

Identification and characterization of candidate *Rlm4* blackleg resistance genes in *Brassica napus* using next-generation sequencing

Reece Tollenaere¹, Alice Hayward¹, Jessica Dalton-Morgan¹, Emma Campbell¹, Joanne R.M. Lee¹, Michal T. Lorenc², Sahana Manoli², Jiri Stiller², Rosy Raman³, Harsh Raman³, David Edwards² and Jacqueline Batley^{1,*}

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

²Australian Centre for Plant Functional Genomics, School of Agriculture and Food Sciences, University of Queensland, Brisbane, Qld, Australia

³NSW Department of Primary Industries, Wagga Wagga Agricultural Institute, Wagga Wagga, NSW, Australia

Received 10 February 2012;

revised 23 April 2012;

accepted 23 April 2012.

*Correspondence

(Tel +61 (0)7 334 69534;

fax +61 (0)7 336 51188;

email j.batley@uq.edu.au)

Summary

A thorough understanding of the relationships between plants and pathogens is essential if we are to continue to meet the agricultural needs of the world's growing population. The identification of genes underlying important quantitative trait loci is extremely challenging in complex genomes such as *Brassica napus* (canola, oilseed rape or rapeseed). However, recent advances in next-generation sequencing (NGS) enable much quicker identification of candidate genes for traits of interest. Here, we demonstrate this with the identification of candidate disease resistance genes from *B. napus* for its most devastating fungal pathogen, *Leptosphaeria maculans* (blackleg fungus). These two species are locked in an evolutionary arms race whereby a gene-for-gene interaction confers either resistance or susceptibility in the plant depending on the genotype of the plant and pathogen. Preliminary analysis of the complete genome sequence of *Brassica rapa*, the diploid progenitor of *B. napus*, identified numerous candidate genes with disease resistance characteristics, several of which were clustered around a region syntenic with a major locus (*Rlm4*) for blackleg resistance on A7 of *B. napus*. Molecular analyses of the candidate genes using *B. napus* NGS data are presented, and the difficulties associated with identifying functional gene copies within the highly duplicated *Brassica* genome are discussed.

Keywords: *Brassica napus*, *Leptosphaeria maculans*, candidate genes, next generation sequencing.

Introduction

Brassica napus is a member of the large and agronomically important Brassicaceae family, which consists of approximately 340 genera and 3350 species. The *Brassica* genus includes many important vegetable and oilseed crops, with six cultivated *Brassica* species: *B. napus* (canola/rapeseed/oilseed rape), *Brassica rapa* (Chinese cabbage/turnip rape), *Brassica oleracea* (including broccoli, cabbage and cauliflower), *Brassica juncea* (Indian mustard), *Brassica nigra* (black mustard) and *Brassica carinata* (Ethiopian mustard). *Brassica napus* is the evolutionary result of an interspecific hybridization between *B. rapa* (AA genome, $2n = 20$) and *B. oleracea* (CC genome, $2n = 18$), generating an amphidiploid genome AAC ($2n = 4x = 38$).

As of 2009, canola was the third largest source of vegetable oil and oil extraction meal worldwide, with 12.1 million metric tonnes of oil produced (Freidt and Snowdon, 2009). However, canola production is often devastated by blackleg disease, caused by the virulent fungal pathogen, *Leptosphaeria maculans* (Hayward *et al.*, 2012). Blackleg is the most serious fungal disease of oilseed Brassicas, in particular *B. napus* and *B. rapa* (Howlett *et al.*, 2001), and canola production in Australia was virtually eliminated by this fungal pathogen in the early 1970s (Howlett, 2004). Today, blackleg remains a major cause of crop loss in Australia and internationally, and

is estimated to be responsible for an average of 15% crop losses per annum worldwide, valued at US\$900M (Fitt *et al.*, 2008).

Many practices are employed in an attempt to reduce the devastating effect of *L. maculans* on rapeseed production. Improvements in agronomic practices, for example sowing a new crop at least 500 m from previous years' stubble, have mitigated crop losses for Australian growers. However, the most effective and sustainable approach to the treatment of blackleg disease is to identify resistance genes in *Brassica* species and apply these in breeding resistant varieties (Hayward *et al.*, 2012).

There are both qualitative and quantitative types of resistance to *L. maculans* (Balesdent *et al.*, 2001, 2002). Qualitative resistance is race-specific and depends on the presence of a single resistance (*R*) gene in the plant corresponding to an avirulence (*Avr*) gene in the pathogen (Ansan-Melayah *et al.*, 1998). Qualitative resistance is expressed from seedlings to adults in cotyledons and leaves (Delourme *et al.*, 2006). In contrast, quantitative resistance is generally thought to be race nonspecific, mediated by many genes and expressed at the adult plant stage, conferring only partial resistance to the pathogen (Delourme *et al.*, 2006; Rimmer, 2006).

