Identification of Genome-Wide DNA Variants and SNP Haplotypes Associated with Avirulence Genes of Leptosphaeria Maculans in Western Canada

Qilin Chen
Saskatoon Research and Development Centre
Gary Peng
Saskatoon Research and Development Centre
Randy Kutcher
University of Saskatchewan
Fengqun Yu (fengqun.yu@canada.ca)
Agriculture and Agri-Food Canada https://orcid.org/0000-0002-7957-7632

Research article

Keywords:

Posted Date: April 30th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-24766/v1

Abstract

Background

Blackleg, caused by *Leptosphaeria maculans*, is a serious disease of canola/ oilseed rape in many parts of the world. An integrated approach is needed to control the disease, with genetic resistance as a key component of the management strategy. Towards this goal, a reliable approach for *L. maculans* race structure assessment becomes essential to gain understanding of the frequency of avirulence genes and race groups of the pathogen, and provide guidance for deployment of resistant canola cultivars.

Results

A total of 162 representative isolates collected in western Canada were selected for genome re-sequencing with an Illumina platform. Assembly of the short reads against the reference genome of *L. maculans* 'brassicae' isolate v23.1.3 led to the discovery of 21,016 DNA variants (SNPs and InDels), 93% SNPs and 7% InDels, with a transition/transversion (Ts/Tv) ratio of 3.1 genome wide. InDels occurred mainly in GC-blocks and the Ts/Tv ratio of SNPs in AT-blocks was > 2 times higher than that in GC-blocks. The number of variants were positively correlated with supercontig size, GC-block size and gene numbers. DNA variants in most avirulence genes were SNPs, except a deletion in *AvrLm1*. The number of SNPs varied from 1–2 in *AvrLm2*, *AvrLmJ1-5-9*, *AvrLm6*, *AvrLm10A*, *AvrLm10B* and *AvrLm11*, 8 in *AvrLm3* and 38 in *AvrLm4-7*. This study is the first report of triallelic SNPs in *AvrLm2* and *AvrLm4-7*, and premature STOP codons in *AvrLm4-7*. Nine SNP haplotypes were identified in *AvrLm4-7*, however, only 2 ~ 3 haplotypes occurred in other avirulence genes, and in total 47 haplotype groups were identified from the isolates. The 47 SNP haplotype groups were translated into 44 protein haplotype groups and then isolates of *L. macualns* collected in western Canada were classified into10 races.

Conclusion

In this study, we document the shortcoming of inferring races from SNP genotyping, and propose the use of SNP haplotyping for more reliable and informative analysis of *L. maculans* race structure.

Background

The fungal pathogen *Leptosphaeria maculans*, a filamentous ascomycete, is the causal agent of blackleg (phoma stem canker) of canola or oilseed rape (*Brassica napus* L.) that often leads to economic losses [1, 2]. Sustainable canola production requires effective management of blackleg using an integrated approach, including crop rotation, fungicide seed treatment, and the development of resistant cultivars [3–6]. The use of resistant cultivars has been considered one of the most effective and economical way to control the disease. However, breeding resistant cultivars is challenging because the pathogen has both sexual and asexual reproduction systems, which enables it to combine the most fitting genotypes and quickly increase their frequencies in the population through clonal reproduction [7–9]. The sexual reproduction system creates a high level of genetic diversity, enabling the pathogen to adapt to the resistance genes used in crop cultivars, such as *Rlm1* in France [10], *LepR3* in Australia [11] and *Rlm3* in Canada [12]. Therefore, the constant search for new resistance genes from closely related *Brassica* species and introgression of these genes into canola is a major objective in blackleg resistance breeding programs [13, 14]. Development of resistant cultivars involving specific resistance genes against *L. maculans* is deemed an on-going process due to host-pathogen co-evolution [15, 16]. To gain an upper hand, the molecular mechanism underlying how the pathogen responds to selection pressure and subsequently overcomes the host resistance should be thoroughly investigated [17, 18]. To this end, fundamental studies have been conducted on the genetic diversity of the pathogen [19, 20] and host-pathogen interaction [21, 22]. The completion of the *L. maculans* genome sequence [15] was a significant development in the study of this fungal pathogen and provides a reference genome to which molecular markers can be physically mapped [23]. In addition, several avirulence (Avr) genes, *AvrLm1* [7], *AvrLm2* [24], *AvrLm3* [25], *AvrLm4-7* [26], *AvrLmJ1-5-9* [27, 28], *AvrLm6* [29], *AvrLm1*0 [30] and *AvrLm11* [31] have been cloned.

