Bean yellow mosaic virus: from Koch’s postulates to next generation sequencing and their use to unravel the cause of black pod syndrome of narrow-leafed lupin.

This thesis is presented for the degree of Doctor of Philosophy of The University of Western Australia

Monica Kehoe

BSc in Microbiology

‘This thesis is presented for the degree of Doctor of Philosophy of The University of Western Australia’

School of Plant Biology

Faculty of Science

March 2014
ABSTRACT

Black pod syndrome (BPS) causes devastating losses in Lupinus angustifolius (narrow-leafed lupin) crops in Australia, and late infection with Bean yellow mosaic virus (BYMV) (genus Potyvirus) was suggested as a possible cause. In 2011, an end of growing season survey of L. angustifolius plants with BPS from six locations in south western Australia was conducted. Tissue samples from different positions on each of these symptomatic plants were tested for BYMV and generic potyvirus by ELISA and RT-PCR. Detection was most reliable when RT-PCR with generic potyvirus primers was used on tissue taken from the main stem of the plant just below the black pods. An initial glasshouse experiment revealed that mechanical inoculation of L. angustifolius plants with BYMV after pods had formed caused pods to turn black. A subsequent experiment in which plants were inoculated at eight different growth stages confirmed that BPS was only induced when L. angustifolius plants were inoculated after first flowering when pods were forming. Thus, BYMV was isolated from a symptomatic L. angustifolius survey sample, inoculated to and maintained in culture hosts, inoculated to healthy L. angustifolius test plants inducing BPS and then successfully re-isolated from them. As such, Koch’s postulates were fulfilled for the hypothesis that late infection with BYMV causes BPS in L. angustifolius plants.

One of the questions remaining was whether any other related plant viruses may also cause BPS and be a threat to Lupinus species. Hardenbergia mosaic virus (HarMV), genus Potyvirus, belongs to the Bean common mosaic virus (BCMV) potyvirus lineage found only in Australia. The original natural host of HarMV is Hardenbergia comptoniana, family Fabaceae, which is indigenous to the South-West Australian Floristic Region (SWAFR), where Lupinus spp. are grown as introduced grain legume crops or exist as naturalized weeds. Two plants of H. comptoniana and one of L. cosentinii, each with mosaic and leaf deformation symptoms, were sampled from a small patch of disturbed vegetation at an ancient ecosystem-recent agroecosystem interface. Potyvirus infection was detected in all three samples by ELISA and RT-PCR. After sequencing on an Illumina HiSeq 2000, three complete and two nearly complete HarMV genomes from H. comptoniana and one complete HarMV genome from L. cosentinii were obtained. Phylogenetic analysis which compared either whole genomes or coat protein genes revealed that three of the complete and one of the nearly complete new genomes
were in HarMV clade I, one of the complete genomes in clade V and one nearly complete genome in clade VI. The complete HarMV genome from *L. cosentinii* differed by only eight nucleotides from one of the HarMV clade I genomes from a nearby *H. comptoniana* plant, with only one of these nucleotide changes being non-synonymous. Recombination analysis revealed evidence of two recombination events amongst the six complete genomes. This is the first report of HarMV naturally infecting *L. cosentinii* and the first example for the SWAFR of virus emergence from a native plant species to invade an introduced plant species.

Next generation sequencing (NGS) is quickly emerging as the go-to tool for plant virologists when sequencing whole virus genomes. A comparison of the genomic and biological properties of BYMV isolates from *L. angustifolius* plants with BPS, systemic necrosis or non-necrotic symptoms, and from two other plant species was conducted. When one *Clover yellow vein virus* (ClYVV) (genus *Potyvirus*) and 22 BYMV isolates were sequenced on the Illumina HiSeq2000, one new CIYVV and 23 new BYMV sequences were obtained. When the 23 new BYMV genomes were compared with 17 other BYMV genomes available on Genbank, phylogenetic analysis provided strong support for existence of nine phylogenetic groupings. I proposed that the current system of nomenclature based on biological properties be replaced by numbered groups (I-IX). This is because previous group names based on coat protein phylogeny and host specialization were not well supported by this study. Biological studies involving seven isolates of BYMV and one of CIYVV with 9 different plant species gave no symptoms or reactions that could be used to distinguish BYMV isolates from *L. angustifolius* plants with BPS from other isolates. NGS proved a valuable tool to obtain complete BYMV and CIYVV genomes. It demonstrated that consideration needs to be given to issues regarding sample preparation, adequate levels of coverage across a genome and methods of assembly. This study also provided important lessons in NGS that will be helpful to other plant virologists in the future.

Recombination was previously suggested to be playing a role in BYMV’s evolution towards host specialization. As such, recombination analysis was conducted on the complete coding regions of 33 BYMV genomes and two genomes of the related *Clover yellow vein virus* (CIYVV). This analysis found evidence for 12 firm recombination events within BYMV phylogenetic groups I-VI, but none
within groups VII-IX or CIYVV. The greatest numbers of recombination events within a sequence (two or three each) occurred in four groups, three of which formerly constituted the single ancestral generalist group (I, II and IV), and group VI. Sequences in groups III and V had one event each. These findings with whole genomes are consistent with recombination being associated with expanding host ranges, and call into question the proposed role of recombination in the evolution of BYMV host specialization. Instead, they indicate that recombination explains the very broad natural host ranges of three BYMV groups (I, II, IV) which infect both monocots and dicots, and that the three groups with narrow natural host ranges (III, V, VI) now have the potential to reduce specificity and broaden their natural host ranges.

The confirmation of BYMV as the cause of BPS in *L. angustifolius* constitutes an important step forward in the quest for a BPS resistant cultivar. The next step should include identification of potential markers for resistance and could make use of NGS technologies (e.g. RNA-seq experiments) to do so. In the meantime there are already established integrated disease management strategies in place for early infection with BYMV in *L. angustifolius* which should be implemented. Particular emphasis should be given to maximizing volunteer clover control within the crop, and avoiding planting next to or downwind of legume pastures or other *L. angustifolius* crops that may contain sources of BYMV. The increased knowledge about the molecular characterization of BYMV obtained in this project will assist in moving forward with BYMV research in the future. The use of a numbered phylogenetic grouping system will prevent confusion over using biological host names to name such groups. Also, a better understanding of the role that recombination has played, and will likely play, in the future evolution of BYMV means that future research on this virus has the potential to reveal exciting insights, especially as new BYMV genomes from more diverse hosts and geographical locations are sequenced.
ACKNOWLEDGEMENTS

I would like to thank my principal supervisor, Professor Roger Jones, from the University of Western Australia (UWA) and the Department of Agriculture and Food Western Australia (DAFWA) for his continued support, guidance and enthusiasm over the course of this PhD research project. Thanks also to my supervisor Dr Bevan Buirchell, also from DAFWA and UWA for his support for this research project, in particular for his generosity in making the next generation sequencing component possible.

A very special thanks to Ms Brenda Coutts for her friendship, support, guidance and insightful discussions in relation to the project (and everything else!), which were invaluable.

I would like to thank the Grains Research and Development Corporation (GRDC) for financial support via a Grains Industry Research Scholarship (GIS), and UWA for an Australian Postgraduate Award (APA).

To the staff at DAFWA, you have ensured that this PhD was conducted in a friendly and fun environment where your encouragement, advice, and assistance were greatly valued. In particular I would like to thank Eva Gajda, Mirjana Banovic and Stuart Vincent for their help and support in the laboratory and glasshouses. Mr Geoff Thomas for helpful discussion and Mr Cliff Staples for providing lupin seed as well as advice about how to grow them. To my fellow DAFWA PhD students – Brenda Coutts, Alison Mackie, Kevin Foster and Dominie Wright – it has been a pleasure sharing this experience with you all! Also thanks to Belinda Cox, a great friend who just happens to share my office – thanks for all the chats and advice.

I would also like to thank the staff at the school of Plant Biology for their administrative support including applications, travel arrangements, purchasing and many other forms that are part of a research project.

To Jonathan Ryan, there are no words – you’ve been wonderful and this would have been so much harder without your love and encouragement. Thanks to my wonderful friends who have been so supportive along the way, it would be impossible to name you all but I would particularly like to thank Nicole Patterson, Colin Davis and Craig Webster for being there and reminding me what it’s all really about. Your friendship is greatly valued.

Finally, to my mum – thanks for everything, you’re a legend!
THESIS DECLARATION

This thesis is presented as a series of scientific papers that includes two papers which have been published or accepted for publication, and two manuscripts which have been prepared for submission for publication.

I declare that this thesis is my own work and the results of my own research. The contributions of the co-authors of publications arising from this thesis is mainly associated with initial research directions, advice on experiments as required, funding and editorial input in various versions of the drafts of each paper and/or chapter. I have the permission of all co-authors to include this work in my thesis.

This thesis was completed during the course of enrolment for the degree of Doctor of Philosophy at the University of Western Australia, and has not previously been accepted for a degree at this or any other institution.

Candidate: Monica Kehoe
Co-ordinating supervisor: Professor Roger Jones
PUBLICATIONS ARISING FROM THIS THESIS AND STATEMENT OF CONTRIBUTION

Four papers have been published from this thesis:

  My contribution to this work: Organising and conducting the end of growing season survey, all experimental work including ELISA and RT-PCR testing, all data collection, analysis and paper writing.

  My contribution to this work: Collection of samples, all laboratory testing, analysis of next generation sequencing data and subsequent phylogenetic and recombination analyses, and paper writing.

  My contribution to this work: All glasshouse inoculations and testing, laboratory testing by ELISA and RT-PCR, analysis of next generation sequencing data and paper writing.

  My contribution to this work: Data collection, all recombination analysis and paper writing.
# TABLE OF CONTENTS

ABSTRACT

ACKNOWLEDGEMENTS

THESIS DECLARATION

PUBLICATIONS ARISING FROM THIS THESIS AND STATEMENT OF CONTRIBUTION

TABLE OF CONTENTS

CHAPTER 1

GENERAL INTRODUCTION

CHAPTER 2

BLACK POD SYNDROME OF LUPINUS ANGUSTIFOLIUS IS CAUSED BY LATE INFECTION WITH BEAN YELLOW MOSAIC VIRUS

CHAPTER 3

HARDBERGIA MOSAIC VIRUS: CROSSING THE BARRIER BETWEEN NATIVE AND INTRODUCED PLANT SPECIES.

CHAPTER 4

PLANT VIROLOGY AND NEXT GENERATION SEQUENCING: EXPERIENCES WITH A POTA VIRUS

CHAPTER 5

SPLIT PERSONALITY OF A POTA VIRUS: TO SPECIALIZE OR NOT TO SPECIALIZE?

CHAPTER 6

SUMMARY AND CONCLUDING REMARKS

APPENDIX

PUBLICATIONS, PRESENTATIONS AND EXTENSION ACTIVITIES ARISING FROM THIS THESIS
Chapter 1

General Introduction
Unravelling the cause of black pod syndrome of narrow-leaved lupins

1. Lupins – where did they come from?

*Lupinus* *L.* is a distinct genus within the family *Fabaceae*, having diverged from the *Genistae*, around 12-14 million years ago (Dunn 1984; Käss and Wink 1997; Gladstones 1998a; Aïnouche et al. 2004). Within the genus *Lupinus*, there is evidence its’ original ancestors were located in the Old World around Turkey, Syria and surrounding countries. It is likely that spread then occurred westwards along the Mediterranean and eastward across northern Asia, then into North America (Käss and Wink 1997; Gladstones 1998a). This theory is supported by more recent phylogenetic evidence that points towards there having been at least two possible divergence events to the New World as recently as 10 million years ago. These phylogenies also show the New World species to consist of two separate clades: one eastern South American mostly lowland clade consisting of around 35 species and the second a western North and South American, mostly highland clade of around 225 species. The exception to this being species with unifoliate leaves of which there are about 19, occurring only in the south-east USA and eastern South America (Drummond 2008; Eastwood et al. 2008; Eastwood and Hughes 2008).

2. Narrow-leaved lupins in Western Australia

Modern breeding of *L. angustifolius* *L.* (narrow-leaved lupin) began in Germany in 1927 when the first sweet (i.e. low alkaloid) lupin types were selected. Lupin breeding focusing on this species in Western Australia (WA) began in 1954 at the University of Western Australia. This work later moved to the then Department of Agriculture in the 1970’s and led to the first fully domesticated *L. angustifolius* cultivars: Uniwhite (1967), Uniharvest (1971) and Unicrop (1973). They were low in alkaloids and therefore sweet enabling them to be fed to livestock and potentially used for human consumption (Gladstones 1982a; Gladstone 1982b). They were also soft seeded, i.e. they had permeable seed coats and so lacked dormancy. Therefore they were able to germinate uniformly when sown into moist soil, and had non-shattering pods so seed wasn’t spilt before harvest (French et al. 2008). Improvements to domesticated *L. angustifolius* cultivars include breeding for disease resistance (e.g., Cowling et al. 1998). For viruses, breeding has focused on reduced levels of seed
transmission of *Cucumber mosaic virus* (CMV) (family *Bromoviridae*, genus *Cucumovirus*) (Jones and Cowling 1995; Jones 2001).

3. **The narrow-leafed lupin market**

*Lupinus* spp. (lupins), are cool season crop legumes with worldwide importance that are grown for their grain as summer annual crops in regions of the world with cool temperate climates, as winter crops mainly in regions with Mediterranean-type climates, or as wet season or irrigated crops at high altitude in tropical regions (Gladstones 1998a). Australia is currently the largest producer of lupins in the world producing around 85% of the global crop. It is followed by the European Union, which produces 5% and countries such as Belarus, Chile, Peru, Bolivia, South Africa and the Russian Federation making up the remaining 10%. In Australia, 90% of lupin production is for export with the main destinations being the Republic of Korea (45%), European Union (27%) and Japan (12%) (Lawrence 2007). South western Australia is responsible for 80% of Australian lupin production, and 95% of this is from *L. angustifolius* (French et. al. 2008). While the majority of production is for stock feed (including feeding sheep on farm), there is growing use of lupins as a partial replacement for fish-meal in the global aquaculture industry (Glencross 2008). *L. angustifolius* is also an important component of crop rotations in which it helps in the management of fungal root disease in cereal crops, soil rehabilitation and improving the yield of cereal crops grown after it due to their nitrogen-fixing capabilities (Chalk 1998; Harries and Peek 2008).

4. **Lupins and potential health benefits for humans**

Lupin seed for human consumption currently accounts for less than 4% of production but there is considerable potential for this to increase as more people make use of it for its health benefits. Lupin flour contains 40-45% protein, 25-30% fibre and negligible sugar and starch. This means that it can be substituted into foods containing high levels of carbohydrate to increase their protein and fibre contents whilst having little effect on the characteristic tastes and textures of these products (Evans et. al. 1993; Lee et. al. 2006). Increasing protein and fibre content at the same time as decreasing refined carbohydrates may also help to decrease the risk of cardiovascular
disease through the lowering of blood cholesterol and blood pressure (Hodgson and Lee 2008). Lupin seeds are also a rich source of lutein, which can help to delay the process of age-related macular degeneration in the eye in individuals at risk of the disease, e.g. people aged 55 and over with insufficient lutein present in the bloodstream (Fryirs et. al. 2008). However, there may also be a correlation between peanut allergies and lupin allergies. This means that foodstuffs derived from lupin need to be appropriately labelled and research is required to properly understand allergic reactions to lupin and the relationship between peanut and lupin allergies (Moneret-Vautrin et. al. 1998; Shaw et. al. 2008; Smith et. al. 2008).

5. Black pod syndrome

Despite being particularly well adapted to acid sandy soils and drought through drought escape mechanisms (including deep rooting), L. angustifolius still yield well below their full potential in the Mediterranean climate of south-west WA (French and Buirchell 2005). There is potential for yields of above 4 t/ha, but they rarely reach above 2 t/ha (Gladstones 1998a,b; Buirchell 2008). One of the main reasons for this is the Black pod syndrome (BPS), particularly in southern higher rainfall areas. BPS is a syndrome where the pods, especially the primary pods on the main stem of a L. angustifolius plant, prematurely turn black or brown and produce few or no seeds (Fig 1). First noticed in the 1990’s, the syndrome contributes significantly to lower and unreliable yields of L. angustifolius especially in the south coastal region of WA and losses of 40% of the potential yield have been recorded in partially infected crops (Gladstones 1998b; Buirchell 2008).

There is little published data on BPS and the remaining knowledge is from the unpublished observations and small-scale experiments of Department of Agriculture and Food Western Australia (DAFWA) researchers. The cause (or causes) of BPS is as yet unknown. Possible causes of BPS that were hypothesised included (i) excessive vegetative growth leading to poor pod set, (ii) nutrient deficiencies and (iii) infection with Bean yellow mosaic virus (BYMV) (family Potyviridae, genus Potyvirus). The excessive vegetative growth leading to poor pod set hypothesis was not supported in field experiments consisting of treatments with or without growth hormone applied to
different *L. angustifolius* genotypes (White *et al.* 2007). Also, nutrient deficiency was disproved when full rates of fertilizer application had no effect on the levels of BPS in different *L. angustifolius* genotypes. BYMV is non-persistently transmitted by aphids (Jones and McLean 1989; Jones 2001). Consequently, when two insecticides were applied every two weeks at double the recommended rates, levels of BPS were diminished providing some support for the suggestion that BYMV vectored by aphids may be playing a role (White and Baker 2009). However, BPS was sometimes present in the primary pods without any of the necrotic stem streaking symptoms usually associated with BYMV infection in *L. angustifolius*. In addition, Enzyme linked immuno-sorbent assay (ELISA) tests failed to detect BYMV from the majority of BPS symptomatic plant samples tested (Jones, R.A.C. unpublished).

![Figure 1](image.png)

**Figure 1.** Symptoms of Black pod syndrome (BPS) seen on a *Lupinus angustifolius* plant in the field.

Between 1996-2003, *L. angustifolius* Crop Variety Trials (CVTs) were rated for BPS at the end of the growing season, and BYMV was considered the most likely cause (Gladstones 1998b; White and Baker 2009). However, annual field resistance screening experiments involving young *L. angustifolius* plants found no difference in susceptibility between BYMV susceptible breeding lines and lines identified in CVTs as being partially resistant to BPS (Jones and Coutts 1998). Thus, rating young plants for BYMV resistance in the field gave different results from rating mature plants for
BPS at the end of the growing season (after pod formation). CVTs in 2005 demonstrated a correlation between high levels of BPS and diminished grain yields, and a genetic basis to partial BPS resistance was suggested. Subsequently, a cultivar with partial BPS resistance was released under the name cv. Jenabillup (White et al. 2007; Buirchell 2008).

6. **Bean yellow mosaic virus**

BYMV has an elongated capsid with helical symmetry (filamentous, flexuous rods) and a particle length of 750nm (ICTVdb Management 2006). It is non-persistently transmitted by aphids and not known to be transmitted in the seed of narrow-leafed lupin (Jones and McLean 1989; Jones 2001; Jones et. al. 2003; Jones and Smith 2005). There is a whole subgroup of viruses within the genus *Potyvirus* called the BYMV group. BYMV itself is divided into seven distinct phylogenetic groupings based on the original hosts that isolates came from. One of these is a general group with a very wide natural host range, including both monocots and dicots. Almost all isolates sequenced from WA are from this more primitive generalist group. The other six groups are more highly evolved specialist groups named after their principal natural hosts, e.g. the lupin group which contains only isolates obtained from broad-leaved lupins outside Australia, and the broad bean group which contains only isolates from broad beans (Wylie et. al. 2008). Recombination has played an important role in host specialization of BYMV. Evidence of recombination was found across all phylogenetic groups, as well as between the closely related *Clover yellow vein virus* (CIYVV) and other BYMV groups (Wylie and Jones 2009).

In addition to the phylogenetic groupings associated with BYMV, there are two biological groupings. These are called the necrotic (BYMV-N) and the non-necrotic (BYMV-NN). However, they are only relevant regarding infection of *L. angustifolius*. The BYMV-N isolates cause a systemic hypersensitive resistance reaction controlled by expression of the *Nbm-1* gene in *L. angustifolius*, whereas the BYMV-NN isolates do not. Isolates of BYMV that come from the lupin phylogenetic group (separate from the general group) are also able to cause a necrotic reaction in two lines of *L. angustifolius* that do not contain the *Nbm-1* gene, suggesting the likely presence of a
second hypersensitivity gene (Cheng and Jones 1999; Cheng and Jones 2000; Jones and Smith 2005; Jones et al. 2008). In the field, when BYMV-N infects *L. angustifolius* plants early in the growing period, its interaction with *Nbm-1* causes young plants to die from systemic necrosis. This removes the plant from the crop so it no longer acts as a source of infection for further within-crop virus spread. If an infection with BYMV-NN occurs, the plant does not die but remains in the field acting as a continuing source of virus for spread by aphid vectors to other plants for the rest of the growing season (Cheng and Jones 1999, 2000; Jones et al. 2003).