Most proteins encoded by *R* genes in plant genomes belong to the nucleotide binding site and leucine-rich repeat (NBS-LRR) domain-containing class of proteins (Baumgarten *et al.*, 2003;

Persson *et al.*, 2009). NBS-LRR proteins have a variable N-terminus, which commonly contains a domain with similarity to the *Drosophila* Toll and mammalian Interleukin-1 receptor (TIR) or a coiled-coil (CC) sequence (Persson *et al.*, 2009). NBS-LRR proteins are located intracellularly and may interact directly with pathogen Avr proteins, known as effectors, or with host factor proteins that enable the indirect identification of pathogen effectors.

Little information is available on the genetic control of quantitative resistance to *L. maculans*. Numerous quantitative trait loci (QTL) studies have identified many regions on both the A and C genomes involved in quantitative resistance (Balesdent *et al.*, 2002; Delourme *et al.*, 2004, 2008; Ferreira *et al.*, 1995; Kaur *et al.*, 2009; Pilet *et al.*, 2001; Raman *et al.*, 2012). Nevertheless, the different QTLs cannot be easily compared because of different marker systems being used, along with differences in the blackleg isolates and inoculum concentrations present at each location. Fifteen specific resistance genes have been genetically mapped in the cultivated *Brassica* species to date. *BLMR1* on A10 has been fine-mapped to within 0.13 cM (Long *et al.*, 2011); however, no *B. napus* blackleg resistance gene has been cloned to date.

Despite the extensive efforts in genetic mapping, the use of map-based cloning to identify genes underlying important traits is very difficult in complex genomes, such as *B. napus*. There are also problems with candidate gene approaches, which employ homology searches with genes identified from closely related, well-characterized model organisms, such as *Arabidopsis*. This is because multiple homologues and paralogues of any given gene of interest can complicate their identification and role in the polyploid species. Next-generation sequencing (NGS) technologies have advanced rapidly in the past few years, providing the ability to sequence reference genomes as well as re-sequence cultivars of interest at relatively low cost (Duran *et al.*, 2010; Edwards and Batley, 2010; Imelfort and Edwards, 2009; Malory *et al.*, 2011; Varshney *et al.*, 2009). Illumina's HiSeq 2000 currently produces more than 600 billion nucleotides per run with a read length of >100 bp (<http://www.illumina.com>), and the sequence volumes and lengths for this technology are under continuous improvement. With advances in genome sequencing technologies (Edwards and Wang, 2012), single nucleotide polymorphism (SNP) discovery from NGS data is now routine (Allen *et al.*, 2011; Duran *et al.*, 2009a,b; Kharabian-Masouleh *et al.*, 2011; Subbaiyan *et al.*, 2012). Re-sequencing is used to identify genetic variation between individuals, which can provide molecular genetic markers and insights into gene function (Imelfort *et al.*, 2009). The process of whole-genome re-sequencing using short-read technologies involves aligning a set of millions of reads to a reference genome sequence (Batley and Edwards, 2009). Once this has been achieved, it is possible to determine variation in nucleotide sequence between the individuals.

Recently, a qualitative (race-specific) trait locus controlling blackleg resistance was identified in a doubled-haploid (DH) population derived from the Australian canola cultivars Skipton and Ag-Spectrum (Raman *et al.*, 2012). This major locus, designated *RlmSkipton*, was genetically mapped within BRMS075-CB10439 marker interval to *B. napus* chromosome A7, within 0.8 cM of the SSR marker *Xbrms075*, and corresponds to *Rlm4* (Resistance to *L. maculans* 4) based on recognition of an *AvrLm4*-containing *L. maculans* isolate (Raman *et al.*, 2012).

The avirulence gene *AvrLm4-7* leads to the dual recognition by two Resistance genes, *Rlm4* and *Rlm7*, and elicits the appropriate defence response (Parlange *et al.*, 2009). Therefore, loci mapping to *Rlm4* remain candidates for both *Rlm4* and *Rlm7*. In this study, the alignment of this QTL region with the newly available reference genome sequence of *B. rapa* (Wang *et al.*, 2011), using sequence-based molecular markers, permitted the physical mapping of the *Rlm4* region and the identification of candidate disease resistance genes. Subsequent mapping of Illumina paired sequence reads from the two parents Skipton and Ag-spectrum to the reference genome, and comparing polymorphisms in and around the candidate genes of interest, allowed the molecular characterization of the candidate genes and potential localization of the causative polymorphism.

This study provides a molecular characterization of two *Rlm4* gene candidates using next-generation DNA sequencing. This approach can be applied to any trait and population of interest for the rapid identification and characterization of candidate genes.

Results and discussion

Previous methods to identify *Rlm* candidate genes in our laboratory used a simple gene homology-based approach, whereby Resistance genes to *L. maculans* from the model organism *Arabidopsis* were compared with the available *Brassica* sequence databases (Lim *et al.*, 2007; Love *et al.*, 2006). However, this proved problematic in that either no gene candidates were identified, or multiple homologous gene candidates were identified that did not map to the known resistance loci on the reference genome. With the advancements and accessibility of NGS, and the recent release of the *B. rapa* genome sequence (Wang *et al.*, 2011), the approach taken in this study was to sequence the parents of a mapping population that shows segregation for the *Rlm4* locus to rapidly identify candidate genes.

