FINAL REPORT*

Plant Health Australia Ltd

Molecular Tagging of Rust Resistance Genes in Eucalypts

Bala Thumma[1], Geoff Pegg[3] Paul Warburton[2], Jeremy Brawner[2], Paul Macdonell[2], Xiaohui Yang[1] and Simon Southerton[1]

[1] CSIRO Plant Industry, GPO Box 1600, Canberra ACT 2601

[2] CSIRO Plant Industry, Queensland Biosciences Precinct, 306 Carmody Rd, St Lucia Qld 4067

[3] Agric-Science Queensland, Ecosciences Precinct, 41 Boggo Road, Dutton Park, Brisbane, Qld 4102

REPORT SUMMARY

Resistance to myrtle rust is quantitative and likely to be controlled by variation in many genes. Using single SNP (single nucleotide polymorphism) marker and SNP:SNP interaction analyses we identified in the vicinity of 10 to 15 SNPs that are likely to play a role in rust resistance. Three out of six highly significant SNPs occur in three different bZIP transcription factors which have been implicated in biotic and abiotic stress responses. The other three SNPs occur in three NB-LRR resistance genes situated on chromosome 3 in the vicinity of the previously identified rust resistance locus *Ppr1.* Individual SNPs explained small proportions (1.5 to 3.4%) of genetic variation in resistance. These findings suggest that the pathogen is unlikely to rapidly evolve virulence to substantially more eucalypt genotypes and that breeding for resistance will be most efficient if more markers are identified and employed in marker-assisted selection. Using a prediction model based on 13 SNPs we were able to predict 70% of resistant and susceptible trees correctly in a population of 300. Using the methods described in this report it should be possible to identify most of the molecular variants controlling rust resistance in eucalypts, which should result in much higher accuracy prediction models.

* See appendix A for Milestone calendar

PROJECT AIM

The aim of the project was to identify molecular variation in eucalypt genes (single nucleotide polymorphisms (SNPs) or insertions/deletions (indels)) that confers resistance to the myrtle rust fungus.

INTRODUCTION

Myrtle rust in Australian forests

Myrtle rust (*Puccinia psidii*) is a disease of South American origin that attacks a wide range of myrtaceous species including eucalypts. It is a serious disease threat to Australian native flora and plantation forests in tropical, sub-tropical, and possibly temperate regions in Australia (TOMMERUP *et al.* 2003). Myrtle rust attacks the growing tips of eucalypt foliage. It does not appear to attack mature foliage or juvenile foliage over three weeks in age. As of July 2012 the pathogen had spread to Florida, Hawaii and Japan. In April 2010, *P. psidii*, initially described as *Uredo rangelii*, was detected for the first time in Australia on the central coast of New South Wales. Both *Corymbia* and *Eucalyptus* have subsequently been shown under quarantine conditions to be susceptible to the rust (Louise Morin, pers. comm.). This research has shown that within most eucalypt species there are resistant and susceptible genotypes. Forest pathologists are now convinced, based on host range, morphology, and molecular evidence, that myrtle rust is a pathotype of the broader guava rust complex. Consequently, the impact of the disease caused by this new rust fungus is likely to be similar to that of guava rust in other countries (CARNEGIE *et al.* 2010).

In Brazil where *Eucalyptus grandis*-based hybrids are extensively used, guava rust causes severe damage to the plantation forestry. While *E. globulus* and *E. nitens* are particularly susceptible to the rust (ALFENAS *et al.* 2003; TELECHEA *et al.* 2003; TOMMERUP *et al.* 2003), preliminary bioclimatic modelling suggests that most temperate eucalypt plantations in Australia are located in regions of relatively low risk (GLEN *et al.* 2007). However, these predictions are based on limited data and it is quite possible that the pathogen will attack plantations outside of the currently predicted higher risk regions (GLEN *et al.* 2007). It is also likely that once the pathogen starts spreading from the region of where the incursion was

first detected its range may extend beyond the predicted range. Climate change is also likely to affect future risk distribution patterns.

