International Journal of Biology; Vol. 7, No. 3; 2015
ISSN 1916-9671 E-ISSN 1916-968X
Published by Canadian Center of Science and Education

Population Diversity of *Leptosphaeria maculans* in Australia

Dhwani A. Patel[1], Manuel Zander[1], Angela P. Van de Wouw[2], Annaliese S. Mason[3], David Edwards[4] & Jacqueline Batley[4]

[1] School of Agriculture and Food Sciences and Centre for Integrative Legume Research, University of Queensland, Brisbane, Australia

[2] School of BioSciences, University of Melbourne, Parkville, Australia

[3] Department of Plant Breeding, Land Use and Nutrition, Justus Liebig University, Giessen, Germany

[4] School of Plant Biology, University of Western Australia, Perth, Australia

Correspondance: Jacqueline Batley, School of Plant Biology, University of Western Australia, Crawley, WA 6009, Australia. Tel: 61-8-6488-5929. Email: jacqueline.batley@uwa.edu.au

Received: April 21, 2015 Accepted: April 27, 2015 Online Published: May 3, 2015

doi:10.5539/ijb.v7n3p18 URL: http://dx.doi.org/10.5539/ijb.v7n3p18

Abstract

The fungal pathogen *Leptosphaeria maculans*, causal agent of blackleg disease, is a primary cause of canola (*Brassica napus*) crop loss in Australia. Expanding our knowledge of the occurrence of this pathogen in Australia will provide valuable insights into developing methods of resistance against it. In this study, we examine the population diversity of *L. maculans* in Australia using single nucleotide polymorphisms (SNPs). An Illumina GoldenGate 384 SNP assay was developed and used to genotype 59 blackleg isolates collected from across Australia, in different years and from different stubble sources. Limited linkage disequilibrium, absence of significant clustering in the principal component analysis and a mixed dendrogram suggest that the Australian *L. maculans* population as a whole is panmictic. Some evidence of clonality concentrated in each state was also observed. There was a lack of correlation between SNP haplotypes, stubble cultivar and year of collection. These results suggest a high rate of sexual reproduction and evolutionary diversification in the pathogen. These features could enable the pathogen to overcome resistance and continue to cause disease in *Brassica* crops. Analysis of these fungal population isolates will help shed some light on evolution and pathogenicity questions in this important crop pathogen.

Keywords: *Leptosphaeria maculans*, blackleg, Single Nucleotide Polymorphisms, goldengate, genetic diversity

1. Introduction

The ubiquitous fungal pathogen *L. maculans* is the causal agent of phoma stem canker (blackleg) in *Brassica napus*, *B. juncea*, *B. rapa* and *B. oleracea*: canola, vegetable and mustard crops (West, Kharbanda, Barbetti, & Fitt, 2001). This ascomycete was first described in 1791 by Tode and 1849 by Desmaziéres (Gout, Eckert, Rouxel, & Balesdent, 2006). Severe epidemics of blackleg disease occurred in Australia during the 1970s, wiping out the nascent canola industry (Rouxel & Balesdent, 2005). The 1972 epidemic in Australia caused almost 90% crop losses, highlighting the susceptibility of canola to *L. maculans*. Annually, this pathogen causes an average loss of AUD $100 million to the Australian economy (Zander et al., 2013). Furthermore, *L. maculans* isolates present in Australia are classified as highly virulent, able to cause disease even in the more resistant *Brassica* species: *B. juncea*, *B. nigra* and *B. carinata* (Purwantara, Salisbury, Burton, & Howlett, 1998).

Daverdin et al. (2012) describe that rapid evolution in pathogens gives rise to new strains to combat crop defences. Maintenance of crop resistance directly depends on the field population size of the pathogen, its evolutionary potential and cropping practices that directly affect its reproductive system (Daverdin et al., 2012). *L. maculans* can survive as a saprobe in the stubble of infected plants for many years and this is usually favoured by dry hot summers and cold winters (West, Kharbanda, Barbetti, & Fitt, 2001). During this period, it produces sexual inoculum (ascospores), which can travel from several hundred metres to several hundred kilometres (Travadon et al., 2011) and infect plants followed by asexual spore (conidia) production at the site of infection (Rouxel & Balesdent, 2005).

Recent genome sequencing revealed that the *L. maculans* genome has an isochore-like structure (Rouxel et al., 2011), where the genome is divided into AT and GC-rich blocks, probably caused by the amplification of transposable elements and repeat-induced point (RIP) mutations. The RIP mechanism causes nucleotide substitutions from C to T and G to A and is a premeiotic repeat-inactivation mechanism specific to fungi that creates genetic diversification in the fungal genome (Rouxel et al., 2011). Fudal et al. (2009) reported that RIP affects the *AvrLm6* locus, causing gene inactivation and leading to virulence. Furthermore, isolates have been found to undergo continuous deletions and mutations apart from RIP that lead to further genome diversity. Current approaches to establish blackleg resistance in canola have not been successful in fully controlling this pathogen (Hayward, McLanders, Campbell, Edwards, & Batley, 2012). Therefore, understanding how this fungus has evolved, diversified and spread in Australia is important in providing information for the breeding and sowing of improved resistant varieties of *Brassica*.

Population diversity in *L. maculans* has been examined using a variety of markers. Genetic differences between eastern and western Australian isolates have previously been found using microsatellites and minisatellites, which were attributed to the presence of arid desert between the coasts (Hayden, Cozijnsen, & Howlett, 2007). Minisatellite markers used to analyse four field populations in France found high levels of gene and genotypic diversity within populations and high gene flow between populations, consistent with randomly mating populations (Gout et al., 2006) Dilmaghani et al. (2012) also used minisatellite markers to show that the *L. maculans* population in Western Canada comprises two genetically distinct populations. A further study implementing fourteen minisatellite markers also found clonal sub populations of this pathogen on *B. oleracea* in Mexico (Dilmaghani et al., 2013). However, Travadon et al. (2011) found the French *L. maculans* population to be panmictic. This study also employed minisatellite and microsatellite markers. Other investigations into blackleg population structure have also been conducted using amplified fragment length polymorphisms (AFLPs) (Purwantara, Barrins, Cozijnsen, Ades, & Howlett, 2000) and restricted fragment length polymorphisms (RFLPs) (Barrins, Ades, Salisbury, & Howlett, 2004). Single Nucleotide Polymorphisms (SNPs) have recently become a popular choice of molecular marker for population diversity studies, and offer significant benefits in terms of abundance in genomes and ease of high-throughput assessment. SNPs are single base-pair differences between two individuals at a particular locus (Appleby, Edwards, & Batley, 2009). SNPs can be classified as transitions (C to T, G to A), transversions (C to G, A to T, T to G or C to A) and insertions/deletions (indels) of a single base pair. Such molecular markers are good tools to analyse the various processes encompassing the population genetics and evolutionary processes of an organism. These include mating systems, patterns of speciation, dispersal, mutation, migration and selection etc. (Giraud, Enjalbert, Fournier, Delmotte, & Dutech, 2008; Gout et al., 2006).

The Illumina GoldenGate genotyping assay can be used to simultaneously analyse 384-3072 SNP loci across multiple individuals (Tindall et al., 2010). Previous studies using the Illumina GoldenGate assay have shown that it can be used to reliably score SNPs for genetic analysis (Durstewitz et al., 2010). Furthermore, it is cost-effective and flexible for analysing large numbers of SNPs (Appleby et al., 2009). We applied 384 previously developed *L. maculans* SNPs (Zander et al., 2013) in a GoldenGate assay to analyse 59 Australian *L. maculans* population isolates collected from different years, regions and cultivars, assessing the diversity of this pathogen across Australia.

2. Materials and Methods

2.1 Fungal Samples

A total of 59 fungal isolates were analysed using the Illumina GoldenGate assay, this comprised of 96 samples including replicates and controls (Table 1). Isolates were carefully selected to cover a wide range of parameters including region of collection, cultivar grown at collection site, year isolated and *Avr* gene complement (not shown) (Table 1, Appendix A). The isolates received were either stored in liquid form (agar piece in water) or filter form (filter discs in silica beads). The isolates were grown and genomic DNA extracted as in Zander et al. (2013). The extracted DNA was quantified using a Qubit Fluorometer (Life Technologies, 2013). The reference isolate v23.1.3, (for which the genome sequence is available) (Table 1) (Rouxel et al., 2011) was used for data analysis.