Regular profiling of Avr genes in *L. maculans* populations provides key information on the deployment of effective specific resistance genes that may be used for blackleg resistance breeding [32]. The Avr profile can be determined by using differential lines carrying known resistance genes based on the gene-for-gene theory [33, 34]. This approach is time- and resource-consuming, and its accuracy can be affected by impurity in the seed stock of differential hosts and subjectivity of researchers. Therefore, alternative methods have been explored to simplify and improve the accuracy of Avr profiling, and molecular markers are potentially more efficient and objective options. Generations of molecular markers, including random amplified polymorphic DNAs (RAPD), restriction fragment length polymorphisms (RFLP), simple sequence repeats (SSRs) and DNA variants (SNPs and InDels) have been used for the pathogen differentiation and monitoring. For example, Goodwin and Annis (1991) [35] used random decamer primers to differentiate three Canadian *L. maculans* isolate groups. RAPD markers have been used to assess the aggressiveness of *L. maculans* isolates [36, 37]. Polymerase chain reaction (PCR) was also used to detect DNA variation in *AvrLm1*, *AvrLm6* [38] and *AvrLm4-7* [20]. PCR based methods can usually detect polymorphism in terms of presence/absence or size variation at a specific locus, but cannot capture DNA variants that potentially impact gene function. Brown (1996) [39] also indicated the importance of selecting molecular markers for plant pathogen research. Recent advances in modern DNA sequencing technologies, represented by Next Generation Sequencing (NGS), significantly reduce the cost of discovering DNA variants, including SNPs. Although NGS has found its broad application in fundamental genome and genetic research [40–42], there are limited reports on the identification of genome-wide DNA variants in plant pathogens, including *L. maculans*. Zander et al. (2013) [43] reported 21,814 genome-wide SNPs in *L. maculans* based on a study of two Australian isolates.

The majority of SNPs are typically biallelic with lower information content as compared to other multiallelic markers [44], and single SNPs alone are inadequate for genetic diagnosis [45, 46]. Use of SNP haplotypes was suggested to be more effective shortly after the discovery of high-density SNP markers from studies on human genetic diversity [47] and crop genetics of quantitative traits [48]. Haplotype is a term contracted from 'haploid genotype', and refers usually to a combination of alleles on the same chromosome, which are transmitted together, although the concept has been used in different contexts [49–

51]. SNP haplotypes have been increasingly applied in genetic studies. For example, a SNP haplotype can be constructed based on linkage disequilibrium for genetic analysis of crop genome [44]. Haplotypes are often inferred from various algorithms, including parsimony [52], maximum likelihood [53] and Bayesian [54]. The SNP haplotype inference by a statistical algorithm, however, may also result in an incorrect outcome [54]. Because a haplotype carries more information than each individual SNP [46], SNP haplotypes have been used broadly in biomedical research on human diseases [55–57], association mapping in plants [45, 56, 59, 60], marker-selection for resistance genes in crop species [60] and crop yield improvement [44]. Studies using SNP haplotypes have not been reported on plant pathogen race profiling.

Cost-effective genotyping technologies that capture sequence variation at ultra-high resolution are now available and commercial genotyping platforms that can generate thousands or millions of data points per experiment have become almost routinely used for genetic research [61]. In this study, we describe re-sequencing of 162 representative *L. maculans* isolates, which were collected in western Canada with the Avr profile characterized originally using a set of differential hosts, for the identification of DNA variants. The objectives were to: i) identify and characterize genome-wide DNA variants with a special focus on cloned Avr genes; ii) identify SNP haplotypes in these Avr genes and their protein isoforms; and iii) explore the possibility of *L. maculans* race structure determination through the analysis of haplotypes. Ultimately, we were interested to know whether SNP haplotypes are more reliable than individual SNPs for genotyping *L. maculans* isolates, especially for Avr profiling.