Genetic control of rust resistance in Eucalyptus

Prior to the commencement of the project a major locus for myrtle rust resistance (*Ppr*) had been identified on linkage group3 (≈ chromosome 3) of the *Eucalyptus* genome using several full-sib families of *E. grandis* (JUNGHANS *et al.* 2003; MAMANI *et al.* 2010). This suggested that resistance to myrtle rust was likely to be controlled by variation in a small number of genes of relatively large effect. However, recent genetic studies indicate that rust resistance in *Eucalyptus* is under relatively complex genetic control, meaning that resistance may be under the control of many genes. Genetic control of rust resistance in *Eucalyptus* is influenced by both additive and epistatic effects (ALVES *et al.* 2012).

As a first step towards breeding *Eucalyptus grandis* for resistance to guava rust, Mamani et al (2010) recently identified several microsatellite markers linked to the major locus controlling rust resistance (Ppr1) on linkage group 3. One of the microsatellite markers linked to the Ppr1 was developed at CSIRO and has been mapped on a *E. globulus* linkage map (THAMARUS *et al.* 2002). However, these quantitative trait locus (QTL) markers are only useful for screening progeny from specific crosses involving resistant parents. Forest tree breeding populations generally contain numerous unrelated families. Markers identified using QTL studies are unlikely to be useful in other families due to the break down in linkage between markers and the disease resistance locus. This was shown by Mamani et al (2010) who did not find any association with the resistance locus when these markers were tested in a breeding population consisting of 96 unrelated trees. Population-based association studies are therefore needed to identify markers that are directly linked to the disease locus and which can be used across different genetic backgrounds.

RESEARCH METHODS

In light of previous research, we hypothesised that resistance to myrtle rust was oligogenic, or under the control of a small number of genes. We aimed to discover these genes using a bulked segregant analysis approach in natural populations.

Eucalypt populations, inoculation and assessment

Seed of *E. nitens*, *E. globulus* and *E. grandis* were obtained from the Australian Tree Seed Centre. Approximately 300 seedlings of each species were screened for myrtle rust resistance in collaboration with Geoff Pegg at the Queensland DAFF facilities at Boggo Rd.

Puccinia psidii urediniospores were originally collected from *Rhodamnia sessiliflora* and multiplied in the glasshouse using *Syzygium jambos* and *Rhodamnia rubescens*. Urediniospores were collected and placed into a desiccator for 48 hours before being placed into Nunc tubes and stored at -80^{o}C until required. Urediniospores were removed from -80^{o}C storage and allowed to warm to room temperature prior to being added to sterile distilled water (SDW). The surfactant 'Tween 20' was added at a rate of two drops per 100ml of SDW and the spore suspension stirred to reduce clumping. Spore counts were then conducted using a haemocytometer and the suspension adjusted to a concentration of 1 x 10^5 spores/ml.

Seedlings were inoculated using a fine mist spray (2.9 kPa pressure), generated by a compressor driven spray gun (Iwata Studio series 1/6 hp; Gravity spray gun RG3, Portland, USA), applied to the upper to seedlings at maintained at 18^{o}C in the dark for 24 hours before being placed into a shade-house.

Seedlings were assessed 12-20 days after inoculation using a disease rating scale: 1 = presence of yellow flecking with no pustule development; 2= presence of a hypersensitive reaction (HR) with fleck or necrosis but no pustule development; 3 = small pustules, <0.8mm diameter, with one or 2 uredinia; 4 = medium-sized pustules, 0.8 to <1.6mm diameter with about 12 uredinia; 5 = large pustules, >1.6mm diameter, with 20 or more uredinia on leaves, petioles and/or shoots (Junghans *et al*., 2003).