7. **Are other viruses involved in BPS?**

While infection with BYMV is currently the most likely candidate for the main cause of BPS, it should be noted that it is possible that other plant viruses present in WA may also be playing a role. The most common source from which BYMV invades *L. angustifolius* crops in WA is from nearby infected annual, self-regenerating legume pastures dominated by *Trifolium subterraneum* (subterranean clover). In years when there are high numbers of aphids present following high summer and early autumn rainfall associated with high temperatures early in the growing season, infection levels can be very high in pastures and so subsequent spread to and losses in *L. angustifolius* crops can be considerable (Jones 2001). BYMV strains present in WA are not known to be seed-borne in *L. angustifolius*. However, they are seed-borne in annual pasture clover species and this is the method by which the virus survives the hot dry summer period which separates successive growing seasons (Jones 1993; McKirdy et al. 2000). A number of other viruses are found amongst perennial pasture species growing in irrigated pastures which persist all year round, including CIYVV, the closely related potyvirus found with incidences of up to 23% within *T. repens* (white clover) in WA (McKirdy and Jones 1997; Jones 2013). CIYVV has also been reported to infect *L. angustifolius* crops in south-eastern Australia (Jones and McLean 1989; Jones 2013). Other legume viruses found include *White clover mosaic virus* (WCMV), *Alfalfa mosaic virus* (AMV) and *Subterranean clover redleaf virus* (SCRLV) (McKirdy and Jones 1997; Coutts and Jones 2002; Jones 2013). The movement of BYMV from pastures into *L. angustifolius* crops is well documented (Cheng et al. 2002; Thackray et al. 2002; Jones 2005) but there is no record of other pasture infecting viruses, apart
from CIYVV, causing necrosis in *L. angustifolius* crops, especially not at damaging levels. CIYVV is of interest, being readily present in close proximity to lupin crops where they coincide with *T. repens* pastures (Jones and McLean 1989). As mentioned before it has been reported to have undergone recombination with BYMV (Wylie and Jones 2009). It also causes a pod necrosis disease on snap bean (another grain legume) in the United States (Larsen *et al.* 2008).

8. **The interface between an ancient and a recent agroecosystem**

As well as often being grown in close proximity to annual pastures in the south-west of WA, *L. angustifolius* crops (part of the recent agroecosystem), also tend to be grown in equally close proximity to native bushland (an ancient ecosystem). This part of WA is within the South West Australian Floristic Region (SWAFR) which is one of the worlds’ 25 Global Biodiversity hotspots, the only one in Australia. Recorded so far are 7380 native vascular plants, 49% of them endemic and 2500 of conservation concern (Brooks *et al.* 2002; Hopper and Gioia 2004). The 302, 627km² of the SWAFR includes much of the lupin growing areas in the WA grain-belt and coastal areas where *L. angustifolius* crops are absent but ild naturalised lupins grow as weeds (Fig 2a,b). The SWAFR is also unique in that it was only recently that plant cultivation began in the region following colonization by Europeans in 1829. This introduction of agriculture threatened all forms of native plants and animals. Such threats as excessive land clearing has led to rising salinity, and habitat loss, wood collection and too frequent fires, and damage from introduced weeds and plant pathogens (Brooks *et al.* 2002; Hopper and Gioia 2004). Recent work on indigenous and introduced Australian potyviruses in WA has taught us that there is much yet to discover. In the WA environment, indigenous viruses have the potential to emerge from the natural vegetation and infect introduced plants and therefore crops, and introduced viruses can spread from introduced plants (e.g. *L. angustifolius*) to damage native plants (Webster *et al.* 2007; Jones 2009; Coutts *et al.* 2011; Vincent *et al.* 2014). Of particular interest is the indigenous virus *Hardenbergia mosaic virus* (HarMV) which caused necrosis in *L. angustifolius* plants in the glasshouse and in a controlled field situation (Webster *et al.* 2007; Jones 2009; Coutts *et al.* 2011, Luo *et al.* 2011).
Figure. 2 a) Boundaries of the Southwest Australian Floristic Region, b) lupin production zones in Western Australia consisting of 1 – high rainfall, north; 2 – medium rainfall, north; 3 – low rainfall, north; 4 – high rainfall, central and great southern; 5 – medium rainfall, central; 6 – medium rainfall, central and great southern; 7 – low rainfall, east; 8 – south coast; blank coastal area from Perth to Albany is not recommend for growing lupins but contains naturalised *Lupinus angustifolius*, *L. luteus* and *L. cosentinii* growing as wild plants. Source: DAFWA.

9. Next generation sequencing

Often referred to as “Next-gen sequencing” or NGS, this is fast becoming a popular method of obtaining whole plant-virus genomes in one reaction. As the technology advances, the sizes of the fragments increase and so does the amount of genome coverage in each reaction leading to a better resolved genome. More importantly, the costs continue to decrease as the technology is improved and uptake increases. As an example, in just nine years one of the most popular platforms (Illumina) has seen great improvement in terms of the amount of data that can be generated at any one time. It has grown from the first Solexa 1G machine which was capable of producing 1Gb of data, to the recently released NextSeq 500 which is capable of sequencing a 3GB Genome (e.g. Human genome) at 30 times coverage in just 29 hours (Mardis 2013; Illumina 2014).

NGS is changing the face of modern medicine, with applications extending far beyond whole genome sequencing and marching on into personalised medicine. The applications of this technology include rare variant detection, *de novo* mutant identification, somatic variant detection and the study of rare mendelian disorders.
Many of these applications occur in the field of cancer genomics which is leading the way in the use of these technologies (Koboldt et al. 2013).

For plant virologists, and plant pathologists generally, the uptake of these technologies has been somewhat slower. However, without a doubt these smaller fields of research have benefited greatly from the advances in technology and analysis made their colleagues in the medical sciences. As the technologies become cheaper and easier to access, virus discovery and even multiplex virus diagnostics are set to benefit greatly. The challenges for plant virologists lie not in accessing and using these technologies, but in obtaining the skills required to analyse and interpret the very large datasets they are suddenly able to access (Boonham et al. 2013).

It is clear that no one method (biological or molecular) is comprehensive enough to stand alone in terms of characterization of novel viruses. The relatively recent explosion in molecular technologies and the vast amounts of data retrievable from a single sample constitute fantastic advances. However, there is a cautionary tale to be told when making statements based on sequence data alone – as tempting and as easy as that may seem. This was pointed out in a publication entitled “98% identical, 100% wrong: percent nucleotide identity can lead plant virus epidemiology astray” (Duffy and Seah 2010). Discovery of new virus or virus-like sequences needs to be handled with care as new disease incursions can have implications for biosecurity and trade. Some have gone so far as to make recommendations regarding the identification of plant viruses from sequence data alone (with no biological data associated to back it up), and the potential plant biosecurity issues associated with them. These include the use of the term “uncultured virus” along with a tentative name for any plant virus sequence not associated with a recognised virus infection (MacDiarmid et al. 2013). Without collection and equal consideration of biological data, combined with studies on the epidemiology of a virus in different environments, the true value of the molecular data cannot be understood. There is room for both classical plant virology and new molecular methods to co-exist, and the combination of the two is something to be encouraged. These are important discussions and distinctions, desperately needed so
that researchers can maintain an open dialogue with risk analysts, biosecurity agencies and policy makers (MacDiamid et al. 2013).

10. Metagenomics, bioinformatics and plant virology

With the advent of NGS technologies, whole metagenomic analysis has become a practical approach for use in plant virology particularly with regards to cost. By analyzing the nucleotide sequence content of a plant, sequences of potential pathogens are produced without requiring previous knowledge of what is present, thereby avoiding having to narrow the search field, as might occur using generic PCR primers. This approach has worked successfully in both confirming the presence of known plant viruses in a sample, and in successfully identifying an unknown and previously undescribed virus, with one of the first examples being Gayfeather mild mottle virus from a sample of the plant *Liatris spicata* (Adams et. al. 2009). There are now multiple examples of plant and plant community metagenomic studies that have happened in the context of plant viruses (e.g. Wylie and Jones 2011; Roossinck 2012; Wylie et al. 2012, 2014).

There are many applications for the data obtained from NGS and downstream analyses. Phylogenetic analysis of whole genomes allows for inferences to be made regarding the evolutionary history of a gene, set of genes or a species and their relationships to others. The rate of evolution can also be estimated and the phylogeography of genes or species assessed. Having access to complete genomes of plant virus also makes recombination analysis more feasible. Molecular characterisation is an important component of new virus discovery, which is now more accessible than ever thanks to these technologies (Boonham et al. 2013).

11. Outline and specific aims of this thesis

It is important to discover the cause or causes of BPS so that appropriate disease management strategies can be devised, and to help local lupin breeders in their quest to produce BPS-resistant *L. angustifolius* cultivars. BPS-resistant cultivars would permit growers in south-west Australia who have ideal lupin growing conditions and currently suffer high incidences of BPS to grow high yielding lupin crops on a regular
basis. They could then capitalize on the economic potential that lupins have and the niche markets the industry has already begun to foster. The specific aims of this project are as follows:

- identify whether BYMV causes black pod syndrome and satisfy Koch’s postulates to establish whether it is the causal agent
- obtain BYMV isolates from lupin plants with BPS and determine their molecular and biological properties
- investigate whole genomes of BYMV isolates from plants with BPS through the use of next generation sequencing
- compare whole BYMV genomes from plants with BPS with genomes isolated from other sources
- explore the potential threat to *Lupinus* spp. from indigenous potyviruses
- deliver a set of recommendations and management plans to growers, advisers and the lupin industry
Literature cited


Gladstones, J.S., Atkins, C. and Hamblin, J. (Eds). Lupins As Crop Plants:
Biology Production and Utilisation. CABI International, Wallingford/New
York.


from Australia: biological properties and comparison of coat protein nucleotide


western New World: Derived evolution of perennial life history and

15. Duffy S. and Seah, Y.M. (2010) 98% identical, 100% wrong: percent
nucleotide identity can lead plant virus epidemiology astray. Philo. Trans. Roy.
Soc. B. 365, 1891-1897.

species. In: Proceedings of the 3rd International Lupin Conference, June 4-8,
La Rochelle, France, pp. 67-85.

17. Eastwood, R.J., Drummond, C.S., Schifino-Wittmann, M.T. and Hughes, C.E.
(2008) Diversity and evolutionary history of lupins – insights from new
wealth’ Proceedings of the 12th International Lupin Conference, 14-18
September 2008, Fremantle, Western Australia. International Lupin
Association, Canterbury, New Zealand.

health and wealth’ Proceedings of the 12th International Lupin Conference, 14-
18 September 2008, Fremantle, Western Australia. International Lupin
Association, Canterbury, New Zealand.


specific disease, pp 65-68. Department of Agriculture and Food, Western Australia. South Perth Australia.


Chapter 2

Black pod syndrome of Lupinus angustifolius is caused by late infection with Bean yellow mosaic virus

Accepted for publication in Plant Disease:


Note: A minor erratum will be published for this paper. Table 2, Growth stage 6, Mandelup control should read “0”, not “6”.
Black Pod Syndrome of *Lupinus angustifolius* Is Caused by Late Infection with Bean yellow mosaic virus

M. A. Kehoe, B. J. Buirchell, B. A. Coutts, and R. A. C. Jones, School of Plant Biology and Institute of Agriculture, Faculty of Science, University of Western Australia, Crawley, WA 6009, Australia; and Crop Protection and Lupin Breeding Branches, Department of Agriculture and Food Western Australia, Bentley Delivery Centre, Perth, WA 6983, Australia

**Abstract**


Black pod syndrome (BPS) causes devastating losses in *Lupinus angustifolius* (narrow-leaved lupin) crops in Australia, and infection with *Bean yellow mosaic virus* (BYMV) was suggested as a possible cause. In 2011, an end-of-growing-season survey in which *L. angustifolius* plants with BPS were collected from six locations in southwestern Australia was done. Tissue samples from different positions on each of these symptomatic plants were tested for BYMV and generic potyviruses by enzyme-linked immunosorbent assay and reverse-transcription polymerase chain reaction (RT-PCR). Detection was most reliable when RT-PCR with generic potyvirus primers was used on tissue taken from the main stem of the plant just below the black pods. Partial coat protein nucleotide sequences from eight isolates from BPS-symptomatic *L. angustifolius* plants all belonged to the BYMV general phylogenetic group. An initial glasshouse experiment revealed that mechanical inoculation of *L. angustifolius* plants with BYMV after pods had formed caused pods to turn black. This did not occur when the plants were inoculated before this growth stage (at first flowering) because BYMV infection caused plant death. A subsequent experiment in which plants were inoculated at eight different growth stages confirmed that BPS was only induced when *L. angustifolius* plants were inoculated after first flowering, when pods had formed. Thus, BYMV was isolated from symptomatic *L. angustifolius* survey samples, inoculated to and maintained in culture hosts, inoculated to healthy *L. angustifolius* test plants inducing BPS, and then successfully reisolated from them. As such, Koch’s postulates were fulfilled for the hypothesis that late infection with BYMV causes BPS in *L. angustifolius* plants.

The genus *Lupinus* (lupin) contains several annual species grown as cool-season grain legumes. These species are planted as rain-fed summer crops in regions of the world with cool temperate climates, as rain-fed winter crops in regions with Mediterranean-type climates, and as rain-fed wet season or irrigated crops at high altitude in tropical regions (13). Australia is currently the largest lupin producer, growing around 85% of the global crop (11,25). The southwestern Australian grainbelt is responsible for 95% of Australia’s production. 95% of this being *Lupinus angustifolius* (narrow-leaved lupin). *L. angustifolius* is also an important component of Australian crop rotations because it helps in the management of fungal root diseases of cereals, soil rehabilitation, and, due to its nitrogen-fixing capabilities, improving the yield of cereal crops grown in rotation with it (3,13).

Southwestern Australia has a Mediterranean-type climate with predominantly sandy soils suitable for *L. angustifolius* crops (2,11). Although yields are commonly 1 to 2 t/ha, the potential exists for 4 t/ha. Currently, one of the major reasons for poor yields is black pod syndrome (BPS), which was first noticed in the 1990s (12). *L. angustifolius* plants affected by BPS are seen late in the growing season, after first flowering, when they develop characteristic flat, black pods that have little or no seed. Damage to crops can be extensive, with losses of 40% reported (2). Possible causes of BPS that were hypothesized included (i) excessive vegetative growth leading to poor pod set, (ii) nutrient deficiencies, and (iii) infection with *Bean yellow mosaic virus* (BYMV) (family Potyviridae, genus Potyvirus). The excessive vegetative growth leading to poor pod set hypothesis was not supported in field experiments consisting of treatments with or without growth hormone applied to different *L. angustifolius* genotypes (32). Also, nutrient deficiency was disproved when full rates of fertilizer application had no effect on the levels of BPS in different *L. angustifolius* genotypes. BYMV is nonpersistently transmitted by aphids (16,21). Consequently, when two insecticides were applied every 2 weeks at double the recommended rates, levels of BPS were diminished, providing some support for the suggestion that BYMV vectored by aphids may be playing a role (31). However, BPS symptoms were sometimes present in the primary pods without any of the necrotic stem streaking symptoms usually associated with BYMV infection in *L. angustifolius*, and enzyme-linked immunosorbent assay (ELISA) tests failed to detect BYMV from the majority of BPS-symptomatic plant samples tested (R. A. C. Jones, unpublished).

In 1996 to 2003, *L. angustifolius* crop variety trials (CVTs) were rated for BPS at the end of the growing season, and BYMV was considered the most likely cause (12,31). However, annual field resistance screening experiments involving young *L. angustifolius* plants found no difference in susceptibility between BYMV-susceptible breeding lines and lines identified in CVTs as being partially resistant to BPS (18). Thus, rating young plants for BYMV resistance in the field gave different results from rating mature plants for BPS at the end of the growing season (after pod formation). CVTs in 2005 demonstrated a correlation between high levels of BPS and diminished grain yields, and a genetic basis to partial BPS resistance was suggested. Subsequently, a cultivar with partial BPS resistance was released under the name ‘Jenabillup’ (2.22). A small proportion of BYMV isolates found in the field overcame this resistance (4.5). When other BYMV isolates infect *L. angustifolius* plants early in the growing period, their interaction with *Nbr-I* causes young plants to die from systemic necrosis. This
removes the infected plant from the crop, so that it no longer acts as a source of infection for further within-crop virus spread (4.5,19). The most common BYMV source for virus spread to L. angustifolius crops in southwestern Australia is nearby infected legume pastures dominated by Trifolium subterraneum (subterranean clover). However, BYMV is not seedborne in L. angustifolius (16,21).

Other sources include neighboring infected L. angustifolius, virus-infected legume crops, and native legumes (16,24,26). Except for one isolate from a monocrop plantation breeding site, all BYMV isolates from southwestern Australia that have been sequenced are in the "general" phylogenetic group (12). Improvements to domesticated L. angustifolius cultivars include breeding for disease resistance (8). For viruses, breeding has focused on reducing levels of seed transmission of Pea enation mosaic virus (PeMV) (family Bromoviridae, genus Cucumovirus: 16,20).

Although the NbSn-I gene is present in all cultivars of L. angustifolius evaluated thus far, this is not the consequence of active plant breeding (4.5,18,22). Finding the genetic basis of partial resistance to BYMV is of great interest to L. angustifolius breeding programs (12). This article describes research to establish the cause of partial resistance to BYMVMI by testing the hypothesis that BYMV is not transmitted by late infection of BYMV-MI. It also describes research to identify the most reliable approach for detecting BYMV within BPS-symptomatic plants, and where BYMV isolated from BPS-symptomatic plants fit within their phylogenetic groupings.

Materials and Methods

Glasshouse grown plants, inoculations, and virus isolate culture. L. angustifolius plants were cultivated in a glasshouse. Plants of L. angustifolius were grown in washed river sand and plants of Nicotiana benthamiana and T. subterraneum ‘Woogonup’ in a sterile-soil mixture. For mechanical inoculation to maintain culture isolates as part of experiments, virus-infected leaves from recently systemically infected plants were ground in 0.1 M phosphate buffer, pH 7.2, and the infective sap was filtered through a needled sieve. Cultures of virus isolates were maintained by serial mechanical inoculation of infective sap to plants of N. benthamiana or T. subterraneum. Isolate BYMVMI was obtained from freeze-dried leaf material from one of the original plants used in the present study. The isolate was used as the positive control in ELISA and reverse-transcription polymerase chain reaction (RT-PCR), and as inoculum for the experiments. An isolate of Cucumber yellow vein virus (CYVV) (family Potyviridae, genus Potyvirus) from New South Wales, Australia, was from freeze-dried leaf material obtained from the West Australian Plant Pathogen Culture Collection (WAC10102). It was mechanically inoculated to plants of T. subterraneum and maintained in this species, as with the BYMV cultures, and used as a positive control for ELISA and RT-PCR testing.

ELISA. For testing by ELISA, samples were extracted (1 g per 20 ml) in phosphate-buffered saline (10 mM potassium phosphate, 150 mM sodium chloride [pH 7.4], Tween 20 at 0.05 ml/liter, and polyvinyl pyrrolidone at 20 g/liter) using a mixer mill (Retsch). Sample extracts were tested for BYMV or CYVV by double-antibody sandwich ELISA, as described by Clark and Adams (6). For generic potyvirus testing, samples were extracted in 0.05 M sodium carbonate buffer, pH 9.6, and tested using the antigen-coated indirect ELISA protocol of Torrance and Pedley (29). The polyclonal antisera to BYMV was from DSMZ, to CYVV from Agdia, and to generic potyvirus from Agdia. All samples were tested in duplicate wells in microtiter plates. Sac from BYMV or CYVV-infected and healthy T. subterraneum leaf samples was incubated in paired wells to provide positive and negative controls. The substrates was p-nitrophenyl phosphate at 1.0 mg/ml in diethanolamine, pH 9.8, at 100 ml/liter. Values for absorbance at 405 nm were measured in a microplate reader (Bio-Rad Laboratories). Absorbance values of positive samples were always more than three times those of the healthy sap control.

RT-PCR and sequencing. For testing by RT-PCR, total RNA was extracted from each sample using a Spectrum Plant Total RNA kit (Sigma-Aldrich), according to the manufacturer’s instructions. Reverse transcription was performed using Improm-II reverse transcriptase (Promega), according to the manufacturer’s instructions, using the random primers provided. PCR was performed with the GoTaq green master mix (Promega), according to the manufacturer’s instructions. Primers used for BYMV were specific to the BYMV-MI genome. Primers to identify BYMV were used by Wylie et al. (33), generic potyvirus identification primers LepPotF and LepPotR from Webster (30), and CYVV primers CYVV-F and CYVV-R from Larsen et al. (23). The PCR products of eight samples were sequenced using an Applied Biosystems 3730 DNA analyzer with Big Dye Terminator 3.1 chemistry. Sequence trimming and analysis was done using Geneious Pro 5.6.2. Partial coat protein (CP) sequences were submitted to GenBank with the accession numbers KF823008 to KF823015. In addition to the two new CPs, 15 other CP sequences selected to represent each of the seven different phylogenetic groups of BYMV, were retrieved from GenBank (33). Sequences were aligned with Clustal W, and neighbor-joining phylogenetic trees with a bootstrap value of 1,000 were constructed using MEGA 5.0 (28).

