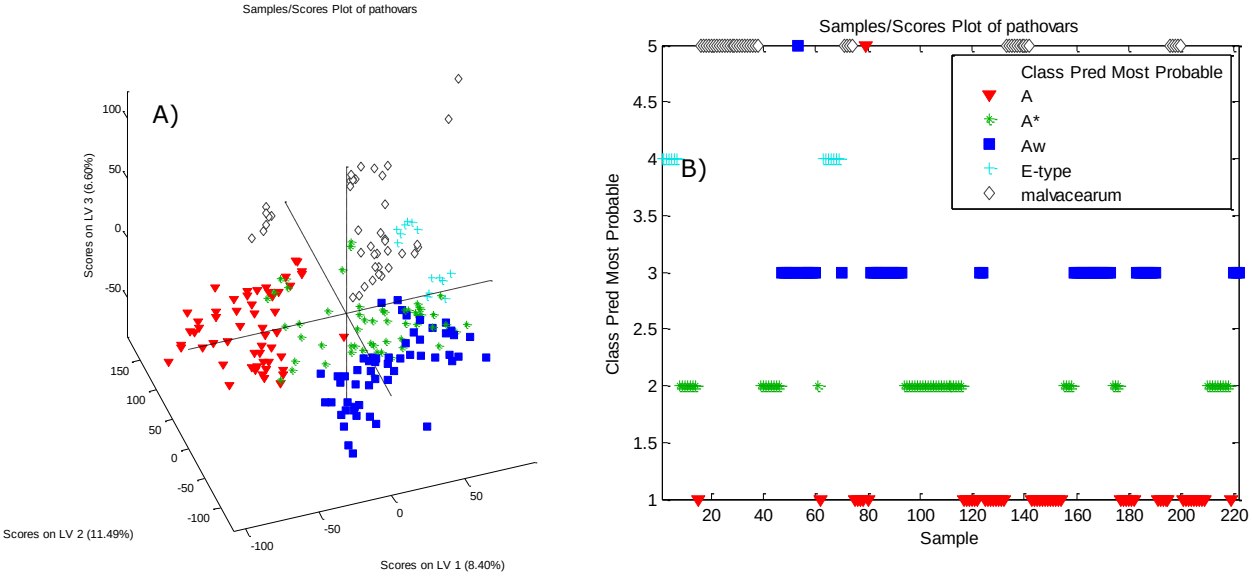


Figure 14: Loadings plot for PCA PC1. Resonances due to betaine are the major influence on PC1.

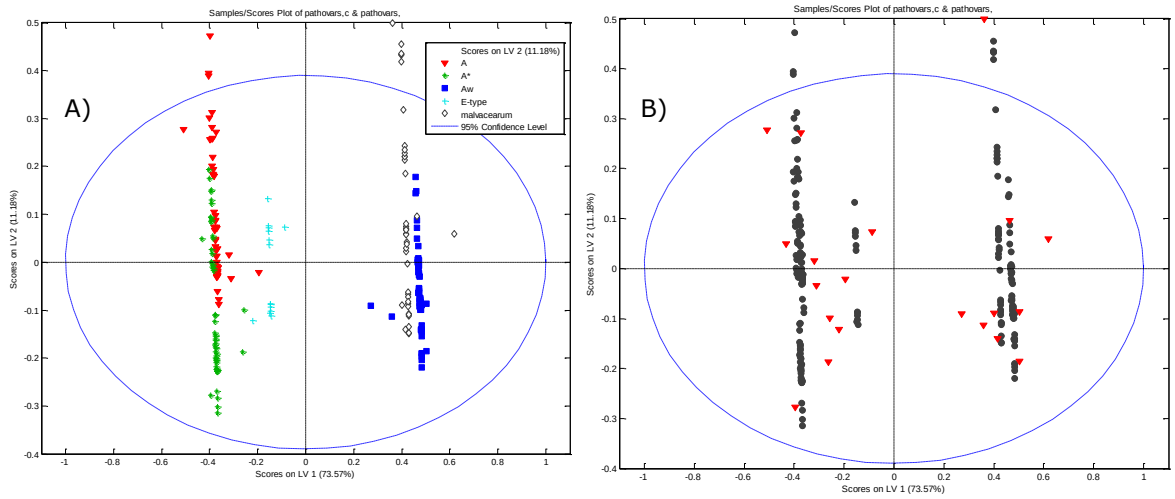
To test the predictive power of the NMR data in a PLSDA model the dataset was analysed by PLSDA by the five pathovars. Clustering by pathovars was observed and the class predictions were good with an error rate of 0.5% to 8% with cross validation (Figure 15).



Class. Err (CV): 0.06 0.09 0.03 0.005 0.06

Figure 15: PLDA model for pathovars A) scores plot and B) class predictions.

To test whether the method could possibly to predict the identity of unknown samples, the original data set was split into a 'Model' group and a 'Test' group. The initial PLSDA model was generated using the model or calibration data set and then the test group was loaded in to determine if the correct class for the test pathovars could be determined (Figure 16 and Appendix 2.1).



Class. Err (Pred): 0.13 0.225 0.03 0.31 0.03

Figure 16: PLDA model for pathovars A) scores plot by class and B) scores plot by calibration or test sample.

Examination of the loadings plots revealed that again betaine had a major influence on the grouping (Figure 17).

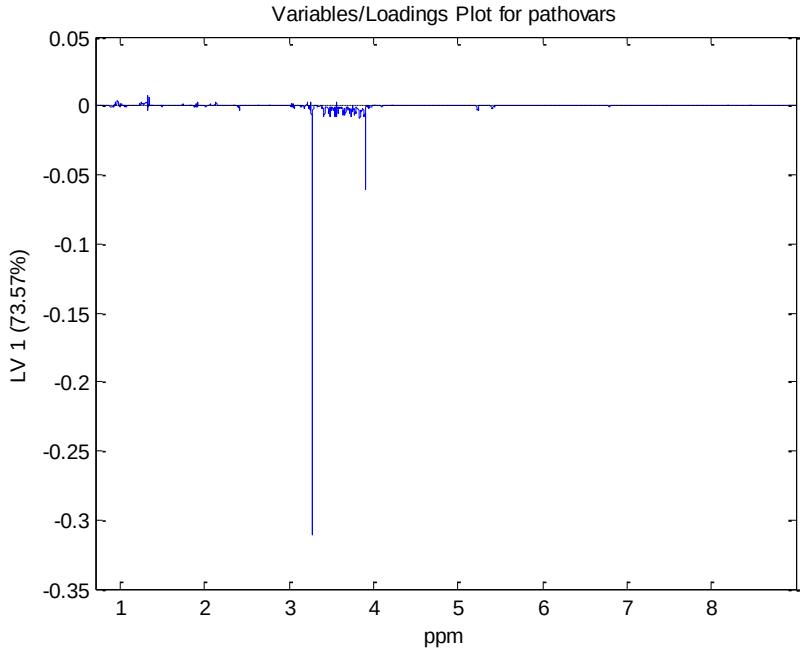


Figure 17: PLDA model loadings plot for level 1.

The class error of prediction varied between 3% and 30% and analysis of the most probable class for showed there were errors in both the calibration and test sets (Figure 18).