Identification of candidate genes for *Rlm4*

A 2.75-Mbp region of *Brassica* chromosome A7 corresponding to a major locus for *Rlm4* was identified by genetic mapping in a *B. napus* cv. Skipton × Ag-Spectrum DH population. Flanking markers for *Rlm4* (Raman *et al.*, 2012) were physically mapped on the *B. rapa* A genome sequence (Wang *et al.*, 2011). Selection of this sequence region and comparison with known *Arabidopsis* disease resistance genes identified eighteen candidate genes, from over 2000 predicted genes, with disease resistance characteristics (here designated *BLR1* (*Brassica Leptosphaeria Resistance 1*) to *BLR18*). These candidate genes were then characterized in the Skipton and Ag-Spectrum parental lines that differ in *Rlm4*-based resistance following Illumina paired-end sequencing of these parents and alignment to the reference genome.

The average depth of Illumina sequence read coverage across the reference A genome was ~10× for both Skipton and Ag-Spectrum; though, read coverage varied from 0% to 100%, with an average of 67%, across the candidate gene regions (genic region plus 1 kb either side; Table 1, Figure 1). Sequence read coverage is determined by sequencing data volumes and also sequence conservation between the reads and the reference, and with an average of 10× read depth, reference regions with zero coverage are likely to be due to significant sequence variation or loss between the sequenced parent and the

Table 1 Coverage and single nucleotide polymorphism (SNP) analysis of 18 candidate genes underlying the *Rlm4* quantitative trait loci

Gene name	Cultivar	Short read coverage (%)	Gene region SNP number			Gene region SNP density/kb	SNP type
			1 kb 5'	Genic	1 kb 3'		
<i>BLR1</i>	Skipton	53	0	1 CDS	0	0.25	1 non-synonymous
	Ag-Spectrum	45					
<i>BLR2</i>	Skipton	100	N/A	N/A	N/A	N/A	N/A
	Ag-Spectrum	0					
<i>BLR3</i>	Skipton	83	0	8 (2 CDS, 6 intronic)	2	2.31	2 synonymous
	Ag-Spectrum	47					
<i>BLR4</i>	Skipton	50	7	0	0	2.72	N/A
	Ag-Spectrum	42					
<i>BLR5</i>	Skipton	69	5	13 (9 CDS, 4 intronic)	0	3.93	5 non-synonymous 4 synonymous
	Ag-Spectrum	83					
<i>BLR6</i>	Skipton	71	0	6 CDS	1	2.16	2 non-synonymous 4 synonymous
	Ag-Spectrum	71					
<i>BLR7</i>	Skipton	62	0	0	0	0.00	N/A
	Ag-Spectrum	77					
<i>BLR8</i>	Skipton	91	0	2 CDS	0	0.79	1 non-synonymous 1 synonymous
	Ag-Spectrum	73					
<i>BLR9</i>	Skipton	8	0	2 CDS	0	0.75	1 non-synonymous 1 synonymous
	Ag-Spectrum	62					
<i>BLR10</i>	Skipton	60	0	0	0	0.00	N/A
	Ag-Spectrum	10					
<i>BLR11</i>	Skipton	91	7	6 CDS	1	4.12	4 non-synonymous 2 synonymous
	Ag-Spectrum	82					
<i>BLR12</i>	Skipton	92	0	0	0	0.00	N/A
	Ag-Spectrum	83					
<i>BLR13</i>	Skipton	100	1	0	0	0.40	N/A
	Ag-Spectrum	64					
<i>BLR14</i>	Skipton	42	1	3 CDS	0	0.97	1 non-synonymous 2 synonymous
	Ag-Spectrum	35					
<i>BLR15</i>	Skipton	94	0	1 intronic	0	0.35	N/A
	Ag-Spectrum	82					
<i>BLR16</i>	Skipton	88	0	0	0	0.00	N/A
	Ag-Spectrum	88					
<i>BLR17</i>	Skipton	50	0	0	0	0.00	N/A
	Ag-Spectrum	60					
<i>BLR18</i>	Skipton	96	0	0	2	0.61	N/A
	Ag-Spectrum	100					
Average		66.8	1.2	2.5 (1.9 CDS)	0.4	1.2	

Genes chosen as best candidates for further analysis are highlighted.

reference. Given this, the NGS data revealed *BLR2* as a priority candidate for *Rlm4*, as none of the Ag-Spectrum reads mapped to the gene, whilst 100% coverage was observed with reads from the resistant parent Skipton (Table 1).

The number of SNPs in each candidate gene region ranged from 0–18 based solely on the short-read data, with an average density of 1.2 SNPs per kilobase across the 18 gene regions. On average, a similar number of SNPs between the parent cultivars was found within the 1 kb upstream regions as for the coding regions of each predicted gene, with fewer SNPs noted in the 1 kb downstream. Where coverage was present for both cultivars, SNP density was not directly related to coverage but rather was gene-specific, providing a good basis for narrowing the candidate gene selection (Table 1). Those genes that contained no predicted SNPs or

SNPs which were synonymous were thought to be low-priority candidates. Following SNP analysis, a second gene, *BLR11*, was prioritized along with *BLR2* as a potential *Rlm4* candidate based on the good gene coverage and high SNP density. Therefore, sequence comparison using the NGS data was effective for rapidly narrowing the list of *Rlm4* QTL candidates.