Survey for BPS and comparison of virus testing methods. A late-growing-season survey to obtain samples of BPS-symptomatic L. angustifolius plants was conducted in southwestern Australia in September and October 2011. Survey locations were six farms: three at Gibson and one at each at Arthur River, Fingelly, and Wongan Hills (Fig. 1). Samples were from (i) crops of ‘Jemabilup’ at two farms at Gibson and (ii) L. angustifolius CVs at experimental plots for screening L. angustifolius accessions for various traits at the remaining farms. No BPS-symptomatic plants were found at Wongan Hills. In total, 72 symptomatic whole plants were collected from the other five locations. To assess the most appropriate sampling position for reliable detection of BYMV in these samples, two to five positions on the shoots of each symptomatic whole plant were sampled. Every plant was sampled in the middle of the main stem (typically still green but BPS-symptomatic at the top of the main stem near the primary black buds. Only 48 of the symptomatic plants had not yet shed their youngest leaves and these were tested when still present. Seven plants from Gibson 6 BPS with minimal or no stem streaking still had their older leaves present and these, along with their petioles, were sampled. In total, 206 samples from the 72 symptomatic plants were tested by ELISA and RT-PCR for generic potyvirus, BYMV, and CYVV.

Induction of BPS by BYMV inoculation. An initial greenhouse experiment was conducted to determine (i) whether late BYMV infection could induce BPS in plants of L. angustifolius ‘Mandelpit’ (BPS susceptible) and (ii) whether BYMV from a source other than BPS-symptomatic L. angustifolius plants could induce BPS. Five plants were mechanically inoculated with BYMV-MI at first flowering (i.e., flowering on the main stem) and another five plants 2 weeks later, as primary pods were forming on the main stem. Five plants were left uninoculated as healthy controls. Once symptoms had formed, young leaf samples from each plant were tested individually by ELISA for BYMV.

Effect of plant growth stage at time of infection. A further greenhouse experiment was done with L. angustifolius ‘Mandelpit’ (BPS susceptible) and Jemabilup (partially BPS resistant) to assess the importance of growth stage at the time of inoculation with a BYMV isolate from a BPS-symptomatic plant. Growth stage at the time of inoculation was determined based on the descriptions given by Dracup and Kirby (9). Eight plants were mechanically inoculated with BYMV isolate AR 535C at 2-week intervals beginning at 740 Plant Disease | Vol. 98 No. 6
6 weeks after sowing. At the time of inoculation, the first growth stage had only leaflets present. The subsequent growth stages at each 2-week interval consisted of stem elongation; bad formation; buds and initial flowers; first flowering and primary pods; full primary pods and secondary flowers; full primary and secondary pods; with tertiary flowers; and the inoculations finished after 20 weeks, with the final growth stage, where pods were forming on tertiary shoots but pods on the primary and secondary shoots were fully formed (9). At each stage and for each cultivar, eight plants were mock inoculated with healthy leaf sap from a T. suberrealis plant to act as controls. Young leaf samples were taken from plants inoculated at growth stages 1 to 5 and tested individually by ELISA for BYMV and generic potyvirus. Stem material from just below the black pods on the main stem was sampled from plants inoculated at growth stages 6 to 8 and tested individually by RT-PCR for BYMV and generic potyvirus. Testing occurred once symptoms appeared, with this being 2 to 3 weeks after inoculation for stages 1 to 5, where the symptoms presented as typical early BYMV infection on both the leaf and stem, and 4 to 6 weeks after inoculation for stages 6 to 8, where the symptoms presented as black pod symptoms.

Results

Survey for BPS, ELISA, and RT-PCR testing and isolate sequencing. Regardless of sampling location on BPS-symptomatic plants, the greatest numbers of virus detections were achieved using generic potyvirus primers by RT-PCR (Table 1). In particular, the main stem just below the black pods was the most reliable sampling location, giving a positive result with 71 of 72 samples. When the same samples were tested by RT-PCR using BYMV-specific primers, 59 of 72 gave positive results, including the single sample that was not detected with generic potyvirus primers. ELISA testing of the same samples was less reliable, giving positive results for 54 of 72 samples with generic potyvirus antibodies and 60 of 72 samples with BYMV-specific antibodies. The remaining sampling locations were not as reliable or suitable for testing because detections varied greatly, depending on whether generic potyvirus antibodies for ELISA, BYMV-specific antibodies for ELISA, generic potyvirus primers for RT-PCR, or specific BYMV primers for RT-PCR were used. When the midstem samples were tested, positive detections were 24 to 44 of 72 for ELISA and 42 to 64 of 72 for RT-PCR. When testing younger leaves in plants in which they had not yet been shed, positive detections were 1 to 9 of 48 for ELISA and 23 to 46 of 48 for RT-PCR. Testing of young leaf petioles gave positive detections in 0 to 1 of 7 for ELISA and 2 to 5 of 7 for RT-PCR. With older leaves, there were no detections by ELISA and 1 to 7 of 7 by RT-PCR. CTVV was never detected by ELISA or RT-PCR in any of the samples tested.

Direct nucleotide sequencing was performed on RT-PCR products from six samples that gave a positive result by RT-PCR with generic potyvirus primers only and two others which gave positive results with both generic potyvirus primers and BYMV-specific primers. When the eight new partial CP sequences were compared with partial sequences of 15 isolates from GenBank, a neighbor-joining 1,000 bootstrap phylogenetic tree placed all eight into the BYMV general group (Fig. 2).

Inoculation of BPS by BYMV inoculation. In the initial experiment, L. angustifolius ‘Mandelp’ plants inoculated with BYMV-MI at first flowering produced symptoms of bending of the shoot

<table>
<thead>
<tr>
<th>Sample location on plant, format</th>
<th>Virus target</th>
<th>Detected/total (×)</th>
<th>Detection (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stem just below black pods</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELISA</td>
<td>Potyvirus</td>
<td>54/72</td>
<td>75</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>Potyvirus</td>
<td>71/72</td>
<td>99</td>
</tr>
<tr>
<td>BYMV</td>
<td></td>
<td>59/72</td>
<td>82</td>
</tr>
<tr>
<td>Midstem</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELISA</td>
<td>Potyvirus</td>
<td>24/72</td>
<td>33</td>
</tr>
<tr>
<td>BYMV</td>
<td></td>
<td>44/72</td>
<td>61</td>
</tr>
<tr>
<td>Old leaf</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELISA</td>
<td>Potyvirus</td>
<td>0/7</td>
<td>-</td>
</tr>
<tr>
<td>BYMV</td>
<td></td>
<td>1/7</td>
<td>-</td>
</tr>
<tr>
<td>Young leaf petiole</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELISA</td>
<td>Potyvirus</td>
<td>0/7</td>
<td>-</td>
</tr>
<tr>
<td>BYMV</td>
<td></td>
<td>1/7</td>
<td>-</td>
</tr>
<tr>
<td>RT-PCR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BYMV</td>
<td></td>
<td>1/7</td>
<td>-</td>
</tr>
</tbody>
</table>

* In total, 72 BPS-symptomatic L. angustifolius plants were collected from five farms in southwestern Australia and each plant was tested individually at two to five sample locations on the plant. Testing was by enzyme-linked immunosorbent assay (ELISA) with BYMV-specific and generic potyvirus antibodies and by reverse-transcription polymerase chain reaction (RT-PCR) with BYMV-specific and generic potyvirus primers.

* Number of samples with BYMV detected per number of samples tested.
apex (i.e., a 'shepherd's crook' appearance), leaf death, severe stem streaking, and necrosis of the stem and growing tips. All five plants inoculated were dead 2 weeks after inoculation. In contrast, four of five plants inoculated when their primary pods had formed developed blackening of their primary pods (i.e., typical BPS symptoms) and necrotic stem streaking (Fig. 3). The remaining plant developed very small pods on the main stem (because the pods were just forming rather than fully formed at the time of inoculation), and these became necrotic and aborted. BPS appeared within 4 weeks of inoculation. Leaf samples taken from all inoculated plants tested positive to BYMV by ELISA. The five un inoculated plants developed no symptoms and no BYMV was detected by ELISA.

Effect of growth stage at time of BYMV inoculation. Depending on growth stage at the time of inoculation with BYMV-AR93C, two to eight of eight plants for *L. angustifolius* 'Jenabilup' and four to eight of eight plants of 'Mandelup' became infected (Table 2). With both cultivars, the first five growth stages were all inoculated before pods were present on the main stem. All infected plants developed BYMV symptoms as described for inoculations at first flowering in the initial experiment. At growth stage 5 (first flowering), any pods that developed after inoculation became necrotic and never formed properly. All infected plants from inoculations at growth stages 1 to 5 died. Inoculation at growth stages 6 to 8 produced symptoms of BPS, including black pods and necrotic stem streaking. It took longer for systemic symptoms to appear in these plants than at growth stages 1 to 5, and they were not always killed by the infection (Fig. 4). Although BPS symptoms were the same for plants of 'Mandelup' and 'Jenabilup', they appeared sooner after inoculation in 'Mandelup' than in 'Jenabilup'.

Discussion

This study shows that late infection with BYMV causes BPS in *L. angustifolius* plants. The survey results satisfy the first condition of Koch’s postulates by confirming that “the pathogen is regularly associated with the disease”. The second condition states that “the pathogen should be isolated from diseased plants and grown in culture”. This was accomplished successfully by culturing isolates in *T. subterraneum* plants. The combined results of the initial greenhouse experiment and the experiment investigating the effect of growth stage at the time of inoculation satisfy the third condition, “that the disease should be reproduced when a culture of the pathogen is used to introduce disease to a healthy susceptible host”. Late inoculation with BYMV-MI, which was first isolated from a non-*L. angustifolius* source (*Mellotia indica*), or BYMV-

Fig. 2 Neighbor-joining relationship phylogram obtained from the alignment of eight partial coat protein sequences (431 bp in length) belonging to eight bean yellow mosaic virus (BYMV) isolates collected from black pod syndrome (BPS)-symptomatic *Lupinus angustifolius* plants in southwestern Australia in 2011. Isolates from this study are highlighted in gray. The other 15 isolate sequences shown were selected from GenBank to represent the seven different BYMV phylogenetic groups (33). The alignment was generated using ClustalW and tree branches were bootstrapped with 1,000 replications. The tree was rooted with Clover yellow vein virus (CYVV) accession number NC003536. Superscript “a” denotes isolates that were amplified using generic potyvirus primers only and superscript “b” denotes isolates that were amplified using both generic potyvirus primers and specific BYMV primers.

742 Plant Disease / Vol. 96 No. 6
AR93C isolated from a L. angustifolius plant with BPS, both induced BPS in healthy L. angustifolius plants when inoculated after pot inoculation. A positive result in 72% of BYMV inoculation, the presence of BYMV was confirmed in L. angustifolius plants by ELISA or RTPCR, thus satisfying the fourth condition, that “the pathogen should be detected in or reisolated from inoculated plants”.

From the 72 BPS-symptomatic L. angustifolius survey samples, 12 were positive only using RTPCR with generic potyvirus primers. Five of these isolates were sequenced and confirmed to be BMYV only. These represent four of the five locations where plants were sampled. Although not sequenced, it seems reasonable to assume here that the remaining seven isolates were also BYMV and therefore represent a single species present. Three other isolates from samples that gave positive results using both BYMV and generic potyvirus primers were also sequenced and confirmed to be BMYV only. The single sample which gave a positive result with BYMV-specific primers but not with generic potyvirus was a sample of midstream material. It was not chosen for sequencing because the sample taken from just below the black pods on the same plant returned a positive result with both generic potyvirus and specific BMYV primers. We found no evidence of the related Potyvirus CIYVV or any other virus occurring in conjunction with BPS-symptomatic plants in the crops and CVTs surveyed here. However, the possibility that other Potyvirus species might also induce BPS occasionally cannot be discounted (e.g., when CIYVV spreads from T. repens [white clover] pastures to L. angustifolius in other parts of Australia; 17,21). The method used here involves carefully peridial inoculation into six BYMV groups evolved during plant domestication in different parts of the world and are specialist groupings adapted to distinct types of cultivated plant hosts; for example, the “lapin” group is adapted to broad-leaved lupin species domesticated in the Mediterranean region, such as L. latifolius (yellow lupin) and L. albus (white lupin) (33). Interestingly, all the isolates of BYMV collected from BPS-symptomatic L. angustifolius plants were from the general group and none of the specific groups, such as the lupin group. BPS symptoms from this general group are not known to be seedborne in L. angustifolius, which is not surprising because the plants are killed (16,21). If possible, future research should focus on isolates from these specialist groupings in similar BPS-BYS experiments, along with isolates of the general group that overcome the Nbn-1 gene. It would also be of interest to compare the complete genomes

<table>
<thead>
<tr>
<th>Growth stage</th>
<th>Control or inoculated</th>
<th>Symptoms</th>
<th>Jenahillp</th>
<th>Mandelupp</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Leaflets</td>
<td>Control</td>
<td>NS</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Inoculated</td>
<td>SC, NSS, LD, SN</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>2. Stem elongation</td>
<td>Control</td>
<td>NS</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Inoculated</td>
<td>SC, NSS, LD, SN</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>3. Bud formation</td>
<td>Control</td>
<td>NS</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Inoculated</td>
<td>SC, NSS, LD, SN</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>4. Buds and initial flowers</td>
<td>Control</td>
<td>NS</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Inoculated</td>
<td>SC, NSS, LD, SN</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>5. First flowering and primary pods</td>
<td>Control</td>
<td>NS</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Inoculated</td>
<td>SC, NSS, LD, SN</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>6. Full primary pods, secondary flowers</td>
<td>Control</td>
<td>NS</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Inoculated</td>
<td>NSS, BPS, LD, SN</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>7. Full primary and secondary pods, tertiary flowers</td>
<td>Control</td>
<td>NS</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Inoculated</td>
<td>NSS, BPS, LD, SN</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>8. Fully podded primary and secondary, and some tertiary pods</td>
<td>Control</td>
<td>NS</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Inoculated</td>
<td>NSS, BPS, LD, SN</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

* Eight plants each of 'Jenahillp' and 'Mandelupp' were mechanically inoculated at 2-week intervals with BYMV isolate AR93C at eight growth stages, beginning with growth stage 1 at 6 weeks after sowing. New leaf samples from each plant were tested individually by enzyme-linked immunosorbent assay (ELISA) for growth stages 1 to 5. Samples taken from the main stem just below the primary pods were tested by reverse-transcription polymerase chain reaction for growth stages 6 to 8. Eight control plants of each cultivar were inoculated with healthy sap at each growth stage.

* Growth stages based on descriptions given in Drew and Kirby (9).

* Cooded symptom descriptions: SC = shepherds crook, M = mosaic, LD = leaf drop, NSS = necrotic stem streaking, SN = systemic necrosis and plant death, BPS = black pod syndrome, BPSN = new pods become necrotic and aborted, and NS = no symptoms.

* Number of plants out of eight inoculated plants in which BYMV was detected by ELISA testing with BYMV-specific antibodies and generic potyviruses antibodies.

Table 2. Symptoms and detection of Bean yellow mosaic virus (BYMV) in plants of two cultivars of Lupinus angustifolius inoculated at eight growth stages*
of such isolates with the complete genomes of BYMV isolates from BPS-symptomatic *L. angustifolius* plants collected as part of this study.

Although ‘Jenabilup’ proved less susceptible to BPS than ‘Mandelup’ in experimental field situations (2), this was not reproduced in the glasshouse studies. However, symptoms developed more rapidly after inoculation in ‘Mandelup’ plants than in ‘Jenabilup’ plants inoculated at the same time but the final BPS symptoms and their severity were the same. The differences between the field and the glasshouse findings may be due to the use of mechanical inoculation instead of natural inoculation by aphids. Alternatively, the differences may be due to the presence or absence of various modifier genes influencing *Nbm-1* (5,22). Another possibility is that greater amounts of aphid-inoculation pressure may be required to infect some cultivars than others, such as occurs with BWYV transmission by aphids to different cultivars of *Brassica napus* (7). Aphids might also have a feeding preference for ‘Mandelup’ over ‘Jenabilup’ but further research involving aphid inoculations would be required to investigate these scenarios.

Late-season infection of *L. angustifolius* plants with BYMV resulting in BPS seems likely to result from an interaction between the *L. angustifolius* hypersensitive resistance gene *Nbm-1* and BYMV. Thus, differences in expression of necrotic symptoms in young (plant death) versus old (BPS) plants probably result from plant growth stage at time of infection in the presence of *Nbm-1*. When late infection occurs, not all mature plants display necrotic symptoms on all branches (16,21). Possibly, this could be a consequence of physiological changes occurring in the plant at and after flowering or to mature plant resistance in the presence of *Nbm-1* slowing systemic movement in mature plants (1).

There are already integrated disease management recommendations in place for BYMV in *L. angustifolius* in Australia, such as using perimeter nonhost barriers, avoiding fields with large perimeter/area ratios, retaining stubble ground cover, promoting early

---

**Fig. 3.** A, Portion of stem of a *Lupinus angustifolius* ‘Mandelup’ plant with black pod syndrome (BPS) in primary pods and necrotic stem streaking following mechanical inoculation with bean yellow mosaic virus (BYMV) isolate MR3C once the primary pods had formed; B, comparison of the apical portions of a *L. angustifolius* ‘Mandelup’ plant mechanically inoculated with BYMV-MR (left) and an un inoculated control plant (right); C, comparison of opened pods and seed within them produced by the plants shown in B (BPS [left] and healthy [right]); D, comparison of cross sections through seed from the pods shown in C (BPS [left] and healthy [right]).

**Fig. 4.** Apical portion of a *Lupinus angustifolius* ‘Mandelup’ plant inoculated with bean yellow mosaic virus isolate MR3C once primary pods had formed. Note black pods and stem necrosis typical of black pod syndrome (left) compared with healthy (right).
crop canopy development, generating high plant densities, close row spacing, maximizing weed control, and using crop rotation (16). However, due to the critical importance of growth stage at the time of infection (i.e., podding) for BIPS, in this instance, we place particular emphasis on maximizing weed control within the crop (especially cover pasture volunteers) and avoiding planting next to or downstream of legume pastures, other L. angustifolius crops, other grain legume crops, and native legumes that are likely to contain sources of BYMV.

Acknowledgments

This research was funded by an Australian Postgraduate Award (APA), and an Australian Chicks Research and Development Corporation (GHRDC) Stipendship. Project number GIS100039. It was undertaken using the facilities at the Department of Agriculture and Food Western Australia. We thank P. White for helpful discussions about his earlier studies on BIPS, and E. Gajko, S. Vincent, and M. Banovic for glasshouse and laboratory support.

Literature Cited


Chapter 3

Hardenbergia mosaic virus: crossing the barrier between native and introduced plant species.

Published paper:
Short communication

Hardenbergia mosaic virus: Crossing the barrier between native and introduced plant species

M.A. Kehoe, B.A. Coutts, B.J. Buirchell, R.A.C. Jones

*School of Plant Biology and Institute of Agriculture, Faculty of Science, University of Western Australia, Crawley, WA 6009, Australia
+Crop Protection and Lupin Breeding Branch, Department of Agriculture and Food Western Australia, Bentley Delivery Centre, Perth, WA 6902, Australia

ABSTRACT

Hardenbergia mosaic virus (HarMV), genus Potyvirus, belongs to the bean common mosaic virus (BCMV) potyvirus lineage found only in Australia. The original host of HarMV, Hardenbergia comptoniata, family Fabaceae, is indigenous to the South-West Australian Floristic Region (SWAFR), where Lupinus spp. are grown as introduced grain legume crops, and exist as naturalised weeds. Two plants of H. comptoniata and one of Lupinus condensatus, each with mosaic and leaf deformation symptoms, were sampled from a small patch of disturbed vegetation at an ancient ecosystem–recent agroecosystem interface. Potyvirus infection was detected in all three samples by ELISA and RT-PCR. After sequencing on an Illumina HiSeq 2000, three complete and two nearly complete HarMV genomes from H. comptoniata and one complete HarMV genome from L. condensatus were obtained. Phylogenetic analysis which compared (i) the four new complete genomes with the three HarMV genomes on Genbank (two of which were identical), and (ii) coat protein (CP) genes from the six new genomes with the 38 HarMV CP sequences already on Genbank, revealed that three of the complete and one of the nearly complete new genomes were in HarMV clade I, one of the complete genomes in clade V and one nearly complete genome in clade VI. The complete HarMV genome from L. condensatus differed by only eight nucleotides from one of the HarMV clade I genomes from a nearby H. comptoniata plant, with only one of these nucleotide changes being non-synonymous. Fairwise comparison between all the complete HarMV genomes revealed nucleotide identities ranging between 82.2% and 100%. Recombination analysis revealed evidence of two recombination events amongst the six complete genomes. This study provides the first report of HarMV naturally infecting L. condensatus and the first example for the SWAFR of virus emergence from a native plant species to invade an introduced plant species.

Crown Copyright © 2014. Published by Elsevier B.V. All rights reserved.
isolated from native plants or naturalised weed species introduced as potential pasture species, apart from passionfruit woodiness virus (PWV) from cultivated Passiflora spp. (Gibbs et al., 2008; Coutts et al., 2011). HarMV is the best studied member of this lineage. At the CP level, HarMV is most closely related to six other potyviruses found only in Australia: clitoria chlorosis virus (CChV), hibbertia virus Y (HyVY), siratro 1 virus Y (Si1VY), siratro 2 virus Y (Si2VY), passiflora mosaic virus (PpMV) and PWV (Coutts et al., 2011).