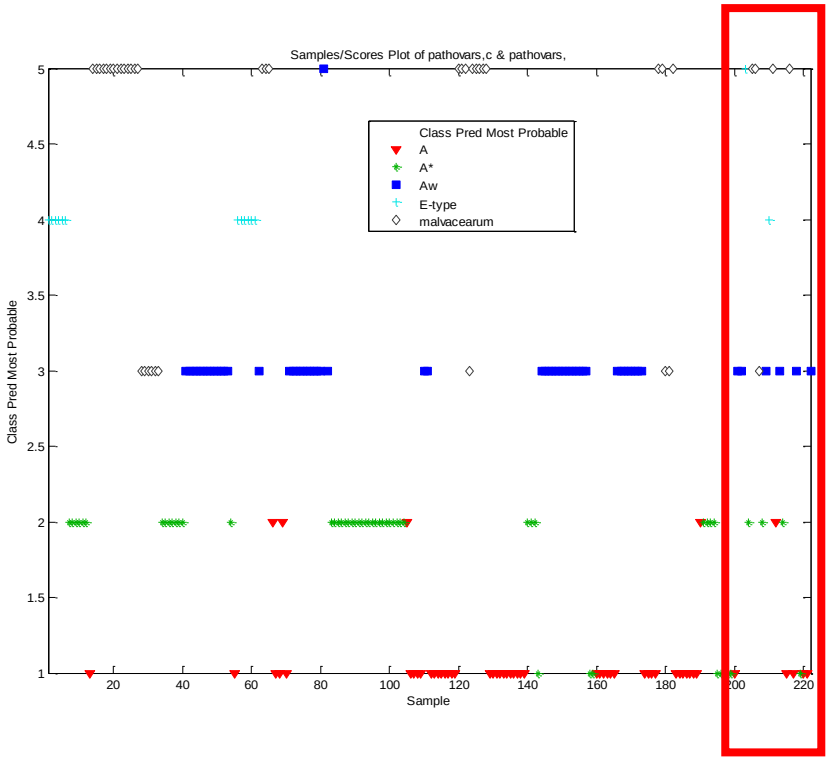
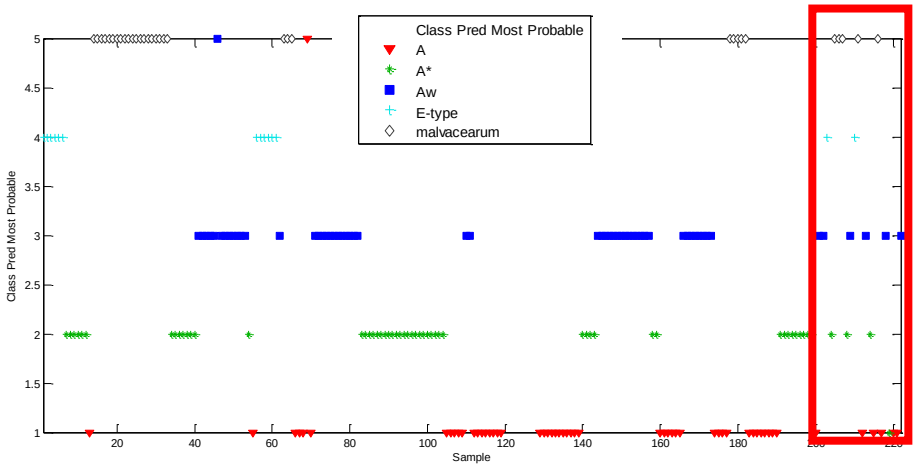


Figure 18: PLDA class predictions. Test set are in the red box.

The above model was constructed using mean centering (see Appendix 2.2 for detail) and it was found that a much more robust model could be constructed using autoscaling in the preprocessing step (Figure 19, Appendix 2.3 for detail). Error of prediction ranged from 0% to 19%.



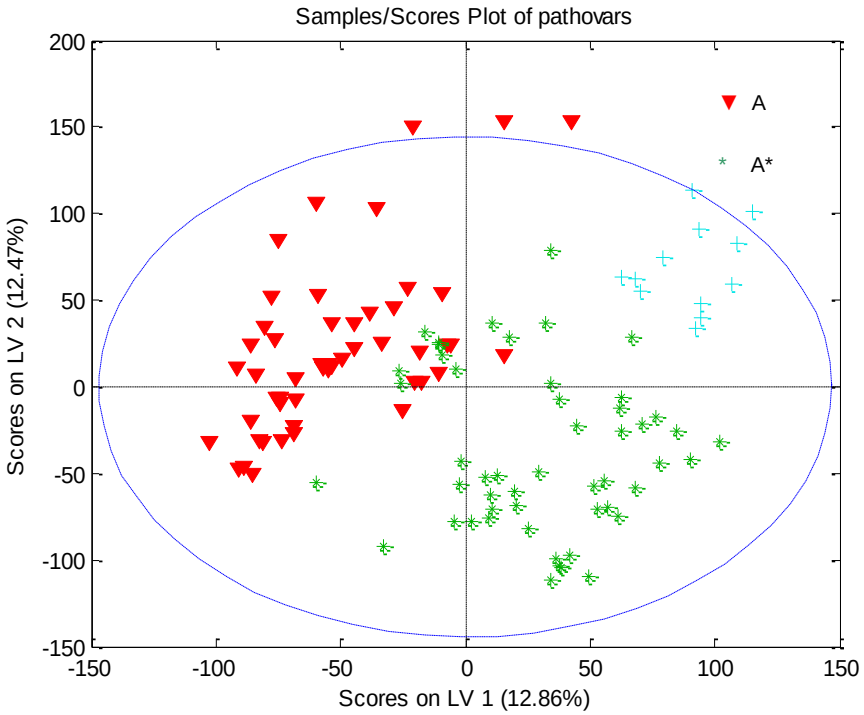
Class. Err (Pred): 0.07 0.19 0 0 0.07

Figure 19: PLDA class predictions. Test set are in the red box.

Although improved it was thought that by splitting the pathovars into two groups (*X. citri* subsp. *citri* pathotypes A and A* and *X. alfalfae* pv *citrumelonis* isolates – "group 1"), and *X. citri* subsp. *citri* pathotype Aw and *X. citri* subsp. *malvacearum* isolates – "group 2") that the predictive power of the model would improve. The group 1 isolates tended to cluster together in PCA and so in a mixed model it is harder to get good prediction between these pathovars. New datasets for both calibration (model building) and testing were created by separating the groups.

Focus on group 1 (*X. citri* subsp. *citri* pathotypes A, A* and *X. alfalfae pv citrumelo)*

PLSDA of all the isolates in group 1 (A, A* and E-Type) data clustered well with class prediction errors of 0% to 2% (Figure 20, Appendix 2.4).



Class. Err (CV): 0.02 0.02 0

Figure 20: PLDA scores plot.

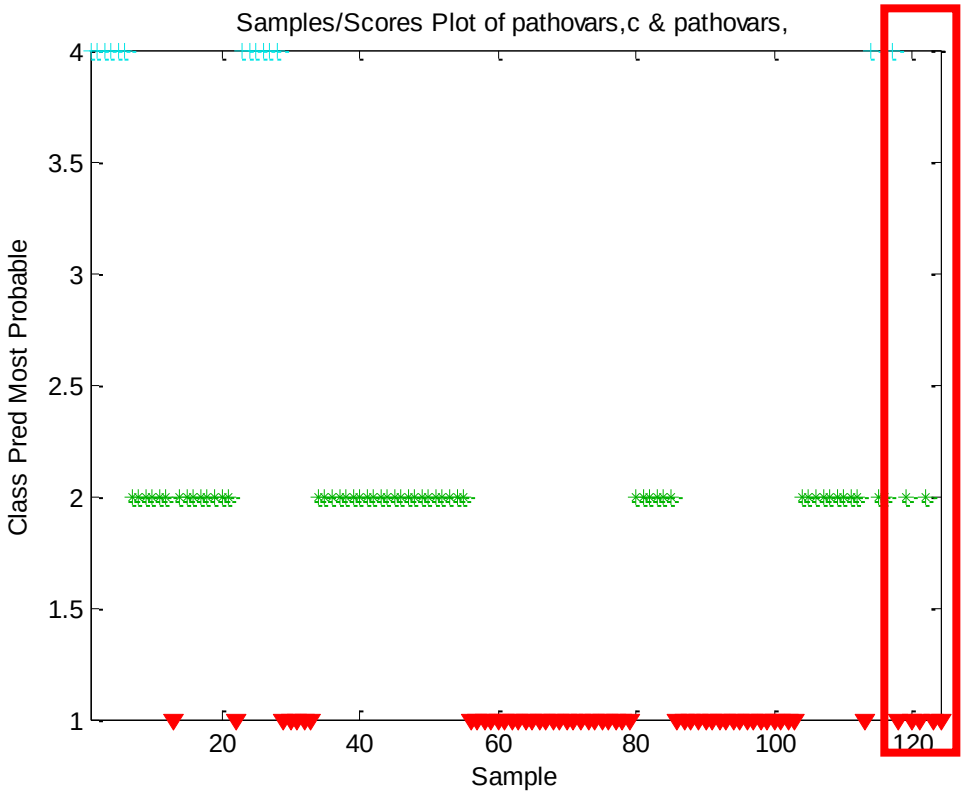


Figure 21: PLSDA class predictions. Test set are in the red box.