Myrtle rust candidate genes

Myrtle rust resistance genes are most likely to belong to a class of receptor proteins that contain a nucleotide binding domains (NB) and a leucine rich repeat (LRR). Plant NB-LRR proteins confer resistance to diverse pathogens, including fungi, oomycetes, bacteria, viruses and insects (DODDS and RATHJEN 2010). Candidate NB-LRR resistance genes for this study were identified by searching the *Eucalyptus* reference genome. In total, 556 NB-LRR genes were identified in the eucalypt genome. In addition to the NB-LRR genes, 485

candidate genes known to play a role in pathogen defense responses in other plants were identified including MAP Kinases (29), glutathione S-transferases (72), superoxide dismutases (11), peroxidases (53), chitinases (33), PR proteins (47), lignin biosynthesis (56) and a number of transcription factors belonging to the MYB (70), WRKY (45), AP2 (35), ERF (11) and bZIP (23) families.

Bulked segregant analysis

DNA from resistant trees was mixed together in equal proportions to make a "resistant bulk," and DNA from susceptible trees was mixed together to make a "susceptible bulk." We performed high throughput Illumina sequencing of these bulks. SNP markers putatively associated with disease resistance were identified by comparing allele frequencies of the selected candidate genes in resistant and susceptible bulks.

Genotyping and association analysis

Primer sequences for candidate rust resistance SNPs were used to design Fluidigm SNP assays. These SNPs were genotyped across the entire population of 300 seedlings using the Fluidigm EP1 genotyping platform and association analysis was carried out to identify the SNP markers associated with resistance. Logistic regression was used to identify the SNPs associated with disease resistance. SNP:SNP interactions were analysed using the 'RandomForest' package.

RESULTS AND DISCUSSION

Preliminary marker studies in Eucalyptus globulus

In an effort to identify the markers linked to rust resistance we initially focussed on disease resistance genes on linkage group 3 (LG3) that were identified in the recently published *Eucalyptus grandis* reference genome sequence, as this linkage group contains the *Ppr* locus.

To identify markers linked to rust resistance we first used seedlings from two sub-species of *Eucalyptus globulus*, *E. globulus bicostata* and *E. globulus globulus* that had been screened for rust resistance by Louis Morin (CSIRO). Disease severity was assessed by the method of Junghans et al. (2003). Disease severity scores for *E. globulus* are shown in Figure 1. DNA

was isolated from 18 resistant and susceptible seedlings from each of the two sub-species. Forty disease resistance (NB-LRR) genes mapping to the *Ppr* locus region of LG3 were selected from the *E. grandis* reference genome for sequencing and identification of putative rust resistance markers. DNA from 18 resistant and susceptible seedlings in each species was pooled into two bulks. The forty selected candidate resistance genes were PCR amplified in the two DNA pools and sequenced with Illumina high throughput sequencing.

Figure 1. Disease symptoms in *E. globulus*. Seedling were divided into 5 classes based on the severity of the disease: 1 = immune plants, 2 = hypersensitive plants; 3 = plants with few pustules; 4 = plants with a small number of pustules and 5 = plants with many pustules.

High throughput sequencing reads were aligned against the *E. grandis* reference genome sequence to identify candidate markers. Rust resistance markers i.e., single nucleotide polymorphisms (SNPs) and insertion-deletions (INDELS) were identified by comparing allele frequencies between resistant and susceptible pools. Analysis of gene sequence data revealed a candidate SNP marker and a 3bp INDEL (deletion) in the Eucgr C02718 gene

(disease resistance protein: TIR-NBS-LRR class) that appeared to be linked to rust resistance. Sequence analysis showed that the deletion was present exclusively in the susceptible pools while the insertion only occurred in resistant pools in both the sub-species. However when we genotyped the INDEL in the individual resistant and susceptible seedlings, we observed the deletion in both resistant as well as susceptible seedlings. The most likely explanation for this result is that the PCR amplification step had introduced errors in the DNA sequence data.