*Lupinus* spp. were introduced to the SWAFR around the early 1900s where they were used as sheep feed initially, and later domesticated as grain legume crops for rotation with cereals (French et al., 2008). *L. angustifolius* (narrow-leafed lupin), *L. cosentini* (sandplain lupin), *L. luteus* (yellow lupin) and *L. mutabilis* (pearl lupin) became infected by HarMV experimentally in a glasshouse environment (Webster et al., 2007). Also, naturally occurring aphids spread HarMV from introduced *H. comptoniae* infected plants to *L. angustifolius* plants growing in experimental field plots (Luo et al., 2011). However, except within these field plots, HarMV has not been found infecting any *Lupinus* spp. or other introduced plant species naturally. We therefore investigated an introduced lupin-*H. comptoniae* interface scenario involving an ancient ecosystem (i.e. native Australian plants) and a recent agroecosystem (i.e. introduced species) in the SWAFR (Webster et al., 2007; Jones, 2009; Alexander et al., 2014; Vincent et al., 2014). As such, we report the first detection of HarMV infecting *L. cosentini* naturally, and the first example of an indigenous virus effectively crossing the barrier between native and introduced species in the SWAFR. We also present four new complete and two nearly complete HarMV genome sequences, including one complete sequence from *L. cosentini* and the remaining sequences from *H. comptoniae*.

Leaf tissue from two *H. comptoniae* and one *L. cosentini* plant all showing leaf mosaic and deformation symptoms, were collected at the agro-ecological interface from a patch (50 m × 5 m) of disturbed native vegetation surrounding experimental field plots at Medina near Perth, Western Australia (WA). Within the patch, the *L. cosentini* plant sampled was growing as a naturalised weed in close proximity (4 m) to the sampled *H. comptoniae* plants.

The samples were tested with generic potyvirus monoclonal antibodies (Agdia, USA) using the antigen-coated indirect ELISA protocol of Torrance and Read, 1985. Absorbance values (A405) were regarded as positive when more than three times those of the healthy sap control. For testing by RT-PCR, total RNA was extracted using a Spectrum Total RNA kit (Sigma-Aldrich, Australia) according to manufacturers' instructions. Reverse transcription was performed with Improb-II reverse transcriptase (Promega, Australia) using the random primers provided according to manufacturer's instructions. PCR was performed with the GoTaq green master mix (Promega, Australia). PCR primers for generic potyvirus identification were from Webster, 2008.

Total RNA from each potyvirus positive sample was sent to the Australian Genome Research Facility (AGRF) for library preparation and barcoding (24 samples per lane) before 100 bp paired-end sequencing on an Illumina HiSeq2000. For each sample, reads were first trimmed using CLC Genomics Workbench 6.5 (CLC bio) with the quality scores limit set to 0.01, maximum number of ambiguities to two and removing any reads with <30 nt. Contigs were assembled using the de novo assembly function of CLC Genomics Workbench with automatic word size, automatic bubble size, minimum contig length 500, mismatch cost two, deletion cost three, length fraction 0.5 and similarity fraction 0.9. Contigs were sorted by length and the longest subjected to a BLAST search (Altschul et al., 1990). In addition, for sample MD2 and MD3 reads were also imported into Geneious 6.1.6 (Biomatters) and provided with a HarMV reference sequence obtained from Genbank.
Fig. 1. Neighbour-joining relationship phylogenies obtained from alignment of complete and nearly complete hadendorphia mosaic virus (HarMV) genomes trimmed to the length of the shortest nearly complete HarMV genome from this study (MD4-B with 9,006 nt). The alignments were generated in MEGA 5.2.1 using ClustalW and tree branches were bootstrapped with 1000 replications. The trees were rooted with a sequence of passionfruit woodiness virus (PWV) accession number NC01478, which is the closest relative to HarMV from the Australian potyviruses group with a complete genome. Complete HarMV genomes with isolates retrieved from Genbank shown in bold text and new isolates from this study shown in grey.

[Table and Figure Description]

(NCO15394). Mapping was performed with minimum overlap 10%, minimum insertion/deletion gap 4 and a dynamic programming set to iterate up to 10 times. A consensus between the contig of interest from CLCGW and the consensus from mapping in Geneious was created in Geneious by alignment with Clustal W. Open reading frames (ORFs) were predicted and annotations made using Geneious.

The new sequences were aligned by Clustal W in MEGA 5.2.1 with those retrieved from Genbank prior to phylogenetic analysis (Tamura et al., 2011). Phylogenetic analysis compared (i) the four new HarMV genomes with the three HarMV genomes available on Genbank (two of which were identical), (ii) all available genomes after trimming to the length of the shortest genome (MD4-B with 8,868 nt), (iii) complete CP genes from the five of the new genomes with the 38 HarMV CP sequences already on Genbank, and (iv) all available CP genes after trimming to the length of the shortest nearly complete HarMV CP from this study (MD4-B with 404 nt). Neighbour-joining trees were made using the number of differences model with a bootstrap value of 1000, Maximum Likelihood trees using the Tamura-Nei model with a bootstrap value of 1000, and Minimum Evolution trees using the number of differences model with a bootstrap value of 1000. Tables of percentage nt differences were calculated for the complete genomes using the pairwise comparison function with the number of differences model. Final sequences were submitted to Genbank with their Accession numbers KJ152152-KJ152157 (Table 1). The BDP4 package (Martin et al., 2010) was used to detect recombination between six HarMV whole genomes and one PWV whole genome. Default parameters were used for the six programmes implemented within BDP: BDP (Martin and Rzhetsky, 2000), GENECONV (Fadda et al., 1999), Bootscans (Martin et al., 2005), Macchi (Maynard Smith, 1992), Chi- maera (Posada and Crandall, 2001), Tseng (Boni et al., 2007) and SiScans (Gilbert et al., 2000). A recombination pattern was considered acceptable if detected by four or more of these programmes (Oshina et al., 2002; Wylie and Jones, 2009).

Two samples from H. comptonii (MD2 and MD4) and one from L. coventrini (MD3) returned a generic potyvirus positive result from testing by ELISA using generic potyvirus antibodies and RT-PCR using generic potyvirus primers.

The overall numbers of 100 nt reads obtained from sequencing on the Illumina Hiseq2000 were 17,616,843 (MD2), 13,154,290 (MD3) and 12,538,704 (MD4) (Table 1). After sequences were trimmed the remaining reads were 17,182,534 (MD2), 12,840,928 (MD3) and 12,239,175 (MD4). The numbers of contigs produced after de novo assembly were 1,681, 677 and 1,806 for samples MD2, MD3 and MD4, respectively. Blastn analysis for MD2 revealed a contig of 9672 nt, with an average coverage of 5072 times and constructed from 499,921 reads which was most closely related to HarMV accession no. HQ161081. MD3 Blastn analysis revealed a contig of 9665 nt, with an average coverage of 3516 times and constructed from 347,261 reads which was most closely related to HarMV (HQ161081). For MD4, Blastn analysis revealed four contigs of interest (i) MD4-A with a contig of 9648 nt, with average coverage 2573, constructed from 282,188 reads which was most closely related to HarMV (HQ161081); (ii) MD4-B with a contig of 8975 nt, with average coverage of 194, constructed from 17,748 reads which was most closely related to HarMV (HQ161081); (iii) MD4-C with a contig of 9635 nt, with average coverage of 345, constructed from 33,716 reads which was most closely related to HarMV (HQ161081); and (iv) MD4-D with a contig of 9600 nt, with average coverage of 216, constructed from 20,120 reads which was most closely related to HarMV (HQ161080). MD2 and MD3 were subjected to reference assembly against HarMV accession number NC015394 obtained from Genbank. MD2 produced a contig of 9751 nt with average coverage of 5354 times and constructed from 522,746 reads. MD3 produced a contig of 9726 nt with average coverage 3612 times and constructed from 352,433 reads. These results are comparable with those from de novo assembly of these two samples. The final sequences obtained were complete genomes with lengths of 9647–9688 nt, with the exception of MD4-A (missing 114 nt from the 5’ UTR region) and MD4-B (missing 91 nt from the 5’ UTR region and 675 nt from the 3’ UTR region). Final sequences for MD2 and MD3 consisted of a consensus of the contig from de novo assembly and the mapped consensus sequence. Final sequences for MD4-A to MD4-D consisted of their de novo contigs only.

Phylogenetic analysis with complete genomes alone placed MD2, MD3 and MD4-C into HarMV clade I while MD4-D proved to be the first whole genome representing HarMV clade V. When all genomes were clustered either to trim to the length of the shortest genome (8906 nt), the groupings were the same with MD4-B placed in clade I and MD4-A in clade VI (Fig. 1). When their complete CP genes were analysed, MD2, MD3 and MD4-C were placed in clade I, MD4-D in clade V and MD4-A in clade VI. When all CP genes were analysed after trimming to the length of the shortest sequence (404 nt), the groupings were the same with MD4-B placed in clade I (Fig. 2). One exception was HQ161080, which was placed in clade IV when complete CPs were analysed, but in clade V for partial CPs. When all genomes or coat proteins were analysed using Maximum Likelihood or Minimum Evolution methods, the tree topologies shown were the same as the Neighbour-joining method. A pairwise comparison of the nt percentage identities of the complete genomes revealed identities ranging between 82.2% and 100%. When the PWV sequence used as an outgroup in the phylogenetic analysis was included in the comparison, percentage
Fig. 2. Neighbour-joining relationship phylogenies obtained from alignment of complete and partial humbenjia mosaic virus (HarMV) coat protein (CP) sequences trimmed to the length of the shortest nearly complete HarMV CP from this study (MD4-B with 404 nt). The alignments were generated in MEGA 5.2.1 using ClustalW and tree branches were bootstrapped with 1000 replicates. The trees were rooted with a sequence of passionfruit woody mosaic virus (PWM) accession number NC_004575, which is a close relative of HarMV from the Australian psyllid group. Isolates retrieved from Genbank that were part of a complete genome sequence shown in bold text and new isolates from this study shown in grey.

identities between PWV and the HarMV sequences ranged from 68.6–69.1%. Sequence MD3 from *L. coventinii* was 99.9% identical to that of MD2, which was from a *L. coventinii* plant growing nearby (approx. 4 m). MD3 had just eight separate nt differences from MD4, only one of which was non-synonymous with a Serine instead of the Glycine at the position 6799 in the nt sequence (positioned within the Nia-Pro gene). The other HarMV sequences from this same site had sequence identities 82.2–98.8% to each other.

Recombination analysis conducted on the complete HarMV genomes (four from this study and three from Genbank) revealed two separate recombination events (Fig. 3). The sequence of HQ161080 was predicted as a recombinant, with MD4-D as a major parent and MD3 the minor parent in all programmes within RDPL4.
except for Geneconv, with the range in P-values from 1.978 × 10⁻⁶ to 1.083 × 10⁻²⁵. The breakpoint was estimated at a position between 9085 and 9173 nt. The second event was located in H1610801 where the major parent was predicted to be MD4-C and the minor parent MD4-D. This was predicted in all programmes with the exception of 3Seq, with the range in P-values from 7.395 × 10⁻³ to 3.354 × 10⁻². The breakpoint was estimated to be at a position between 9084 and 9160 nt.

The addition of four complete and two nearly complete new genomes of HarMV adds to the molecular knowledge about this recently characterised indigenous Australian virus. When analysed in conjunction with the three sequences available in Genbank (two identical), we found that the groupings resembled those seen previously in analyses involving the CP alone. There are now whole genome representatives for four of the nine suggested clades for HarMV (Webster et al., 2007; Coutts et al., 2011). Previously, there were just three sequences representing two clades (Wylie and Jones, 2011). Comparison of HarMV with PWV at whole genome level revealed the extent of divergence between these two species. At the CP level, PWV diverged from HarMV by 24.1–28.7% (Webster et al., 2007; Coutts et al., 2011), but with whole genomes the divergence was >30%. This is a much higher than the species demarcation borderline of 23–24% suggested for potyviruses at the nt level (Adams et al., 2005).

Recombination is a very successful method of viral evolution with many well studied examples of economically important viruses, including potyviruses such as bean yellow mosaic virus (BYMV), potato virus Y (PVY) and turnip mosaic virus (TuMV) (e.g., Ohshima et al., 2002; Wylie and Jones, 2008; Karasev and Gray, 2013; Ohshima, 2013). Recombination analysis revealed recombination amongst HarMV whole genomes with patterns that matched their phylogenetic groupings. Following the addition of six new CP sequences to the sequences from Webster et al., 2007; Coutts et al., 2011 suggested that the HarMV phylogenetic clades IV and V should be combined into one clade. However, based on our recombination analysis, we suggest that they should remain separate because MD4-D (which represents the first genome sequence from clade V) was not a recombinant, whilst H1610801 from clade IV was recombinant (Fig. 3). Moreover, as next generation sequencing techniques are becoming cheaper, more accessible and more widely used, the number of complete plant virus genomes available in public databases is rising rapidly. Given the readiness of potyviruses to undergo recombination both within species and with other species, it is becoming increasingly important to use recombination analysis to provide a thorough understanding of the genetic makeup and phylogenetic placement of a potyvirus. Addition of further complete HarMV genome sequences to those studied here would allow other phylogenetic HarMV groupings defined from CP sequences to be confirmed for whole genomes. Although the recombination analysis conducted here only included HarMV and PWV, it would be interesting to include whole genomes of some of the more closely related Australian potyviruses such as PaMV, CVC, HbIV, SIYV and S2VY, once they become available. Ongoing study of the evolution of such a geographically distinct group of viruses provides an exciting prospect!

This study provides the first report of HarMV naturally infecting L. cosentinii, a lupin species that become naturalised after its introduction to the SWAFR. It also constitutes the first example for the SWAFR of an indigenous virus that has made a host species jump (Woodhouse et al., 2005), successfully crossing the ancient ecosystem-recent agro-system interface. Luo et al., 2011 had previously demonstrated HarMV spread to L. angustifolius in an artificial field plot situation with deliberately introduced H. contortum infector plants.

So far, the natural host range of HarMV is limited to H. contortum and H. villosa, with only the former being native to the SWAFR (Webster et al., 2007). So it seems that the nt and amino acid changes required to facilitate this jump may be few, given there were just eight nt and one amino acid differences between the HarMV sequence from L. cosentinii (MD3) and the closest HarMV sequence from H. contortum at the same site (MD2). Little is known regarding the feeding preferences associated with the natural aphid vectors of HarMV. It is possible they have a feeding preference for H. contortum, or that their transmission efficiencies are lower on their non-preferred hosts (e.g., L. cosentinii). Further experimentation would provide insight and understanding regarding the limited natural host range for HarMV. Its artificial host range includes species in four additional families (Webster et al., 2007; Coutts et al., 2011; Vincent et al., 2014).

Our research provides the first example for the SWAFR of virus emergence from native plants to invade an introduced plant species. The symptoms in the L. cosentinii plant infected with HarMV were leaf mosaic and leaf deformation, but more severe symptoms such as plant death and stunting were recorded with experimental HarMV infection of L. cosentinii (Webster et al., 2007). The natural distribution of H. contortum in the SWAFR covers much the same area as the distribution of L. cosentinii. Moreover, this region borders a larger area of south west Australia where most of the grain legume crop species L. angustifolius is produced (Western Australian Herbarium, 1998). Further studies are
warranted to determine whether HarMV infection is cause for concern for the grain legume industry. Viruses of native plant communities are sometimes ignored or poorly researched, even though viruses co-evolved with wild plants well before plant domestication and these communities are likely to contain potentially damaging viral pathogens. Here we have further evidence that as new contact between native plants and introduced crops or weeds increases due to man’s activities and climate change, the threat of emerging viruses from indigenous plants to introduced plants is set to increase (Cooper and Jones, 2006; Webster et al., 2007; Jones, 2009; Jones and Barbetten, 2012; Alexander et al., 2014; Vincent et al., 2014).

Acknowledgements

This research was funded by an Australian Postgraduate Award (APA), and an Australian Grains Research and Development Corporation (GRDC) Studentship, Project number GRS100039. It was undertaken using the facilities at the Department of Agriculture and Food Western Australia. This study forms part of a PhD project by the first author at the University of Western Australia.

References

Chapter 4

Plant virology and next generation sequencing: experiences with a *Potyvirus*

Published paper:

Plant Virology and Next Generation Sequencing: Experiences with a Potyvirus

Monica A. Kehoe1,2, Brenda A. Coutts1,2, Bevan J. Buichell1,2, Roger A. C. Jones1,2

1 School of Plant Biology and Institute of Agriculture, Faculty of Science, University of Western Australia, Crawley, WA, Australia; 2 Crop Protection and Crop Breeding Branches, Department of Agriculture and Food Western Australia, Perth, WA, Australia

Abstract

Next generation sequencing is quickly emerging as the go-to tool for plant virologists when sequencing whole virus genomes, and undertaking plant metagenomic studies for new virus discoveries. This study aims to compare the genomic and biological properties of Bean yellow mosaic virus (BYMV) (genus Potyvirus), isolates from Lupinus angustifolius plants with black pod syndrome (BPS), systemic necrosis or non-necrotic symptoms, and from two other plant species. When one Clover yellow vein virus (CYVV) (genus Potyvirus) and 22 BYMV isolates were sequenced on the Illumina HiSeq2000, one new CYVV and 23 new BYMV sequences were obtained. When the 23 new BYMV genomes were compared with 17 other BYMV genomes available on Genbank, phylogenetic analysis provided strong support for existence of nine phylogenetic groupings. Biological studies involving seven isolates of BYMV and one of CYVV gave no symptoms or reactions that could be used to distinguish BYMV isolates from L. angustifolius plants with black pod syndrome from other isolates. Here, we propose that the current system of nomenclature based on biological properties be replaced by numbered groups (I-IX). This is because use of whole genomes revealed that the previous phylogenetic grouping system based on partial sequences of virus genomes and original isolation hosts was unsustainable. This study also demonstrated that, where next generation sequencing is used to obtain complete plant virus genomes, consideration needs to be given to issues regarding sample preparation, adequate levels of coverage across a genome and methods of assembly. It also provided important lessons that will be helpful to other plant virologists using next generation sequencing in the future.

Introduction

Next generation sequencing (NGS) technologies are fast becoming a popular method to obtain whole plant virus genomes in a relatively short period of time [1]. Their uptake by plant virologists has been slower than by their counterparts in the medical sciences where the applications are extending much further, rapidly approaching the concept of personalized medicine. Such a situation was impossible before the advent of NGS and its rapid evolution into an affordable and accessible technology now appearing on laboratory bench-tops throughout the world [2,3]. Because the ability to now total RNA extractions for NGS, it is becoming increasingly common to use it to sequence complete genomes of plant viruses and still obtain excellent results [1-9]. The challenge now lies not in accessing and using NGS technology, but its analyzing and interpreting the very large datasets suddenly at our disposal [1].

Bean yellow mosaic virus (BYMV) (family Potyviridae, genus Potyvirus) is a single-stranded positive sense RNA virus that occurs worldwide. It is a virus with an extensive natural host range that encompasses monocots and dicots, and both domesticated and wild plant species [10,11]. It is transmitted non-persistently by many different aphid species [12]. BYMV causes severe diseases and losses in many cultivated plant species worldwide. For example, early BYMV infection, which causes serious losses, normally results in systemic necrosis and plant death [13-15]. In contrast, late infection with BYMV causes black pod syndrome (BPS) in Lupinus angustifolius (narrow-leaved lupin) also resulting in damaging losses [16]. Plants with BPS develop characteristic flat, black pods that have little or no seed [17]. It seems likely that both the BPS and systemic necrosis responses are related to presence of hypersensitivity Noha-1 gene and another similar resistance gene [15,18-20].

Wiley et al. [21] provided evidence for existence of seven BYMV phylogenetic groupings based on coat protein (CP) sequences and the original hosts of the isolates sequenced: one generalist group with a broad host range including monocots and dicots called the general group, and six other specialist groups each named after the original hosts of the isolates within them (broad bean, canna, lupin, monocot, pea, W). Partial CP sequences from BYMV isolates originally from L. angustifolius plants with BPS,
systemic necrosis or non-necrotic symptoms placed all of them into the general group [16,21].

This study aims to compare the genomic and biological properties of BYMV isolates from L. angustifolius plants with BPS, systemic necrosis or non-necrotic symptoms, and from two other plant species. NGS was used to sequence 22 BYMV isolates, obtained as part of a study conducted in 2011 and from previous studies in south-west Australia [16,19]. Here, we present the results of genome comparisons with the resulting 23 new BYMV genomes and one Clover yellow vein virus (CYVV) genome with 17 genomes retrieved from Genbank, and biologically diverse studies with seven BYMV and one CYVV isolates. We also made recommendations based on the lessons learned from our NGS studies which will be useful to plant virologists employing this approach to obtain whole genomes of other plant viruses.

Materials and Methods

Isolates and host plants

Seventeen BYMV isolates were collected from L. angustifolius plants with BPS (i.e. systemic necrotic stem streaking with black pods) (11) and systemic necrosis (no black pods) (6), and two from L. corymbifera plants with mosaic and leaf deformation as part of a 2011 study in south-western Australia [16]. The remaining three BYMV isolates (FB, LMBNN, and LP) were from previous studies [19]. They had been maintained as freeze-dried leaf material obtained from the West Australian Plant Pathogen Culture Collection (FB - WAC10015, LMBNN - WAC10084 and LP - WAC10026). The CYVV isolate was from the same culture collection (WAC10102).