Permutation testing was carried out to test the validity of the model and investigate for over fitting. The results of these tests (Appendix 2.4) demonstrate that the model is robust and hence could be used to predict unknowns using the sample preparation and data analysis methodology described here.

Focus on group 2 (*X. citri* subsp. *citri* pathotype Aw and *X. citri* subsp. *malvacearum*)

Analysis of the PLSDA scores plots for the calibration and test samples (Figure 22, Appendix 2.5) show that excellent prediction of class is possible.

Permutation testing was carried out to test the validity of the model and investigate for over fitting.

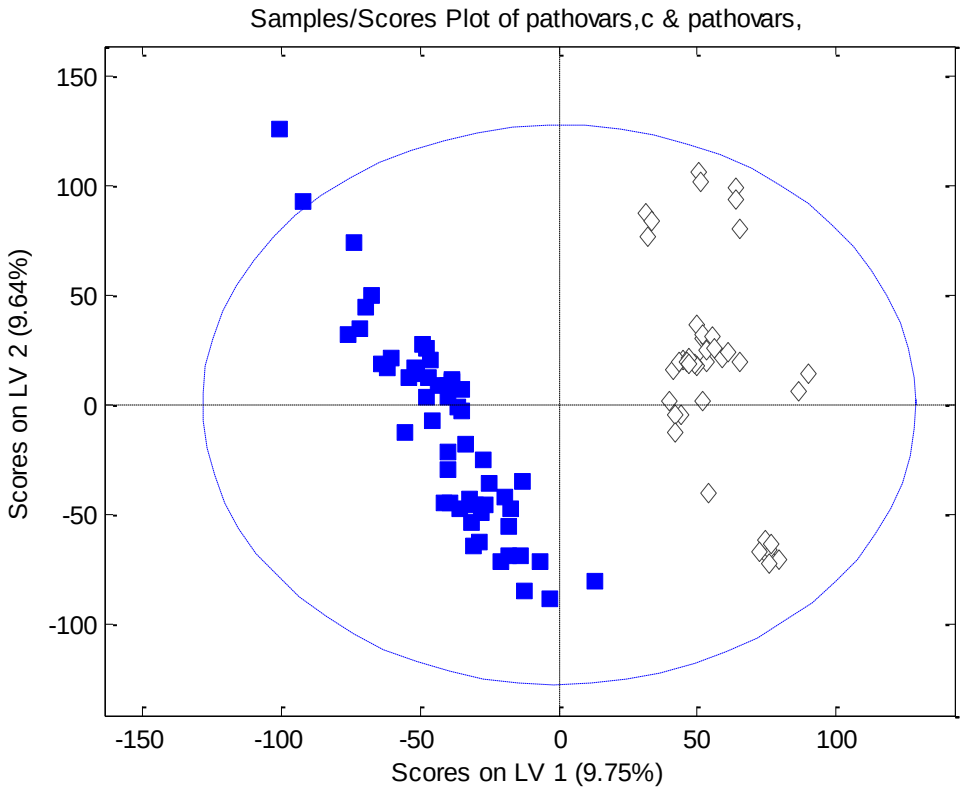


Figure 22: PLDA scores plot.

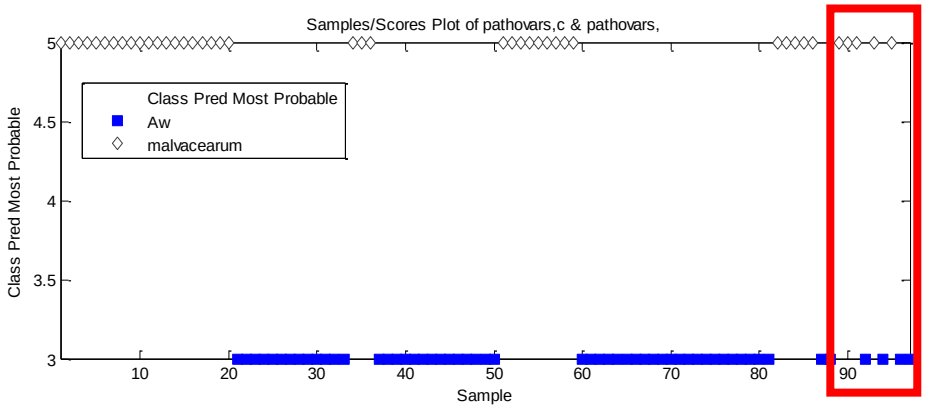


Figure 23: PLDA class predictions. Test set are in the red box.

The results of these tests (Appendix 2.6) demonstrate that the model is robust and hence could be used to predict unknowns using the sample preparation and data analysis methodology described here.

3.4 Conclusions

The proteomic component of this project has been highly successful. The expression of proteins differed sufficiently between operative pairs of different pathovars to allow these differences to be used to effectively guide the development of novel diagnostic assays. The genes encoding those differentially expressed proteins have been identified by reference to whole genome sequences stored in international databanks, and new DNA-based diagnostic assays designed. Using this approach we have developed and validated two novel molecular tests that deliver a higher level of specificity than has previously been available. These could for example be multiplexed to provide a single-step reaction to identify a Xanthomonad found on citrus in a cotton-growing area as either *X. citri* subsp. *citri*, X*. citri* subsp*. malvacearum* or *X. alfalfae pv citrumelonis*.

Notwithstanding these positive outcomes, the formulation of these novel diagnostics was not a simple process. Proteomic analyses were technically demanding and required mastery of many new technologies, from hardware to software. Even once the differentially expressed proteins had been identified and the genes encoding them determined, it remained a significant task to filter through the options to find molecules that were reliable in their provision of sufficient, stable discriminatory power. We evaluated more than thirty different primer pairs in PCR to successfully design two pairs that gave the specificity and reliability required. This means that even though we had selected proteins spots that were absent in one pathovar and present in another, the lack of protein expression observed was not due to changes in gene sequence (leading to premature termination or altered sequence of the encoded protein), but must in stead have been due to changes in the regulatory regions of those genes.

Both of the successful diagnostic assays target outer member proteins (OMPs), which are required in bacteria for normal growth and resistance to environmental stressors and play a role in pathogenicity. Within the cell outer membrane of Xanthomonads, there are three dominant proteins - OmpW, MopB and SodM, the first two of which were identified by our proteomics analyses. The precise function of OmpW in Xanthomonads is not yet known but in *Escherichia coli* it has been demonstrated to be involved the maintenance of membrane integrity during oxidative stress (Asakura *et al*. 2008) and in *Vibrio alginolyticus* (Xu *et al*. 2004) and *Photobacterium damsela* (Wu *et al*. 2006) it has a role as an osmosensitive protein. MopB is the most abundant outer membrane protein and mopB, the gene that encodes it, is highly conserved amongst Xanthomonads(Chen *et al*. 2010). Mutation studies indicate MopB is involved in the production of xanthan, possibly regulating the expression of the operon for xanthan gum synthesis. Perhaps more excitingly, mutation has also shown surface deformation, increased sensitivity to environmental stress (temperature, pH), reduced adhesion and other defects in pathogenesis (Chen *et al*. 2010). This clearly indicates that the protein we have identified using proteomics has an important in pathogenicity, confirming the suitability of the gene for this OMP as a target in diagnostic assays.

During this project the team has also generated other nucleotide sequence data that could now also be used to develop more efficient and rapid diagnostics to separate the pathotypes within *X. citri* subsp. *citri*. For example, more subtle difference between *X. citri* subsp. *citri* pathotype A and the pathotypes A* and Aw that have significantly restricted pathogenicity, could be amenable to more sophisticated methods such as detection by real time PCR or the use of high resolution melt analysis. Molecular diagnostics are not currently available to provide differentiation at this level.