In addition to the small number of *E. globulus* seedlings screened above, we scored rust resistance in a population of 300 *E. globulus* seedlings derived from 30 native provenance families. The rust score data for 6 seedlings per family was used to estimate the heritability of rust resistance. The heritability of rust resistance was estimated to be 0.73 ± 0.23, which is high in comparison to other traits such as wood properties and growth. In *E. grandis* the heritability estimates of rust resistance ranged from 0.30 to 0.86 (MIRANDA *et al.* 2013)

Marker discovery in E. grandis

We used the approach described in the Research Methods section to discover candidate SNPs controlling rust resistance in *E. grandis*. We inoculated 300 seedlings of an *E. grandis* population with myrtle rust in collaboration with Geoff Pegg (QLD DEEDI). The seedlings were derived from 21 provenances distributed along the coast of northern NSW through to northern Queensland. Disease severity scores for *E. grandis* are shown in Figure 2.

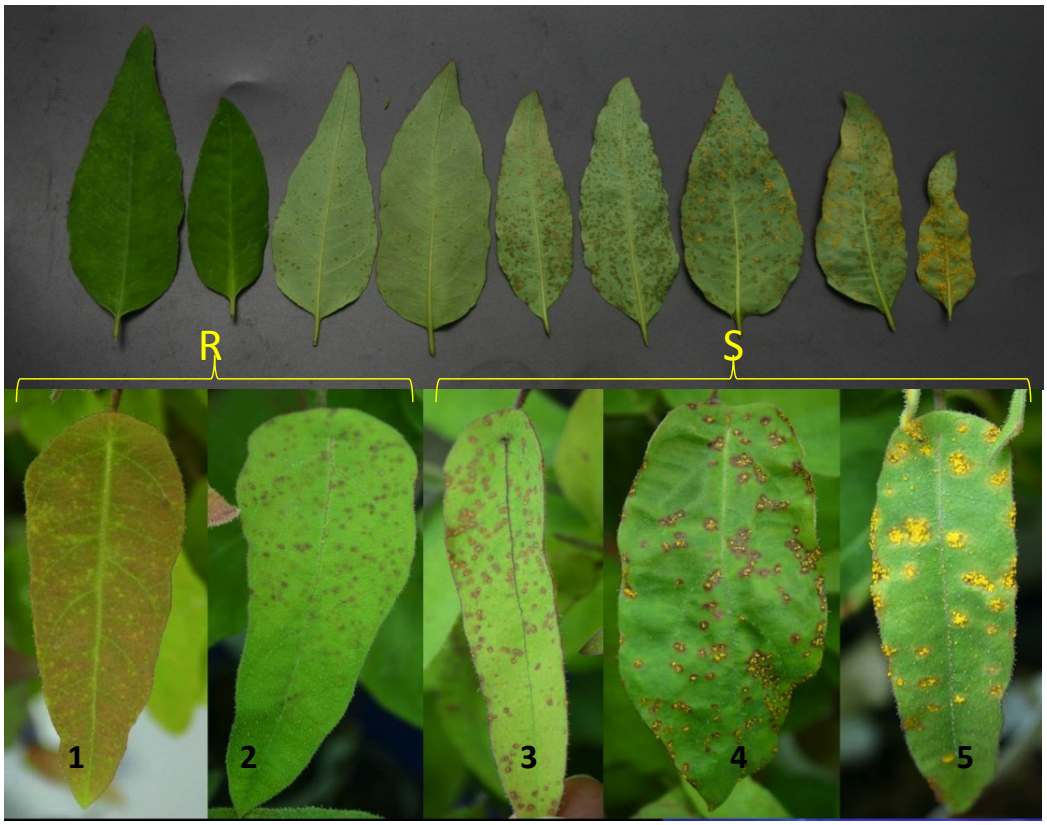

Figure 2. Disease symptoms in *E. grandis*. See Figure 1 for a description of the disease severity scores.