Table 1. List of *L. maculans* population isolates used in this study

As referred to in text	Isolate	Year Cultured	Species isolated from	Stubble cultivar	Stubble collection site	Country/State	Replicates	Reference
Ref	Reference (v23.1.3)	Mid-1990		-	-	Europe	N/A	(Rouxel et al., 2011)
Lm-1	04MGPS021 (21)	2004	*B. napus*	AG-Emblem	Eyre Peninsula	SA	2	
Lm-2	06MGPP041 (41)	2006	*B. napus*	Skipton	Lake Bolac	Vic	2	
Lm-3	04MGPP003	2004	*B. napus*	TI1 Pinnacle	Geelong	VIC	N/A	
Lm-4	04MGPP008	2004	*B. napus*	Unknown	Wonwondah	VIC	N/A	
Lm-5	04MGPP016	2004	*B. napus*	AG-Emblem	Bordertown	SA	N/A	
Lm-6	04MGPP022	2004	*B. napus*	Grace	Moyhall	SA	N/A	
Lm-7	04MGPP026	2004	*B. napus*	Grace	Moyhall	SA	N/A	
Lm-8	04MGPP035	2004	*B. napus*	TI1 Pinnacle	Geelong	VIC	N/A	
Lm-9	04MGPP041	2004	*B. napus*	Grace	Wonwondah - Pymers	VIC	N/A	
Lm-10	04MGPP043	2004	*B. napus*	Grace	Wonwondah - Pymers	VIC	N/A	
Lm-11	04MGPP045	2004	*B. napus*	Grace	Wonwondah - Pymers	VIC	N/A	
Lm-12	04MGPP046	2004	*B. napus*	TI1 Pinnacle	Laharum	VIC	N/A	
Lm-13	04MGPP049	2004	*B. napus*	TI1 Pinnacle	Laharum	VIC	N/A	
Lm-14	04MGPS006	2004	*B. napus*	Surpass 400	Eyre Peninsula	SA	N/A	
Lm-15	04MGPS016	2004	*B. napus*	Surpass 603CL	Bordertown -Ballinger	SA	N/A	
Lm-16	04MGPS024	2004	*B. napus*	ATR-Beacon	Bordertown - Ivan	SA	N/A	
Lm-17	05MGPP002	2005	*B. napus*	ATR-Beacon	Woseley	SA	N/A	
Lm-18	05MGPP033	2005	*B. napus*	Skipton	Yeelana	SA	N/A	
Lm-19	06MGPP019	2006	*B. napus*	ATR-Beacon	Wagga Wagga	NSW	N/A	
Lm-20	06MGPP025	2006	*B. napus*	ATR-Beacon	Wagga Wagga	NSW	N/A	
Lm-21	06MGPS032	2006	*B. napus*	Surpass 501TT	Keith	SA	N/A	
Lm-22	07VTJH002	2007	*B. juncea*	JC05002	Horsham	Vic	N/A	
Lm-23	07VTJH020	2007	*B. juncea*	JC05007	Horsham	Vic	N/A	
Lm-24	D13	2009	*B. napus*	Hyola50	Cummins	SA	N/A	(Marcroft et al., 2012)
Lm-25	09SMJ087	2009	*B. juncea*	EXCEED OasisCL	Kaniva	VIC	N/A	
Lm-26	10SMJ041	2010	*B. juncea*	EXCEED OasisCL	Tamworth	NSW	N/A	
Lm-27	LM300	2002	*B. napus*	TI1 Pinnacle	Mt Barker	WA	1	
Lm-28	LM580	2003	*B. napus*	ATR-Beacon	Wonwondah	Vic	N/A	
Lm-29	LM592	2003	*B. napus*	TI1 Pinnacle	Mt Barker	WA	N/A	
Lm-30	LM659	2003	*B. napus*	Hyden	Wongan Hills	WA	N/A	
Lm-31	LM661	2003	*B. napus*	Hyden	Wongan Hills	WA	N/A	
Lm-32	IBCN13	1991	B. napus	Unknown	Mt Barker	WA	1	(Balesdent et al., 2005)
Lm-33	IBCN15	1988	*B. napus*	Unknown	Streatham	Vic	2	(Purwantara et al., 2000)
Lm-34	IBCN16	1988	*B. napus*	Unknown	Mt Barker	WA	2	(Purwantara et al., 2000)
Lm-35	IBCN17	1988	*B. napus*	Unknown	Millicent	SA	1	(Balesdent et al., 2005)
Lm-36	IBCN18	1988	*B. napus*	Unknown	Penshurst	Vic	2	(Purwantara et al., 2000)
Lm-37	IBCN75	1987	*B. napus*	Unknown	Mt Barker	WA	2	(Purwantara et al., 2000)
Lm-38	IBCN76	1987	*B. napus*	Unknown	Mt Barker	WA	2	(Purwantara et al., 2000)

Lm-39	D8 (M)	2005	*B. napus*	Surpass 501TT	Mt Barker	WA	2	(Marcroft et al., 2012)
Lm-40	D9 (M)	2005	*B. napus*	ATR-Beacon	Mt Barker	WA	2	(Marcroft et al., 2012)
Lm-41	PHW1223	1987	*B. napus*	Unknown	Mt Barker	WA	2	(Purwantara et al., 2000)
Lm-42	V4	1988	*B. napus*	Unknown	Numurkah	Vic	N/A	(Van de Wouw et al., 2010)
Lm-43	35	1988	*B. napus*	Unknown	Penshurst	Vic	N/A	(Van de Wouw et al., 2010)
Lm-44	80	1988	*B. napus*	Unknown	Millicent	SA	N/A	(Van de Wouw et al., 2010)
Lm-45	89	1988	*B. napus*	Unknown	Millicent	SA	N/A	(Van de Wouw et al., 2010)
Lm-46	535	2003	*B. napus*	TI1 Pinnacle	Lake Bolac	Vic	1	
Lm-47	1245	1988	*B. napus*	Unknown	Galong	NSW	N/A	(Van de Wouw et al., 2010)
Lm-48	04S012	2004	*B. napus*	Surpass603CL	Bordertown	SA	1	
Lm-49	04S005	2004	*B. napus*	Surpass400	Eyre Peninsula	SA	1	
Lm-50	04P042	2004	*B. napus*	Grace	Wonwondah	Vic	N/A	
Lm-51	05P032	2005	*B. napus*	Skipton	Yeelanna	SA	N/A	
Lm-52	06P039	2006	*B. napus*	Skipton	Lake Bolac	Vic	1	(Van de Wouw et al., 2010)
Lm-53	06S014	2006	*B. napus*	Surpass 501TT	Bordertown	SA	N/A	
Lm-54	06S012	2006	*B. napus*	ATR-Beacon	Bordertown	SA	1	
Lm-55	06S039	2006	*B. napus*	Hyola60	Lake Bolac	Vic	N/A	(Van de Wouw et al., 2010)
Lm-56	06J085	2006	*B. juncea*	Unknown	Horsham	VIC	1	
Lm-57	06J095	2006	*B. juncea*	Unknown	Horsham	Vic	N/A	
Lm-58	06J112	2006	*B. juncea*	Unknown	Horsham	Vic	N/A	
Lm-59	04MGPP029	2004	B. napus	TI1 Pinnacle	Geelong	VIC	N/A	

Note: VIC-Victoria; NSW-New South Wales; WA-Western Australia; SA-South Australia; Isolate v23.1.3 is the result of a series of in vitro crosses between European field isolates (Balesdent et al., 2001); Not all data on these isolates was available, "-" denotes an unknown variable. IBCN numbers represent the IDs of "International Blackleg Collection Network" isolates (Marcroft et al., 2012).

2.2 Illumina GoldenGate assay

A total of 384 SNPs were selected for the Illumina GoldenGate assay. The SNPs were chosen to cover a range of the 76 supercontigs on which SNPs were predicted, from the list of 21,814 SNPs described in Zander et al. (2013). A designability assessment conducted using the Illumina Assay Design Tool (ADT) scored the 384 SNPs at 0.4 or above, which is deemed a good score for the Illumina GoldenGate assay (Durstewitz et al., 2010). Sample preparation for the Illumina GoldenGate assay was performed according to the Illumina GoldenGate Genotyping Assay guide (According to manufacturer's instructions). The software "Genome Studio" (Illumina Inc., 2013) was used to manually cluster the SNPs into one of the two possible genotype clusters (A and B) for this haploid organism. SNPs that clustered confidently were selected for future data analyses and monomorphic and non-clustering SNPs (did not clearly separate into either the 'A' group or the 'B' group) were eliminated from further analyses, resulting in 214 high-quality SNPs. A sub- set of 193 SNPs was used for linkage disequilibrium (LD) analysis (SNPs monomorphic in all isolates except Lm-1 and Lm-2, the isolates which were used to identify polymorphic SNPs for the assay, were omitted).

2.3 Data Analysis

The data set of 214 SNPs was sorted according to predicted positions on the *L. maculans* supercontigs, as outlined in Supplementary Figure S1 of Rouxel et al. (2011). In order to look for potential SNP blocks relating to a parameter, the isolates were sorted individually, based on each parameter eg. State, stubble species or stubble cultivar (Table 1). Manhattan plots generated using the R package 'Gapit' (Lipka et al., 2012), were used to visualise any possible association between SNPs and these parameters.

All isolates were considered to be part of one population for the statistical analyses. The SNP positions were given 1 and 0 values (for the dendrogram and PCA analyses) or 'A/A' and 'B/B' (for LD analysis) for each genotype call

and 'NA' was assigned to missing values. A binary distance matrix was generated and used to create a phylogenetic dendrogram. The R package "pvclust" (Suzuki & Shimodaira, 2006) was used to generate a dendrogram with 1000 bootstrap iterations, binary distance and complete clustering. Population LD was calculated using the R package 'genetics' (Warnes, Gorjanc, Leisch, & Man, 2012). R^2 LD values were used to generate the heatmap using the R package 'LDHeatmap' (Shin, Blay, McNeney, & Graham, 2006) to visualise LD. PCA was performed using the R packages 'ade4' (Dray & Dufour, 2007) and 'maptools' (Lewin-Koh et al., 2012). SNPs with possible null and private alleles were checked against the reference (Rouxel et al., 2011) using the alignment tool in Geneious Pro version 5.6 (Biomatters Ltd., 2015; Kearse et al., 2012).

3. Results

3.1 Illumina GoldenGate results

The results from the GoldenGate assay supported the SNP prediction of Zander et al. (2013). The data generated was sorted for SNPs that had high confidence clusters. 2.6% of SNPs had missing values (NA) for all isolates, 29.9% were monomorphic and 11.7% were non-clustering SNPs. SNPs belonging to these three categories were eliminated from further analyses. No correlation between eliminated SNPs and SNP score or supercontig on which they were positioned could be observed. Filtering for quality polymorphic data resulted in a dataset of 214 SNPs. A subset of 193 SNPs was used for linkage disequilibrium (LD) after elimination of a further 21 SNPs. Reproducibility less than 100% was due only to missing data in one or other of the replicates.