Results

Selection of isolates for re-sequencing

Cluster analysis was conducted using the phenotypic data from bioassays on the differential hosts (Table S1) inoculated individually with the 1590 *L. maculans* isolates from western Canada. Six clusters containing 125, 325, 179, 143, 195, and 623 isolates respectively, were identified (Fig. 1). A total of 162 isolates (Table S2), 27, 31, 19, 23, 14, and 48 isolates from different clusters (Fig. 1) were selected to represent different geographic regions and Avr profiles for DNA sequencing.

NGS short reads were assembled to the reference genome of *L. maculans*. Template coverage (TC) and median coverage (MC) were examined to evaluate the quality of the NGS sequences. Both TC and MC varied with isolates and supercontigs (SCs), and four isolates (12CC-357, 12CC-77, 13CCMB03-06, 13CCSK05-01) with lower TC and MC were removed for further analysis. The remaining 158 isolates showed TC > 80% (Fig. S1A), and 5 SCs (SC 27, 28, 33, 34, 36) carried lower TC ranging from 49.6–73.8% (Figure S1B). For MC, 88% of the isolates displayed

values greater than 20. Except that SC 30 had a MC as high as 621, all the other SCs exhibited MC ranging from 12 to 26. In summary, the average TC and MC of the isolates were 91% and 36, respectively.

Variant discovery

When compared with the reference genome of *L. maculans*, 37,947 variants were detected from the 158 *L. maculans* isolates. After removing monomorphic variants and filtering the data against the criteria of phred quality score (Q > 30) and minimum allele frequency (MAF > 5), eventually 21,016 polymorphic variants were confirmed and used in subsequent analysis. The number of variants in each isolate ranged from 8,591 to 14,997 with an average of 9,575.

Distribution of genome-wide DNA variants

To understand the distribution of the genome-wide DNA variants, variant numbers in genomic regions, including the sizes of SCs, GC and AT blocks, were investigated (Table 1). Variant numbers were highly correlated with SC size (R^2 = 0.9472, Fig. S2A &Table 1) and gene number (R^2 = 0.9053, Fig. S2B & Table 1) in SCs. The R^2 for the correlation of variant number with GC block size (R^2 = 0.9231, Fig. S2C &Table 1) was greater than the R^2 with AT blocks (R^2 = 0.8361, Fig. S2D &Table 1). The AT- and GC-blocks accounted for 35.2% and 64.8% of the genome size (Table 2), containing 28.6% and 71.4% of variants; this indicated that variants were proportionally distributed in the two blocks.

Table 1

Variant distribution and density in supercontigs and different genomic regions of *L. maculans*.