A limitation of the proteomic approach is its dependence on the availability of whole genome sequence data to identify proteins that are differentially expressed. Only one sequence for *X. citri* subsp*. citri* was available to us, so our analyses had to concentrate on proteins that were expressed in *X. citri* subsp*. citri* and absent in other pathovars. We have examples of proteins that are expressed in *X. citri* subsp*. malvacearum* but not *X. citri* subsp*. citri* but had no reference genome for *X. citri* subsp*. malvacearum* against which we could map those proteins. As a result, in the later stages of this project the team has also moved to undertake some whole genome sequencing to allow us to mine these data further in future. These genomes have undergone primarily annotation and we have located the positions of the genes (*mopB* and *ompP1*) targeted by our novel diagnostic assays. They also provide a foundation for further metabolomics work in these species, for example to investigate whether sequence variation determines the differences in betaine glycine production that may influence the host range. This work provides a basis for, and will continue in, Program 2's exemplar project in the PBCRC.

An extensive 'library' of 2D gel profiles was generated from the hydrophobic protein fractions during the course of this work. Once the skills were honed, our operator was able to identify certain patterns and motifs that distinguished isolates pathogenic to citrus, from bacteria pathogenic to other hosts. In theory this approach might also offer a potential new approach to determining the identity of bacteria isolated from plants in an incursion. In reality, however, the practicality of running the 2D gels may preclude the general usefulness of proteomic profiling as the actual diagnostic test.

The proteins that we have determined are differentially expressed, and/or peptides derived from them, could also be used directly as antigens to generate novel antibodies by creating a new phage-displayed scFv library. Novel antibodies would then be adapted, singly or in combination, to laboratory or field situations (lateral flow devices, 'dipsticks' etc) for routine diagnostics. The one-off task of creating the new library provides a value-added resource that could subsequently be 'mined' by projects screening for antibodies of other useful specificities. This might for example deliver antibodies to discriminate other pathovars within the same species, to detect all members of the species or to detecting bacteria from other genera that attack the same host.

The metabolomics analyses have also shown success in the ability to distinguish and predict the identity of different functional and physiological ecotypes of bacterial pathogens. Metabolite profiles from cell pellets have shown that the pathovars of *Xanthomonas citri* can be distinguished by chemometric modelling. The distinguishing metabolites are not unique to any pathovars but are distinguishing due to the different amounts present in each cell culture. This likely represents underlying metabolic differences between the pathovars and so represents a new way to classify these pathovars. There are some visible differences between isolates of the same pathovars but this did not interfere with over all prediction in this case. Metabolomic methods can therefore classify pathogens on the basis of their expressed physiology, even when it is not possible to infer a direct mechanistic link to specific virulence factors.

A key molecule identified in this way is glycine betaine (N,N,N-trimethylglycine), an osmoprotectant that allows bacteria to survive high-osmolarity environments. Whilst its biosynthetic pathway is conserved within bacteria, there is some divergence in the enzymes involved. In *X. citri* subsp. *citri* the biosynthesis of glycine betaine is from choline in a two-step pathway with betaine aldeyhde as the intermediate, involving three key genes *betA*, *betB* and *betT* (these genes have been mapped to our whole genome sequences). The variations in the concentration of glycine betaine observed between these *X.citri* subsp. *citri* and *X. citri* subsp. *malvacearum* may reflect their differential ability to survive within host plants with different osmolority.

Preliminary analysis of infected versus uninfected plants also suggested that spectral differences can be observed, providing chemometric models of infection that could with validation be used to rapidly identify infected plants. It may also lead to the identification of specific biomarkers of infection. Such biomarkers may allow detection of infection before symptoms become evident, through the development of diagnostic NMR tests that detect spectral differences for high throughput analysis. These biomarkers could also be assessed for suitability for use in a simpler assay of infection (e.g. incorporation into microsensors such as molecularly imprinted polymers, ELISA kit etc) that could be deployed by field staff including quarantine officers and state regulatory officers.

This project has provided the first use of both proteomics and metabolomics to plant pathogenic bacteria. Both systems have proven to be technically demanding, but have delivered very specific and very robust new methodologies to diagnose organisms to levels not previously possible. Of the two methods, metabolomics provided some early trouble in the preparation of extracts but once this was overcome, proficient operators needed little time to generate results. The drawback however was its inability to identify amongst our samples, molecules that were unique to any particular pathovar – metabolomics only identified molecules that were expressed at different levels of expression. Proteomic extractions were simple but the techniques to display and analyse the protein profiles were complicated and time-consuming to master. Once differentially expressed proteins were evident, there followed a multistage process to determine their identity and further computational analyses to identify the genes

that encode those proteins. The evaluation of different DNA-based assays to find those discriminating between the relevant pathovars also involved many 'blind alleys'. That said, the proteomic analyses have delivered robust new biomarkers for use in novel diagnostics. In short, both proteomics and metabolomics have strong potential to deliver further diagnostic capacity in future, and their use as a component of other projects should be supported. Further research is warranted.

3.5 Achievements against objectives

Evaluation of new functional platforms for use as tools to discover biomarkers that effectively differentiate closely related pathovars within specific organisms used as models.

- Completed. Platforms technically demanding but results indicate once optimised they have capacity to differentiate organisms to new levels of specificity.

Identification of novel diagnostic targets for national and international validation, which can be fed into projects evaluating different delivery platforms.

- Completed. Two new methods developed and available for uptake into other delivery platforms.

Increased national incursion response capability through improved specificity and turnaround time of diagnostic tests.

- Completed. New methods satisfy these criteria.

Reduced possibility of misdiagnosis and false positive diagnosis.

- Completed. New methods satisfy these criteria.

Methods provided to improve tracking of bacterial pathovars in incursion.

- Completed. New methods satisfy these criteria.

To foster new national and international scientific partnerships.

- Completed – new partnerships within this project and new partnerships with DOA Thailand. This project has also underpinned the development of one of the PB CRC's exemplar projects, which will involve members of this project team and Prof Jim Stack at KSU.

Linkages to other projects enhancing plant bacteriology capacity through the training of specialist scientists and postgraduate students.

- Completed through:
    - New Bacteriologist at NSW DPI, Dr Toni Chapman. This position would not have been created without this project as evidence of need expressed by industry and the plant biosecurity community.
    - Extensive training of Dr Nuttima Kositcharoenkul, bacteriologist at Dept of Agriculture Bangkok Thailand, through the associated project CRC20093.
    - Honours student DPI Victoria (not funded by CRC but work stemmed directly from preliminary work in metabolomics conducted during this project).

3.6 References

Ansong C, Yoon H, Norbeck AD, Gustin JK, McDermott JE, Mottaz HM, Rue J, Adkins JN, Heffron F and Smith RD. 2008 'Proteomics analysis of the causative agent of Typhoid fever' *Journal Proteomics Research* 7: 546-557.

Asakura H, Kawamoto K, Haishima Y, Igimi S, Yamamoto S, and Makino S. 2008 'Differential expression of the outer membrane protein W (OmpW) stress response in enterohemorrhagic *Escherichia coli* O157:H7 corresponds to the viable but non-culturable state' *Research in Microbiology* 159: 709-717.

Bereswill S, Pahl A, Bellemann P, Zeller W, Geider K. 1992 'Sensitive and species-specific detection of *Erwinia amylovora* by polymerase chain reaction analysis' *Appl. Environ. Microbiol* 58:3522–3526.

Berg T, Tesoriero LA and Hailstones DL. 2005 'PCR-based detection of *Xanthomonas campestris* pathovars in Brassica seed' *Plant Pathology* 54: 416–427.

Bundy JG, Willey TL, Castell RS, Ellar DJ and Brindle KM. 2005 'Discrimination of pathogenic clinical isolates and laboratory strains of Bacillus cereus by NMR-based metabolomic profiling' *FEMS Microbiology Letters* 242: 127-136.

Buttner D, Noel L, Thieme F and Bonas U. 2003 'Genomic approaches in *Xanthomonas campestris* pv *vesicatoria* allow fishing for virulence genes' *Journal of Biotechnology* 106: 203-214

Chen YY, Wu CH, Lin JW, Weng SF and Tseng YH. 2010 'Mutation of the gene encoding a major outer-membrane protein in *Xanthomonas campestris pv. campestris* causes pleiotropic effects, including loss of pathogenicity' *Microbiology* 156: 2842-2854.