Table 2**.** Percent reproducibility of replicates used in the assay

Isolate name	Replicates	Reproducibility (%)
Lm-1	2	99.22
Lm-2	2	98.96
Lm-27	1	98.96
Lm-32	1	99.74
Lm-33	3	97.92
Lm-34	3	92.52
Lm-35	1	99.22
Lm-36	3	97.20
Lm-37	3	96.26
Lm-38	3	96.26
Lm-39	3	93.93
Lm-40	3	92.06
Lm-41	3	94.39
Lm-46	1	95.57
Lm-48	1	98.44
Lm-49	1	98.96
Lm-52	1	97.40
Lm-54	1	98.96
Lm-56	1	99.22

3.2 General Marker and Population Statistics

The SNP data was mined for possible private alleles (Table 3). Private alleles are unique alleles in isolates that denote genetic distinctiveness. Of the 21 private allele SNPs, eight occurred in the isolate Lm-1 and seven occurred in the isolate Lm-2, both of which were used for SNP discovery. Of the other four alleles occurring at low frequency, four were in Lm-2 and Lm-55 only, and two were present in only Lm-2 and Lm-24 and Lm-2 and Lm-30. SNPs private in Lm-2 and Lm-55 were the only ones that consistently occurred in intergenic regions of the genome. The SNP polymorphic information content (PIC) scores (See Appendix B) ranged from 0.3-0.5, indicative of relatively high polymorphism. General statistics can be found in Appendix B and association analysis of SNPs to the state the isolates were collected from can be found in Figures C1-C4 of Appendix C.

Table 3. Possible private alleles identified and their location in the *L. maculans* genome

SNP name	SuperContig	Location (bp)	Polymorphic in	Location in genome
SNP 55	SuperContig_2	1120505	Lm-1	Intergenic
SNP 81	SuperContig_13	1254623	Lm-1	Intergenic
SNP 85	SuperContig_13	1424634	Lm-1	Intergenic
SNP 115	SuperContig_8	774901	Lm-2 and Lm-55	Intergenic
SNP 125	SuperContig_10	992433	Lm-2 and Lm-55	Intergenic
SNP 131	SuperContig_6	638539	Lm-2 and Lm-30	Intergenic
SNP 135	SuperContig_11	578821	Lm-2	Intergenic
SNP 137	SuperContig_11	970997	Lm-2 and Lm-55	Intergenic
SNP 140	SuperContig_11	1524610	Lm-1	End of SC
SNP 141	SuperContig_3	95443	Lm-2	In gene similar to peroxisomal membrane protein CBX93615.1
SNP 151	SuperContig_4	346457	Lm-2 and Lm-55	Intergenic
SNP 153	SuperContig_4	873687	Lm-1	Downstream of gene product similar to lipolytic protein G-D-S-L family CBX93216.1
SNP 160	SuperContig_4	1358095	Lm-2	Intergenic
SNP 161	SuperContig_4	1362010	Lm-2	Intergenic
SNP 181	SuperContig_9	796976	Lm-1	Intergenic
SNP 188	SuperContig_14	264857	Lm-2	Hypothetical protein CDS CBX98019.1
SNP 194	SuperContig_14	1268037	Lm-1	Hypothetical protein CDS CBX98389.1
SNP 198	SuperContig_16	447349	Lm-1	Exon of gene whose product is similar to epoxide hydrolase CBX97267.1
SNP 203	SuperContig_17	350439	Lm-2	Hypothetical protein CDS CBX96839.1
SNP 204	SuperContig_17	523022	Lm-2 and Lm-24	Hypothetical protein CDS CBX96888.1
SNP 205	SuperContig_17	757759	Lm-2	Intergenic

3.3 Population Analysis

No significant correlation was observed in the data between the SNPs and the isolate collection site, the stubble cultivar, resistance complement of the cultivar or the year the isolates were collected (Table 1). A low level of association was observed between SNP73 on SC 12 and *B. napus* (as the stubble species) compared to *B. juncea* cultivars as the stubble species (Figure 1). No association to state, stubble cultivar, year collected or place collected could be found (Appendix C Figures C1-C4; other data not shown).

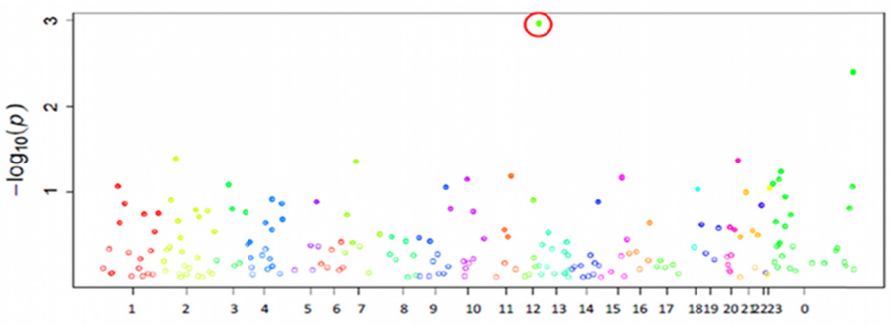


Figure 1. Manhattan plot of all isolates showing association between SNPs and *B. napus* species

Note: SNP 73 circled in red; x-axis supercontigs; y-xis –log10 values of association between stubble species and SNPs.

3.4 Phylogenetic Tree

In order to visually analyse the relationships between the isolates, a dendrogram was generated. The isolates used for SNP prediction, Lm-1 and Lm-2 are on separate clades as can be discerned from the resulting phylogenetic tree (Figure 2). Based on this analysis, a number of isolates appeared to be genetically identical. DNA replicates used in the assay were noted to be the same (Table 2).

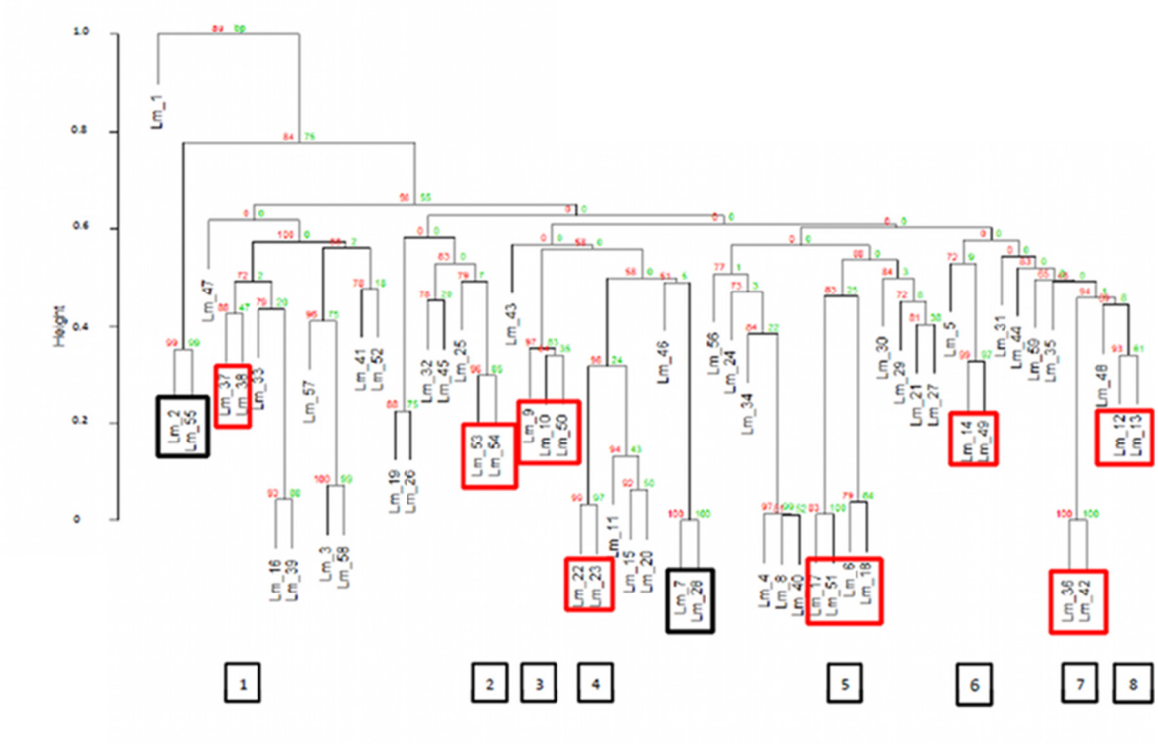


Figure 2. Phylogenetic tree based on all isolates

Note: Bootstrap values: red-Approximately Unbiased (AU) p-values calculated by multiscale bootstrap resampling; green-Bootstrap Probability (BP) p-values calculated by normal bootstrap resampling; AU values >95 % strongly supported by data (Suzuki & Shimodaira, 2006); Replicates not shown; Left to right red boxes 1-8; Black boxes indicate similar isolates.

From the dendrogram it can be seen that the isolate Lm-55 was most similar to Lm-2 (Figure 2). Both isolates were collected in the same year (2006) and the same region of the state (Victoria) but from different stubble cultivars of *B. napus*. Isolates Lm-7 and Lm-28 were also seen to be similar. However, Lm-7 was collected in 2004 and Lm-28 in 2003 from different stubble cultivars (Grace and ATR-Beacon respectively) and from different states (Wonwondah in Victoria and Moyhall in South Australia respectively, both close to a common border about 120 km apart). Boxes 1-8 denote isolates that group locally based on year collected, site of collection, state, stubble species and/or stubble cultivar (Figure 2), the details for which are seen in Table 1. The majority of clusters were from Victoria and South Australia. The most common groupings were based on state and stubble cultivar.

Some differences between certain isolates grouping together were also noticeable. Lm-16 and Lm-39 were isolated in different years (2004, 2005), from different stubble cultivars (ATR-Beacon, Surpass501TT) and in different places (Bordertown, SA and Mt. Barker, WA). The same differences could be seen for Lm-41 and Lm-52, Lm-32 and Lm-45, Lm15 and Lm-20, Lm-4, Lm-8 and Lm-40, Lm-6 and Lm-18 and Lm-21 and Lm-27. Therefore, small groupings based on the parameters listed in Table 1 were seen throughout the tree along with larger differences. No other patterns could be elicited from the positioning of sub-clades in this tree.

3.5 Principal Component Analysis (PCA)

The Principal Component Analysis (PCA) primarily showed a random distribution of isolates along the two principal component axes (Figure 3). Isolates Lm-1 and Lm-2 were completely different to each other, as expected

based on their use for identification of polymorphic SNPs for designing the assay. Genetically identical isolates were plotted at the same point or grouped together in the same area, validating the results of the dendrogram (Figure 2). Isolates Lm-51 and Lm-17 were both collected in 2005 from South Australia. Isolates Lm-9 and Lm-10 were collected in 2004 isolated from "Grace" in Wonwondah. No other conclusive correlations could be elucidated.