Super-Contigs	Size (base)	Gene numbers	AT-region rate (%)	Coding region coverage (%)	Variant number	Variant density (per kb)			
						whole SC	AT-blocks	GC-Blocks	AT-blocks/GC-Blocks
SC0	4258568	1263	25.4	45.67	1919	0.45	1.31	0.16	8.3
SC1	3378610	1080	20.6	45.96	1545	0.46	1.80	0.11	16.8
SC2	2939989	916	19.6	46.09	1185	0.40	1.59	0.11	14.0
SC3	2348246	565	37.2	34.61	1395	0.59	1.02	0.36	2.9
SC4	1918205	662	22.5	46.87	794	0.41	1.45	0.18	8.1
SC6	1888674	510	37.0	33.79	894	0.47	0.79	0.54	1.5
SC5	1869450	633	20.5	50.65	945	0.51	2.00	0.12	16.4
SC8	1809296	618	18.4	45.84	682	0.38	0.56	0.37	1.5
SC9	1772623	474	28.3	39.7	970	0.55	1.14	0.16	6.9
SC7	1769547	567	25.1	44.69	821	0.46	0.74	0.48	1.5
SC10	1758670	434	33.2	35.57	791	0.45	0.89	0.29	3.1
SC13	1634580	492	27.2	42.29	782	0.48	0.59	0.58	1.0
SC12	1631710	439	28.1	38.01	769	0.47	0.70	0.60	1.2
SC11	1590160	285	39.3	25.52	655	0.41	0.82	0.49	1.7
SC15	1560629	442	29.5	39.64	692	0.44	1.07	0.16	6.7
SC14	1533332	512	20.2	51.78	633	0.41	1.04	0.12	8.9
SC17	1445693	415	23.7	41.99	447	0.31	0.61	0.17	3.7
SC16	1397653	353	34.4	39.83	963	0.69	0.63	0.07	8.5
SC18	1351976	366	30.5	40.36	925	0.68	0.32	0.09	3.8
SC19	1186800	288	31.9	36.34	463	0.39	0.07	0.09	0.7
SC20	1087932	292	30.4	40.47	665	0.61	0.01	0.04	0.2
SC21	1020521	296	29.1	40.2	450	0.44	1.69	0.37	4.6
SC22	731443	35	75.0	5.53	277	0.38	0.23	1.33	0.2
SC23	521426	157	25.6	42.97	267	0.51	2.36	0.64	3.7
SC24	475869	116	32.5	40.12	256	0.54	1.06	0.57	1.9
SC25	318058	99	21.7	48.58	152	0.48	1.80	0.94	1.9
SC26	261540	74	27.8	43.28	106	0.41	1.99	1.11	1.8
SC27	250629	3	55.9	0.38	100	0.40	0.40	0.67	0.6
SC28	236098	1	55.5	0.45	75	0.32	0.14	0.35	0.4
SC29	200940	7	62.7	3.22	129	0.64	0.50	0.51	1.0
SC30	154863	35	96.7	12.49	6	0.04	0.25	11.04	0.0
SC31	143268	37	36.0	40.50	53	0.37	0.06	0.22	0.3
SC32	87679	2	53.8	3.16	53	0.60	0.68	0.52	1.3
SC35	79158	0	43.5	0.00	20	0.25	0.29	0.22	1.3
SC33	65326	0	68.9	0.00	17	0.26	0.22	0.34	0.6
SC34	58596	1	66.8	0.69	47	0.80	0.74	0.93	0.8
SC37	52193	0	28.2	0.00	9	0.17	0.48	0.05	8.9
SC36	35372	0	58.5	0.00	28	0.79	0.53	1.16	0.5
SC38	23239	0	96.9	0.00	7	0.30	0.18	4.23	0.0

SC39	22454	0	66.4	0.00	18	0.80	0.67	1.06	0.6
SC40	21590	0	86.5	0.00	11	0.51	0.43	1.03	0.4

Variants composition in different genomic regions

Genome-wise, transitions (Ts), transversions (Tv) in SNPs, insertions and deletions in InDels were 70.1, 22.8, 3.0 and 4.1%, respectively (Table 2); the majority of DNA variants were SNPs (Ts and Tv) with a Ts/Tv ratio of 3.1. Ts/Tv, however, varied with genomic regions, for instance, it was 7.5 in AT-blocks, much higher than 2.3 in GC-blocks. Similarly, non-coding region possessed a ratio of 3.7, higher than 2.2 in coding regions including small secreted protein (SSP) genes and non-SSP genes.

InDels made up approximately 7% of total genome-wide variants, and were unevenly distributed in AT-blocks and GC-blocks. InDels occurred much more frequent in GC-blocks (9.5%) as compared with AT-blocks (1.2%). Deletions were found in nearly equal proportion with insertions in AT-blocks, or slightly higher than insertions in GC-block at whole genome level. Comparatively, deletions appeared more frequently, 3.8 times as many as insertions in SSP genes (Table 2).