Analysis of the disease severity in 300 seedlings indicated that the resistant plants are distributed among all the provenances and there is no spatial pattern in disease severity (Figure 3). Approximately 48% of *E. grandis* seedlings were resistant (disease severity score 1 or 2) to myrtle rust. In comparison, 59% of 300 *E. globulus* seedlings derived from 30 Flinders Island families were resistant to the rust fungus and 58% of 300 *E. dunnii* seedlings derived from 20 provenance bulks were resistant to the rust fungus (unpublished data).

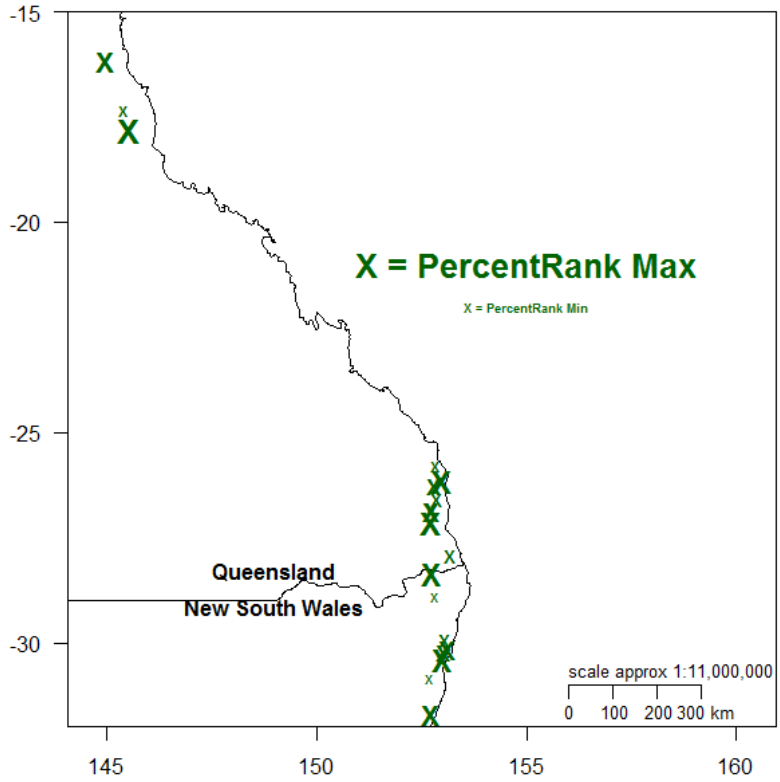

Figure 3. Myrtle rust disease severity scores for 21 provenances of *E. grandis.* The position of the "X" indicates the location of the provenance and the size of the "X" indicates the severity of the disease score.

From the disease severity screen of 300 *E. grandis* seedlings we identified 20 highly resistant (immune) and 20 highly susceptible seedlings. DNA was isolated from these seedlings and equal amounts of DNA were pooled from resistant and susceptible seedlings to create a resistant and a susceptible DNA bulk. The two bulked DNA samples were subjected to whole genome Illumina high throughput sequencing. High throughput sequencing reads were then aligned against the *E. grandis* reference sequence to determine allele frequencies in the rust resistance candidate genes.

Candidate rust resistance markers were identified by comparing SNP allele frequencies from 1041 disease related genes in the resistant and susceptible bulks. A total of 192 SNPs with the highest differences were selected for genotyping and association analysis. Approximately half of the SNPs lay in NB-LRR genes and half were in pathogen defense genes. Interestingly, the deletion identified earlier in *E. globulus* was present in both

resistant and susceptible pools in *E. grandis*. This suggests that the results from whole genome sequencing may be less biased compared to sequencing results from PCR amplification of candidate genes.