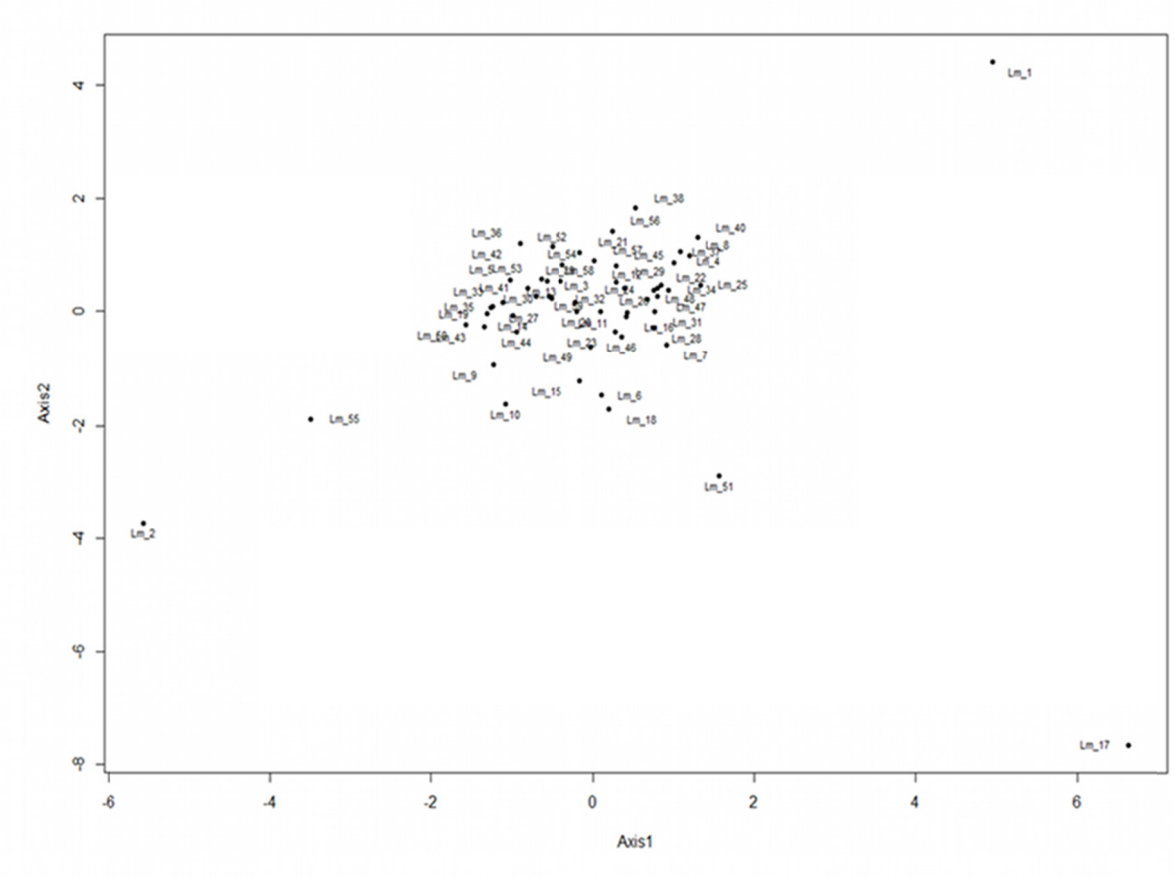


Figure 3. PCA displaying correlation between Australian *L. maculans* population isolates across 214 SNP loci isolates

Note: Axis1-Principal Component 1, Axis2-Principal Component 2; SNP prediction based on Lm-1 and Lm-2; replicates not shown.

3.6 Linkage Disequilibrium

The sub-data set of 193 SNPs, after elimination of 21 SNPs considered private alleles, was analysed to measure pairwise linkage disequilibrium (LD) between SNPs. The heatmap in Figure 4 displays results of the LD calculation. Based on the R^2 values displayed in the heatmap, little significant LD was observed in the *L. maculans* population. R^2 =1 indicates no recombination, and thus high LD, and R^2 =0 indicates considerable recombination thus no LD. Only 0.83% of p-values associated with pairwise comparisons were significant. Furthermore, the heatmap failed to show any noticeable patterns or blocks of LD.

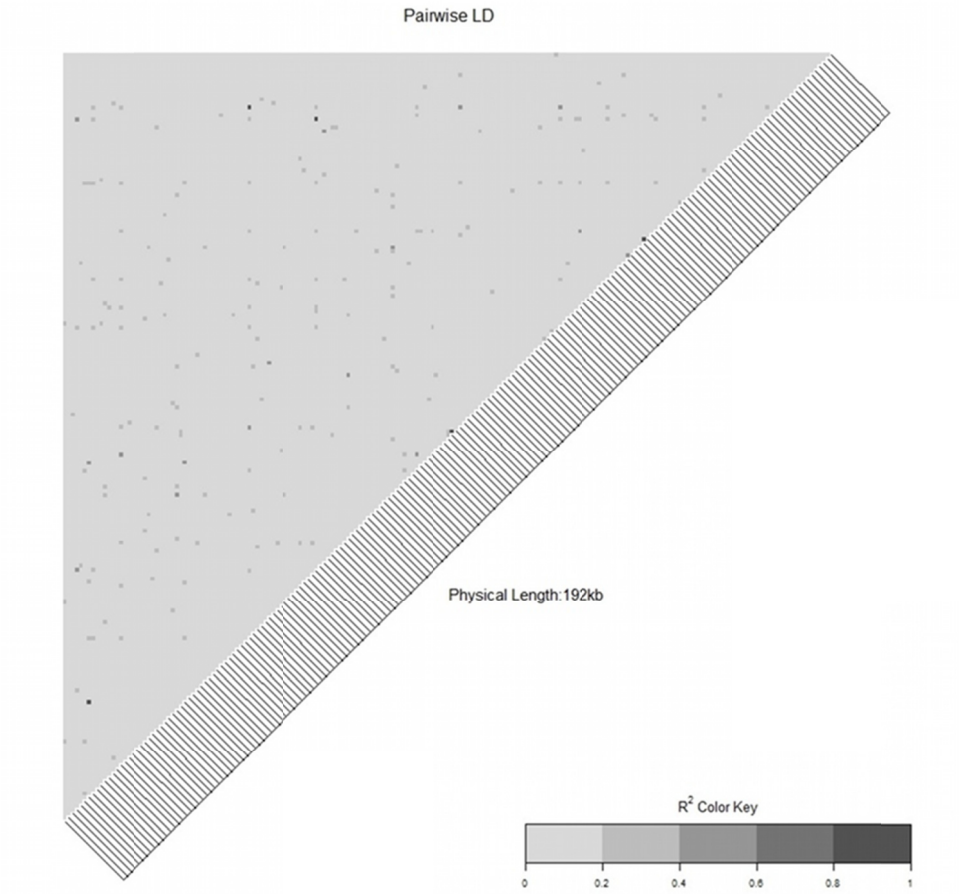


Figure 4. Heatmap displaying Linkage disequilibrium in Australian *L. maculans* population isolates across 193 SNP loci

Note: R^2=0 considerable recombination; R^2=1 no recombination.

4. Discussion

The isolates used in this study were collected from all around Australia. The absence of any obvious genetic variance specific to a certain parameter suggests that the Australian *L. maculans* isolates comprise a single population, but that some possible subpopulations localised to the state or site of collection do occur. Cultivar stubble may assist in maintaining the large population size (Daverdin et al., 2012; Travadon et al., 2011). No noticeable patterns or haplotypes were detected in the data, suggesting that this fungus evolves rapidly under selection pressure from the host. Sexual reproduction in this pathogen facilitates the production of ascospores which is its primary inoculum (Rouxel & Balesdent, 2005). Human transport may aid in transporting infected material to different regions (Travadon et al., 2011). This in turn leads to random mating between isolates from different regions, creating genetic variance at the avirulence loci and assisting the pathogen to overcome host resistance (Dilmaghani et al., 2012).

It is expected that the population of a sexual reproductively active pathogen will be panmictic. In a panmictic population, members may interact with one another at random, which creates extensive recombination and genetic diversity (Polk & Peek, 2010). Previous *L. maculans* population studies have reached the same conclusion of panmixia (Barrins et al., 2004; Travadon et al., 2011). We assumed the null hypothesis of a panmictic blackleg population across Australia while conducting population analyses, which was supported by the results of the manhattan plots, dendrogram, principal component analysis and linkage disequilibrium analysis.

Overall, the phylogenetic tree and PCA analysis suggested that *L. maculans* possesses a high evolutionary potential, indicative of populations able to overcome genetic resistance (McDonald & Linde, 2002). Large population size, high rate of mutation, high genotype flow and mixed reproduction all confer high evolutionary potential to the pathogen, putatively enabling it to overcome host genetic resistance (McDonald & Linde, 2002). The length of the tree branches indicates genetic similarity between isolates. Based on this, isolates Lm-28 and

Lm-7 appear to be genetically identical with 78% shared alleles (22% missing values). These isolates have been collected from Moyhall, South Australia in 2003 and Wonwondah, Victoria in 2004 respectively. Van de Wouw et al. (2010) also used isolates Lm-6, Lm-17, Lm-18 and Lm-51 for genotyping at the *AvrLm1* and *AvrLm6* loci. They classified Lm-18 and Lm-51 as haplotype 24, Lm-17 as haplotype 10 and Lm-6 as haplotype 4 based on their *Avr* genotype. These haplotypes displayed a completely different association in their study as compared to the dendrogram. This highlights the efficacy of using a large number of SNPs chosen from across the genome to classify the relationships between isolates.