On the basis of the NGS reads that mapped to the gene regions of interest, PCR primers were designed for amplification of the candidate genes in both parents to validate the results of the Illumina read mapping and to assess complete sequence divergence between the parents. For those genes where no reads mapped to coding regions in Ag-Spectrum, such as *BLR2*, primers were designed to amplify this region based on the nearest mapping genomic read pairs.

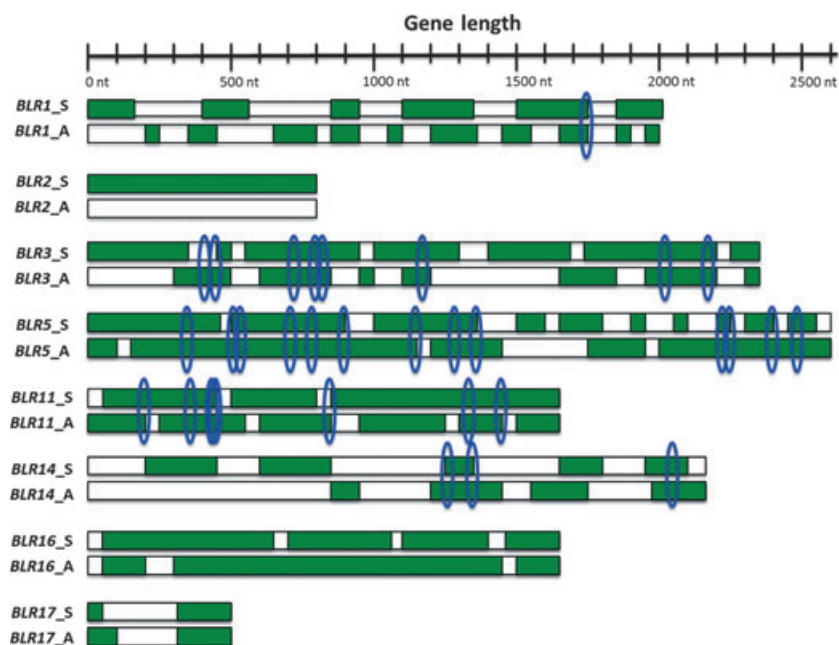


Figure 1 Read coverage and single nucleotide polymorphism (SNP) locations for eight selected SNAP predicted genes. Green boxes represent read coverage in that area of the gene; white represents lack of coverage. SNPs between the Skipton and Ag-Spectrum cultivars are circled in blue. S, Skipton, A, Ag-Spectrum.

BLR2

The complete absence of Ag-Spectrum Illumina reads mapping to the *BLR2* region (Figure 1) gave an early indication that there may be substantial sequence variation between the Ag-Spectrum and Skipton DNA in this region. Successful amplification of the *BLR2* gene region, including some predicted 5' and 3' sequence, in Skipton and Ag-Spectrum with the F3/R3 primer pair (Table 2) revealed an obvious difference in amplicon lengths.

Sequencing these amplicons revealed a 437 nucleotide (-nt) difference in size between Skipton and Ag-Spectrum. This difference included insertions of 405-nt and 6-nt within 1 kb downstream of the Ag-Spectrum gene relative to Skipton, as well as a 26-nt insertion and 4-nt deletion in the Ag-Spectrum 5' region (positions-530 and-105, respectively) and a 4-nt and 3-nt deletion in the Ag-Spectrum coding region. All amplicons sequenced per cultivar were identical.

Within the amplified genomic region, Skipton showed 98% DNA sequence identity to the *B. rapa* reference genome whilst Ag-Spectrum shared just 79.4% identity. Given the necessary

stringency in whole-genome NGS read mapping from polyploid genomes, this degree of divergence from the reference sequence in Ag-Spectrum explains the lack of read coverage for this gene region (Table 1). Within the putative coding sequence of the *BLR2* gene, Skipton and Ag-Spectrum shared 99.8% and 90.3% identity with *B. rapa*, respectively.

The predicted amino acid sequence for Skipton BLR2 was found to be almost identical to that for the reference *B. rapa* protein (99.5%). In contrast, the predicted Ag-Spectrum sequence shared 83.7% amino acid identity with *B. rapa* from position 1 to 44, at which point a stop codon truncates the putative protein. These results suggest that the genomic environment of *BLR2* is substantially different between Skipton and Ag-Spectrum, and that the Ag-Spectrum BLR2 protein is likely to be severely truncated and thus non-functional relative to the conserved Skipton protein.

Genetic mapping of this gene in the Skipton/Ag-Spectrum population (Raman *et al.*, 2012), further supports it as a candidate. The gene was located within the QTL region, and the mapping showed a clear and strong association between resis-

Table 2 Details of primers used in this project

SNAP gene	Primer pair	Primer sequence 5'–3'	Annealing temperature (°C)	Product size (bp)	
				Skipton	Ag Spectrum
<i>BLR2</i>	F3 R3	TCATAAACTGCTTTTCATGG TGCCTTACGTAACCAAACC	60	2402	2839
<i>BLR11</i>	FI RI	GTTCCCCCGTCCATAGG CTAAATCGTTACGGGCAAGG	60	2019	2012
Other primers:	M13 F primer (universal)	GTTTCCAGTCACGACGTT	50		
	M13 R primer (universal)	GACCATGATTACGCCAAGCTA			

Primer pairs, annealing temperatures and amplicon lengths are also shown.