All plants were maintained at 18–22°C in an insect-proof, air-conditioned glasshouse. Plants of L. angustifolius cv. Jenkins (partially resistant to BPS), Mandelup (susceptible to BPS) and germplasm accession PS6079 (Nmm-1 gene absent) were grown in washed river sand. Plants of Nicotiana benthamiana, Typhodium subterraneum cv. Woongella (subterranean clover), Cheneiodium inanum (inaudi), C. quinoa, Pium sativum cv. Greenleaf pea, and Vicia faba cv. Coles early dwarf (faba bean) were grown in steam-sterilised potting mix. Cultures of virus isolates were maintained by serial mechanical inoculation of infective sap to plants of N. benthamiana or T. subterraneum. For inoculations to maintain cultures, or as part of experiments, virus-infected leaves from systemically infected plants were ground in 0.1M phosphate buffer, pH 7.2, and the infective sap mixed with cattle before being rubbed onto leaves.

For testing by ELISA, leaf samples were extracted (1 g per 20 ml) in phosphate-buffered saline: 10 mM potassium phosphate, 150 mM sodium chloride, pH 7.4. Twenty at 5 ml/liter, and polyvinyl pyrrolidone at 20 g/liter) using a mixer mill (Retsch, Germany). Sample extracts were tested for BYMV or CYVV by double-antibody sandwich ELISA based on a modified protocol described by Clark and Adams [22] and according to manufacturer’s recommendations. For generic Potyviridae testing, samples were extracted in 0.05 M sodium carbonate buffer, pH 9.5, and tested using the antigen-coated indirect ELISA protocol of Tonracion and Peal [23]. The polyclonal antisera to BYMV was from DSMZ (AS-0717), Germany, to CYVV from Nenek Phytodiagnostics – formerly Ardenz, UK (1102-5) and to generic potyviruses from Agria, USA (SRA27300). All samples were tested in duplicate wells in microtitre plates. SLP from BYMV or CYVV infected and healthy T. subterraneum leaf samples was included in paired wells to provide positive and negative controls. The substrate was p-nitrophenyl phosphate at 1.0 mg/ml in diethanolamine, pH 9.0, at 100 ml/liter. Absorbance values at A405 were measured in a microplate reader (Bio-Rad laboratories, USA). Absorbance values of positive samples were always more than three times those of the healthy sap control.

Sequence data

Twenty two BYMV and one CYVV sample were sent for NGS on an Illumina HiSeq 2000 (Table 1). For BYMV in total there were 11 samples from L. angustifolius plants with BPS, six from L. angustifolius plants with systemic necrosis and one from a L. angustifolius plant with non-necrotic symptoms. The remaining samples consisted of isolates from other Lupinus spp., or were isolates from other hosts representing other phylogenetic groups based on Wylie et al. [21], including two samples from L. corymbifera, one from L. paludosus, and one from V. faba. The single CYVV sample was from T. refracta (white clover). Total RNA was extracted from each sample using a Spectrum Plant Total RNA kit (Sigma-Aldrich, Australia). Following extraction, total RNA was sent to the Australian Genome Research Facility (AGRF) for library preparation and barcoding (24 samples per lane) before 100 bp paired-end sequencing on an Illumina HiSeq2000. For each sample, reads were first trimmed using CLC Genomics Workbench 6.5 (CLCGW) (CLC bio) with the quality scores limit set to 0.01, maximum number of ambiguities to two and removing any reads with <30 nucleotides (nt). Contigs were assembled using the de novo assembly function of CLCGW with automatic word size, automatic bubble size, minimum contig length 500, mismatch cost two, insertion cost three, deletion cost three, length fraction 0.5 and similarity factor 0.5. Contigs were sorted by length and the longest subjected to a BLAST search [24]. In addition, reads were also imported into Geneious 6.1.5 (Biomatters) and provided with a reference sequence obtained from Genbank (KJX173278 for BYMV and KX003556 for CYVV). Mapping was performed with minimum overlap 10%, minimum overlap identity 80%, allowing 10% and fine tuning set to iterate up to 10 times. A consensus between the contigs of interest from CLCGW and the consensus from mapping to Geneious was created in Geneious by alignment with Clustal W. Open reading frames (ORFs) were predicted and annotations made using Geneious. Finalized sequences were designated as “complete” based on comparison with the reference sequences used in the mapping process, “nearly complete” if some of the 3’ or 3’ UTR was missing but the coding region was intact, and “partial” if all of the 3’ or 3’ UTR and some of the P1 or CP genes were missing.

Phylogenetic analysis

The new sequences were aligned with the 17 retrieved from Genbank using Clustal W in MEGA 5.2.1, prior to phylogenetic analysis [25]. Phylogenetic analysis compared (i) coding regions of all BYMV genome sequences and (ii) coding regions of all BYMV genome sequences except seven with average coverage of 10 times or less. Neighbour-joining trees were made using the number of differences model with a bootstrap value of 1000, Maximum Likelihood trees using the Tamura-Nei model with a bootstrap value of 1000, and Minimum Evolution trees using the number of differences model with a bootstrap value of 1000. Tables of nucleotide (nt) percentage differences were calculated for the complete genomes using the pairwise comparison function with the number of differences model. Final sequences were submitted to the European Nucleotide Archive (ENA) with accession numbers HCO700417-HCO700470 (Table 1).

Biological data

For host range studies, seven isolates of BYMV and one of CYVV were mechanically inoculated onto leaves of L. angusti-
<table>
<thead>
<tr>
<th>Plant/Host ID</th>
<th>Symptoms*</th>
<th>No. of reads obtained</th>
<th>No. of reads after removing duplicates (CLC)</th>
<th>No. of Contigs produced (CLC)</th>
<th>Sample sequence ID</th>
<th>Accession number</th>
<th>Contig length (nt) (CCWG)</th>
<th>Average coverage (CCWG)</th>
<th>No. of contigs mapped to contig of interest (CCWG)</th>
<th>Length of consensus (nt) (Geneious)</th>
<th>Averate coverage (Geneious)</th>
<th>No. of reads mapped to ref. sequence (Geneious)</th>
<th>Length of Geneious consensus (nt) (Geneious)</th>
<th>Genome completeness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lupinus cassinii M</td>
<td>M</td>
<td>12,684,310</td>
<td>12,402,361</td>
<td>387</td>
<td>MD1</td>
<td>HG970847</td>
<td>9,547</td>
<td>10,173</td>
<td>987,972</td>
<td>9,381</td>
<td>10,562</td>
<td>1,002,513</td>
<td>9,285</td>
<td>partial</td>
</tr>
<tr>
<td>L. angustifolius SS, SC</td>
<td>31,131,660</td>
<td>29,497,124</td>
<td>1,851</td>
<td>MD5</td>
<td>HG970848</td>
<td>968–2,578 (5)</td>
<td>11–14 (5)</td>
<td>111–359 (5)</td>
<td>9,541</td>
<td>9</td>
<td>894</td>
<td>9,287</td>
<td>partial^4</td>
<td></td>
</tr>
<tr>
<td>L. angustifolius SS, SC</td>
<td>14,542,928</td>
<td>13,695,123</td>
<td>897</td>
<td>MD6</td>
<td>HG970849</td>
<td>2,625,1,430</td>
<td>6,7</td>
<td>161,202</td>
<td>9,344</td>
<td>7</td>
<td>713</td>
<td>9,285</td>
<td>partial^4</td>
<td></td>
</tr>
<tr>
<td>L. canadensis ML, LD</td>
<td>12,026,256</td>
<td>11,701,675</td>
<td>472</td>
<td>MD7</td>
<td>HG970850</td>
<td>9,633</td>
<td>1,780</td>
<td>173,242</td>
<td>9,636</td>
<td>1,821</td>
<td>175,858</td>
<td>9,530</td>
<td>nearly</td>
<td></td>
</tr>
<tr>
<td>L. angustifolius SS, SC</td>
<td>10,841,638</td>
<td>10,582,230</td>
<td>802</td>
<td>SP1</td>
<td>HG970851</td>
<td>9,524</td>
<td>446</td>
<td>43,220</td>
<td>9,344</td>
<td>437</td>
<td>43,877</td>
<td>9,528</td>
<td>nearly</td>
<td></td>
</tr>
<tr>
<td>L. angustifolius SS, SC</td>
<td>11,348,648</td>
<td>11,076,092</td>
<td>1,908</td>
<td>GB17A</td>
<td>HG970852</td>
<td>9,655</td>
<td>1,723</td>
<td>169,883</td>
<td>9,983</td>
<td>12,273</td>
<td>954,307</td>
<td>9,530</td>
<td>complete</td>
<td></td>
</tr>
<tr>
<td>L. angustifolius SS, BP</td>
<td>30,708,142</td>
<td>29,877,467</td>
<td>2,498</td>
<td>GB42C</td>
<td>HG970853</td>
<td>701–1,557 (8)</td>
<td>6–14 (8)</td>
<td>49–191 (8)</td>
<td>9,544</td>
<td>10</td>
<td>1,045</td>
<td>9,190</td>
<td>nearly^4</td>
<td></td>
</tr>
<tr>
<td>L. angustifolius SS, BP</td>
<td>11,780,826</td>
<td>11,532,853</td>
<td>360</td>
<td>GB32A</td>
<td>HG970854</td>
<td>9,593</td>
<td>5,241</td>
<td>511,164</td>
<td>9,915</td>
<td>5,317</td>
<td>511,902</td>
<td>9,538</td>
<td>complete</td>
<td></td>
</tr>
<tr>
<td>L. angustifolius VC</td>
<td>11,166,630</td>
<td>10,924,778</td>
<td>476</td>
<td>LMBNN</td>
<td>HG970855</td>
<td>9,533</td>
<td>2,826</td>
<td>278,099</td>
<td>9,544</td>
<td>2,925</td>
<td>279,855</td>
<td>9,531</td>
<td>nearly</td>
<td></td>
</tr>
<tr>
<td>L. angustifolius BP, NvSS</td>
<td>14,710,190</td>
<td>14,415,378</td>
<td>820</td>
<td>ES69C</td>
<td>HG970856</td>
<td>554–2,300 (9)</td>
<td>3–11 (9)</td>
<td>19–211 (9)</td>
<td>9,544</td>
<td>8</td>
<td>850</td>
<td>9,269</td>
<td>partial^4</td>
<td></td>
</tr>
<tr>
<td>L. angustifolius M, BP</td>
<td>14,308,420</td>
<td>13,991,683</td>
<td>907</td>
<td>ES67C</td>
<td>HG970857</td>
<td>534–1,413 (9)</td>
<td>3–8 (9)</td>
<td>18–101 (9)</td>
<td>9,544</td>
<td>6</td>
<td>660</td>
<td>9,479</td>
<td>nearly^4</td>
<td></td>
</tr>
<tr>
<td>L. angustifolius M, BP</td>
<td>15,468,144</td>
<td>15,091,935</td>
<td>837</td>
<td>ES55C</td>
<td>HG970858</td>
<td>9,587</td>
<td>1,631</td>
<td>159,066</td>
<td>9,567</td>
<td>1,656</td>
<td>159,704</td>
<td>9,514</td>
<td>complete</td>
<td></td>
</tr>
<tr>
<td>L. angustifolius M, BP</td>
<td>13,667,576</td>
<td>13,370,007</td>
<td>796</td>
<td>ES11A</td>
<td>HG970859</td>
<td>799–1,417 (9)</td>
<td>767–2,207</td>
<td>8,193–31,962 (9)</td>
<td>10,324</td>
<td>2,708</td>
<td>280,928</td>
<td>9,530</td>
<td>complete</td>
<td></td>
</tr>
<tr>
<td>L. angustifolius SS, BP</td>
<td>14,890,148</td>
<td>14,565,374</td>
<td>855</td>
<td>PN89A</td>
<td>HG970860</td>
<td>9,556</td>
<td>7,219</td>
<td>704,254</td>
<td>10,262</td>
<td>7,382</td>
<td>715,798</td>
<td>9,530</td>
<td>complete</td>
<td></td>
</tr>
<tr>
<td>L. angustifolius SS, BP</td>
<td>15,739,358</td>
<td>15,445,952</td>
<td>795</td>
<td>PN89A</td>
<td>HG970861</td>
<td>9,532</td>
<td>4,625</td>
<td>468,607</td>
<td>9,625</td>
<td>4,934</td>
<td>475,451</td>
<td>9,530</td>
<td>complete</td>
<td></td>
</tr>
<tr>
<td>L. angustifolius SS, BP</td>
<td>16,826,134</td>
<td>16,427,670</td>
<td>919</td>
<td>PN77C</td>
<td>HG970862</td>
<td>646–926 (5)</td>
<td>4–5 (5)</td>
<td>27–50 (5)</td>
<td>9,544</td>
<td>4</td>
<td>471</td>
<td>9,274</td>
<td>partial^4</td>
<td></td>
</tr>
<tr>
<td>L. angustifolius SS, BP</td>
<td>14,713,078</td>
<td>14,380,894</td>
<td>1,643</td>
<td>AR87C</td>
<td>HG970863</td>
<td>9,561</td>
<td>2,600</td>
<td>252,909</td>
<td>9,912</td>
<td>2,587</td>
<td>256,835</td>
<td>9,530</td>
<td>complete</td>
<td></td>
</tr>
<tr>
<td>L. angustifolius SS, BP</td>
<td>12,557,974</td>
<td>12,252,296</td>
<td>927</td>
<td>AR98C</td>
<td>HG970864</td>
<td>676–1,204 (5)</td>
<td>6–7 (5)</td>
<td>56–88 (5)</td>
<td>9,544</td>
<td>6</td>
<td>617</td>
<td>9,447</td>
<td>nearly^4</td>
<td></td>
</tr>
<tr>
<td>L. angustifolius SS, BP</td>
<td>16,337,855</td>
<td>15,905,779</td>
<td>1,023</td>
<td>AR98C</td>
<td>HG970865</td>
<td>9,547</td>
<td>5,856</td>
<td>588,049</td>
<td>9,990</td>
<td>6,004</td>
<td>576,414</td>
<td>9,530</td>
<td>complete</td>
<td></td>
</tr>
<tr>
<td>L. pilosus VC, SM, LD, Y</td>
<td>17,060,028</td>
<td>16,402,537</td>
<td>889</td>
<td>LP</td>
<td>HG970866</td>
<td>9,521</td>
<td>1,159</td>
<td>113,369</td>
<td>9,034</td>
<td>737</td>
<td>6,866</td>
<td>9,520</td>
<td>nearly</td>
<td></td>
</tr>
<tr>
<td>Viciea faba VC, SM, LD</td>
<td>16,293,018</td>
<td>15,728,171</td>
<td>519</td>
<td>FB</td>
<td>HG970867</td>
<td>9,464</td>
<td>1,035</td>
<td>190,666</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>9,417^4</td>
<td>nearly</td>
<td></td>
</tr>
<tr>
<td>Trifolium repens</td>
<td>Not recorded</td>
<td>11,265,996</td>
<td>11,121,357</td>
<td>149</td>
<td>CIVV</td>
<td>HG970870</td>
<td>9,439</td>
<td>6,195</td>
<td>6,195</td>
<td>9,385</td>
<td>65</td>
<td>6,295</td>
<td>9,459</td>
<td>nearly</td>
</tr>
</tbody>
</table>

*Assigned symptom descriptions: BP, black pods; LD, distortion; M, mosaic; M, mild necrotic stem streaking; NvSS, no visible stem streaking; SS, necrotic stem streaking; SM, severe mosaic; SC, shephers crook appearance (i.e. bending over and apical tip necrosis); VC, vein clearing; Y, yellowing.

^Numbers in parenthesis represent the total number of contigs for the sample with lengths indicated by the preceding range.

^CCWG genomics workshop.

^Indicates that the genome is draft only, meaning less than or equal to ten times average coverage.

^Indicates that the final sequence is derived entirely from CCWG de novo assembly.

doi:10.1371/journal.pone.0104560.t001

| August 2014 | Volume 9 | Issue 8 | e104560 | 3 |

Page 57

Next Generation Sequencing Experiments with a Purpose
Results

Sequence data

From the single CiYVV and 22 BYMV samples, the numbers of raw reads obtained from NGS were 10,281,130; 31,131,660, but these numbers were reduced to 10,382,250; 29,677,476 after trimming (Table 1). Following de novo assembly of each individual sample using CLC GW, the numbers of contigs produced were 149,249. Contig of interest lengths were 534-9,655 nt with average coverage 3-10,173 times and the numbers of reads mapped to each contig were 10,907,972. After mapping to a reference genome in Genoscope, the lengths of the consensus sequences were 9,034-10,324 nt, with average coverage of 4-12,313 times and the numbers of reads mapped to the consensus sequences were 471,1,002,513. Final sequence lengths consisted of the consensus of the coming from CLG GW and the consensus from Genoscope, and were 9,274-9,530 nt. All samples yielded one sequence of interest, with the exception of FB, which contained a second BYMV sequence which we called “LPesFB”. In all cases, except for CiYVV, the contents of interest were most closely related to BYMV after being subjected to Blastn analysis. CiYVV was most closely related to the only other CiYVV complete genome available on Genbank. In total, there were nine complete genomes, ten nearly complete genomes (including CiYVV) and five partial genomes.

Phylogenetic analysis

Phylogenetic analysis comparing the coding regions of 23 new complete or nearly complete BYMV genomes and one nearly complete CiYVV genome with those of 17 BYMV and one CiYVV genome retrieved from Genbank provided 100% bootstrap support for eight of nine phylogenetic groups (I, II, IV-IX). The remaining group (III) had 98% bootstrap support. Seven of the new genomes had average coverages of less than or equal to ten times (MD5, MD6, BG42G, E690C, E676C, P276C, and AR89C) and five of these (MD5, MD6, E676C, E690C and P276C) did not fit well within groups I and II; although they appear to belong to them, genomes such as MD6 and P276C sit out on their own, almost completely separate from the other sequences, leaving groups I and II poorly resolved (Figure 1a). In contrast, when sequences of the seven genomes with poor average coverage (≤10 times) were removed, phylogenetic analysis gave the same results but with much greater resolution between groups I and II and improved bootstrap support for groups I-IX (Figure 1b). Those removed were designated as “draft” genomes because all had low coverage and/or small gaps. When all the genomes, including those with poor coverage were analyzed using Maximum Likelihood or Minimum Evolution methods, the tree topologies shown were the same as the Neighbor-Joining method. The range of original isolation hosts within each grouping varied (Table 2). Group I consisted of nine sequences from two
dicot, and two monocot species. Group II consisted of seven sequences from two dicot and one monocot species. Group III consisted entirely of three sequences from one monocot species. Group IV was made up of three sequences from an unknown original host or hosts, as well as two from a monocot and one from a dicot species. Groups V-IX consisted entirely of dicot species belonging to a single family, and were represented by up to three sequences. All dicot species were from families Fabaceae or Gentianaceae, and all monocot species were from families Orchidaceae or Iridaceae.

Sequence analysis

When the coding regions of the 16 new BYMV genomes (draft genomes excluded) and one CiYVV genome were analyzed against those retrieved from Genbank, the nt percentage identities within each phylogenetic group were ≥98.6% (I), ≥98.6% (II), ≥93.9% (III), ≥94% (IV), ≥90.7% (V), ≥90.8% (VI), ≥97.6% (VII) and ≥97.3% for CiYVV (Table S1). When the six sequences from L. angustifolius plants with BPS were compared to each other, their percentage nt identities were ≥98.8%. When the sequences from all L. angustifolius plants were compared to each other, their percentage nt identities were also ≥93.8%. Across all 33 BYMV sequences used in this analysis, the nt identities were ≥75.6%. When the CiYVV sequences were compared to the BYMV sequences, overall the percentage nt identities were 66.4-67.9%.

Biological data

All seven BYMV isolates and one CiYVV isolate inoculated to plants caused systemic symptoms of varying severity in N. benthamiana, T. subterreanum and V. faba (Table S). However, apart from CiYVV and BYMV isolate BG17A in V. faba, none of them induced systemic necrotic symptoms, which were severe only with CiYVV. In C. amaranthoides, CiYVV and five BYMV isolates caused obvious systemic symptoms, while infection was restricted to inoculated leaves with the isolate originally from L. angustifolius plants with BPS and another originally from an L. angustifolius plant with non-necrotic symptoms. In C. quinoa, although all isolates infected inoculated leaves, only CiYVV caused systemic invasion. In contrast, in P. sativum, only BYMV isolate LP caused any infection.

In L. angustifolius cv. Jenahillip and Mandalup, three BYMV isolates caused systemic non-necrotic symptoms. These were originally from plants of this species with non-necrotic symptoms (LBNN) or systemic necrotic symptoms (ES1A) and L. cosentinii (MD7) from a plant with mosaic and leaf distortion. All other BYMV isolates and the CiYVV caused systemic necrotic symptoms in cv. Jenahillip and Mandalup. In accession P26697, with CiYVV and four BYMV isolates for which symptom data are available, the reactions resembled those in cv. Jenahillip, with the exception of MD5 which produced severe mosaic (i.e., non-necrotic) symptoms instead of systemic necrosis. Isolates LBNN and ES1A caused non-necrotic symptoms, while CiYVV and LP caused systemic necrosis. Failure of isolates AR89C and MD7 to infect P26697 probably represents escapes, but there was no seed left of P26697 for further testing. Isolate LP did not infect L. angustifolius cv. Mandalup on two separate occasions by sap inoculation, but further inoculations using grafting or self pollination would be needed to establish if this is a resistance reaction.