192 large effect SNPs were genotyped in *E. grandis* seedlings screened for disease severity to identify the markers associated with rust resistance. 172 SNPs were genotyped successfully across 300 seedlings and association analysis carried out. Seventeen SNPs were significantly associated with disease resistance. We then binned the p-values into the 10 possible classes (Figure 4). We observed a bias away from the expectation of 17 SNPs per bin in the distribution of the p-values. The highest number of SNPs was observed in the lowest p-value bin, and a significant proportion of these SNPs had very low p-values (<0.01). This strongly suggests that several of the SNPs in the lowest p-value class are genuinely associated with rust resistance.

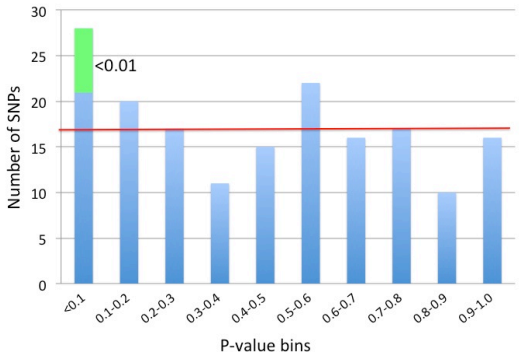

Figure 4. Binned p-values of 172 candidate SNPs. The expected number of SNPs in each bin (17) is shown by the red line. The number of highly significant SNPs (6) is illustrated by the green shading.

Table 1 lists 17 SNPs that are significantly associated with rust resistance ($p<0.05$). About 65% of SNPs lie in NB-LRR genes, which is similar to the proportion of NB-LRR genes among the candidate genes (53%). Interestingly, three SNPs (18%), including the most strongly associated SNP, lie in bZIP transcription factors, which among other roles are known to function in biotic stress responses. bZIP genes represented only 2% of the candidate genes we investigated. Other interesting SNPs included three SNPs that occur in NB-LRR genes in relatively close proximity (within about 20 million base pairs) to the *Ppr1* locus (Figure 5)

previously identified by Mamani et al (2010). The most striking result from the association analysis was the very low proportion of genetic variation (R^2) explained by each SNP. The SNP of largest effect lies in the NB-LRR 3 gene and is the closest SNP to the *Prp1 locus* on chromosome 3. This SNP only explains about 3.4% of variation. This is compelling evidence that resistance to myrtle rust is quantitative, and likely to be controlled by variation in many genes.

Table 1. Candidate SNPs significantly associated with rusts resistance in *E. grandis*. Highlighted genes include three bZIP transcription factors and three NB-LRR genes that occur in close proximity to the Ppr1 locus.

Chromosome	Gene location	P-value	R^2
1	**bZIP 1**	0.01	0.029
8	NB-LRR	0.01	0.028
5	Alcohol dehydrogense	0.01	0.027
5	NB-LRR	0.01	0.027
9	NB-LRR	0.01	0.024
3	**NB-LRR 1**	0.01	0.022
3	**NB-LRR 3**	0.02	0.034
1	**bZIP 2**	0.02	0.021
11	Glutathione peroxidase	0.03	0.016
8	NB-LRR	0.03	0.018
10	MYB 1	0.04	0.016
9	**bZIP 3**	0.04	0.016
3	NB-LRR	0.04	0.020
5	NB-LRR	0.04	0.016
3	**NB-LRR 2**	0.04	0.016
5	NB-LRR	0.05	0.015
3	NB-LRR	0.05	0.016

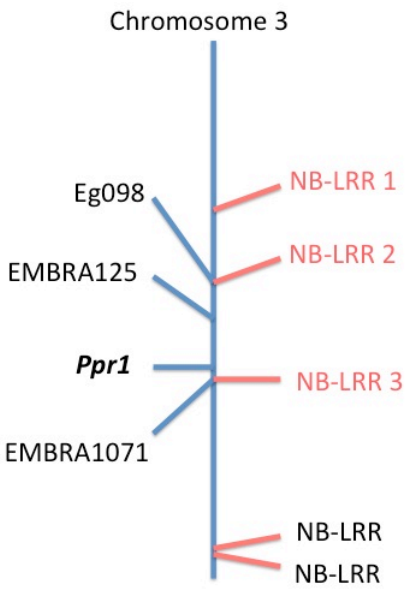

Figure 5. Position of SNPs significantly associated with rust resistance in *E. grandis* that occur on chromosome 3. The three NB-LRR in red text were also found to be significantly associated with rust resistance using SNP:SNP analysis (see Table 2).