F, forward primer; R, reverse primer.

tance phenotype and the flanking SSR and SNP markers known to reside on chromosome A7.

BLR11

BLR11 was selected for further analysis because of the high read coverage and presence of numerous SNPs between Skipton and Ag-Spectrum identified from the short-read mapping data (Table 1, Figure 1). A total of six SNPs were identified within the predicted gene, with a further eight SNPs identified 1 kb up- and downstream. These SNPs were predicted to have four non-synonymous effects on the putative protein creating further interest in this candidate.

Amplification of the *BLR11* candidate gene using the F1/R1 primer pair for both Skipton and Ag-Spectrum resulted in similar-sized amplicons, encompassing the predicted coding region and 100–200 nt up- and downstream. The Ag-Spectrum product, however, contained an 11-nt deletion in the predicted 5' region upstream of the gene and a 4-nt insertion at position 605 of the putative coding region, giving 98.4% CDS identity to Skipton.

The predicted amino acid sequence for Skipton BLR11 was found to share 99.3% identity with *B. rapa* and 100% identity with Ag-Spectrum until residue 201, at which point the 4-nt insertion in Ag-Spectrum resulted in a frameshift mutation and subsequent stop codon at position 203. Therefore, the predicted protein product of BLR11 is truncated in Ag-Spectrum and may be non-functional. Thus, BLR11 also remains a promising candidate for *Rlm4*-mediated resistance in Skipton.

In all, predictions of SNPs and indel polymorphisms between NGS reads for the *B. napus* parents of a QTL mapping population segregating for *Rlm4* resistance has enabled the prioritization of candidate genes for further analysis. This approach allows the rapid identification of candidate genes for traits of interest, without the requirement for high-resolution map-based cloning. Whilst the success of the method has been demonstrated using whole-genome shotgun data, other sources of sequence data, such as RNA-seq, could also be utilised if data were available; however, this would limit the identification of genes to those expressed in the tissues sequenced and may have complications in distinguishing between paralogues.

Experimental procedures

Genetic and physical mapping of the *Rlm4* QTL

The genetic map position for *Rlm4* was determined in a Skipton cv. × Ag-Spectrum cv. mapping population, segregating for blackleg resistance (Raman *et al.*, 2012). The 12 SSR markers underlying the QTL on chromosome A7 were compared with the public *B. rapa* A genome (Wang *et al.*, 2011) reference sequence using BLAST (Altschul *et al.*, 1990). This physical region was annotated with predicted genes using SNAP gene prediction software (Korf, 2004) tailored to Brassica. Genes with homology to known resistance genes were selected.

Identification of *B. napus* orthologues of the *B. rapa* candidate genes

Whole-genome, paired-end short-sequence reads (>10× coverage) for cv. Skipton and Ag-Spectrum were generated using the Illumina Genome Analyser IIx (GAIIx) according to manufacturer's instructions (Illumina, San Diego, CA). These were mapped to assembled *B. rapa* and *B. oleracea* genome

sequences using SOAP2 (Li *et al.*, 2009). SNPs were identified as sequence differences between the two varieties supported by at least two reads from each variety. Sequence reads matching the genomic regions can be downloaded from http://www.brassicagenome.net/downloads/R_XA_0003.zip.

Amplification, cloning and sequencing of *Rlm4* candidate genes

PCR primers were designed with the program AmplifX version 1.5.4 (Jullien, 2008) using the Skipton and Ag-Spectrum short-read information mapped to the *B. rapa* reference. Table 1 lists the primers and PCR conditions used. PCR primers were compared with the A and C reference genomes to increase genome specificity of the primers.

Candidate genes were amplified from Skipton and Ag-Spectrum genomic DNA using PCR with the primers listed in Table 2. A total reaction volume of 40 µL consisted of 1 × iTaq PCR buffer (containing 100 mM Tris-HCl (pH 8.3) and 500 mM KCl) (Scientific), 200 µM each dNTP (Scientific), 0.5 µM each primer, 1.5 U iTaq DNA polymerase (Scientific), RNase- and DNase-free water (Gibco) and 50 ng DNA. Thermocycling conditions for the reaction were 94 °C for 2 min, followed by 35 cycles of 94 °C for 30 s, annealing for 1 min and extension at 72 °C. Final extension was performed at 72 °C for 8 min. Gel electrophoresis on a 1% (w/v) agarose gel in 1 × TAE buffer (Sambrook and Russell, 2001) containing ethidium bromide resolved products of appropriate size, which were excised and purified using a silica method based on Boyle and Lew (1995).