Discussion

Before this study was conducted, there were only 17 complete BYMV genomes on Genbank. The ten complete and eight nearly complete genomes from this study doubled available BYMV
Figure 1. Neighbor-joining relationship phylograms obtained from alignment of the coding regions of bean yellow mosaic virus (BYMV) genomes. The alignments were generated in MEGA 5.2.1 using ClustalW and tree branches were bootstrapped with 1000 replications. The trees were rooted with a sequence of Clover yellow vein virus (CYVV), the closest relative to BYMV. New isolates from this study shown in grey, isolates obtained from *Lupinus angustifolius* plants with BPS are denoted by *, and isolates with genomes designated as “draft” are denoted by _. a) Complete coding regions of BYMV genomes, including draft sequences, with isolates retrieved from Genbank. b) The same sequences as in a) but with draft sequences removed from the analysis.

doi:10.1371/journal.pone.0104580.g001

Table 2. Original hosts of isolates within each phylogenetic grouping.

<table>
<thead>
<tr>
<th>Phylogenetic group (old name)</th>
<th>Accession numbers</th>
<th>Dicot</th>
<th>Monocot</th>
</tr>
</thead>
<tbody>
<tr>
<td>I (general)</td>
<td>FA4626,1, 2X1727,1, HJ970847, HJ970851, HJ970851-52, HJ970856-57, HJ970860-62, HJ970864-65, HJ970865</td>
<td>Lupinus angustifolius* (6)*, L. coffeina (1)</td>
<td><em>Diuris magnifica</em> (1), <em>Fremphothera</em> sp. (1)</td>
</tr>
<tr>
<td>II (general)</td>
<td>JX956423, HJ970848, HJ970849, HJ970850, HJ970854-55, HJ970858-59, HJ970863</td>
<td>L. angustifolius (S), L. coffeina (1)</td>
<td><em>Diuris</em> sp. (1)</td>
</tr>
<tr>
<td>III (monocot)</td>
<td>ABO59888, ABO5988B, ABO439729</td>
<td>-</td>
<td><em>Gleocaulis</em> hybrid (3)</td>
</tr>
<tr>
<td>IV (general)</td>
<td>ABO59888B, ABO5988B, AB534942*, ABO59888, ACO04810, ABO55196, AY19396</td>
<td><em>Ectostruma russellianum</em> (1), <em>Gleocaulis</em> sp. (1), <em>Gleocaulis</em> hybrid (1)</td>
<td></td>
</tr>
<tr>
<td>V (Fabaceae)</td>
<td>ABO439732, U470733</td>
<td><em>Trifolium pratense</em> (1), <em>Vicia faba</em> (1)</td>
<td>-</td>
</tr>
<tr>
<td>VI (lupin)</td>
<td>HJ970866, HJ970868</td>
<td>L. pilosula* (1), <em>Vicia faba</em> (1)</td>
<td>-</td>
</tr>
<tr>
<td>VII (Fabaceae)</td>
<td>ABO439731, HJ970867</td>
<td><em>V. faba</em> (2)</td>
<td>-</td>
</tr>
<tr>
<td>VIII (W)</td>
<td>DO4614</td>
<td>L. elbus (1)</td>
<td></td>
</tr>
<tr>
<td>IX (pea)</td>
<td>ABO13130</td>
<td><em>Pisum sativum</em> (1)</td>
<td></td>
</tr>
</tbody>
</table>

*Species from *Lupinus*, *Vicia* and *Trifolium* are from family Fabaceae. *Ectostruma* is from family Gentianaceae. *Gleocaulis* and *Fremphothera* are from family Iridaceae. *Diuris* is from family Orchidaceae.

*Numbers in parentheses represent the number of genomes with this original isolation host.

*Denotes an unknown original host for that accession number.

doi:10.1371/journal.pone.0104580.t002
Table 3. Responses of seven plant species to inoculation with eight different isolates of Bean yellow mosaic virus (BYMV) and one of Clover yellow vein virus (CYVV).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CYVV</td>
<td>T. repens</td>
<td>Not recorded</td>
<td>LNS, SCS, SVC</td>
<td>LNS, SCS, SVC</td>
<td>S, N, S, Y, LD</td>
<td>SS, SN, DR</td>
<td>SSS, R, B, M</td>
<td>SM</td>
<td>NI</td>
<td>VC, SM, LD, St</td>
<td>SS, SM, LD, SN</td>
<td>n/a</td>
</tr>
<tr>
<td>LAnBn</td>
<td>L. angustifolius</td>
<td>VC</td>
<td>LCS</td>
<td>LMCS</td>
<td>SM, DR, B</td>
<td>B, M</td>
<td>B, M</td>
<td>MM</td>
<td>NI</td>
<td>MM</td>
<td>VC, M, LD, St</td>
<td>MM</td>
</tr>
<tr>
<td>APR9C</td>
<td>L. angustifolius</td>
<td>SS, BP</td>
<td>LCS</td>
<td>LNS</td>
<td>MSS, SN, Y</td>
<td>SSS, DR, SN</td>
<td>SSS, DR, SN</td>
<td>NF</td>
<td>M</td>
<td>NI</td>
<td>VC, M, LD</td>
<td>M</td>
</tr>
<tr>
<td>MDS</td>
<td>L. angustifolius</td>
<td>SS, SC</td>
<td>LNS, SCS, SVC</td>
<td>LNS, LCS</td>
<td>SSS, SN</td>
<td>SSS, DR, SN</td>
<td>SSS, DR, SN</td>
<td>MM</td>
<td>NI</td>
<td>VC, M, LD</td>
<td>M</td>
<td>I</td>
</tr>
<tr>
<td>G81T</td>
<td>L. angustifolius</td>
<td>SS, SC</td>
<td>LCS, SCS</td>
<td>LCS, LNS</td>
<td>M, DR, Y, SSS, N</td>
<td>M</td>
<td>MM</td>
<td>NI</td>
<td>MM</td>
<td>MVC, SM, LD</td>
<td>SS, M</td>
<td>I</td>
</tr>
<tr>
<td>ES11A</td>
<td>L. angustifolius</td>
<td>SS, SC</td>
<td>LCS, SCS, SVC</td>
<td>LVC, LCS, NSH</td>
<td>M, DR, Y</td>
<td>B, M</td>
<td>B, M</td>
<td>MM</td>
<td>NI</td>
<td>MVC, MM</td>
<td>SM</td>
<td>I</td>
</tr>
<tr>
<td>MED7</td>
<td>L. caesalpinii</td>
<td>MI, DR</td>
<td>LCS, SCS, SVC</td>
<td>LCS</td>
<td>SM, DR, B</td>
<td>PSN, M, B</td>
<td>NF</td>
<td>MM</td>
<td>NI</td>
<td>MVC, M</td>
<td>SM</td>
<td>I</td>
</tr>
<tr>
<td>LP</td>
<td>L. pilosus</td>
<td>VC, SM, LD, Y</td>
<td>SCS, SVC</td>
<td>LNS</td>
<td>SSS, DR, SN</td>
<td>SSS, SN</td>
<td>M</td>
<td>MM</td>
<td>NI</td>
<td>VC, M, LD, St</td>
<td>M</td>
<td>I</td>
</tr>
</tbody>
</table>

Unoinoculated: n/a
Mock: n/a

Leaves were inoculated with infective sap. Samples from inoculated and tip leaves were tested by ELISA for BYMV, CYVV and potyviruses in general.

*Locations where isolates were collected from: CYVV, New South Wales (NSW); LAnBn, Me. Barker, Western Australia (WA); APR9C, Arthur River, WA; MDS and MED7, Medina, WA; G81T, Gibson, WA; ES11A, Esperance, WA; LP, South Perth, WA.

**Coded symptom descriptions:** B: bunchy new growth; BP: black pod; DR: leaf drop; LDC: local chlorotic spots; LD: leaf distortion; LMCS: local mild chlorotic spots; LNS: local necrotic spots; M: moss; MM: mild mossy necrotic stem streaking; MVC: mossy local vein clearing; N: not infected; NSH: local necrotic spots with halos; nt: not tested; PSN: partial necrotic stem streaking (the infection in uninoculated leaves); P: reddening; SC: shepherds crook appearance (i.e., bending over and apical tip necrotic); SS: necrotic stem streaking; SSs: severe necrotic stem streaking leading to plant death; SCS: systemic chlorotic spots; SM: severe mossy; SN: severe necrosis of uninoculated leaves and new growth; St: stunting; SVC: systemic vein clearing; VC: localized vein clearing; Y: yellowing.

*Denotes that the initial round of inoculations failed to infect plants of this species, and due to a lack of seed the inoculations were not repeated.

**Denotes that two rounds of inoculation failed to infect L. angustifolius cv. Mandelup. Further testing involving sap, aphid and graft inoculations are required.

doi:10.1371/journal.pone.0104580.t003
genomic data in the database. Moreover, the five additional partial genomes we obtained will be useful in future studies. Our genome results enabled the phylogenetic makeup of BYMV to be examined thoroughly, revealing presence of nine distinct groups, including the subdivision of the former specialist group into three new groups. We recommend replacing the phylogenetic groupings of Wylie et al. [21] with numbered group names I-IX. We have not included one former specialist group based on CP genes, the cauna group, in our analysis because it was not represented by any whole genome sequence. Use of whole genomes revealed that the previous phylogenetic grouping system based on partial genome sequences and original isolation hosts was unsustainable. This is because genome sequences from broad bean are present in two former specialist groups (now V and VII), from various Lupinus species in two former specialist groups (now VI and VIII), and two former specialist groups (now I and II). Moreover, although we have not re-analyzed sequences of CP genes, Wylie et al. [21] had previously placed a CP sequence from the closest species (Eustoma russellianum family Gentianaceae), in the former monocot group (now III). Numbering of groups prevents such confusion arising from use of the original isolation hosts. Our results highlight the importance of using complete genomes wherever possible to define phylogenetic groupings. The results also highlight the need for further sequencing and analysis of BYMV isolates likely to belong to former specialist phylogenetic groupings, which will provide greater insight into the genetic makeup of BYMV.

Close examination of the nt percentage sequence identities between BYMV and CIVV genomes revealed that the divergence between them is greater than previously thought. Overall, BYMV percentage nt identities ranged from 73.6 to 99.3%. The species demarcation for potyviruses is currently 23-24% divergence at the nt level [26], and some of the BYMV isolates compared came close to this. The two CIVV genomes shared 97.5% nt identity, but when compared to all the BYMV genomes, nt identities were 66.4-67.9%, well beyond the species demarcation point for potyviruses. CIVV was originally considered an isolate of BYMV but was later shown to be a distinct virus [26,27,29]. Our percentage identities support that distinction. However, some BYMV phylogenetic groups were more closely related to CIVV than others. For example when compared with all other BYMV sequences, the single sequences from groups VIII and IX had percentage identities of just 70.4-70.8% and 73.4-76.0% to BYMV respectively, whereas when compared to CIVV their nt percentage identities were 67.0-67.2% (Table S1). Again, further genome sequences from these groups and CIVV are required for a more conclusive analysis.

Based on our phylogenetic and sequence analyses, BYMV isolates associated with BPS in L. angustifolius were not different phylogenetically from other BYMV isolates we sequenced from L. angustifolius, L. coriaceus, or other hosts within the same phylogenetic groups I and II. Also, from the host data from our inoculations, there was no least reaction that could be used to distinguish a particular isolate as causing BPS. However, there were some other interesting differences. Although isolate ES11A behaved in a similar manner to isolate LMBNN, which overcomes the Nm-1 hypersensitivity gene in L. angustifolius plants [19,20], it was isolated from a L. angustifolius plant originally displaying systemic necrosis. CIVV behaved like isolate LP, but whether CIVV interacts with both Nm-1 and the second putative BYMV hypersensitivity genes, or unknown CIVV-specific genes in L. angustifolius, is not clear [19]. CIVV and all group I and II isolates failed to infect P. sativum cv. Greenash although the group VI isolate LP did cause infection. This may be due to the fact that this cultivar, like many commercial pea cultivars, may contain the BYMV resistance gene mlo and CIVV resistance genes cys or cys-2 [29,30] and their responses are strain specific. Induction of severe necrotic symptoms in V. faba by CIVV but not the BYMV isolates is expected, as this is the classical method for distinguishing BYMV from CIVV [10,20].

In this study, we used NGS to obtain complete virus genomes and it proved both an advantage and a disadvantage over traditional sequencing methods. It allowed large amounts of data to be generated quickly, but analysis of the data proved a major challenge. Many free programs exist for the assembly of NGS data (e.g. Velvet, SOAP de novo, Abyss and bowtie) but they all require the researcher to be proficient in the use of command line driven applications. As so-called “benchtop biologists”, the use of Generics and CLEGW was easy to learn and their cost was acceptable in view of the time saved in learning the use of command line driven programs. That said, our success was probably attributable to the small genome sizes of plant viruses, particularly BYMV and CIVV, which are both c. 9533 nt long. Larger genomes, from unpurified RNA samples would undoubtedly be much harder to piece together, but not impossible. We found that in most cases (17 out of 23) there was sufficient average coverage to be confident of good genome representation for the isolate sequenced. These sequences had average coverages as low as 65 and 4x7 with remaining average coverages being greater than 737 and up to 12,013 times when mapped back to a reference sequence using Generics. Currently, sequencing a human genome of approximately 300 MB on an Illumina platform requires 30 times coverage to be adequate [31]. Therefore, it seems reasonable to designate our virus genomes with less than 30 times coverage as draft sequences. Although not meeting minimum requirements for average coverage, they are still valuable data sets, particularly given the low numbers of complete or nearly complete BYMV genomes available (now 32 including those from this study).

The settings used in de novo assembly are sufficient to distinguish between more than one strain or group of a plant virus when present in the same sample, as previously demonstrated by Recho et al. [9]. In our case, the sample from a V. faba BYMV isolate (FB) retrieved from the culture collection also contained a nearly complete LP isolate genome. The contamination probably occurred more than ten years ago when they were maintained next to each other in the same glasshouse prior to freeze-drying and storage in the collection. In such instances, if we had only been using Generics to map to a reference genome, we would have likely missed the second sequence. It is therefore important to perform de novo assembly, as well as mapping to a reference genome. In cases where either the mapping or the de novo sequence had a gap, it was usually resolved after alignment with the sequence from the second program. However, for genomes with coverage less than ten (i.e. the draft genomes) this method was ineffective.

The uptake of NGS amongst plant virologists is increasing as the cost associated with it decreases [1]. The relatively small genome size of plant viruses allows us the opportunity to extract complete or nearly complete genomes using commercial packages. Use of NGS does raise concerns regarding the consequences of an increase in the discovery of virus or virus-like sequences. As such, MacDiarmid et al. [32] made recommendations regarding the identification of plant viruses through NGS, and the potential biosecurity issues associated with this. One of the recommendations was that the term “uncultured virus” should be used with any plant virus sequence not associated with a recognized virus infection. We support this recommendation whole-heartedly.

We know of no recommendation regarding requirements for depth of coverage for plant virus genomes, particularly ones
involving new virus discoveries. Until such time as an appropriate set of comparative studies are done, we would recommend following in the path of our Human Genome colleagues by requiring a minimum coverage of at least 30 times, but this would likely lead to many nearly complete or draft plant virus genomes. However as with BIMV for example, we required coverage well into the 1000s to ensure a complete genome (including 5’ and 3’ UTRs, a constant challenge for plant virologists). Our samples sent for sequencing were total RNA, so different methods of sample preparation might have increased the numbers of virus reads. For example, use of subtractive hybridization [4], or extracting for cDNA first, followed by random cDNA synthesis [131]. Despite this, there is no doubt that NGS has been an exceedingly useful tool for our study.

References

Supporting Information
Table S1 Nucleotide percentage similarities of the coding regions of thirty three Bean yellow mosaic virus and two Clover yellow vein virus isolates, calculated in MEGA 5.2.1 using a pairwise comparison with the number of differences model. (DOCX)

Acknowledgments
We thank E. Gajd, S. Vincenzi and M. Barovic for glasshouse and laboratory support.

Author Contributions
Conceived and designed the experiments: MAK RAG BJB. Performed the experiments: MAK. Analyzed the data: MAK. Contributed reagents/materials/analysis tools: MAK. Contributed to the writing of the manuscript: MAK RAG BJB.
Chapter 5

Split personality of a *Potyvirus*: to specialize or not to specialize?

Published paper:

Split Personality of a Potyvirus: To Specialize or Not to Specialize?

Monica A. Kehoe1,2, Brenda A. Coutts1,2, Bevan J. Buichell1,2, Roger A. C. Jones1,2

1School of Plant Biology and Institute of Agriculture, Faculty of Science, University of Western Australia, Crawley, WA, Australia; 2Crop Protection and Lupin Breeding Branches, Department of Agriculture and Food Western Australia, Bentley Delivery Centre, Perth, WA, Australia

Abstract

Bean yellow mosaic virus (BYMV), genus Potyvirus, has an extensive natural host range encompassing both dicots and monocots. Its phylogenetic groups were considered to consist of an ancestral generalist group and six specialist groups derived from this generalist group during plant domestication. Recombination was suggested to be playing a role in BYMV’s evolution towards host specialization. However, in subsequent phylogenetic analysis of whole genomes, group names based on the original hosts of isolates within each of them were no longer supported. Also, nine groups were found and designated I–IX. Recombination analysis was conducted on the complete coding regions of 33 BYMV genomes and two genomes of the related Clover yellow vein virus (CYVV). This analysis found evidence for 12 firm recombination events within BYMV phylogenetic groups I–VI, but none within groups VII–IX or CYVV. The greatest numbers of recombination events within a sequence (two or three each) occurred in four groups, three of which had a generalist group (I, II, and IV), and group VI. The individual sequences in groups III and V had one event each. These findings with whole genomes are consistent with recombination being associated with expanding host ranges, and call into question the proposed role of recombination in the evolution of BYMV, where it was previously suggested to play a role in host specialization. Instead, they indicate that recombination explains the very broad natural host ranges of the three BYMV groups which infect both monocots and dicots (I, II, IV), and (I) suggest that the three groups with narrow natural host ranges (III, V, VII) which also showed recombination now have the potential to reduce host specificity and broaden their natural host ranges.

Introduction

Bean yellow mosaic virus (BYMV), genus Potyvirus, occurs worldwide, and has an extensive natural host range that encompasses domesticated and wild plants species, including both monocots and dicots. It causes serious diseases in a wide range of crops [1–4]. e.g. recent studies found that late infection with BYMV causes black pod syndrome (BPS) in Lapsana communis (narrow-leaved lupin) and substantial yield losses [5,6]. BYMV is transmitted non-persistently by many different aphid species [1,7]. It consists of a RNA single stranded plus sense genome of about 10 kb. Its genome comprises two open reading frames (ORFs). There is one large polyprotein which is processed into ten proteins (biological characteristics linked to each in parentheses): PI (symptomatology); HCC-Pro (aphid transmission, systemic movement, suppression of gene silencing, self-interaction); P3 (plant pathogenicity); 6K1; CI (cell to cell movement); 6K2 (membrane attachment); VPg (genome replication); Nia-Pro (protein-protein interaction, cellular localization); Nib (RNA-dependent RNA polymerase, involved in replication); CP (aphid transmission, virus assembly, movement) [8,9]. The second ORF, called Pipo, is embedded within P3, is around 180 nucleotides in length and translated in the +2 reading frame relative to the polyprotein [10]. Pipo has been linked to virulence determinacy in potyvirus resistant plants of Phaseolus vulgaris (pea) and long distance virus movement [11,12].

Wylie et al. [13] analyzed the coat protein (CP) gene sequences of 64 BYMV isolates and banded the names of the phylogenetic groups found on the types of original plant hosts that the isolates within each group came from. They proposed that these groups consisted of an ancestral generalist group with a wide natural host range and six specialist groups with narrow natural host ranges derived from the generalist group. They suggested that host specialization of BYMV had arisen within isolated crop domestication centers in different parts of the world [14,15]. When Kehoe et al. [16] analyzed 40 whole BYMV genomes, they found nine phylogenetic groups which they named I–IX. The former ancestral group (called the general group) was split into three separate groups (I, II and IV). The genera the original isolation
### Table 1. Bean yellow mosaic virus and Clover yellow vein virus genomes analyzed for recombination.