	Genome percentage (%)	Variant percentage in whole genome (%)	Variant composition (%)				
			Transition (Ts)	Transversion (Tv)	Insertion	Deletion	Ts/Tv
Whole genome	-	-	70.1	22.8	3.0	4.1	3.1
AT-blocks	35.2	28.6	87.1	11.6	0.6	0.6	7.5
GC-blocks	64.8	71.4	63.2	27.3	4.0	5.5	2.3
Non-coding regions	59.8	69.7	72.7	19.9	3.2	4.2	3.7
Coding regions	40.2	30.3	63.9	29.6	2.6	3.9	2.2
SSP genes	0.2	1.0	62.8	28.4	1.9	7.0	2.2
Non-SSP genes	40	29.2	63.9	29.6	2.7	3.8	2.2

Table 2 Variants identified in different genomic regions of *L. maculans*

Characterization of DNA variants in the cloned Avr genes

As fore-mentioned, *AvrLm1*, *AvrLm2*, *AvrLm3*, *AvrLm4-7*, *AvrLm6*, *AvrLmJ1-5-9*, *AvrLm10A*, *AvrLm10B* and *AvrLm11* of *L. maculans* have been cloned. The variants in these genes can be categorized by SeqMan Pro software into four groups: non-synonymous, nonsense, frameshift and synonymous. Non-synonymous variants occurred in all of the Avr genes except *AvrLm1*, *AvrLm10A* and *AvrLm10B*. *AvrLm3* and *AvrLm4-7* carried 6 and 32 non-synonymous variants, respectively, which were greater than in the other Avr genes. Frameshift variant caused by an InDel was found in *AvrLm1* only, although previously reported also in *AvrLm11* [31] and *AvrLm4-7* [62]. Synonymous variants occurred in all Avr genes except *AvrLm1*, *AvrLmJ1-5-9* and *AvrLm6* (Fig. 2).

The majority of the SNPs identified in *L. maculans* were biallelic; however, there were also 236 triallelic DNA variants detected in this study (Table S3). Of these triallelic variants, 41 were located in genes, including a triallelic SNP (A/G/C) site in *AvrLm2* at 1,887,678 bp in SC6 (Fig. 3A) and a SNP (C/G/T) site in *AvrLm4-7* at position 1,374,707 bp in SC12 (Fig. 3B). The presence of the triallelic SNP detected by NGS was verified by Sanger-sequencing the *AvrLm4-7* gene from 3 isolates following PCR amplification and TOPO-TA cloning (Fig. S3).

Nonsense mutations might produce truncated proteins, so unlike other nonsynonymous SNPs, most nonsense mutations theoretically result in non-functional proteins. In this study at least 40 nonsense SNPs were identified in 32 genes, six of them located in *AvrLm4-7*, but none in any of the other Avr genes (Table S4, Fig. 3C). The presence of the nonsense SNP in *AvrLm4-7* was also confirmed by Sanger-sequencing (Fig. S3).

AvrLm11 was identified in a dispensable chromosome, and its presence/absence was previously reported [31]. However, the full length of *AvrLm11* was detected in all isolates sequenced in the current study despite of three SNPs. Similarly, a deletion was previously reported in *AvrLm4-7* when PCR was utilized [62], but no deletion was found in any of the isolates employed in this study.

SNP haplotypes in Avr genes and their corresponding protein isoforms

The number of SNP haplotypes varied with Avr genes. For ease of description, the absence and presence of *AvrLm1* were treated as two haplotypes. Two haplotypes were also found for *AvrLm6*, *AvrLmJ1-5-9*, *AvrLm10A* and *AvrLm10B*, and three haplotypes for *AvrLm2*, *AvrLm3* and *AvrLm11*, and nine SNP haplotypes for *AvrLm4-7* (Table 3). The number of SNP haplotypes was only partially dependent on the number of SNPs in each gene because of SNP linkage. For example, due to the lack of recombination among the first four SNPs, GCGT or TTAA in *AvrLm3*, nine SNPs in *AvrLm3* constituted only three haplotypes, the same number of haplotypes as that in *AvrLm2*, which only three SNPs were found. Concatenation of SNP haplotypes of Avr genes in individual isolates led to the identification of 48 haplotype groups (Table S5). The two most popular haplotype groups, SNP-1:A-2:1–3:2–47:2–6:1-J59:1-10A:1-10B:1–11:1 and SNP-1:A-2:1–3:1–47:1–6:1-J59:1-10A:1-10B:1–11:1, had similar frequency and were present in over 40% of the isolates (Table 4).