PCR-purified DNA was ligated into the pGEM-T Easy vector (Promega) overnight at 4°C according to the manufacturer's protocol. Heat shock competent *E. coli* XL1-Blue cells were transformed with ligated plasmid, then plated onto an LB agar medium containing 100 µg/mL ampicillin, 50 mg/mL X-gal and 50 mg/mL IPTG and grown overnight at 37°C. Positive clones (white colonies) were screened by PCR using the universal M13F and M13R primers (see Table 1). Colonies testing positive for PCR inserts were selected for plasmid purification using the High pure DNA plasmid isolation kit (Roche) according to the manufacturer's instructions.

Purified plasmid DNA was used as the template for Sanger sequencing reactions performed by the Australian Genome Research Facility (AGRF). Nucleotide and protein alignments and assemblies were performed using Geneious Pro v5.4.6 (Drummond *et al.*, 2012) with a cost matrix of 65%, a gap open penalty of 6 and a gap extension penalty of 3. Protein sequence prediction and alignments were performed using the Geneious alignment tool with default parameters.

Conclusions

This study highlights the applicability of NGS data for rapid and efficient discovery of QTL candidates within physically mapped genome regions. Using whole-genome shotgun reads for parents of a *B. napus* population segregating for *Rlm4* resistance, two strong candidates for *Rlm4* in *B. napus* cv. Skipton were identified. Candidates were selected based on the sequence identity with known resistance genes and sequence divergence in the NGS reads for the susceptible parent. These genes, BLR2 and BLR11, were both found to encode truncated proteins in the susceptible Ag-Spectrum parent, with a particularly severe truncation noted for BLR2. Analyses are now underway to

confirm the functional *Rlm4* locus in cv. Skipton including expression analyses and transformation into susceptible lines. Furthermore, the NGS data for these two lines has enabled the identification of 400 SNPs within the existing *Rlm4* QTL, a subset of which have been included in a custom 6K Illumina Infinium chip for further fine mapping of the *Rlm4* QTL. This method of candidate gene identification and applicability to fine mapping through the use of molecular markers flanking a QTL region, which map to specific regions of the assembled genome sequence, overcomes the problems associated with candidate gene approaches based purely on sequence similarity from model species in polyploid genomes. Combining this with PCR amplification, genetic mapping and sequencing validation of single specific gene products further confirms the genes as candidates. Whilst this has been applied for blackleg resistance here, the method is transferable to any trait.

Acknowledgements

The authors would like to acknowledge funding support from the Australian Research Council (Projects LP0882095, LP0883462, LP0989200, LP110100200 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