Cordwell SJ, Len ACL, Touma RG, Scott NE, Falconer L, Jones D, Connolly A, Crossett B and Djordjevic SP. 2008 ' Identification of membrane-associated proteins from *Campylobacter jejuni* strains using complementary proteomics technologies' *Proteomics* 8: 122-139.

Cubero J and Graham J H. 2002 'Genetic Relationship among Worldwide Strains of Xanthomonas Causing Canker in Citrus Species and Design of New Primers for Their Identification by PCR' *Applied and Environmental Microbiology* 68(3): 1257-1264.

da Silva ACR, Ferro JA, Reinach FC, Farah CS, Furlan LR, Quaggio RB, Monteiro-Vitorello CB, Van Sluys MA, Almeida NF Jr., Alves LMC, do Amaral AM, Bertolini MC, Camargo LEA, Camarotte G, Cannavan F, Cardozo J, Chambergo F, Ciapina LP, Cicarelli RMB, Coutinho LL, Cursino-Santos JR, El-Dorry H, Faria JB, Ferreira AJS, Ferreira RCC, Ferro MIT, Formighieri EF, Franco MC, Greggio CC, Gruber A, Katsuyama AM, Kishi LT, Leite RP Jr., Lemos EGM, Lemos MVF, Locali EC, Machado MA, Madeira AMBN, Martinez-Rossi NM, Martins EC, Meidanis J, Menck CFM, Miyaki CY, Moon DH, Moreira LM, Novo MTM, Okura VK, Oliveira MC, Oliveira VR, Pereira HA Jr., Rossi A, Sena JAD, Silva C, de Souza RF, Spinola LAF, Takita MA, Tamura RE, Teixeira EC, Tezza RID, Trindade dos Santos M, Truffi D, Tsai SM, White FF, Setubal JC and Kitajima JP. 2002 'Comparison of the genomes of two Xanthomonas pathogens with differing host specificities' *Nature* 417(6887): 459-463.

Gerlach RG and Hensel M. 2007 'Protein secretion systems and adhesins: the molecular armory of Gram-negative pathogens' *International Journal of Medical Microbiology* 297: 401-415

Hancock, REW, Egli C, Benz R, Siehnel RJ. 1992 'Overexpression in *Escherichia coli* and Functional Analysis of a Novel PP- selective Porin, OprO, from *Pseudomonas aeruginosa' Journal of Bacteriology* 174: 471-476.

Koeuth T, Versalovic J, and Lupski JR. 1995 'Differential subsequence conservation of interspersed repetitive *Streptococcus pneumoniae* BOX elements in diverse bacteria' *Genome Research* 5: 408-418.

Mongkolsuk S, Whangsuk W, Vattanaviboon P, Loprasert S, Fuangthong M. 2000 'A Xanthomonas alkyl hydroperoxide reductase subunit C (ahpC) mutant showed an altered peroxide stress response and complex regulation of the compensatory response of peroxide detoxification enzymes' *Journal of Bacteriology* 182: 6845-6849.

Nandi B, Nandy RK, Sarkar A. 2005 'Structural features, properties and regulation of the outer-membrane protein W (OmpW) of *Vibrio cholerae*' *Microbiology* 151: 2975-2986.

Perera MR, Taylor SP, Vanstone VA and Jones MGK. 2009 'Protein biomarkers to distinguish oat and lucerne races of the stem nematode, Ditylenchus dipsaci, with quarantine significance for Western Australia' *Nematology* 11: 555-563

Qian W, Jia Y, Ren SX, He YQ, Feng JX, Lu LF, Sun Q, Ying G, Tang DJ, Tang H, Wu W, Hao P, Wang L, Jiang BL, Zeng S, Gu WY, Lu G, Rong L, Tian Y, Yao Z, Fu G, Chen B, Fang R, Qiang B, Chen Z, Zhao GP, Tang JL and He C. 2005 'Comparative and functional genomic analyses of the pathogenicity of phytopathogen *Xanthomonas campestris* pv. *campestris' Genome Research* 15: 757-767.

Ratcliffe L, Mian S, Slater K, King H, Napolitano M, Aucoin D and Mobasheri A. 2009 'Proteomic identification and profiling of canine lymphoma patients' Veterinary and Comparative Oncology 7: 92-105.

Richard SB, Ferrer JL, Bowman ME, Lillo AM, Tetzlaff CN, Cane DE, and Noel JP. 2002 'Structure and mechanism of 2-C-Methyl-D-erythritol 2,4-Cyclodiphosphate synthase' *The Journal of Biological Chemistry* 277: 8667-8672.

Rodoni B, Merriman P, McKirdy S, Wittwer G. 2006. 'Costs associated with fire blight incursion management and predicted costs of future incursions'. 10th International workshop on fire blight. Bologna, Italy, July 5-9, 2004. *Acta Horticulturae.* 704:55-61pp

Schadd NW, Postnikova E, Lacy GH, Sechler A, Agarkova I, Stromberg PE, Stromberg VK, and Vidaver AK. 2005 'Reclassification of *Xanthomonas campestris pv. citri* (ex Hasse 1915) Dye 1978 forms A/B/C/D, and E as X*. smithii subsp. citri* (ex Hasse) sp. nov. nom. rev. comb. nov., *X. fuscans* subsp. *aurantifolii* (ex Gabriel 1989) sp. nov. nom. rev. comb. nov., and X*. alfalfae* subsp*. citrumelo* (ex Riker and Jones) Gabriel et al., 1989 sp. nov. nom. rev. comb. nov.; X*. campestris pv malvacearum* (ex Smith 1901) Dye 1978 as X*. smithii subsp. smithii* nov. comb. nov. nom. nov.; *X. campestris pv. alfalfae* (ex Riker and Jones, 1935) Dye 1978 as *X. alfalfae subsp. alfalfae* (ex Riker et al., 1935) sp. nov. nom. rev.; and "var. *fuscans*" of *X. campestris pv. phaseoli* (ex Smith, 1987) Dye 1978 as X*. fuscans subsp. fuscans* sp. nov' *Systematic and Applied Microbiology* 28: 494-518.

Schaad N, Postnikova E, Lacy G, Sechler A, Agarkova IV, Stromberg PE, Stromberg VK, Vidaver PE. 2006 'Emended classification of xanthomonad pathogens on citrus' *Systematic and Applied Microbiology* 29: 690-695.

Shevchenko A, Wilm M, Vorm O and Mann M. 1996 'Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels.' *Anal. Chem.* 68: 850-858.

Shevchenko A, Tomas H, Havlis J, Olsen JV, and Mann M. 2006 'In-gel digestion for mass spectrometric characterization of proteins and proteomes' *Nat. Protoc.* 1: 2856-2860.

Sun X, Stall RE, Jones JB, Cubero J, Gottwald TR, Graham JH, Dixon WN, Schubert TS, Chaloux PH, Stromberg VK, Lacey GH and Sutton BD. 2004. 'Detection and characterization of a new strain of Citrus canker bacteria from Key/Mexican lime and Alemow in South Florida.' *Plant Disease* 88: 1179-1188.

Thieme F, Koebnik R, Bekel T, Berger C, Boch J, Buttner D, Caldana C, Gaigalat L, Goesmann A, Kay S, Kirchner O, Lanz C, Linke B, McHardy AC, Meyer F, Mittenhuber G, Nies DH, Niesbach-Klosgen U, Patschkowski T, Ruckert C, Rupp O, Schneiker S, Schuster SC, Vorholter FJ, Weber E, Puhler A, Bonas U, Bartels D and Kaiser O. 2005 'Insights into genome plasticity and pathogenicity of the plant pathogenic bacterium *Xanthomonas campestris pv. vesicatoria' J. Bacteriol.* 187: 7254-7266.

Ullrich M, Bereswill S, Volksch B, Fritsche W, and Geider K. 1993 ,Molecular Characterization of field isolates of *Pseudomonas syringae* pv. *syringae* differing in coronatine production' *Journal of General Microbiology* 139: 1927-1937.

Vauterin L, Hoste B, Kersters K, and Swings J. 1995 'Reclassification of *Xanthomonas' International Journal of Systematic Bacteriology* 45: 472-489.