Table 3 SNP haplotypes and protein isoforms identified for *L. maculans* avirulence genes

	SNP haplotype names	codes	Frequency	SNP haplotypes	Protein isoform names	Amino changes
AvrLm1*	AvrLm1_presence	1:1	17.3%	Presence	Presence	
	AvrLm1_absence	1:2	82.1%	Absence	Absence	
AvrLm2	AvrLm2SNP_Haplo_1	2:1	81.4%	GGG	AvrLm2PRO_Haplo_1	WT
	AvrLm2SNP_Haplo_2	2:2	16.0%	CAC	AvrLm2PRO_Haplo_2	$G^{133}H,D^{146}Q$
	AvrLm2SNP_Haplo_3	2:3	2.6%	AAC	AvrLm2PRO_Haplo_3	$G^{133}N,D^{146}Q$
AvrLm3	AvrLm3SNP_Haplo_1	3:1	49.4%	GCGTCGTTC	AvrLm3PRO_Haplo_1	WT
	AvrLm3SNP_Haplo_2	3:2	37.2%	TTAAGATTT	AvrLm3PRO_Haplo_2	$S^{51}N,L^{78}F,I^{85}L,H$
	AvrLm3SNP_Haplo_3	3:3	13.5%	TTAACACGT	AvrLm3PRO_Haplo_3	$S^{51}N,I^{58}L,L^{78}F,I$
AvrLm4-7	AvrLm47SNP_Haplo_1	47:1	36.5%	Haplotype A	AvrLm47PRO_Haplo_1	$N^{41}K$
	AvrLm47SNP_Haplo_4	47:2	10.3%	Haplotype B	AvrLm47PRO_Haplo_4	WT
	AvrLm47SNP_Haplo_6	47:3	3.2%	Haplotype C	AvrLm47PRO_Haplo_6	$R^{32}S,N^{41}K$
	AvrLm47SNP_Haplo_3	47:4	12.2%	Haplotype D	AvrLm47PRO_Haplo_3	$N^{41}K,D^{86}N,G^{120}$
	AvrLm47SNP_Haplo_7	47:5	1.9%	Haplotype E	AvrLm47PRO_Haplo_7	$N^{41}K,G^{120}R$
	AvrLm47SNP_Haplo_2	47:6	23.1%	Haplotype F	AvrLm47PRO_Haplo_2	Amino change
	AvrLm47SNP_Haplo_5	47:7	10.3%	Haplotype G	AvrLm47PRO_Haplo_5	Amino change
	AvrLm47SNP_Haplo_8	47:8	1.3%	Haplotype H	AvrLm47PRO_Haplo_8	Amino change
	AvrLm47SNP_Haplo_9	47:9	1.3%	Haplotype I	AvrLm47PRO_Haplo_9	Amino change
AvrLm6	AvrLm6SNP_Haplo_1	6:1	98.1%	CT	AvrLm6PRO_Haplo_1	WT
	AvrLm6SNP_Haplo_2	6:2	1.9%	TC	AvrLm6PRO_Haplo_2	$M^{2}V,D^{126}K$
AvrLm5-9	AvrLmJ59SNP_Haplo_1	J59:1	92.3%	G	AvrLmJ59PRO_Haplo_1	WT
	AvrLmJ59SNP_Haplo_2	J59:2	7.7%	A	AvrLmJ59PRO_Haplo_2	$R^{55}K$
AvrLm10A	AvrLm10ASNP_Haplo_1	10A:1	83.3%	AA	AvrLm10APRO_Haplo_1	WT
	AvrLm10ASNP_Haplo_2	10A:2	16.7%	GC	AvrLm10APRO_Haplo_1	WT
AvrLm10B	AvrLm10BSNP_Haplo_1	10B:1	84.0%	GA	AvrLm10BPRO_Haplo_1	WT
	AvrLm10BSNP_Haplo_2	10B:2	16.0%	AG	AvrLm10BPRO_Haplo_1	WT
AvrLm11	AvrLm11SNP_Haplo_1	11:1	84.0%	TGT	AvrLm11PRO_Haplo_1	$D^{31}T$
	AvrLm11SNP_Haplo_2	11:2	9.6%	CGT	AvrLm11PRO_Haplo_2	$D^{31}T,K^{34}D$
	AvrLm11SNP_Haplo_3	11:3	6.4%	TTC	AvrLm11PRO_Haplo_3	WT
*	no SNP was identified for avirulence gene *Avr1*, for the sake of description, presence and absence of the gene was considered as 2 different ha					
**	ud denotes undetermined. AvrLm11 SNP haplotype association with avirulence/virulence was not evaluated for lack of an appropriate differen					
Haplotype A	CCGCGCCCGCCTCCGGCCCCGGCCCCCGGCCCCCGGCG					
Haplotype B	CCGCGCCCGCCTCCGGCCCCGGCCCCCGGCCCCGGGCG					
Haplotype C	CCGCGCCCGCCTCCGGCCCCGGCCCCCGGCCCCCGTCG					
Haplotype D	CCGCGCCCGCCTGCGGCTCCGGCCCCCGGCCCCCGGCG					
Haplotype E	CCGCGCCCGCCTGCGGCCCCGGCCCCCGGCCCCCGGCG					
Haplotype F	CCATATTTATTATTAACCTTAATCTCTAATTTTCATTA					
Haplotype G	TTATGTTTGTTATCGATCTTAGTTTTTAGTCTTCATTG					
Haplotype	CCATATTTATTATTAACCTTAATCTCTAGTCTTCATTA					