Given that the functional basis of plant traits can involve gene:gene interactions and the earlier detection of epistatic effects in guava rust resistance in *E. grandis* (ALVES *et al.* 2012), we tested for gene:gene (SNP:SNP) interactions among our SNPs using the random forests (RF) algorithm. While association analysis with logistic regression reveals SNPs with major effects, RF analysis reveals both major effect SNPs as well as inter acting effects among SNPs. Table 2 lists the 13 most important SNPs identified from RF analysis. Interestingly, the list included the three bZIP SNPs and the three NB-LRR SNPs that occur in genes in relatively close proximity to the *Ppr1* locus, that were identified by single SNP association analysis (see Table 1). This is further evidence that these 6 SNPs genuinely play a role in rust resistance. We used the 13 important SNPs to develop a disease prediction model. Using this model we were able to predict 70% of resistant and susceptible trees correctly, an improvement of 40%.

We explored the degree to which SNPs associated with resistance were conserved in other eucalypt species. In other research we sequenced the entire genome of several *E. dunnii* individuals. Since *E. dunnii* is distantly related to *E. grandis,* the occurrence of the same SNP

in *E. dunnii* would suggest that it is functionally important. We observed that about 70% of *E. grandis* SNPs in NB-LRR genes were present in *E. dunnii,* including the three NB-LRR genes in the vicinity of the *Ppr1* locus. However, the three SNPs in bZIP transcription factors were not conserved in *E. dunnii.* This is consistent with the highly specific role NB-LRR genes play in pathogen recognition.

Table 2. SNP:SNP interactions among candidate SNPs for rust resistance in *E. grandis.* Highlighted genes include the three bZIPs and three NB-LRR genes that were identified in single SNP association analysis.

Chromosome	Gene location	Importance
8	Glutathione S-transferase	6.8
9	**bZIP 3**	6.3
3	**NB-LRR 3**	5.8
2	MYB 2	4.6
7	NB-LRR	4.3
1	**bZIP 1**	3.9
3	**NB-LRR 1**	3.6
8	NB-LRR	3.6
3	**NB-LRR 2**	3.5
1	**bZIP 2**	3.3
11	NB-LRR	3.3
1	WYRKY	3.1
6	NB-LRR	3.1

CONCLUSIONS

At the commencement of the project we hypothesised that resistance to myrtle rust is oligogenic. In the light of our findings we consider this hypothesis to be incorrect. Our results strongly suggest that resistance to myrtle rust is controlled by variation in many genes. Using single SNP and SNP:SNP interaction analyses we identified in the vicinity of 10 to 15 SNPs that are likely to play a role in rust resistance. Individually, SNPs explained small proportions of genetic variation in resistance. Given the modest amount of variation

explained by single SNPs, we expect that many dozens of genes are likely to contribute to myrtle rust resistance.

There are important implications arising from these results. The good news is that eucalypts, and probably many other Myrtaceous species that have some resistance, are unlikely to suddenly become a lot more susceptible. This is because in order to become significantly more virulent the rust would need to overcome several barriers (genes). It's more likely that the rust will gradually evolve different forms (one gene at a time) that attack a few more genotypes. Another important consequence of the quantitative nature of rust resistance is that breeding for resistance will be slower. Phenotypic screening will take quite a few generations, because selection has to take place on many genes. The most efficient way to build resistance rapidly would be to use marker-assisted selection with many more markers controlling rust resistance. With the markers we have identified we are confidant that we could improve the level of resistance in *E. grandis* from 48% to over 70%. Increasing resistance to 100% would require the identification of many more markers. We consider that using refinements to the methods we have used here that it should be possible to identify most of the molecular variants controlling rust resistance in eucalypts and other Myrtaceous species.