Vauterin L, Rademaker J, and Swings J. 2000 'Synopsis on the taxonomy of the genus *Xanthomonas' Phytopathology* 90: 677-82.

Verniere C, Hartung J, Pruvost O, Civerolo EL, Alvarez AM, Maestri P and Luisetti J. 1998 'Characterisation of phenotypically distinct strains of *X. axonopodis* pv*. citri* from Southwest Asia' *European Journal of Plant Pathology* 104:477-487.

Versalovic J, Koeuth T, and Lupski JR. 1991 'Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes' *Nucleic Acids Research* 19(24): 6823-6831.

Xu C, Ren H, Wand S, and Peng X. 2004 'Proteomic analysis of salt-sensitive outer membrane proteins of *Vibrio parahaemolyticus' Res Microbiol*. 155: 835-842.

Wu L, Lin X, Wang F, Ye D, Xiao X, Wang S, and Peng X. 2006 'OmpW and OmpV are required for NaCl regulation in *Photobacterium damsela*' J. Proteome Res. 5: 2250-2257.

4. Implications for stakeholders

The end-users are predominantly the scientific community with a stake in diagnostics, particularly those affected by the availability of human capacity and scientific resources. The primary stakeholder is SPHDS as the group with responsibility for National Diagnostic Protocols (NDPs) and other standards underpinning Australia's diagnostic capacity. Other key end-users are the other diagnostic laboratories State Departments and DAFF etc. The plant biosecurity research community more broadly will benefit from the evaluation of novel platforms, work which should also assist the PBCRC in considering new proposals for funding.

Specific implications for these next-users and end-users are:

- Improvements to the NDP for citrus canker. Development of a molecular test that is specific to this organism and does not cross react with the closely-related organism from cotton. This removes the need to conduct bioassays to confirm identity and host specificity of Xanthomonads under these conditions.

- Molecular method to differentiate between isolates of Xanthomonas that are highly pathogenic to citrus and the closely-related organisms that are benign. Similarly underpins the NDP in providing tools to ensure correct diagnosis.

- Nucleotide sequences, including whole genome sequences, have been generated. These will be subjected to on-going data mining to identify new and better diagnostic targets and to design diagnostic tests to even greater levels of specificity, including others not currently available (eg A vs A* vs Aw strains).

- Robust and specific biomarkers identified in this project could subsequently be adapted for use in point-of-care devices for use by non-specialist staff for surveillance, in the field, at ports of entry etc.

Industry is also a stakeholder. In particular the citrus industry and to a lesser extent the cotton industry will benefit directly from the newly developed diagnostic assays, through increased national incursion response capability delivered by the improved specificity and turnaround time, increased confidence in diagnostic results and options to improve tracking of bacterial pathovars in incursions. The favourable evaluation of these newer platforms means that other industries could also have confidence in commissioning research to apply these approaches to discriminate equally difficult taxa.

Australia's plant industries will benefit from the new, specialist capacity that has been fostered during this project. NSW DPI has appointed Dr Toni Chapman to a new position in plant bacteriology and Dr Chapman has been integral to this project over the last two years. NSW DPI only created this position because this project stood as evidence of need for both the work and a bacteriologist position, as expressed by industry and the plant biosecurity community. This project has also linked to the development of international, professional capacity (aka 'off-shore intelligence'), through the training of Dr Nuttima Kositcharoenkul, senior bacteriologist at Dept of Agriculture Bangkok Thailand, through the associated project CRC20093. DPI Victoria also hosted an Honours student through LaTrobe University, who, whilst not funded by the CRC, conducted work that stemmed directly from preliminary work in the metabolomics component of this project.

5. Recommendations

1. Prioritise publications detailing new tests for national and international peer review. Incorporate these methods into the National Diagnostic Protocol for citrus canker, submit for endorsement by Plant Health Committee via SPHDS.

2. Complete lodgement of nucleotide sequences in international database.

3. The PBCRC should consider application of proteomics and metabolomics to other problem taxa. Now that detailed technical methodologies are in place, results are likely to be fast-tracked to allow rapid development of diagnostics with new specificities.

4. Conduct workshop at EMAI for end-users, as an activity of the National Plant Health Diagnostic Network. This would provide training in the use of the NDPs, including new molecular methods for *X. citri* subp. *citri*. Potentially this could be run as a week-long exercise and also include in this workshop training on the NDPs for other citrus EPPs, HLB and Mal Secco.

5. Extend this work through :

    - Ongoing data-mining from this project – specifically finalising the whole genome sequencing for three pathovars used here, and evaluation of additional molecular diagnostics with new specificities, for delivery through more complex delivery platforms such as Real Time PCR, High Resolution Melt etc.

    - Expand genomic sequencing, by conducting whole genome sequencing of an extended panel of Xanthomonas species, and design of new diagnostic methodologies.

    - Expansion of metabolomic analyses to include other techniques. For example, analysis by LCMS may allow the identification of pathovars specific molecules or add more power to the predictive models.

6. Abbreviations/glossary

Insert list of abbreviations of acronyms (for example)

ABBREVIATION	FULL TITLE
bp	Base pair
CRCNPB	Cooperative Research Centre for National Plant Biosecurity
DAFF	Department of Agriculture, Fisheries and Forestry
DNA	Deoxyribonucleic acid
EPP	Emergency plant pest
NMR	Nuclear magnetic resonance
PLSDA	Partial Least Squares Discriminate Analysis
PBCRC	Plant Biosecurity Cooperative Research Centre

7. Plain English website summary

CRC project no:	CRC20054
Project title:	Platforms to differentiate exotic pathovars of plant bacteria
Project leader:	Dr Deborah Hailstones
Project team:	Dr Jo Luck, Dr Simone Rochford, Mr Dave Berryman, Dr Toni Chapman, Ms Celia Smuts, Ms Michelle Berg, Ms Anna Englezou, Mr Vilnis Ezernieks, Dr Nuttima Kositcharoenkul, Dr Mike Jones, Dr Mehmet Cakir
Research outcomes:	This project has demonstrated the effectiveness of proteomics and metabolomics, two novel approaches that identify functional molecules, to discover biomarkers that differentiate closely related bacteria. Novel diagnostic targets have been identified and validated that differentiate organisms to the levels of specificity required, but not previously available. Novel diagnostics will translate to improved incursion response because their greater specificity and reliability reduce turn around time. New specialist capacity and partnerships developed.
Research implications:	This project has delivered positive results in terms of scientific resources and human capacity. The finding that these platforms can fast-track the identification of biomarkers with new specificities, and at the level required, can inform investment decisions for the PBCRC in future. New diagnostic technologies have been developed that will be incorporated into the National Diagnostic Protocol (NDP) and could be adapted to delivery through a range of other platforms. This provides security to industry and those responsible for surveillance and disease management programs. Extensive data has been generated that is yet to be mined, but is likely to deliver further tests in future, without significant further investment. Significant specialist capacity in bacteriology has been fostered both nationally and internationally, through recruitment and professional development, which will benefit Australia's plant industries through local expertise and trusted 'off-shore intelligence'.
Research publications:	Smuts et al. 'Using 2D gels to distinguish between *Xanthomonas axonopodis* pathovars from citrus and cotton' In preparation for submission to *Journal of Proteomics* Hailstones et al. 'Development of novel diagnostics that differentiate citrus canker from other closely related Xanthomonads' In preparation for submission to *Plant Pathology* Rochford et al. re use of metabolomics to differentiate

	pathovars of Xanthomonas. Title to be determined Chapman et al. re comparison of whole genome sequences of citri to malvacearum
Acknowledgements:	Gary Kong and the CRC NPB Board for funding a 'blue sky' project. Leanne Bringolf for excellent technical assistance. Aneesha Deanensen for initial technical work at SABC.