Some local groupings based on state, year collected, stubble species, stubble cultivar and/or site of collection were also observed. Isolates in Box 1-8 (Figure 2) all clustered for certain parameters, with each cluster group sharing a single state of collection. Local groupings such as these may indicate the presence of small clonal subpopulations of this pathogen within Australia, such as were observed by Dilmaghani et al. (2013) on *B. oleracea* in Mexico. Clusters in our study were spread across the tree, indicating genetic differences between these possible subpopulations. Different conditions particular to each state such as weather, cultivars grown and stubble resistance all cumulatively affect the evolution of this pathogen. Therefore, it may be that conditions particular to each state promote asexual reproduction rather than sexual reproduction leading to less diversity within each subpopulation. Dilmaghani et al. (2013) attributed clonality such as this to moving the pathogen from its native biogeographic range, loss of a mating-type by mutation and culture conditions conducive to large-scale dispersal of conidia. This conclusion was made based on the presence of high linkage disequilibrium. Certain isolates like Lm-16 and Lm-39 also clustered together but were vastly different in the parameters associated with them. It is known that human movement transports and introduces infected seed and plant material from one area to another (Dilmaghani et al., 2012) thereby mixing and changing the population, further attributing to its panmictic nature. The bootstrapping of the phylogenetic tree also supports the theory of a randomly interacting mixed population. Bootstrapping values for the main branches were <95%, which indicates low confidence in the hierarchical cluster analysis. On the other hand, most bootstrapping values within each box were >95%, indicating that they were strongly supported by the data (Suzuki & Shimodaira, 2006). An overall analysis of the tree yielded no particular association to any other parameter. The PCA results also displayed a random positioning of isolates. Certain isolates clustered together, such as Lm-36 and Lm-42 isolates in Box 7, validating the dendrogram and also supporting panmixia. These findings could be attributed to the high evolutionary potential of *L. maculans*. Spore dispersal also plays a role in increasing gene flow and generating a random mix of isolates across the population (Travadon et al., 2011). More samples from each region and year will need to be collected and examined to investigate the possibility of clonal sub-populations of *L. maculans* within Australia.

We hypothesised that a possible cause of LD in this population could be selection of *AvrLm* genes, due to their impact on host plant infection. This would lead to loci in the selected region segregating with each other more often than expected by chance and can be visualised as blocks on the LD heatmap. However, we failed to notice any such patterns. As the rate of recombination between loci increases, there is a greater chance of linkage equilibrium in the population, decreasing LD. Populations that are constantly recombining and have a high cross-over rate will show little LD (McVean, 2008). Xu (2006) stated that only 5% of locus-pairs have significant observed association to those expected in a completely panmictic population. Our LD analysis showed 0.83% of p-values associated with pairwise comparisons, to be significant. SNP73, which was seen to be significantly associated with *B. napus* cultivars and was located near *Avr4-7*, did not display significant LD. The SNP and the gene on SC12 of the *L. maculans* genome are 253.6 kb apart and the GC content of the traversing region is 45.2%. Parlange et al. (2009) reported two PCR markers on the border of the *AvrLm4-7* locus; the GC content between those markers was 35.2%. It has been concluded by Rouxel et al. (2011) that recombination in the *L. maculans* genome occurs more frequently within GC-rich regions than between GC-rich regions. Therefore we believe that recombination events in the GC-rich region between the SNP and the gene may have impacted the association between them in *B. napus* cultivars. The number of samples isolated from *B. napus* cultivars may also be too small to clearly display this association on the heatmap in the form of LD. Furthermore, the majority of Australian cultivars that have been genotyped contain *Rlm4* and therefore there has been strong selection pressure at the *AvrLm4-7* locus for a number of years in Australia (Marcroft et al., 2012). However, detailed studies comprising more isolates derived from *B. juncea* isolates will need to be conducted to confirm the validity of this association to *B. napus* cultivars. In the future, examining associations between SNPs and *Avr* genes in AT-rich regions such as these may prove fruitful in analysing the evolution of avirulence genes.

The results from this SNP genotyping assay successfully validated the work conducted by Zander et al. (2013) using the same SNP resource. The SNP prediction supported transferability of SNPs for use in the GoldenGate assay for the chosen SNPs. Stringent clustering criteria (ensuring that all SNPs visually separated into either the ‘A’

group or the 'B' group) yielded 214 SNPs, which were used for subsequent data analysis. Poor clustering could be a result of additional SNPs in the flanking regions of the predicted SNPs which can be resolved in the future by using more isolates for SNP discovery. Private alleles relating to five particular isolates (Lm-1, Lm-2, Lm-24, Lm-30 and Lm-55) were found. Lm-2 and Lm-55 were isolated from the same year and site of collection in Victoria. Alleles present only in the isolates used for SNP prediction (Lm-1 and Lm-2) are due to ascertainment bias of using these for the SNP prediction. Ascertainment bias is introduced because of the method used for SNP discovery (Albrechtsen, Nielsen, & Nielsen, 2010), which in this case, used two isolates (Lm-1 and Lm-2) to predict SNPs. Ascertainment bias can be corrected in the future by using more isolates for SNP prediction. Being closely related and hence sharing sequence similarity might explain the four alleles that were found only in Lm-2 and Lm-55. The two main clades on the phylogenetic tree separated the isolates Lm-1 and Lm-2. Replicates used in this assay were seen to be genetically identical except for missing values. This was validated by the phylogenetic tree and PCA output, which confirmed the reproducibility of this assay.

The purpose behind conducting a large-scale genotyping assay was to understand the genetic diversity in the Australian *L. maculans* population. Our cumulative analysis of these results supports our null hypothesis of a panmictic Australian *L. maculans* population with possible regional clonality. This contrasts with the conclusions of Hayden et al. (2007) who found two genetically distinct eastern and western blackleg populations in Australia using 6 microsatellite and 2 minisatellite markers in 513 isolates collected over two years. The study also found 85% difference within the 13 subpopulations identified and 10% difference between the coasts. However, our results concur with the panmictic conclusion of Travadon et al. (2011). This study analysed 29 field populations of French *L. maculans* isolates using minisatellite markers and also found low genetic differentiation within populations. Further study concentrated on sampling from each region will help provide insights into this theory.

Overall, the high rate of sexual reproduction, ability of the pathogen to survive on stubble for long periods of time and random mating between isolates likely assist in maintaining a large blackleg population in Australia. This combined with its high evolutionary potential enables it to overcome host resistance quickly and cause infection leading to wide-spread crop losses. It is therefore imperative to attempt to restrict the population size of this pathogen using existing methods such as stubble management and introgression of resistance genes in *Brassica* species. Future work involves identifying new disease-associated genes, which will help in developing novel strategies to further control this devastating pathogen.

Acknowledgments

The authors would like to acknowledge funding support from the Australian Research Council (Projects LP0882095, LP0883462, LP0989200, LP110100200, DE120100668 and DP0985953). Support from the Australian Genome Research Facility (AGRF), the Queensland Cyber Infrastructure Foundation (QCIF) and the Australian Partnership for Advanced Computing (APAC) is gratefully acknowledged.

References

Albrechtsen, A., Nielsen, F. C., & Nielsen, R. (2010). Ascertainment Biases in SNP chips affect measures of Population Divergence. *Molecular biology and Evolution*, *27*(11), 2534-2547. http://dx.doi.org/10.1093/molbev/msq148

Appleby, N., Edwards, D., & Batley, J. (2009). New Technologies for Ultra-High Throughput Genotyping in Plants. *Plant Genomics Methods and Protocols*. UK: Humana Press. http://dx.doi.org/10.1007/978-1-59745-427-8_2

Balesdent, M.H., Attard, A., Ansan-Melayah, D., Delourme, R., Renard, M., & Rouxel, T. (2001) Genetic Control and Host Range of Avirulence toward *Brassica napus* Cultivars Quinta and Jet Neuf in *Leptosphaeria maculans*. *Phytopathology, 91*(1), 70–76. http://dx.doi.org/10.1094/PHYTO.2001.91.1.70

Balesdent, M. H., Barbetti, M. J., Li, H., Sivasithamparam, K., Gout, L., & Rouxel, T. (2005). Analysis of *Leptosphaeria maculans* Race Structure in a Worldwide Collection of Isolates. *Phytopathology, 95*(9), 1061-1071. http://dx.doi.org/10.1094/PHYTO-95-1061

Barrins, J. M., Ades, P. K., Salisbury, P. A., & Howlett, B. J. (2004). Genetic diversity of Australian Isolates of *Leptosphaeria maculans*, the Fungus that causes Blackleg of Canola (*Brassica napus*). *Australasian Plant Pathology, 33*(4), 529-536. http://dx.doi.org/10.1071/AP04061

Biomatters Ltd. (2015). Geneious Pro v5.6. Retrieved from http://www.geneious.com

Daverdin, G., Rouxel, T., Gout, L., Aubertot, J. N., Fudal, I., Meyer, M., . . . Balesdent, M. H. (2012). Genome Structure and Reproductive Behaviour influence the Evolutionary Potential of a Fungal Phytopathogen. *Plos Pathogens, 8*(11), e1003020. http://dx.doi.org/10.1371/journal.ppat.1003020

Dilmaghani, A., Gladieux, P., Gout, L., Giraud, T., Brunner, P. C., Stachowiak, A., . . . Rouxel, T. (2012). Migration Patterns and changes in Population Biology associated with the Worldwide Spread of the Oilseed Rape Pathogen *Leptosphaeria maculans*. *Molecular Ecology, 21*(10), 2519-2533. http://dx.doi.org/10.1111/j.1365-294X.2012.05535.x

Dilmaghani, A., Gout, L., Moreno-Rico, O., Dias, J. S., Coudard, L., Castillo-Torres, N., . . . Rouxel, T. (2013). Clonal Populations of *Leptosphaeria maculans* Contaminating Cabbage in Mexico. *Plant Pathology, 62*(3), 520-532. http://dx.doi.org/10.1111/j.1365-3059.2012.02668.x

Dray, S., & Dufour, A. B. (2007). The ade4 package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software, 22*(4), 1-20.