- Allen, A.M., Barker, G.L.A., Berry, S.T., Coghill, J.A., Gwilliam, R., Kirby, S., Robinson, P., Brenchley, R.C., D'Amore, R., McKenzie, N., Waite, D., Hall, A., Bevan, M., Hall, N. and Edwards, K.J. (2011) Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat (*Triticum aestivum* L.). *Plant Biotechnol. J.* **9**, 1086–1099.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
- Ansan-Melayah, D., Balesdent, M.H., Delourme, R., Pilet, M.L., Tanguy, X., Renard, M. and Rouxel, T. (1998) Genes for race-specific resistance against blackleg disease in *Brassica napus* L. *Plant Breed.* **117**, 373–378.
- Balesdent, M.H., Attard, A., Ansan-Melayah, D., Delourme, R., Renard, M. and Rouxel, T. (2001) Genetic control and host range of avirulence toward *Brassica napus* cultivars Quinta and Jet Neuf in *Leptosphaeria maculans*. *Phytopathology*, **91**, 70–76.
- Balesdent, M.H., Attard, A., Kühn, M.L. and Rouxel, T. (2002) New avirulence genes in the phytopathogenic fungus *Leptosphaeria maculans*. *Am. Phytopath. Society*, **92**, 1122–1133.
- Batley, J. and Edwards, D. (2009) Genome sequence data: management, storage and visualization. *Biotechniques*, **46**, 333–336.
- Baumgarten, A., Cannon, S., Spangler, R. and May, G. (2003) Genome level evolution of resistance genes in *Arabidopsis thaliana*. *Genetics*, **165**, 309–319.
- Boyle, J.S. and Lew, A.M. (1995) An inexpensive alternative to glassmilk for DNA purification. *Trends Genet.* **11**, 8.
- Delourme, R., Pilet-Nayel, M.L., Archipiano, M., Horvais, R., Tanguy, X., Rouxel, T., Brun, H., Renard, M. and Balesdent, M.H. (2004) A cluster of major specific resistance genes to *Leptosphaeria maculans* in *Brassica napus*. *Phytopathology*, **94**, 578–583.
- Delourme, R., Chevre, A.M., Brun, H., Rouxel, T., Balesdent, M.H., Dias, J.S., Salisbury, P., Renard, M. and Rimmer, S.R. (2006) Major gene and polygenic resistance to *Leptosphaeria maculans* in oilseed rape (*Brassica napus*). *Eur. J. Plant Pathol.* **114**, 41–52.
- Delourme, R., Piel, N., Horvais, R., Pouilly, N., Domin, C., Vallée, P., Falentin, C., Manzaneres-Dauleux, M.J. and Renard, M. (2008) Molecular and phenotypic characterization of near isogenic lines at QTL for quantitative resistance to *Leptosphaeria maculans* in oilseed rape (*Brassica napus* L.). *Theor. Appl. Genet.* **117**, 1055–1067.
- Drummond, A.J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., Field, M., Heled, J., Kearse, M., Markowitz, S., Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T. and Wilson, A. (2012) Geneious Pro v5.6. Available from <http://www.geneious.com> (Last accessed 1 February 2012).
- Duran, C., Appleby, N., Vardy, M., Imelfort, M., Edwards, D. and Batley, J. (2009a) Single nucleotide polymorphism discovery in barley using AutoSNPdb. *Plant Biotechnol. J.* **7**, 326–333.
- Duran, C., Appleby, N., Edwards, D. and Batley, J. (2009b) Molecular genetic markers: discovery, applications, data storage and visualisation. *Curr. Bioinform.* **4**, 16–27.
- Duran, C., Eales, D., Marshall, D., Imelfort, M., Stiller, J., Berkman, P., Clark, T., McKenzie, M., Appleby, N., Batley, J., Basford, K. and Edwards, D. (2010) Future tools for association mapping in crop plants. *Genome*, **53**, 1017–1023.
- Edwards, D. and Batley, J. (2010) Plant genome sequencing: applications for crop improvement. *Plant Biotechnol. J.* **8**, 2–9.
- Edwards, D. and Wang, X. (2012) Genome sequencing initiatives. In *Genetics, Genomics and Breeding of Oilseed Brassicas* (Edwards, D., Parkin, I.A.P. and Batley, J., eds), pp. 152–157. USA: Science Publishers.
- Ferreira, M.E., Rimmer, S.R., Williams, P.H. and Osborn, T.C. (1995) Mapping loci controlling *Brassica napus* resistance to *Leptosphaeria maculans* under different screening conditions. *Am. Phytopath. Society*, **85**, 213–217.
- Fitt, B.D.L., Hu, B.C., Li, Z.Q., Liu, S.Y., Lange, R.M., Kharbanda, P.D., Butterworth, M.H. and White, R.P. (2008) Strategies to prevent spread of *Leptosphaeria maculans* (phoma stem canker) onto oilseed rape crops in China; costs and benefits. *Plant. Pathol.* **57**, 652–664.
- Freidt, W. and Snowdon, R. (2009) Oilseed rape. In *Oil Crops, Handbook of Plant Breeding 4* (Vollmann, J. and Rajcan, I., eds), pp. 91–126. New York: Springer Science+Business Media.
- Hayward, A., McLanders, J., Campbell, E., Edwards, D. and Batley, J. (2012) Genomic advances will herald new insights into the *Brassica: Leptosphaeria maculans* pathosystem. *Plant Biol.* **14**, 1–10.
- Howlett, B.J. (2004) Current knowledge of the interaction between *Brassica napus* and *Leptosphaeria maculans*. *Can. J. Plant Pathol.* **26**, 245–252.
- Howlett, B.J., Indurm, A. and Pedras, M.S.C. (2001) *Leptosphaeria maculans*, the causal agent of blackleg disease of *Brassicaceae*. *Fungal Genet. Biol.* **33**, 1–14.
- Imelfort, M. and Edwards, D. (2009) De novo sequencing of plant genomes using second-generation technologies. *Brief. Bioinform.* **10**, 609–618.
- Imelfort, M., Duran, C., Batley, J. and Edwards, D. (2009) Discovering genetic polymorphisms in next generation sequencing data. *Plant Biotechnol. J.* **7**, 312–317.
- Jullien, N. (2008) AmpliX version 1.5.4 (software). Available at: <http://ifjfr.nord.univ-mrs.fr/AmplifX-Home-page>. Accessed June 2011.
- Kaur, S., Cogan, N.O.I., Ye, G., Ballie, R.C., Hand, M.L., Ling, A.E., McGearey, A.K., Kaur, J., Hopkins, C.J., Todorovic, M., Mountford, H., Edwards, D., Batley, J., Burton, W., Salisbury, P., Gororo, N., Marcroft, S., Kearney, G., Smith, K.F., Forster, J.W. and Spangenberg, G.C. (2009) Genetic map construction and QTL mapping of resistance to blackleg (*Leptosphaeria maculans*) disease in Australian canola (*Brassica napus* L.) cultivars. *Theor. Appl. Genet.* **120**, 71–83.
- Kharabian-Masouleh, A., Waters, D.L.E., Reinke, R.F. and Henry, R.J. (2011) Discovery of polymorphisms in starch-related genes in rice germplasm by amplification of pooled DNA and deeply parallel sequencing. *Plant Biotechnol. J.* **9**, 1074–1085.
- Korf, I. (2004) Gene finding in novel genomes. *BMC Bioinform.* **5**, 59.
- Li, R., Yu, C., Li, Y., Lam, T.-W., Yiu, S.-M., Kristiansen, K. and Wang, J. (2009) SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* **25**, 1966–1967.
- Lim, G.A.C., Jewell, E.G., Li, X., Erwin, T.A., Love, C., Batley, J., Spangenberg, G. and Edwards, D. (2007) A comparative map viewer Integrating Genetic Maps for *Brassica* and *Arabidopsis*. *BMC Plant Biol.* **7**, 40.
- Long, Y., Wang, W., Sun, Z., Fernando, D.W.G., McVetty, P.B.E. and Li, G. (2011) Identification of two blackleg resistance genes and fine mapping of