7. Appendices

Appendix 1 Genbank Accession numbers

Appendix 2 Metabolomics
Appendix 2.1 Details of PLSDA modelling

This is a model of type: PLSDA
Developed 25-Nov-2011 12:08:047.13
Author: sr60@EDA101724
X-block: pathovars 222 by 27020 (sr60@EDA101724@20110609T132431.53189196 m:20110610121930.500)
Included: [1-222] [1-27020]
Included (in axis units): [n/a] [0.700593-8.99997]
Preprocessing: OSC (Orthogonal Signal Correction), Autoscale
Y-block: y 222 by 5 (sr60@EDA101724@20111125T120741.64046772 m:20111125120741.921)
Included: [1-222] [1-5]
Preprocessing: Autoscale
Num. LVs: 6
Cross validation: random samples w/ 10 splits and 1 iterations

Statistics for each y-block column:
Modeled Class: 1 2 3 4 5
Sensitivity (Cal): 0.982 1.000 0.982 1.000 1.000
Specificity (Cal): 0.916 0.976 1.000 1.000 0.928
Sensitivity (CV): 0.982 0.873 0.964 1.000 0.952
Specificity (CV): 0.898 0.952 0.976 0.990 0.928
Class. Err (Cal): 0.0510071 0.011976 0.00892857 0 0.0361111
Class. Err (CV): 0.0599891 0.0875885 0.0299053 0.00480769 0.0599206
RMSEC: 0.262975 0.218036 0.199319 0.0941664 0.229456
RMSECV: 0.285285 0.287999 0.238847 0.148892 0.253828
Bias: -1.94289e-016 -2.77556e-017 0 6.93889e-017 -5.55112e-017
CV Bias: -0.00299015 0.0137335 -0.00665539 -0.00525728 0.00116928
R^2 Cal: 0.628931 0.744916 0.789377 0.849926 0.656772
R^2 CV: 0.563528 0.55662 0.699062 0.643155 0.580177

Percent Variance Captured by Regression Model

	-----X-Block-----		-----Y-Block-----	
Comp	This	Total	This	Total
1	8.40	8.40	15.26	15.26
2	11.49	19.90	11.97	27.23
3	6.60	26.50	14.58	41.81
4	7.64	34.14	9.20	51.01
5	2.46	36.60	14.60	65.61
6	2.91	39.51	7.79	73.40

Appendix 2.2 Details of PLSDA modelling

This is a model of type: PLSDA_PRED
Developed 25-Nov-2011 13:23:026.39
Author: sr60@EDA101724
X-block: pathovars 20 by 27020 (sr60@EDA101724@20110609T132431.53189196 m:20111125130026.921)
Included: [1-20] [1-27020]
Included (in axis units): [n/a] [0.700593-8.99997]
Preprocessing: OSC (Orthogonal Signal Correction), Mean Center
Y-block: y 20 by 5 (sr60@EDA101724@20111125T132326.31263703 m:20111125132326.328)
Included: [1-20] [1-5]
Preprocessing: Autoscale
Num. LVs: 5
Cross validation: venetian blinds w/ 10 splits

Statistics for each y-block column:
Sensitivity (Cal): 0.980 0.863 0.981 1.000 0.973
Specificity (Cal): 0.914 0.901 0.940 0.947 0.945
Sensitivity (CV): 0.980 0.863 0.981 1.000 0.838
Specificity (CV): 0.914 0.881 0.947 0.953 0.921
Sensitivity (Pred): 0.800 0.750 1.000 0.500 1.000
Specificity (Pred): 0.933 0.813 0.938 0.889 0.933
Class. Err (Cal): 0.0527632 0.118296 0.0396154 0.0263158 0.0407862
Class. Err (CV): 0.0527632 0.12823 0.0362821 0.0236842 0.120475
Class. Err (Pred): 0.133333 0.21875 0.03125 0.305556 0.0333333
RMSEC: 0.266523 0.277124 0.228001 0.193741 0.255471
RMSECV: 0.280449 0.322438 0.244923 0.201444 0.274952
RMSEP: 0.340113 0.333843 0.249411 0.261802 0.246534
Bias: 2.22045e-016 -1.66533e-016 1.11022e-016 5.55112e-017 1.11022e-016
CV Bias: -0.00199428 0.00537199 -0.00430985 0.00158209 -0.000649945
Pred Bias:-0.0482812 0.0740309 0.058917 -0.0525178 -0.032149
R^2 Cal: 0.61862 0.593084 0.728055 0.328248 0.563785
R^2 CV: 0.578586 0.460271 0.687829 0.275221 0.499247
R^2 Pred: 0.429732 0.342701 0.644771 0.269723 0.710549

Percent Variance Captured by Regression Model

	-----X-Block-----		-----Y-Block-----	
Comp	This	Total	This	Total
1	73.57	73.57	26.28	26.28
2	11.18	84.74	6.88	33.16
3	4.01	88.76	10.07	43.23
4	1.98	90.74	7.51	50.73
5	1.47	92.20	5.90	56.64

Appendix2. 3 Details of PLSDA modelling

This is a model of type: PLSDA_PRED
Developed 25-Nov-2011 13:11:026.42
Author: sr60@EDA101724
X-block: pathovars 20 by 27020 (sr60@EDA101724@20110609T132431.53189196 m:20111125130026.921)
Included: [1-20] [1-27020]
Included (in axis units): [n/a] [0.700593-8.99997]
Preprocessing: OSC (Orthogonal Signal Correction), Autoscale
Y-block: y 20 by 5 (sr60@EDA101724@20111125T131126.31265437 m:20111125131126.328)
Included: [1-20] [1-5]
Preprocessing: Autoscale
Num. LVs: 6
Cross validation: venetian blinds w/ 10 splits
Statistics for each y-block column:
Sensitivity (Cal): 0.980 1.000 0.981 1.000 1.000
Specificity (Cal): 0.921 0.974 1.000 1.000 0.927
Sensitivity (CV): 0.980 0.863 0.962 1.000 0.946
Specificity (CV): 0.921 0.954 0.973 0.984 0.921
Sensitivity (Pred): 1.000 0.750 1.000 1.000 1.000
Specificity (Pred): 0.867 0.875 1.000 1.000 0.867
Class. Err (Cal): 0.0494737 0.013245 0.00961538 0 0.0363636
Class. Err (CV): 0.0494737 0.0918063 0.0325641 0.00789474 0.066421
Class. Err (Pred): 0.0666667 0.1875 0 0 0.0666667
RMSEC: 0.258832 0.218874 0.196575 0.0929395 0.230013
RMSECV: 0.281601 0.284734 0.229261 0.137348 0.254047
RMSEP: 0.303048 0.294113 0.223566 0.162107 0.283652
Bias: -2.77556e-017 -5.55112e-017 5.55112e-017 0 -5.55112e-017
CV Bias: -0.00194777 0.00681353 -0.00565333 -0.000431189 0.00121875
Pred Bias:-0.0424663 0.0769244 -0.0074337 -0.0439764 0.016952
R^2 Cal: 0.640312 0.746169 0.797855 0.845414 0.646393
R^2 CV: 0.574328 0.571392 0.727262 0.678903 0.569266
R^2 Pred: 0.539048 0.496343 0.696127 0.816762 0.582518
Percent Variance Captured by Regression Model

	-----X-Block-----		-----Y-Block-----	
Comp	This	Total	This	Total
1	7.81	7.81	16.00	16.00
2	12.28	20.09	11.17	27.17
3	6.15	26.24	15.32	42.48
4	7.58	33.83	9.78	52.26
5	2.81	36.63	14.13	66.39
6	3.27	39.90	7.14	73.52

Appendix 2.4 Details of PLSDA modelling

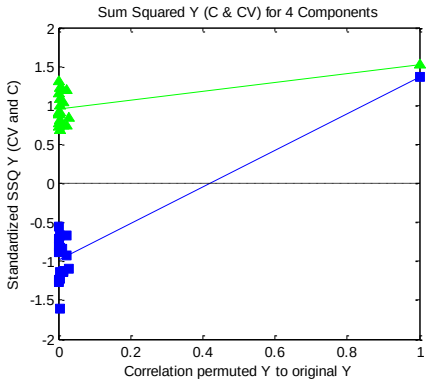
This is a model of type: PLSDA
Developed 25-Nov-2011 13:34:038.42
Author: sr60@EDA101724
X-block: pathovars 113 by 27020 (sr60@EDA101724@20110609T132431.53189196 m:20111125133203.281)
Included: [1-113] [1-27020]
Included (in axis units): [n/a] [0.700593-8.99997]
Preprocessing: OSC (Orthogonal Signal Correction), Autoscale
Y-block: y 113 by 3 (sr60@EDA101724@20111125T133405.68708844 m:20111125133405.703)
Included: [1-113] [1-3]
Preprocessing: Autoscale
Num. LVs: 4
Cross validation: venetian blinds w/ 10 splits