Durstewitz, G., Polley, A., Plieske, J., Luerssen, H., Graner, E. M., Wieseke, R., & Ganal, M. W. (2010). SNP discovery by amplicon sequencing and multiplex SNP genotyping in the allopolyploid species *Brassica napus*. *Genome, 53*(11). http://dx.doi.org/10.1139/G10-079

Fudal, I., Ross, S., Brun, H., Besnard, A.-L., Ermel, M., Kuhn, M.-L., . . . Rouxel, T. (2009). Repeat-Induced Point Mutation (RIP) as an Alternative Mechanism of Evolution Toward Virulence in *Leptosphaeria maculans*. *Molecular Plant-Microbe Interactions, 22*(8), 932-941. http://dx.doi.org/10.1094/MPMI-22-8-0932

Giraud, T., Enjalbert, J., Fournier, E., Delmotte, F., & Dutech, C. (2008). Population Genetics of Fungal Diseases of Plants. *Parasite, 15*(3), 449-454. http://dx.doi.org/10.1051/parasite/2008153449

Gout, L., Eckert, M., Rouxel, T., & Balesdent, M. H. (2006). Genetic Variability and Distribution of Mating Type Alleles in Field Populations of *Leptosphaeria maculans* from France. *Applied and Environmental Microbiology, 72*(1), 185-191. http://dx.doi.org/10.1128/AEM.72.1.185-191.2006

Hayden, H. L., Cozijnsen, A. J., & Howlett, B. J. (2007). Microsatellite and Minisatellite analysis of *Leptosphaeria maculans* in Australia reveals Regional Genetic Differentiation. *Phytopathology, 97*(7), 879-887. http://dx.doi.org/10.1094/PHYTO-97-7-0879

Hayward, A., McLanders, J., Campbell, E., Edwards, D., & Batley, J. (2012). Genomic Advances will herald New Insights into the *Brassica*: *Leptosphaeria maculans* Pathosystem. *Plant Biology, 14*(1), 1-10. http://dx.doi.org/10.1111/j.1438-8677.2011.00481.x

Illumina Inc. (2013). GenomeStudio Software. Retrieved from http://www.illumina.com/applications/microarrays/microarray-software/genomestudio.html

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., & Drummond, A. (2012). Geneious Basic: an Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data. *Bioinformatics, 28*(12), 1647-1649. http://dx.doi.org/10.1093/bioinformatics/bts199

Lewin-Koh, N. J., Bivand, R., Pebesma, E. J., Archer, E., Baddeley, A., Bibiko, H., . . . Turner, R. (2012). maptools: Tools for Reading and Handling Spatial Objects (Version R package version 0.8-20). Retrieved from http://CRAN.R-project.org/package=maptools

Life Technologies (2013). Qubit® 2.0 Fluorometer. Retrieved from http://www.invitrogen.com/site/us/en/home/brands/Product-Brand/Qubit/qubit-fluorometer.html

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., . . . Zhang, Z. (2012). GAPIT: Genome Association and Prediction Integrated tool. *Bioinformatics, 28*(18), 2397-2399. http://dx.doi.org/10.1093/bioinformatics/bts444

Marcroft, S. J., Elliott, V. L., Cozijnsen, A. J., Salisbury, P. A., Howlett, B. J., & Van de Wouw, A. P. (2012). Identifying Resistance Genes to *Leptosphaeria maculans* in Australian *Brassica napus* Cultivars based on Reactions to Isolates with known Avirulence Genotypes. *Crop & Pasture Science, 63*(4), 338-350. http://dx.doi.org/10.1071/CP11341

McDonald, B. A., & Linde, C. (2002). Pathogen Population Genetics, Evolutionary Potential, and Durable Resistance. *Annual Review of Phytopathology, 40*, 349-379. http://dx.doi.org/10.1146/annurev.phyto.40.120501.101443

McVean, G. (2007) Linkage Disequilibrium, Recombination and Selection, in *Handbook of Statistical Genetics*, Third Edition, John Wiley & Sons, Ltd, Chichester, UK. http://dx.doi.org/10.1002/9780470061619.ch27

Parlange, F., Daverdin, G., Fudal, I., Kuhn, M. L., Balesdent, M. H., Blaise, F., . . . Rouxel, T. (2009). *Leptosphaeria maculans* Avirulence Gene *AvrLm4-7* confers a Dual Recognition Specificity by the *Rlm4* and *Rlm7* Resistance Genes of Oilseed Rape, and circumvents *Rlm4*-mediated Recognition through a Single Amino Acid Change. *Molecular Microbiology, 71*(4), 851-863. http://dx.doi.org/10.1111/j.1365-2958.2008.06547.x

Polk, D. B., & Peek, R. M., Jr. (2010). *Helicobacter pylori*: Gastric Cancer and beyond. *Nature Reviews Cancer, 10*(6), 403-414. http://dx.doi.org/10.1038/nrc2857

Purwantara, A., Barrins, J. M., Cozijnsen, A. J., Ades, P. K., & Howle, B. J. (2000). Genetic Diversity of Isolates of the *Leptosphaeria maculans* species complex from Australia, Europe and North America using Amplified Fragment Length Polymorphism analysis. *Mycological Research, 104*(7), 772-781. http://dx.doi.org/10.1017/S095375629900235X

Purwantara, A., Salisbury, P. A., Burton, W. A., & Howlett, B. J. (1998). Reaction of *Brassica juncea* (Indian mustard) lines to Australian isolates of *Leptosphaeria maculans* under Glasshouse and Field Conditions. *European Journal of Plant Pathology, 104*(9), 895-902. http://dx.doi.org/10.1023/A:1008609131695

Rouxel, T., & Balesdent, M. H. (2005). The Stem Canker (blackleg) Fungus, *Leptosphaeria maculans*, enters the Genomic Era. *Molecular Plant Pathology, 6*(3), 225-241. http://dx.doi.org/10.1111/j.1364-3703.2005.00282.x

Rouxel, T., Grandaubert, J., Hane, J. K., Hoede, C., van de Wouw, A. P., Couloux, A., . . . Howlett, B. J. (2011). Effector Diversification within Compartments of the *Leptosphaeria maculans* Genome affected by Repeat-Induced Point Mutations. *Nature Communications*, *2*. http://dx.doi.org/10.1038/ncomms1189

Shin, J. H., Blay, S., McNeney, B., & Graham, J. (2006). LDheatmap: An R Function for Graphical Display of Pairwise Linkage Disequilibria between Single Nucleotide Polymorphisms. *Journal of Statistical Software*, *16*, code snippet 03.

Suzuki, R., & Shimodaira, H. (2006). Pvclust: an R Package for Assessing the Uncertainty in Hierarchical Clustering. *Bioinformatics, 22*(12), 1540-1542. http://dx.doi.org/10.1093/bioinformatics/btl117

Tindall, E. A., Petersen, D. C., Nikolaysen, S., Miller, W., Schuster, S. C., & Hayes, V. M. (2010). Interpretation of Custom Designed Illumina Genotype Cluster Plots for Targeted Association Studies and Next-Generation Sequence Validation. *BMC research notes, 3*, 39. http://dx.doi.org/10.1186/1756-0500-3-39

Travadon, R., Sache, I., Dutech, C., Stachowiak, A., Marquer, B., & Bousset, L. (2011). Absence of Isolation by Distance Patterns at the Regional Scale in the Fungal Plant Pathogen *Leptosphaeria maculans*. *Fungal Biology, 115*(7), 649-659. http://dx.doi.org/10.1016/j.funbio.2011.03.009

Van de Wouw, A. P., Cozijnsen, A. J., Hane, J. K., Brunner, P. C., McDonald, B. A., Oliver, R. P., & Howlett, B. J. (2010). Evolution of Linked Avirulence Effectors in *Leptosphaeria maculans* is affected by Genomic Environment and Exposure to Resistance Genes in Host Plants. *Plos Pathogens, 6*(11). http://dx.doi.org/10.1371/journal.ppat.1001180

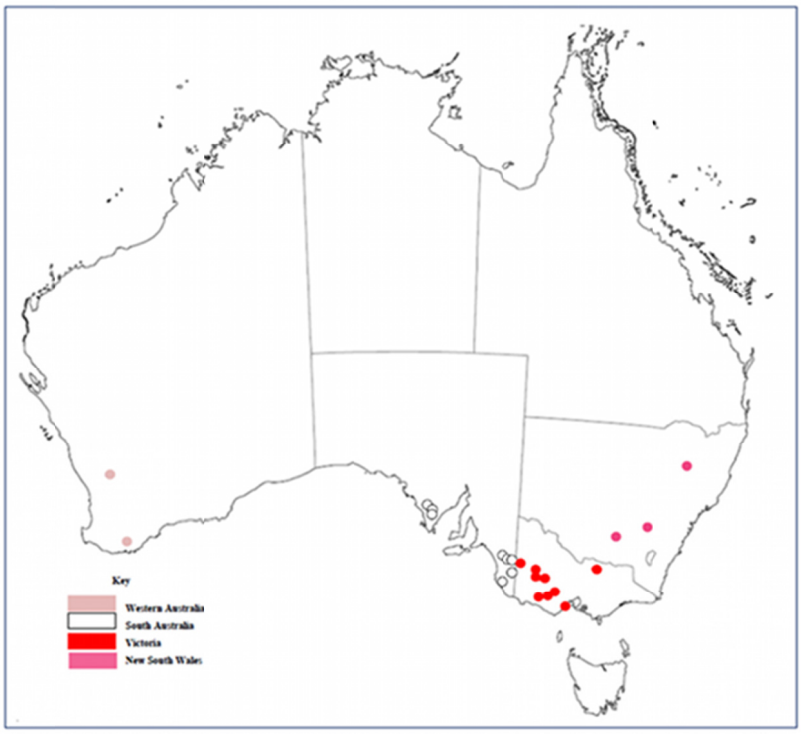
Warnes, G., Gorjanc, G., Leisch, F., & Man, M. (2012). genetics: Population Genetics (Version R package version 1.3.7.). Retrieved from http://CRAN.R-project.org/package=genetics

West, J. S., Kharbanda, P. D., Barbetti, M. J., & Fitt, B. D. L. (2001). Epidemiology and Management of *Leptosphaeria maculans* (phoma stem canker) on Oilseed Rape in Australia, Canada and Europe. *Plant Pathology, 50*(1), 10-27. http://dx.doi.org/10.1046/j.1365-3059.2001.00546.x

Xu, J. R. (2006). Fundamentals of Fungal Molecular Population Genetic Analysis. *Current Issues in Molecular Biology, 8*, 75-89.