- one of these two genes in a *Brassica napus* canola cultivar 'Surpass 400'. *Theor. Appl. Genet.* **122**, 1223–1231.
- Love, C., Logan, E., Erwin, T., Kaur, J., Lim, G.A.C., Hopkins, C., Batley, J., James, N., May, S., Spangenberg, G. and Edwards, D. (2006) Integrating and interrogating diverse *Brassica* data within an Ensembl structured database. *Acta Hort.* **706**, 77–82.
- Malory, S., Shapter, F.M., Elphinstone, M.S., Chivers, I.H. and Henry, R.J. (2011) Characterizing homologues of crop domestication genes in poorly described wild relatives by high-throughput sequencing of whole genomes. *Plant Biotechnol. J.* **9**, 1131–1140.
- Parlange, F., Daverdin, G., Fudal, I., Kühn, M.L., Balesdent, M.H., Blaise, F., Grezes-Beset, B. and 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. *Mol. Microbiol.* **71**, 851–863.
- Persson, M., Staal, J., Oid, S. and Dixelius, C. (2009) Layers of defense responses to *Leptosphaeria maculans* below the *Rlm1* – and Camelexin – dependent resistances. *New Phytol.* **182**, 470–482.
- Pilet, M.L., Duplan, G., Archipiano, M., Barret, P., Baron, C., Horvais, R., Tanguy, X., Lucas, M.O., Renard, M. and Delourme, R. (2001) Stability of QTL for field resistance to blackleg across two genetic backgrounds in oilseed rape. *Crop Sci.* **41**, 197–205.
- Raman, R., Taylor, B., Marcroft, S., Stiller, J., Eckermann, P., Coombes, N., Rehman, A., Lindbeck, K., Lockett, D., Wratten, N., Batley, J., Edwards, D., Wang, X. and Raman, H. (2012) Molecular mapping of qualitative and quantitative loci for resistance to *Leptosphaeria maculans*; causing blackleg disease in canola (*Brassica napus* L.). *Theor. Appl. Genet.* doi: 10.1007/s00122-012-1842-6:1-14 (in press).
- Rimmer, S.R. (2006) Resistance genes to *Leptosphaeria maculans* in *Brassica napus*. *Can. J. Plant Pathol.* **28**, S288–S297.
- Sambrook, J. and Russell, D. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd edn. New York: Cold Spring Harbour Laboratory Press.
- Subbaiyan, G.K., Waters, D.L.E., Katiyar, S.K., Sadananda, A.R., Vaddadi, S. and Henry, R.J. (2012) Genome-wide DNA polymorphisms in elite indica rice inbreds discovered by whole-genome sequencing. *Plant Biotechnol. J.* doi: 10.1111/j.1467-7652.2011.00676.x (in press).
- Varshney, R.K., Nayak, S.N., May, G.D. and Jackson, S.A. (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotechnol.* **27**, 522–530.
- Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., Bai, Y., Mun, J.-H., Bancroft, I., Cheng, F., Huang, S., Li, X., Hua, W., Wang, J., Wang, X., Freeling, M., Pires, J.C., Paterson, A.H., Chalhou, B., Wang, B., Hayward, A., Sharpe, A.G., Park, B.-S., Weisshar, B., Liu, B., Li, B., Liu, B., Tong, C., Song, C., Duran, C., Peng, C., Geng, C., Koh, C., Lin, C., Edwards, D., Mu, D., Shen, D., Soumpourou, E., Li, F., Fraser, F., Conant, G., Lassalle, G., King, G.J., Bonnema, G., Tang, H., Wang, H., Belcram, H., Zhou, H., Hirakawa, H., Abe, H., Guo, H., Wang, H., Jin, H., Parkin, I.A.P., Batley, J., Kim, J.-S., Just, J., Li, J., Xu, J., Deng, J., Kim, J.A., Li, J., Yu, J., Meng, J., Wang, J., Min, J., Min, J., Poulain, J., Wang, J., Hatakeyama, K., Wu, K., Wang, L., Fang, L., Trick, M., Links, M.G., Zhao, M., Jin, M., Ramchiary, N., Drou, N., Berkman, P.J., Cai, Q., Huang, Q., Li, R., Tabata, S., Cheng, S., Zhang, S., Zhang, S., Huang, S., Sato, S., Sun, S., Kwon, S.-J., Choi, S.-R., Lee, T.-H., Fan, W., Zhao, X., Tan, X., Xu, X., Wang, Y., Qiu, Y., Yin, Y., Li, Y., Du, Y., Liao, Y., Lim, Y., Narusaka, Y., Wang, Y., Wang, Z., Li, Z., Wang, Z., Xiong, Z. and Zhang, Z. (2011) The genome of the mesopolyploid crop species *Brassica rapa*. *Nat. Genet.* **43**, 1035–1039.