REFERENCES

ALFENAS, A. C., E. A. V. ZAUZA and T. F. ASSIS, 2003 First record of *Puccinia psidii* on *Eucalyptus globulus* and *E. viminalis* in Brazil. . Australasian Plant Pathology **32:** 325-326.

ALVES, A., C. ROSADO, D. FARIA, L. D. GUIMARÃES, D. LAU *et al.*, 2012 Genetic mapping provides evidence for the role of additive and non-additive QTLs in the response of inter-specific hybrids of Eucalyptus to Puccinia psidii rust infection. Euphytica **183:** 27-38.

CARNEGIE, A., J. WALKER, J. LIDBETTER, L. TESORIERO, M. HORWOOD *et al.*, 2010 *Uredo rangelii*, a taxon in the guava rust complex, newly recorded on Myrtaceae in Australia. Australasian Plant Pathology **39**.

DODDS, P. N., and J. P. RATHJEN, 2010 Plant immunity: towards an integrated view of plant-pathogen interactions. Nature Reviews Genetics **11:** 539-548.

GLEN, M., A. C. ALFENAS, E. A. V. ZAUZA, M. J. WINGFIELD and C. MOHAMMED, 2007 *Puccinia psidii*: a threat to the Australian environment and economy - a review. Australasian Plant Pathology **36:** 1-16.

JUNGHANS, D. T., A. C. ALFENAS, S. H. BROMMONSCHENKEL, S. ODA, E. J. MELLO *et al.*, 2003 Resistance to rust (Puccinia psidii Winter) in Eucalyptus: mode of inheritance and mapping of a major gene with RAPD markers. Theoretical and Applied Genetics **108:** 175-180.

MAMANI, E. C., N. BUENO, D. FARIA, L. S. GUIMARÃES, D. LAU *et al.*, 2010 Positioning of the major locus for Puccinia psidii rust resistance (Ppr1) on the Eucalyptus reference map and its validation across unrelated pedigrees. Tree Genetics & Genomes **6:** 953-962.

MIRANDA, A., M. MORAES, E. TAMBARUSSI, E. FURTADO, E. MORI *et al.*, 2013 Heritability for resistance to Puccinia psidii Winter rust in Eucalyptus grandis Hill ex Maiden in Southwestern Brazil. Tree Genetics & Genomes **9:** 321-329.

TELECHEA, N., M. ROLFO, T. A. COUTINHO and M. J. WINGFIELD, 2003 Puccinia psidii on Eucalyptus globulus in Uruguay. Plant Pathology **52:** 427-427.

THAMARUS, K. A., K. GROOM, J. MURRELL, M. BYRNE and G. F. MORAN, 2002 A genetic linkage map for Eucalyptus globulas with candidate loci for wood, fibre, and floral traits. Theoretical and Applied Genetics **104:** 379-387.

TOMMERUP, I. C., A. C. ALFENAS and K. M. OLD, 2003 Guava rust in Brazil - a threat to Eucalyptus and other Myrtaceae. New Zealand Journal of Forestry Science **33:** 420-428.

APPENDIX A

Milestones Calendar

Project Milestones	Activities to meet objectives	Due date
Milestone 1	- Agreement signed	07/06/2012
Milestone 2	- Seedlings of selected species assessed for Myrtle Rust resistance as described in the methodology in Attachment B.	31/08/2012
Milestone 3	- Candidate resistance genes identified as described in the methodology in Attachment B. - Submission of a Progress Report on activities related to Milestones 2 and 3 to PHA.	30/12/2012
Milestone 4	- Bulked samples sequenced, analysed and resistance gene marker identified, where possible, as described in the methodology in Attachment B. - Submission of a final report on the research to PHA.	30/06/2013