Zander, M., Patel, D. A., Van de Wouw, A., Lai, K., Lorenc, M. T., Campbell, E., . . . Batley, J. (2013). Identifying Genetic Diversity of Avirulence Genes in *Leptosphaeria maculans* using Whole Genome Sequencing. *Functional & Integrative Genomics, 13*(3), 295-308. http://dx.doi.org/10.1007/s10142-013-0324-5

Appendices



Appendix A. Site of collection of *L. maculans* in Australia

Appendix B. General statistics of 214 SNPs

SNP name	Total A	Total B	Total	NA	% A	% B	PIC
SNP1	35	56	91	5	38.46	61.54	0.47
SNP2	27	66	93	3	29.03	70.97	0.41
SNP3	35	57	92	4	38.04	61.96	0.47
SNP4	86	10	96	0	89.58	10.42	0.19
SNP5	23	62	85	11	27.06	72.94	0.39
SNP6	13	76	89	7	14.61	85.39	0.25
SNP7	89	5	94	2	94.68	5.32	0.10
SNP8	5	91	96	0	5.21	94.79	0.10
SNP9	36	54	90	6	40.00	60.00	0.48
SNP10	27	68	95	1	28.42	71.58	0.41
SNP11	82	12	94	2	87.23	12.77	0.22
SNP12	72	18	90	6	80.00	20.00	0.32
SNP13	32	54	86	10	37.21	62.79	0.47
SNP14	36	54	90	6	40.00	60.00	0.48
SNP15	8	64	72	24	11.11	88.89	0.20
SNP16	56	31	87	9	64.37	35.63	0.46
SNP17	40	47	87	9	45.98	54.02	0.50
SNP18	42	49	91	5	46.15	53.85	0.50
SNP19	23	67	90	6	25.56	74.44	0.38
SNP20	48	41	89	7	53.93	46.07	0.50
SNP21	33	55	88	8	37.50	62.50	0.47
SNP22	24	63	87	9	27.59	72.41	0.40
SNP23	37	55	92	4	40.22	59.78	0.48
SNP24	83	12	95	1	87.37	12.63	0.22
SNP25	34	58	92	4	36.96	63.04	0.47

SNP26	19	68	87	9	21.84	78.16	0.34
SNP27	30	56	86	10	34.88	65.12	0.45
SNP28	16	75	91	5	17.58	82.42	0.29
SNP29	5	82	87	9	5.75	94.25	0.11
SNP30	35	58	93	3	37.63	62.37	0.47
SNP31	90	5	95	1	94.74	5.26	0.10
SNP32	52	40	92	4	56.52	43.48	0.49
SNP33	37	48	85	11	43.53	56.47	0.49
SNP34	73	23	96	0	76.04	23.96	0.36
SNP35	23	64	87	9	26.44	73.56	0.39
SNP36	80	12	92	4	86.96	13.04	0.23
SNP37	57	33	90	6	63.33	36.67	0.46
SNP38	35	50	85	11	41.18	58.82	0.48
SNP39	32	60	92	4	34.78	65.22	0.45
SNP40	28	67	95	1	29.47	70.53	0.42
SNP41	31	60	91	5	34.07	65.93	0.45
SNP42	29	56	85	11	34.12	65.88	0.45
SNP43	65	28	93	3	69.89	30.11	0.42
SNP44	35	58	93	3	37.63	62.37	0.47
SNP45	45	44	89	7	50.56	49.44	0.50
SNP46	38	52	90	6	42.22	57.78	0.49
SNP47	7	83	90	6	7.78	92.22	0.14
SNP48	9	84	93	3	9.68	90.32	0.17
SNP49	72	20	92	4	78.26	21.74	0.34
SNP50	74	20	94	2	78.72	21.28	0.33
SNP51	20	71	91	5	21.98	78.02	0.34
SNP52	44	47	91	5	48.35	51.65	0.50
SNP53	30	65	95	1	31.58	68.42	0.43
SNP54	77	16	93	3	82.80	17.20	0.28
SNP55	93	3	96	0	96.88	3.13	0.06
SNP56	20	71	91	5	21.98	78.02	0.34
SNP57	23	71	94	2	24.47	75.53	0.37
SNP58	30	59	89	7	33.71	66.29	0.45
SNP59	68	25	93	3	73.12	26.88	0.39
SNP60	74	20	94	2	78.72	21.28	0.33
SNP61	10	75	85	11	11.76	88.24	0.21
SNP62	74	16	90	6	82.22	17.78	0.29
SNP63	63	28	91	5	69.23	30.77	0.43
SNP64	6	89	95	1	6.32	93.68	0.12
SNP65	7	84	91	5	7.69	92.31	0.14
SNP66	46	43	89	7	51.69	48.31	0.50
SNP67	90	6	96	0	93.75	6.25	0.12
SNP68	76	17	93	3	81.72	18.28	0.30
SNP69	33	61	94	2	35.11	64.89	0.46
SNP70	75	20	95	1	78.95	21.05	0.33
SNP71	61	32	93	3	65.59	34.41	0.45
SNP72	88	7	95	1	92.63	7.37	0.14
SNP73	10	84	94	2	10.64	89.36	0.19
SNP74	29	60	89	7	32.58	67.42	0.44
SNP75	15	67	82	14	18.29	81.71	0.30
SNP76	49	43	92	4	53.26	46.74	0.50
SNP77	22	66	88	8	25.00	75.00	0.38
SNP78	46	43	89	7	51.69	48.31	0.50
SNP79	30	59	89	7	33.71	66.29	0.45
SNP80	17	78	95	1	17.89	82.11	0.29
SNP81	3	93	96	0	3.13	96.88	0.06

SNP82	55	31	86	10	63.95	36.05	0.46
SNP83	45	38	83	13	54.22	45.78	0.50
SNP84	57	39	96	0	59.38	40.63	0.48
SNP85	3	81	84	12	3.57	96.43	0.07
SNP86	6	43	49	47	12.24	87.76	0.21
SNP87	47	44	91	5	51.65	48.35	0.50
SNP88	81	14	95	1	85.26	14.74	0.25
SNP89	66	28	94	2	70.21	29.79	0.42
SNP90	35	58	93	3	37.63	62.37	0.47
SNP91	27	59	86	10	31.40	68.60	0.43
SNP92	14	64	78	18	17.95	82.05	0.29
SNP93	16	80	96	0	16.67	83.33	0.28
SNP94	84	8	92	4	91.30	8.70	0.16
SNP95	48	40	88	8	54.55	45.45	0.50
SNP96	84	11	95	1	88.42	11.58	0.20
SNP97	88	7	95	1	92.63	7.37	0.14
SNP98	36	58	94	2	38.30	61.70	0.47
SNP99	23	69	92	4	25.00	75.00	0.38
SNP100	18	67	85	11	21.18	78.82	0.33
SNP101	36	49	85	11	42.35	57.65	0.49
SNP102	36	49	85	11	42.35	57.65	0.49
SNP103	10	85	95	1	10.53	89.47	0.19
SNP104	42	50	92	4	45.65	54.35	0.50
SNP105	67	24	91	5	73.63	26.37	0.39
SNP106	39	55	94	2	41.49	58.51	0.49
SNP107	47	41	88	8	53.41	46.59	0.50
SNP108	5	91	96	0	5.21	94.79	0.10
SNP109	25	66	91	5	27.47	72.53	0.40
SNP110	24	65	89	7	26.97	73.03	0.39
SNP111	42	49	91	5	46.15	53.85	0.50
SNP112	45	48	93	3	48.39	51.61	0.50
SNP113	64	29	93	3	68.82	31.18	0.43
SNP114	22	69	91	5	24.18	75.82	0.37
SNP115	90	4	94	2	95.74	4.26	0.08
SNP116	46	42	88	8	52.27	47.73	0.50
SNP117	5	89	94	2	5.32	94.68	0.10
SNP118	54	39	93	3	58.06	41.94	0.49
SNP119	90	5	95	1	94.74	5.26	0.10
SNP120	47	32	79	17	59.49	40.51	0.48
SNP121	53	38	91	5	58.24	41.76	0.49
SNP122	28	63	91	5	30.77	69.23	0.43
SNP123	37	57	94	2	39.36	60.64	0.48
SNP124	11	84	95	1	11.58	88.42	0.20
SNP125	4	92	96	0	4.17	95.83	0.08
SNP126	16	74	90	6	17.78	82.22	0.29
SNP127	14	76	90	6	15.56	84.44	0.26
SNP128	7	53	60	36	11.67	88.33	0.21
SNP129	80	11	91	5	87.91	12.09	0.21
SNP130	92	4	96	0	95.83	4.17	0.08
SNP131	4	79	83	13	4.82	95.18	0.09
SNP132	49	39	88	8	55.68	44.32	0.49
SNP133	23	70	93	3	24.73	75.27	0.37
SNP134	66	24	90	6	73.33	26.67	0.39
SNP135	3	83	86	10	3.49	96.51	0.07
SNP136	76	16	92	4	82.61	17.39	0.29
SNP137	4	92	96	0	4.17	95.83	0.08