<table>
<thead>
<tr>
<th>Accession number</th>
<th>Sequence ID</th>
<th>Phylogenetic grouping</th>
<th>Location</th>
<th>Original isolate host</th>
<th>Original host type</th>
<th>Genome Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>H670660</td>
<td>PNA3A</td>
<td>I</td>
<td>WA, Australia</td>
<td><em>Lupinus angustifolius</em></td>
<td>IC, D</td>
<td>[16]</td>
</tr>
<tr>
<td>H670661</td>
<td>PNA64A</td>
<td>I</td>
<td>WA, Australia</td>
<td><em>L. angustifolius</em></td>
<td>IC, D</td>
<td>[16]</td>
</tr>
<tr>
<td>H670652</td>
<td>GB17A</td>
<td>I</td>
<td>WA, Australia</td>
<td><em>L. angustifolius</em></td>
<td>IC, D</td>
<td>[16]</td>
</tr>
<tr>
<td>FJ492961</td>
<td>Fr</td>
<td>I</td>
<td>South Korea</td>
<td><em>fresia sp.</em></td>
<td>IC, D</td>
<td>unpublished</td>
</tr>
<tr>
<td>H670647</td>
<td>MD1</td>
<td>I</td>
<td>WA, Australia</td>
<td><em>L. cenitri</em></td>
<td>NIV, D</td>
<td>[16]</td>
</tr>
<tr>
<td>JX173278</td>
<td>KP2</td>
<td>I</td>
<td>WA, Australia</td>
<td><em>Gleiria magnifica</em></td>
<td>N, M</td>
<td>[42]</td>
</tr>
<tr>
<td>H670651</td>
<td>SF1</td>
<td>I</td>
<td>WA, Australia</td>
<td><em>L. angustifolius</em></td>
<td>IC, D</td>
<td>[16]</td>
</tr>
<tr>
<td>H670665</td>
<td>AP92C</td>
<td>I</td>
<td>WA, Australia</td>
<td><em>L. angustifolius</em></td>
<td>IC, D</td>
<td>[16]</td>
</tr>
<tr>
<td>H670669</td>
<td>NG1</td>
<td>I</td>
<td>WA, Australia</td>
<td><em>L. angustifolius</em></td>
<td>IC, D</td>
<td>[16]</td>
</tr>
<tr>
<td>JX156423</td>
<td>SW3.2</td>
<td>II</td>
<td>WA, Australia</td>
<td><em>Gleiria sp.</em></td>
<td>N, M</td>
<td>[42]</td>
</tr>
<tr>
<td>H670650</td>
<td>MD7</td>
<td>II</td>
<td>WA, Australia</td>
<td><em>L. cenitri</em></td>
<td>NIV, D</td>
<td>[16]</td>
</tr>
<tr>
<td>H670686</td>
<td>AR87C</td>
<td>II</td>
<td>WA, Australia</td>
<td><em>L. angustifolius</em></td>
<td>IC, D</td>
<td>[16]</td>
</tr>
<tr>
<td>H670655</td>
<td>LMBBN</td>
<td>II</td>
<td>WA, Australia</td>
<td><em>L. angustifolius</em></td>
<td>IC, D</td>
<td>[16]</td>
</tr>
<tr>
<td>H670658</td>
<td>ESS5C</td>
<td>II</td>
<td>WA, Australia</td>
<td><em>L. angustifolius</em></td>
<td>IC, D</td>
<td>[16]</td>
</tr>
<tr>
<td>H670654</td>
<td>GE032A</td>
<td>II</td>
<td>WA, Australia</td>
<td><em>L. angustifolius</em></td>
<td>IC, D</td>
<td>[16]</td>
</tr>
<tr>
<td>H670659</td>
<td>ES11A</td>
<td>II</td>
<td>WA, Australia</td>
<td><em>L. angustifolius</em></td>
<td>IC, D</td>
<td>[16]</td>
</tr>
<tr>
<td>ABO79886</td>
<td>M11</td>
<td>III</td>
<td>Japan</td>
<td><em>Gleiria hybrid</em></td>
<td>IC, M</td>
<td>[43]</td>
</tr>
<tr>
<td>ABO79887</td>
<td>IB6</td>
<td>III</td>
<td>Japan</td>
<td><em>Gleiria hybrid</em></td>
<td>IC, M</td>
<td>[43]</td>
</tr>
<tr>
<td>ABO49729</td>
<td>CU3</td>
<td>III</td>
<td>Japan</td>
<td><em>Gleiria hybrid</em></td>
<td>IC, M</td>
<td>[44]</td>
</tr>
<tr>
<td>ABO79888</td>
<td>GE02</td>
<td>IV</td>
<td>Japan</td>
<td>-</td>
<td>-</td>
<td>[45]</td>
</tr>
<tr>
<td>D83749</td>
<td>M8GP</td>
<td>IV</td>
<td>Japan</td>
<td>-</td>
<td>-</td>
<td>[45]</td>
</tr>
<tr>
<td>NCO03492</td>
<td>MB4</td>
<td>IV</td>
<td>Japan</td>
<td>-</td>
<td>-</td>
<td>[45]</td>
</tr>
<tr>
<td>ABO49730</td>
<td>G1</td>
<td>IV</td>
<td>Japan</td>
<td><em>Gleiria hybrid</em></td>
<td>IC, M</td>
<td>[44]</td>
</tr>
<tr>
<td>AWS84180</td>
<td>Lismatatus</td>
<td>IV</td>
<td>Taiwan</td>
<td><em>Eustoma nusslannum</em></td>
<td>IC, D</td>
<td>unpublished</td>
</tr>
<tr>
<td>AV192568</td>
<td>GDD</td>
<td>IV</td>
<td>USA</td>
<td><em>Gleiria sp.</em></td>
<td>IC, M</td>
<td>[46]</td>
</tr>
<tr>
<td>ABO49732</td>
<td>921</td>
<td>V</td>
<td>Japan</td>
<td><em>Trifolium pratense</em></td>
<td>IC, D</td>
<td>[44]</td>
</tr>
<tr>
<td>UA7033</td>
<td>S</td>
<td>V</td>
<td>SA, Australia</td>
<td><em>Vicia faba</em></td>
<td>IC, D</td>
<td>[47]</td>
</tr>
<tr>
<td>H670666</td>
<td>LP</td>
<td>VI</td>
<td>WA, Australia</td>
<td><em>Vicia</em></td>
<td>IC, D</td>
<td>[16]</td>
</tr>
<tr>
<td>H670668</td>
<td>LPxKF</td>
<td>VI</td>
<td>WA, Australia</td>
<td><em>V. faba</em></td>
<td>IC, D</td>
<td>[16]</td>
</tr>
<tr>
<td>ABO49731</td>
<td>90-2</td>
<td>VII</td>
<td>Japan</td>
<td><em>V. faba</em></td>
<td>IC, D</td>
<td>[44]</td>
</tr>
<tr>
<td>H670667</td>
<td>F8</td>
<td>VII</td>
<td>WA, Australia</td>
<td><em>V. faba</em></td>
<td>IC, D</td>
<td>[16]</td>
</tr>
<tr>
<td>D0641248</td>
<td>WLMV</td>
<td>VIII</td>
<td>Malu, USA</td>
<td><em>Vicia</em></td>
<td>IC, D</td>
<td>[44, 49]</td>
</tr>
<tr>
<td>ABO49720</td>
<td>CS</td>
<td>VIII</td>
<td>Japan</td>
<td><em>Pisum sativum</em></td>
<td>IC, D</td>
<td>unpublished</td>
</tr>
<tr>
<td>NCO03536</td>
<td>CYYV</td>
<td>n/a</td>
<td>Japan</td>
<td><em>Phaseoaea vulgar</em></td>
<td>IC, D</td>
<td>[50]</td>
</tr>
<tr>
<td>H670670</td>
<td>CYYV AUS</td>
<td>n/a</td>
<td>NSW, Australia</td>
<td><em>T. repens</em></td>
<td>IC, D</td>
<td>[16]</td>
</tr>
</tbody>
</table>

Notes:
- *IC* indicates the sample originally came from a *L. angustifolius* plant with black pod syndrome.
- *NIV* stands for *introduced cultivated* plant; *M*, *monocott*; *N*, *native plant*; *NW*, *naturalized weed*.
- *IC* indicates the sample originally came from a *L. angustifolius* plant with black pod syndrome.
- *NIV* stands for *introduced cultivated* plant; *M*, *monocott*; *N*, *native plant*; *NW*, *naturalized weed*.

Host species came from within each group were: *Lupinus, Vicia* (Fabaceae), *Fresia* (Iridaceae) and *Diros* (Orchidaceae) in group I; *Lupinus and Diros* in group II; *Gleiria* (Fissicaceae) in group III, *Eustoma* (Gentianaceae) and *Gladiolus* in group IV, *Trifolium* (Fabaceae) and *Vicia* in group V, *Lupinus* in groups VI, VII and VIII; and *Pisum* (Fabaceae) in group IX. Thus, original host species represented in groups I, II and IV (the former general group) were from diverse origins, but those in the other groups were not. Therefore, phylogenetic group names based on natural hosts no longer seemed appropriate.

Recombination is one of the major means by which plant virus evolution and the emergence of new viruses or virus strains occurs [17–21]. There is evidence for high levels of recombination within the *Polyviridae* in particular [22–26]. Wylie and Jones [9] suggested that recombination played an important role in host specialization of BYMV following plant domestication. This suggestion was based on their analysis of seven complete genomes and 64 coat protein (CP) gene sequences. This predicted their general group to be ancestral in 12 out of 19 firm or tentative recombination patterns. However, recombination has been found to reduce host specificity and broaden natural host ranges, such as occurred with the emergence of Maize streak virus as an agricultural pathogen in Africa [21,27,28]. Therefore, given the subsequent availability of many more whole BYMV genomes and
an increase in the numbers of phylogenetic groupings [16], the suggested role of recombination in the evolution of host specialization of BYMV was warranted further analysis.

This research investigated the role that recombination plays in the evolution of BYMV. It examined the hypotheses that (i) recombination is associated with the expansion of natural host ranges; (ii) groups that contain isolates originally from both monocots and dicots (generalists), and that (ii) groups with narrow natural host ranges (specialists) might now be expanding their natural host ranges due to intermingling of strains formerly isolated in different crops. In addition, the study examined the role of recombination in the evolution of BYMV in the crop domestication centres, resulting in recombination events creating groups with broader natural host ranges. To address these hypotheses, we undertook recombination analyses of 33 complete BYMV coding regions and two of the closely related Cercospora lycopersici virus (C GVV). These analyses included 13 BYMV and 13 C GVV genomes obtained as part of research on BPS [16]. As potyviruses frequently undergo recombination (see above), wherever possible, whole genome sequences should be used for recombination analysis. Therefore, our research did not include recombination analysis of BYMV and C GVV CP genes, despite many more CP sequences being available on Genbank. To determine if recombination was playing a role in their symptom expression, our research also examined the example of infection with BYMV causing BPS (late infection) or systemic necrosis (early infection) in L. angustifolius plants [6,29].

Materials and Methods

Thirty-three complete or nearly complete BYMV genomes and two C GVV genomes were retrieved from Genbank (Table 1). They were trimmed to the length of their coding regions, and aligned by Clustal W in MEGA 5.2.1 prior to analysis for recombination [30]. The RDP4 package [31] was used to detect recombination between them. Default parameters were used for the seven programs implemented within RDP: RDP [32], GENECONV [33], bootscan [34], MaxChi [33], Chimera [36], SeqSaudi [37] and SiScan [38] which included using a Bonferroni-corrected P value cutoff of 0.05. A recombination pattern was considered to be a firm event, and genuine evidence of actual recombination, if detected by four or more of these programs, and anything less than four programs was not considered [9,24].

Results

When the complete coding regions of 33 BYMV and two C GVV isolates were analyzed, 12 firm recombination events were identified (Table 2, Fig. 1). The 16 sequences within phylogenetic groups I and II all had two recombination events across their P3, G1, G2, Vfg, Nia-Pro and Nib genes (events 1 and 2). The parental sequences for event 1 were from groups IV and VII. With event 2, one was unknown and the other from group IV. The seven sequences within group II also contained another event, event 3 which occurred across the Vfg, Nia-Pro and Nib genes. It had one unknown parental sequence and one from group I. Two of the sequences from group III (M11 and HSG) contained event 4, which was across the P1 and Hc-Pro genes. It had parental sequences from groups II and V. The third sequence from group III (G1a) contained event 5, located in the P1 gene. Its parental sequences were from groups V and IV. Four of the sequences from group IV (M1G, G1, L1 and GB2) contained recombination events 6 and 7. Event 6 was located across the Hc-Pro and P3 genes and event 7 across the region from P3 to Nib. The parental sequences for events 6 and 7 were groups V and II, and III and VII, respectively. The sequence from GB2 contained an extra event across the region from CI to Nia-Pro. Its parental sequences were from groups VII and IV. The sequence L1 and GB2 had another event across the Nib and CP genes, and its parental sequences were from groups III and IV. The remaining sequence from group IV (GDD) contained event 10, located across the P1 and Hc-Pro genes with parental sequences from groups IV and I. Event 11 was found in both group V and VI sequences, and stretched from the P3 to the Nib regions. Parental sequences for event 11 were an unknown and group III. Group VI sequences (LP and LPxP6) had an additional event (event 12) in the Nib region with parental sequences from group V and an unknown sequence. There was no evidence of recombination in sequences from groups VII to IX, or in the C GVV sequences. The greatest P-values across all 12 recombination events ranged from 6.701×10^-10 to 1.9666×10^-10.

Six of the sequences analyzed from groups I and II (PN83A, PN80A, AR93C, AR97C, ES55C and GB32A) were BYMV isolates from L. angustifolius plants with BPS, but there was no recombination event specific to these sequences. This was also the case for three isolates (GB17A, NG1 and ES11A) from L. angustifolius plants with systemic necrosis within groups I and II.

Discussion

Our research found extensive recombination amongst diverse BYMV genome sequences which is likely to have significant evolutionary implications for the virus. It revealed the presence of extensive recombination within three BYMV phylogenetic groups that include both monocots and dicots as natural hosts, supporting the suggestion that recombination leads to broadening of natural host ranges. It therefore provides evidence for the hypothesis that recombination is responsible for the wide natural host ranges of the BYMV groups that invade both dicots and monocots. It also found recombination events in three BYMV phylogenetic groups with narrow natural host ranges indicating they might now have the potential to broaden their natural host ranges. It therefore provides support for the hypothesis that groups with narrow natural host ranges might now be expanding their natural host ranges due to intermingling of strains formerly isolated from each other within crop domestication centers, resulting in recombination events and broader natural host ranges. Such a scenario would occur as a result of recombination within mixed infections between previously isolated groups. Thus, past expansion of international trade in plants and plant products would have brought BYMV isolates that evolved in isolated crop domestication centers into contact with each other resulting in recombination. These results have broader implications concerning the likely role of recombination in the evolution of plant viruses in general, especially where a distinction exists between specialist and generalist virus groups. Our research also found no indication that recombination is playing a role in producing isolates causing BPS or systemic necrosis in L. angustifolius plants.

Our results resemble those of Wyley and Jones [9] in that the recombination patterns found were similar. However, the dataset from our whole genome analysis was much larger (35 compared to their eight) and revealed four additional firm recombination events. Overall, we detected 12 such events across 33 BYMV and two C GVV genomes, whereas they detected eight events across seven BYMV and one C GVV genome. Their study also identified three tentative recombination events involving BYMV genomes from group IV and an unknown parent within the 3' region of the C GVV genome. In contrast, our analysis, which excluded tentative recombination events, did not reveal any firm events involving
Figure 1. Recombination events between the coding regions of 33 bean yellow mosaic virus (BYMV) and two from Clover yellow vein virus (CYVV) genomes. The locations of genes in the BYMV genome are indicated by the diagram at the top of the Figure. Twelve recombination events were found, labeled 1–12. Each recombination event correlates with the events column in Table 2. Each color represents a phylogenetic group, apart from purple, which represents two such groups, I and II. The phylogenetic groupings of Keohoe et al. 2014a are indicated at the left hand side of the picture. The colour of each event refers to the phylogenetic grouping of its predicted parental sequences, which are detailed within Table 2. The white colour represents events whose parental sequences are unknown. The sequences analyzed were: HG970847, HG970851, HG970852, HG970860, HG970861, HG970865, HG970868, F4920841 and JX173278 (Phylogenetic group I); HG970850, HG970854, HG970853, HG970858, HG970859, HG970863 and JX156423 (II); AB079886, AB079887 and AB439739 (III); D63749, AM688418 and AY192568 (IV); AB439732 and U47033 (V); HG970866 and HG970868 (VI); AB439731 and HG970867 (VII); DQ841248 (VIII); AB317320 (IX); NC003536 and HG970870 (CYVV). No recombination events were detected in sequences from phylogenetic groups VII–IX or within CYVV, but a sequence from group VII is suggested as a parental sequence for one of those from group IV.

doi:10.1371/journal.pone.0105770.g001

either of the two CYVV sequences as a parent. The use of more whole genome sequences gives us greater confidence in the results.

Our results showed eight recombination events within the former general group, now groups I, II and IV (two or three events per genome), and five amongst the former specialist groups where groups III and VI had two events each and group V had one event. Groups VII–IX had no recombination events. Our findings therefore showed that the groups with the most recombination had the broadest natural host ranges that included both monocots and dicots (I, II, IV). They also found recombination within groups III, V and VI (formerly specialist groups) thereby giving them the potential to broaden their natural host ranges and thus generalize. However, caution is required over our interpretation as groups V–IX were only represented by one or two genomes each, so there are likely to be as yet undetected recombination events. Likewise, the limited numbers of sequences in groups V–IX also make deductions difficult regarding (i) the parents of these sequences, or (ii) the roles of these sequences as parents in other recombination events generally. Also, one of the specialist phylogenetic groups based on CP genes reported by Wylie et al. [13] was their calusa group. Isolates from this group were unrepresented by complete genome sequences, so they could not be evaluated.

All three recombination events present in BYMV groups I and II encompass the P3, P1PO, CI and VPg regions of the genome. These regions are responsible for pathogenicity, virus long distance movement, virulence determination towards potyvirus resistance, replication and protein-protein interactions [8,11,12]. However, the recombination events we detected were in isolates originally collected from symptomatic plants in field, greenhouse and experimental situations, so pathogenicity was the only characteristic that could be related to recombination. Moreover, not all viral recombinants will necessarily give rise to viable, fit variants. The nature of potyviruses is such that functions of some genes overlap with others [8]. Recombinant fitness is determined by (i) the degree to which intragenome interactions are disrupted by the event, and (ii) the divergence between the exchanged sequences, where the higher the divergence, the greater the probability that intragenome disruption will occur [39]. Recombinant virus strains or isolates with disrupted intragenome interactions are likely to be removed by negative or purifying selection, e.g. as found within the Geminivirus [39]. Thus, the recombination events detected in our analysis do not reflect overall BYMV recombination rate.

Most of the complete BYMV genomes available for analysis were from Australia or Japan, so there is little scope for deductions based on geography. With the exception of one from a Freista spp. in South Korea, all isolates with genomes that fit into groups I and
Table 2. Recombination events in the coding regions of 33 Bean yellow mosaic virus and 2 Clover yellow vein virus genomes.

<table>
<thead>
<tr>
<th>Event</th>
<th>Phylegetic groupings</th>
<th>Recombinant sequences</th>
<th>Programs detected by</th>
<th>Start position in genome</th>
<th>Genes affected</th>
<th>Parental sequences</th>
<th>Parental phylogenetic group</th>
<th>P-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I, II</td>
<td>PN83A, PN86A, GB17A, Fe, MD1, KP2, SP1, ARB3C, NG1, SW93, MD7, ARB7C, LMBNN, ESS5C, GB32A, ES11A</td>
<td>R, G, B, M, C, S, 3</td>
<td>2947-3069</td>
<td>P3, P10O</td>
<td>GL +90-2</td>
<td>IV × V8</td>
<td>1.30 × 10^{-78} (3)</td>
</tr>
<tr>
<td>2</td>
<td>I, II</td>
<td>PN83A, PN86A, GB17A, Fe, MD1, KP2, SP1, ARB3C, NG1, SW93, MD7, ARB7C, LMBNN, ESS5C, GB32A, ES11A</td>
<td>R, G, B, M, C, S, 3</td>
<td>5203-5457</td>
<td>CI</td>
<td>unknown + MBGP</td>
<td>unknown × IV</td>
<td>3.13 × 10^{-46} (G)</td>
</tr>
<tr>
<td>5</td>
<td>III</td>
<td>Gla</td>
<td>R, G, B, M, C, S, 3</td>
<td>1-191 (undetermined)</td>
<td>3’UTR-P1</td>
<td>S × MBGP</td>
<td>V × IV</td>
<td>7.042 × 10^{-100} (G)</td>
</tr>
<tr>
<td>6</td>
<td>IV</td>
<td>MBGP, G1, Lisanthus, GB2</td>
<td>R, G, B, M, C, S, 3</td>
<td>1721-1929</td>
<td>Hc-Pro</td>
<td>S × ESS11A</td>
<td>V × III</td>
<td>1.105 × 10^{-39} (G)</td>
</tr>
<tr>
<td>7</td>
<td>IV</td>
<td>MBGP, G1, Lisanthus, GB2</td>
<td>R, G, B, M, C, S, 3</td>
<td>2231-2318</td>
<td>Hc-Pro</td>
<td>M11 ×99-2</td>
<td>II × VII</td>
<td>1.773 × 10^{-45} (S)</td>
</tr>
<tr>
<td>8</td>
<td>IV</td>
<td>GB2</td>
<td>R, G, B, M, C, S, 3</td>
<td>5506-5556</td>
<td>CI-6k2</td>
<td>92-2 × G1</td>
<td>VI × V</td>
<td>1.968 × 10^{-190} (G)</td>
</tr>
<tr>
<td>9</td>
<td>IV</td>
<td>Lisanthus</td>
<td>R, G, B, M, C, S, 3</td>
<td>8336</td>
<td>Nb</td>
<td>lbg × G1</td>
<td>II × IV</td>
<td>1.661 × 10^{-78} (S)</td>
</tr>
<tr>
<td>10</td>
<td>IV</td>
<td>GDD</td>
<td>R, G, B, M, C, S, 3</td>
<td>1-191 (undetermined)</td>
<td>3’UTR-P1</td>
<td>MBGP × PN83A</td>
<td>IV × I</td>
<td>7.249 × 10^{-48} (G)</td>
</tr>
<tr>
<td>11</td>
<td>V, VI</td>
<td>LP, LpexFB, S, 92-1</td>
<td>R, G, B, M, C, S, 3</td>
<td>3236-3306</td>
<td>P3</td>
<td>unknown × M11</td>
<td>unknown × III</td>
<td>3.051 × 10^{-67} (G)</td>
</tr>
<tr>
<td>12</td>
<td>VI</td>
<td>Lo, LpexFB</td>
<td>R, B, M, C</td>
<td>7588-7872</td>
<td>Nb</td>
<td>92-1 × unknown</td>
<td>V × unknown</td>
<td>6.701 × 10^{-77} (M)</td>
</tr>
</tbody>
</table>

*Phylegetic grouping determined by Keese et al. (2014a).
†RDP; G, GENECNNY; B, Botetcian; M, Macich; C, Chimonast; S, SScan; L, Seq.
‡Numbers represent nucleotide position in the genome.
§Source of recombinant fragment. Minor parent is listed first, followed by the major parent.
¶The P-value is the greatest value for the event in question.
**The program which detected the greatest P-value.
DOI:10.1371/journal.pone.0105770.t002
II were collected from south-western Australia. Moreover, there are also BMYV isolates from Australia in three other groups (V, VI and VII). These findings reinforce the suggestion that BMYV arrived in Australia at least five different occasions and that international trade, for example of bulbs and seeds, is likely responsible for the worldwide distribution of BMYV [13,40].