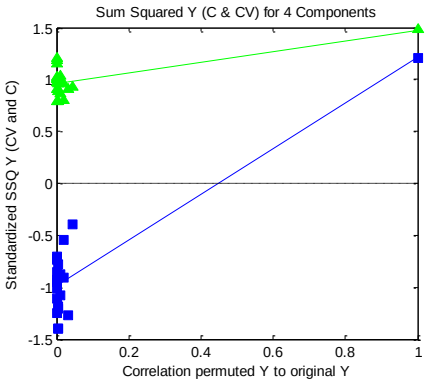
Statistics for each y-block column:
Modeled Class: 1 2 4
Sensitivity (Cal): 1.000 1.000 1.000
Specificity (Cal): 1.000 1.000 1.000
Sensitivity (CV): 0.980 1.000 1.000
Specificity (CV): 0.984 0.952 1.000
Class. Err (Cal): 0 0 0
Class. Err (CV): 0.0179365 0.0241935 0
RMSEC: 0.16912 0.166998 0.100201
RMSECV: 0.218854 0.24889 0.153261
Bias: 5.55112e-017 2.22045e-016 0
CV Bias: -0.00501645 0.0151119 -0.0100955
R^2 Cal: 0.88406 0.887379 0.894221
R^2 CV: 0.80656 0.756484 0.784764

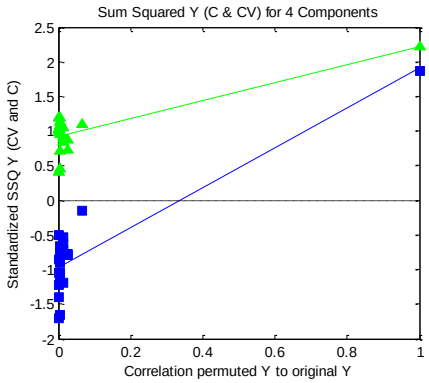
Percent Variance Captured by Regression Model

	-----X-Block-----		-----Y-Block-----	
Comp	This	Total	This	Total
1	12.86	12.86	36.42	36.42
2	12.47	25.32	20.88	57.30
3	4.93	30.25	23.06	80.36
4	4.69	34.94	8.50	88.86

Appendix 2.5 Details of Permutation Testing for PLSDA Model A, A* and E-type







Probability of Model Insignificance vs. Permuted Samples
For model with 4 component(s)

Y-column: 1
Wilcoxon Sign Test Rand t-test
Self-Prediction: 0.000 0.001 0.005
Cross-Validated: 0.000 0.000 0.005

Y-column: 2
Wilcoxon Sign Test Rand t-test
Self-Prediction: 0.000 0.007 0.005
Cross-Validated: 0.000 0.000 0.005

Y-column: 3
Wilcoxon Sign Test Rand t-test
Self-Prediction: 0.000 0.000 0.005
Cross-Validated: 0.000 0.000 0.005

Values less than 0.05 indicate the model is
significant at the 95% confidence level.

Appendix 2.6 Details of PLSDA modelling

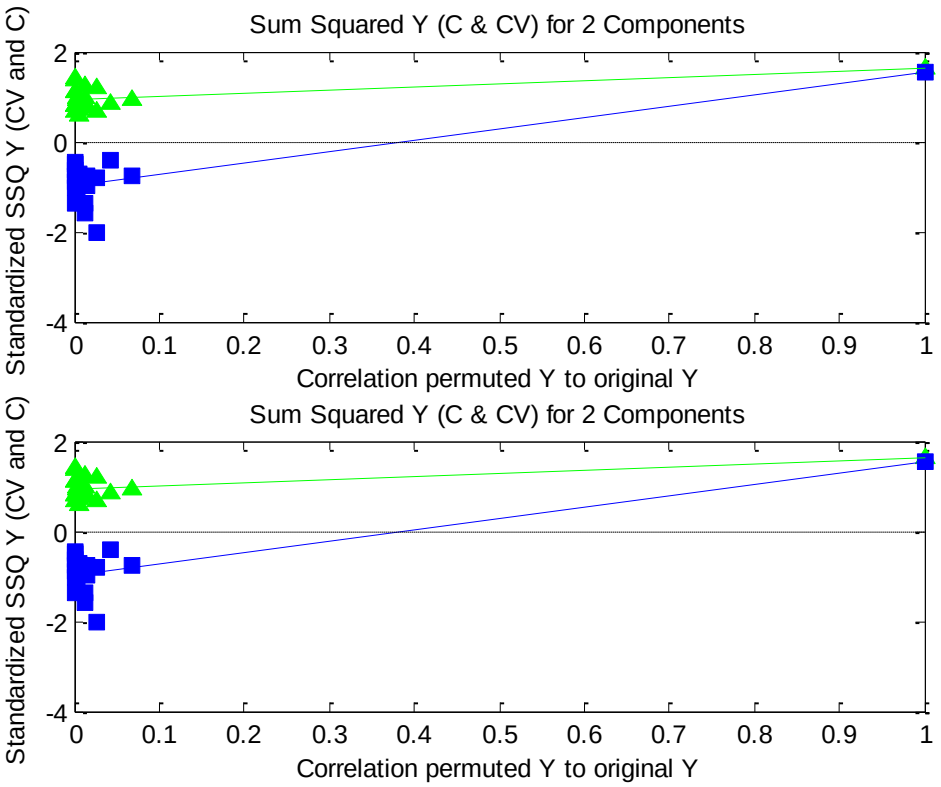
This is a model of type: PLSDA_PRED
Developed 25-Nov-2011 13:56:036.13
Author: sr60@EDA101724
X-block: pathovars 9 by 27020 (sr60@EDA101724@20110609T132431.53189196 m:20111125135253.546)
Included: [1-9] [1-27020]
Included (in axis units): [n/a] [0.700593-8.99997]
Preprocessing: OSC (Orthogonal Signal Correction), Autoscale
Y-block: y 9 by 2 (sr60@EDA101724@20111125T135636.06260286 m:20111125135636.078)
Included: [1-9] [1-2]
Preprocessing: Autoscale
Num. LVs: 2
Cross validation: venetian blinds w/ 9 splits

Statistics for each y-block column:
Sensitivity (Cal): 1.000 1.000
Specificity (Cal): 1.000 1.000
Sensitivity (CV): 1.000 1.000
Specificity (CV): 1.000 1.000
Sensitivity (Pred): 1.000 1.000
Specificity (Pred): 1.000 1.000
Class. Err (Cal): 0 0
Class. Err (CV): 0 0
Class. Err (Pred): 0 0
RMSEC: 0.110649 0.110649
RMSECV: 0.150413 0.150413
RMSEP: 0.163331 0.163331
Bias: -2.22045e-016 3.88578e-016
CV Bias: -0.00756781 0.00756781
Pred Bias:0.00226305 -0.00226305
R^2 Cal: 0.950843 0.950843
R^2 CV: 0.909365 0.909365
R^2 Pred: 0.891981 0.891981

Percent Variance Captured by Regression Model

-----X-Block-----		-----Y-Block-----	
Comp	This	Total	This	Total
1	9.75	9.75	86.82	86.82
2	9.64	19.39	8.26	95.08

Appendix2.6 Details of Permutation Testing for PLSDA Model Aw and malvacearum



Probability of Model Insignificance vs. Permuted Samples
For model with 2 component(s)

Y-column: 1
Wilcoxon Sign Test Rand t-test
Self-Prediction: 0.000 0.000 0.005
Cross-Validated: 0.000 0.000 0.005

Y-column: 2
Wilcoxon Sign Test Rand t-test
Self-Prediction: 0.000 0.000 0.005
Cross-Validated: 0.000 0.000 0.005

Values less than 0.05 indicate the model is
significant at the 95% confidence level.