SNP138	91	5	96	0	94.79	5.21	0.10
SNP139	85	11	96	0	88.54	11.46	0.20
SNP140	92	3	95	1	96.84	3.16	0.06
SNP141	3	89	92	4	3.26	96.74	0.06
SNP142	24	70	94	2	25.53	74.47	0.38
SNP143	51	36	87	9	58.62	41.38	0.49
SNP144	40	47	87	9	45.98	54.02	0.50
SNP145	74	21	95	1	77.89	22.11	0.34
SNP146	73	15	88	8	82.95	17.05	0.28
SNP147	92	4	96	0	95.83	4.17	0.08
SNP148	9	86	95	1	9.47	90.53	0.17
SNP149	38	50	88	8	43.18	56.82	0.49
SNP150	65	29	94	2	69.15	30.85	0.43
SNP151	4	88	92	4	4.35	95.65	0.08
SNP152	31	61	92	4	33.70	66.30	0.45
SNP153	3	89	92	4	3.26	96.74	0.06
SNP154	29	58	87	9	33.33	66.67	0.44
SNP155	45	42	87	9	51.72	48.28	0.50
SNP156	19	71	90	6	21.11	78.89	0.33
SNP157	86	7	93	3	92.47	7.53	0.14
SNP158	66	24	90	6	73.33	26.67	0.39
SNP159	70	23	93	3	75.27	24.73	0.37
SNP160	93	3	96	0	96.88	3.13	0.06
SNP161	3	60	63	33	4.76	95.24	0.09
SNP162	74	20	94	2	78.72	21.28	0.33
SNP163	91	5	96	0	94.79	5.21	0.10
SNP164	75	17	92	4	81.52	18.48	0.30
SNP165	69	25	94	2	73.40	26.60	0.39
SNP166	56	35	91	5	61.54	38.46	0.47
SNP167	61	27	88	8	69.32	30.68	0.43
SNP168	69	24	93	3	74.19	25.81	0.38
SNP169	29	63	92	4	31.52	68.48	0.43
SNP170	30	55	85	11	35.29	64.71	0.46
SNP171	39	50	89	7	43.82	56.18	0.49
SNP172	75	17	92	4	81.52	18.48	0.30
SNP173	17	76	93	3	18.28	81.72	0.30
SNP174	41	47	88	8	46.59	53.41	0.50
SNP175	4	88	92	4	4.35	95.65	0.08
SNP176	26	65	91	5	28.57	71.43	0.41
SNP177	17	76	93	3	18.28	81.72	0.30
SNP178	9	79	88	8	10.23	89.77	0.18
SNP179	59	34	93	3	63.44	36.56	0.46
SNP180	33	59	92	4	35.87	64.13	0.46
SNP181	3	69	72	24	4.17	95.83	0.08
SNP182	25	69	94	2	26.60	73.40	0.39
SNP183	71	17	88	8	80.68	19.32	0.31
SNP184	22	69	91	5	24.18	75.82	0.37
SNP185	48	42	90	6	53.33	46.67	0.50
SNP186	81	14	95	1	85.26	14.74	0.25
SNP187	22	63	85	11	25.88	74.12	0.38
SNP188	3	92	95	1	3.16	96.84	0.06
SNP189	54	35	89	7	60.67	39.33	0.48
SNP190	13	79	92	4	14.13	85.87	0.24
SNP191	15	76	91	5	16.48	83.52	0.28
SNP192	70	21	91	5	76.92	23.08	0.36
SNP193	7	85	92	4	7.61	92.39	0.14

SNP194	3	93	96	0	3.13	96.88	0.06
SNP195	39	49	88	8	44.32	55.68	0.49
SNP196	66	25	91	5	72.53	27.47	0.40
SNP197	7	80	87	9	8.05	91.95	0.15
SNP198	93	3	96	0	96.88	3.13	0.06
SNP199	41	50	91	5	45.05	54.95	0.50
SNP200	13	83	96	0	13.54	86.46	0.23
SNP201	53	39	92	4	57.61	42.39	0.49
SNP202	40	47	87	9	45.98	54.02	0.50
SNP203	3	93	96	0	3.13	96.88	0.06
SNP204	4	82	86	10	4.65	95.35	0.09
SNP205	92	3	95	1	96.84	3.16	0.06
SNP206	73	20	93	3	78.49	21.51	0.34
SNP207	75	18	93	3	80.65	19.35	0.31
SNP208	32	57	89	7	35.96	64.04	0.46
SNP209	60	33	93	3	64.52	35.48	0.46
SNP210	76	19	95	1	80.00	20.00	0.32
SNP211	45	46	91	5	49.45	50.55	0.50
SNP212	33	57	90	6	36.67	63.33	0.46
SNP213	19	72	91	5	20.88	79.12	0.33
SNP214	25	60	85	11	29.41	70.59	0.42

Note: PIC-Polymorphism Information Content (Totals excluding NAs).

Appendix C. Analysis of association between SNPs and state of collection of isolates

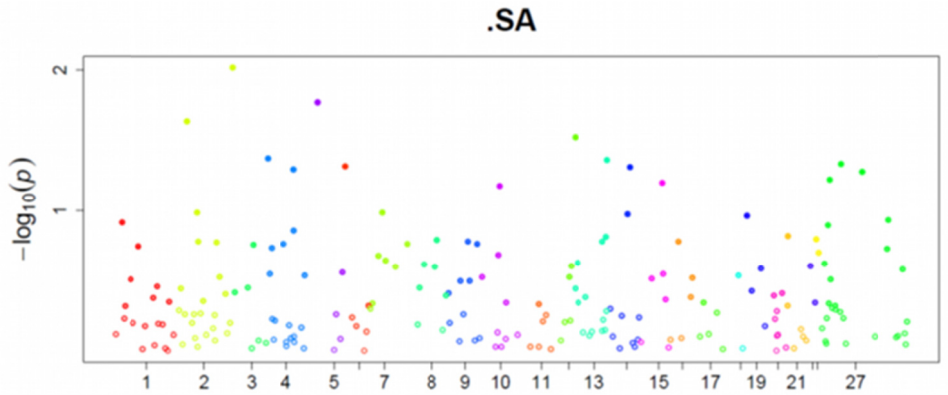


Figure C1: Manhattan plot of all isolates showing association between SNPs and South Australia

Note: x-axis: Supercontigs (27= SC0); y-xis $-\log_{10}$ values of association between stubble species and SNPs.

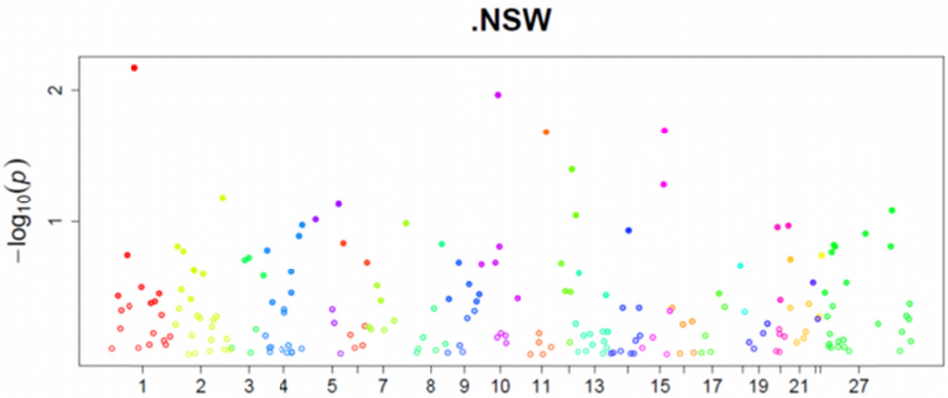


Figure C2: Manhattan plot of all isolates showing association between SNPs and New South Wales

Note: x-axis supercontigs (27= SC0); y-xis $-\log_{10}$ values of association between stubble species and SNPs.

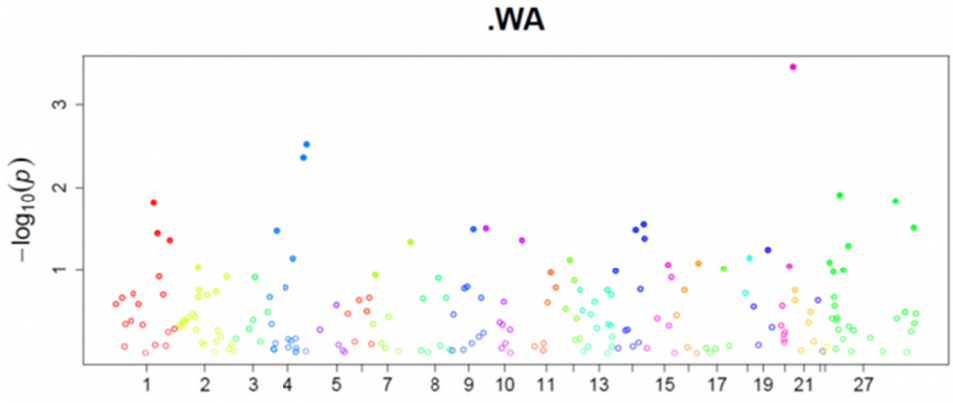


Figure C3 Manhattan plot of all isolates showing association between SNPs and Western Australia

Note: x-axis supercontigs (27= SC0); y-xis $-\log_{10}$ values of association between stubble species and SNPs.

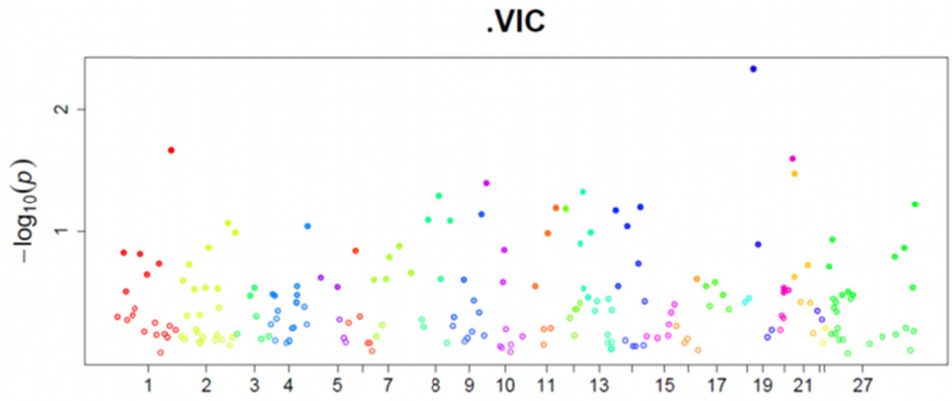


Figure C4 Manhattan plot of all isolates showing association between SNPs and Victoria

Note: x-axis supercontigs (27= SC0); y-xis $-\log_{10}$ values of association between stubble species and SNPs.