It appears unlikely that any of the recombinant events detected in groups I and II (events 1, 2 and 3) were responsible for the emergence of BPS as a significant disease of L. angustifolius caused by BMYV. No recombination event was specific to the BMYV isolates originally from plants with BPS. Moreover, they did not differ from the four BMYV isolates originally from L. angustifolius plants with systemic necrosis, and one other from a plant with a susceptible reaction (non-necrotic symptoms) [6,29,41]. Furthermore, recombination analysis did not distinguish sequences of these L. angustifolius isolates from those of any other hosts in groups I or II.

References


As more whole genomes sequences are submitted to datastores, particularly from regions of the world in which BYMV specialist groups may have originated, or where crop domestication has occurred, the picture should become clearer and we will be better able to answer the question for BYMV: to specialize or not to specialize?

Acknowledgments

We thank E. Gao, S. Vincent and M. Banovic for glasshouse and laboratory support.

Author Contributions

Conceived and designed the experiments: MAK RAC BJ. Performed the experiments: MAK. Analyzed the data: MAK. Contributed reagents/materials/analysis tools: MAK. Contributed to the writing of the manuscript: MAK RAC BJ.
Chapter 6

Summary and concluding remarks
Summary and concluding remarks

6.1 Introduction

South western Australia has a Mediterranean-type climate with predominantly sandy soils suitable for growing *L. angustifolius* crops (Buirchell 2008; Gladstones 1998a). While yields are commonly 1-2t/ha, the potential exists for 4t/ha. Currently, one of the major reasons for poor yields is BPS, which was first noticed in the 1990’s (Gladstones 1998b). *L. angustifolius* plants affected by BPS are seen late in the growing season after first flowering when they develop characteristic flat, black pods that have little or no seed. Damage to crops can be extensive with losses of up to 40% reported for partially infected stands (Buirchell 2008). Possible causes of BPS that were hypothesised included (i) excessive vegetative growth leading to poor pod set, (ii) nutrient deficiencies and (iii) late infection with BYMV. The first two hypotheses were not supported following a series of field experiments (White *et al.* 2007; White and Baker 2009). An added challenge to the acceptance of the third hypothesis was that (i) BPS symptoms were sometimes present in the primary pods without any of the necrotic stem streaking symptoms usually associated with BYMV infection in *L. angustifolius* and (ii) ELISA tests failed to detect BYMV from the majority of plant samples with BPS tested (Jones, R.A.C. unpublished). Also, early infection with BYMV normally causes systemic necrosis and plant death and so is associated with different symptoms to BPS (Jones and McLean 1989; Jones and Coutts 1998; Jones 2001).

6.2 Summary of key findings

6.2.1 Black pod syndrome of *L. angustifolius* is caused by late infection with BYMV

In 2011, an end of growing season survey of *L. angustifolius* crops was conducted across six locations in south western Australia (Chapter 2). A total of 72 *L. angustifolius* plants with BPS were collected, sampled at different positions on each plant and tested for BYMV, the closely related virus CIYVV and generic potyvirus by both ELISA and RT-PCR. Detection was most reliable when RT-PCR was used to test tissue taken from just below black pods on the main stem of the plant. All plants with
BPS were positive for BYMV when tested at this location on the plant. In glasshouse experiments, plants at different growth stages were inoculated with BYMV, and Koch’s postulates were satisfied for the hypothesis that late infection with BYMV causes BPS in *L. angustifolius* plants. The growth stage at time of infection was critical for the expression of BPS, as infection had to occur after first flowering and during pod set on the primary stem.

### 6.2.2 Potential risks to *L. angustifolius* from other plant viruses

In parallel to the survey conducted for *L. angustifolius* plants with BPS in 2011, a sample of a *L. cosentinii* plant showing leaf mosaic and deformation was obtained from a strip of vegetation at the agro-ecological interface between managed and native vegetation which was surrounding experimental field plots (Chapter 3). Also, sampled at the same time were two *H. comptoniana* plants showing mosaic and leaf deformation that were growing in close proximity to the *L. cosentinii* plant sampled. All three plant samples were positive for generic potyvirus, but negative for BYMV. Following analysis by NGS of the total RNA obtained from these three samples, four complete and two nearly complete genomes of the Australian native potyvirus *Hardenbergia mosaic virus* (HarMV) were obtained. Four of these were distinct strains infecting the same *H. comptoniana* plant, and the other two were one sequence each from *L. cosentinii* and *H. comptoniana*. The sequence obtained from *L. cosentinii* constituted the first report of HarMV naturally infecting *L. cosentinii*. It was also the first report for the SWAFR of virus emergence occurring naturally from a native plant to invade an introduced species. Previously in glasshouse studies and experimental field plots HarMV was found to cause similar systemic necrosis in *L. angustifolius* to that caused by early infection with BYMV (Webster *et al.* 2007, Luo *et al.* 2011). Further research is needed to determine if the spread of HarMV to lupins might be of concern for the *L. angustifolius* industry in the future.

### 6.2.3 Biological and molecular analysis of BPS causing isolates of BYMV

After sequencing 22 isolates of BYMV (including 12 from plants with BPS) and one of ClYVV by NGS, a total of 23 BYMV and one ClYVV genomes were obtained (Chapter 4). Phylogenetic analysis of these new complete genomes along
with the 17 already available on genbank revealed that the previous phylogenetic
group names of Wylie et al. (2008) based on the original hosts each isolate came from
were no longer appropriate. This was because there were now genomes from *V. faba*
in two groups, genomes of *Lupinus* spp. in four groups and from monocots in four
groups. I therefore adopted a numbered grouping system (I-IX) for the nine
phylogenetic groups. Phylogenetic analysis showed that all isolates from south
western Australia belonged to groups I or II (with isolate LP the only exception),
including those from symptomatic *L. angustifolius* and *L. cosentinii* plants. When
seven different BYMV isolates were inoculated to a range of crop legumes and
indicator plants in a glasshouse study, there were no discernable differences between
isolates taken from *L. angustifolius* plants with BPS and those from other sources. An
important point to note is that almost all of the BYMV whole genome sequences
available now are either from Japan or Australia, which makes comments based on
BYMV geographical origins of limited value.

6.2.4 Next generation sequencing and plant virology.

The uptake of NGS by plant virologists (and plant pathologists generally) has
been slower than by their counterparts in the medical sciences. However, this
technology has rapidly evolved into an affordable and accessible tool now appearing
on laboratory bench-tops throughout the world (Mardis 2013, Koboldt et al. 2013).
Plant virologists are increasingly turning to NGS to obtain whole plant virus genomes
in a relatively short time frame (Boonham et al. 2014). This study found that the use of
CLC genomics workbench for de novo assembly was sufficient to successfully
distinguish multiple strains, or phylogenetic groups, of a virus in a single sample
(Chapters 3 and 4). I also found that the use of mapping NGS reads to a reference
sequence in Geneious, followed by the creation of a consensus sequence between the
mapped consensus with the de novo assembly could often help to fill in any gaps.
There was also a minimum level of coverage required across a genome to ensure an
accurate representation was obtained. In the medical field, a minimum 30 times
coverage across a human genome is considered adequate (Wetterstrand 2014), and so
this could serve as the suggested minimum until such time as a comprehensive
comparison across plant virus datasets can be carried out. When the new BYMV or
HarMV genomes were considered, they often had average coverages across their genome well into the 1000’s of times, which was necessary to ensure that the 3’ and 5’ UTR regions of the genome were adequately captured. Those with lower average coverage often did not have all of their UTR regions. All genomes with lower than ten times average coverage in our study were designated as partial genomes, because they contained gaps.

6.2.5 Recombination analysis of BYMV genomes, including those from plants with BPS

The coding regions of 33 BYMV and 2 CIYVV genomes (including those sequenced as part of Chapter 4) were analysed for evidence of recombination. Good evidence for 12 recombination events was found, and none were unique to sequences from plants with BPS. The events were also aligned with a particular phylogenetic group or groups (Chapter 5). When these results are considered with regards to BYMV evolution, I found that most recombination events were in the three groups which made up the former general group which infect both monocots and dicots indicating that recombination is associated with the expansion of host ranges. Also, recombination within two former specialist groups may be resulting in groups with broader natural host range (re-generalisation). These findings do not support the suggestion of Wylie and Jones (2009) that recombination was involved in the host specialisation of BYMV. However, phylogenetic groups V-IX were only represented by one or two sequences each, and one of the former specialist groups, the canna group, had no full genome representative. Also, most of the sequences available came from either Japan or Australia. Therefore, this study only provides preliminary data. Further analysis of complete genome sequences is required to provide more comprehensive information about the role of recombination in the evolution of BYMV. In particular, more isolates from poorly represented groups need to be included in such analyses.

6.3 Future directions

It is clear from this research that the solution for BPS in its host L. angustifolius should include breeding for BPS resistance, in combination with use of
integrated disease management strategies already available for BYMV in *L. angustifolius* (Jones 2001). Unfortunately, the nature of the infection, i.e. a late growing season infection, means that any glasshouse studies or resistance screening exercises would likely be awkward and time consuming. I found that it took 5-6 months to successfully carry out a glasshouse experiment with BPS, and this was with < 100 plants. I also found no differences in symptom expression between BYMV isolates from *L. angustifolius* plants with BPS, systemic necrosis, plants with non-necrotic BYMV symptoms or from other species when they were inoculated to host ranges that we might have expected to enable differentiation of BYMV isolates that came from plants with BPS. Field screening for BPS resistance in single row plots with BYMV infector plants introduced later rather than early as has been done in the past (Jones and Coutts 1998) might be more likely to succeed, rather than relying on unpredictable late natural infection of CVT trials year-to-year to assess breeding material for potential BPS resistance.

Modern plant breeding makes use of marker-assisted selection and so the identification of markers for BPS in *L. angustifolius* should be a first goal. There is considerable potential for a set of RNA-seq experiments using NGS technology. This is where the transcriptome of the plant is sequenced and assembled, and shows which genes are functionally active in the tissue sequenced (Künster *et al.* 2010). This comparative genomics approach is a suitable method to gain understanding of the genetic basis of phenotypic variation (Ellegren and Sheldon 2008). It is being used more and more as the technology becomes cheaper and more accessible (e.g. Elmer *et al.* 2010; Künster *et al.* 2010). For BPS, an RNA-seq experiment would entail an initial glasshouse experiment involving *L. angustifolius* plants each with and without BPS, which would then be sampled and the material extracted before being subjected to NGS. The comparison between plants with and without BPS should provide a series of candidate genes in *L. angustifolius* that could be further investigated as potential markers for BPS resistance. This is possible now thanks to the existence of the draft genome of *L. angustifolius* (Nelson *et al.* 2010; Yang *et al.* 2013), and a recent publication detailing just such an approach with Plum pox virus in *Prunus* spp. (Rodamilans *et al.* 2014) The panel of markers that result from such a study would be
useful as resistance to BPS is likely to be a quantitatively inherited trait, not single gene resistance and so multiple gene candidates are needed. Follow up with phenotyping in glasshouse experiments would still be required.

Careful consideration is required however, in determining what form a BPS resistant cultivar should take. Cv. Jenabillup has partial BPS resistance, but this is really no more than an ability to perform better than other cultivars (e.g. Mandelup) under low disease pressure. In years when BYMV infection is at high levels in crops, it develops similar incidences of BPS to those in other cultivars (G.J. Thomas pers. comm.) The best-case scenario would be a cultivar of *L. angustifolius* with extreme resistance to BYMV, but this type of resistance has never been found in *L. angustifolius* during many years of field screening for BYMV resistance in diverse accessions, breeding lines and cultivars (Jones and Coutts 1998; Jones 2001). Also, such single gene resistance to viruses in other crops sometimes breaks down (Harrison 2002; Astier *et al.*, 2007). There are examples of other grain legume crops with single gene resistance to BYMV present in them, such as the *mo* gene for *P. sativum* (Schroeder and Provvidenti 1971) and the *Nbm-1* gene within *L. angustifolius* itself (Cheng and Jones 1999, 2000), although both are strain specific.

If *L. angustifolius* were bred to be without the *Nbm-1* gene resulting in the expression of non-necrotic BYMV symptoms (mottle and plant stunting), that might bring with it further problems. Indeed, yield losses were greater when non-necrotic strains of BYMV were present in *L. angustifolius* stands than when necrotic strains were present (Jones *et al.* 2003). The explanation was that in the presence of the systemic necrosis causing *Nbm-1* gene, when the plant is infected at an early growth stage with most BYMV isolates, it will die and so remove the source of BYMV infection for further spread from the field. Thus, when susceptible cultivars are used, *Nbm-1* no longer operates which would create an even bigger problem than currently occurs with BPS. This is because there would be more BYMV infected source plants within crop for a longer period of time, so more plants become infected overall. Also, BYMV is not known to be seed-borne in *L. angustifolius*, but this is maybe because early infected plants do not produce seed. It is seed-borne in some pasture legumes in
south western Australia, and the BYMV strains that infect lupin in Europe and North America (e.g. from group VI) are seed-borne in *L. luteus* and *L. albus* (McKirdy *et al.* 2004; Jones and Mclean 1989; Jones 2001). It seems unlikely that removal of the *Nbm-1* gene would create another CMV seed transmission situation in *L. angustifolius* (Jones 1988, 2000; Jones and Cowling 1995), as no seed transmission was found with BYMV non-necrotic strains in previous studies (Cheng and Jones 2000; Cheng and Jones 2002; Jones *et al.* 2003).

Another avenue for investigation is resistance to aphids feeding on, and transmitting BYMV to, *L. angustifolius*. Are the differences in the field seen between the cultivars Jenabillup and Mandelup in years when virus levels are relatively low due to aphid feeding behaviour perhaps? Or is it more to do with the speed with which BYMV is able to move systemically through the plant itself? Do aphids perhaps require more or fewer probes per feed on either cultivar to actually induce infection in the plant? Could *L. angustifolius* be bred to be less attractive to its aphid vectors? However, any solution to do with the aphid side of the problem is likely to be troublesome, and breeding for resistance to aphid vectors has rarely been effective at controlling virus spread with other pathosystems. The aphids that transmit BYMV in *L. angustifolius* crops are mostly non-colonizing and transmit the virus non-persistently making the problem particularly difficult (Berlandier *et al.* 1997). Breeding for BYMV infection resistance (i.e. breeding cultivars which require more infective aphids to probe them before they become infected), or resistance to BYMV systemic movement in the plant itself might be more worthwhile. Such resistance types have been used previously with other pathosystems, such as *Potato leafroll virus* in potato (e.g. Beemster 1987; Wilson and Jones 1992, 1993a, b).

One further suggestion could be to further explore a transgenic approach to BYMV resistance in *L. angustifolius*. However, efforts in this area have not proven successful in the past. Early results were promising when a synthetic ‘hairpin’ replicase gene from BYMV was generated. Three independent transgenic lines showed promise at generations T2 and T3 with what appeared to be resistance to BYMV isolate MI, but unfortunately subsequent progeny from these lines were susceptible. The resistance
had been silenced and was no longer expressed (Jones 2001; Jones et al. 2008). These results suggest that further transgenic resistance studies might not provide an effective solution to BPS readily.

Future efforts should also refocus on improving the diagnostic methods used to test BPS symptomatic plants for BYMV. Given the nature of the symptoms (late infections, black pods, stem necrosis), visual diagnosis with a potential follow up with a generic potyvirus RT-PCR can be used to confirm late infections in *L. angustifolius*. However, for this, it would be worthwhile to redesign the BYMV specific primers studies to increase their specificity using the new full length genome sequences obtained in these. For large-scale testing of samples, replacement of conventional RT-PCR through development of a real-time PCR would be helpful as this technique is likely to be not only more effective and potentially labour saving as a routine test, but also more sensitive.

Regarding the molecular analysis of BYMV, it is apparent that to gain a better understanding of the phylogenetic groupings and genomic evolution of the virus more sequences are needed from isolates across a wider range of hosts. Currently, the majority of BYMV whole genome sequences available on genbank come from just two countries: Australia or Japan, and there are few from anywhere else. We have reached the point now with NGS technology and analysis that whole genome sequences are a feasible option. With *Potyviruses*, CP gene sequences are not as reliable for analysis of phylogenetic placement or recombination, as has been found with a number of other potyviruses such as PVY, TuMV and now HarMV and BYMV (Karasev and Gray 2013; Ohshima 2013, Chapters 3, 4, and 5).

### 6.4 Closing remarks

We now know that late infection with BYMV causes BPS in *L. angustifolius*. What we don’t know or understand is the response to the virus in the mature plant itself. I did not investigate whether mature plant resistance is involved. A better understanding of the plant response is what is required next in the quest to breed a BPS resistant cultivar. The type of the BYMV resistance that should be pursued needs
to be carefully considered, to ensure that in doing so we don’t exacerbate the problem that already exists. For the last 15 years or so the focus has been on field screening for resistance to early BYMV infection, and evaluation of BPS late in the growing season in large-scale CVT trials. This has led to better performing cultivars such as Jenabillup, but as mentioned previously it has partial BPS resistance rather than extreme resistance. If a cultivar with extreme resistance to BYMV has not been found in *L. angustifolius* in the last 15 years, it is unlikely to eventuate in the near future by the same methods.

Switching the search to genomic comparison methods might provide a faster path to a suite of potential molecular markers than the field based largely phenotypic methods currently used. Of course, the plants with these molecular markers would still need to be phenotyped to determine their ultimate usefulness.

In the meantime, the management of BPS is best implemented through use of already established integrated disease management recommendations for BYMV in *L. angustifolius* in Australia. These measures include:

- using perimeter non-host barriers
- avoiding fields with a large perimeter:area ratio
- retaining stubble ground cover
- promoting early crop canopy development
- generating high plant densities
- using close row spacing
- maximising weed control
- using crop rotation

Given the critical importance of the growth stage at the time of infection (i.e. late stage at podding) for BPS, the particular emphasis needs to be on maximising volunteer clover control within the crop. Equally important is avoiding planting next to or downwind of legume pastures, other *L. angustifolius* crops, other grain legume crops or native legumes likely to contain sources of BYMV.
Literature cited


Appendix

Publications, presentations and extension activities arising from this thesis
Primary authored manuscripts:


Conference Presentations:


Presentations at local symposium and events
2012 Rottnest Postgraduate Summer School, School of Plant Biology, UWA
Oral presentation: Unravelling the cause of black pod syndrome of narrow-leafed lupin and developing a control solution.


2013 Rottnest Postgraduate Summer School, School of Plant Biology, UWA Oral presentation: Next Generation Sequencing of Bean Yellow Mosaic Virus


2014 Rottnest Postgraduate Summer School, School of Plant Biology, UWA Oral presentation: Bean yellow mosaic virus – next generation sequencing data and recombination analysis.

2014 PhD exit seminar. Held at Department of Agriculture and Food Western Australia, South Perth. Oral presentation: Bean yellow mosaic virus: from Koch’s postulates to next generation sequencing and their use to unravel the cause of black pod syndrome of narrow-leafed lupin.


Media/extension articles and interviews

2012 SeedQuest, article. Japanese research to help Western Australia’s lupin growers. 13 April 2012. Available at: http://www.seedquest.com/solutions.php?type=solution&id_article=26076&id_region=&id_category=&id_crop

2012 Countryman (WA), article. Monica has yen to study lupin pods. 19 April 2012.

2012 ABC Goldfields and Midwest Wheatbelt, radio interview. Rural and resources report, 20 April 2012, 6.40am, 4 mins 47s.