H	
Haplotype I	CCATATTTATTATTAATCTTAGCCTCTAACTTTCATCA
Amino change A	$D^{6}K,R^{32}S,Q^{35}X,N^{41}K,D^{43}K,D^{47}N,D^{57}N,Q^{60}X,Q^{63}X,W^{66}X,M^{70}I,C^{79}Y,D^{83}N,W^{85}X,S^{113}L,G^{116}R,G^{120}S,Q^{121}L,G^{124}S,H^{127}Y,D^{128}N,D^{130}K,W^{131}X,M^{135}I,R^{140}$
Amino change B	$D^{6}K,R^{32}S,Q^{35}X,N^{41}K,D^{43}K,D^{47}N,D^{57}N,Q^{63}X,W^{66}X,G^{67}S,M^{70}I,C^{79}Y,D^{83}N,W^{85}X,V^{87}I,G^{120}S,Q^{121}L,G^{124}S,D^{128}N,D^{130}K,W^{131}X,M^{135}I,R^{140}C,C^{142}Y,D^{143}N$
Amino change C	$D^{6}K,R^{32}S,Q^{35}X,N^{41}K,D^{43}K,D^{47}N,D^{57}N,Q^{63}X,W^{66}X,M^{70}I,C^{79}Y,D^{83}N,W^{85}X,S^{113}L,G^{116}R,G^{120}S,Q^{121}L,G^{124}S,H^{127}Y,D^{128}N,D^{130}K,W^{131}X,M^{135}I,R^{140}C$
Amino change D	$R^{32}S,Q^{35}X,N^{41}K,D^{43}K,D^{47}N,Q^{60}X,Q^{63}X,W^{66}X,M^{70}I,D^{83}N,W^{85}X,V^{87}I,S^{113}L,G^{116}R,G^{120}S,Q^{121}L,G^{124}S,H^{127}Y,D^{128}N,D^{130}K,W^{131}X,M^{135}I,R^{140}C$

Table 4
Prevalence of SNP haplotype groups and races of L. maculans

SNP haplotype	Frequency	Race structure*
SNP-1:A-2:1–3:2–47:2–6:1-J59:1-10A:1-10B:1–11:1**	22.1%	*Avr 2,3,5,6,9,10*
SNP-1:A-2:1–3:1–47:1–6:1-J59:1-10A:1-10B:1–11:1	20.1%	*Avr 2,4,5,6,7,10*
SNP-1:A-2:1–3:2–47:5–6:1-J59:1-10A:1-10B:1–11:1	9.1%	*Avr 2,3,5,6,9,10*
SNP-1:A-2:1–3:1–47:3–6:1-J59:1-10A:1-10B:1–11:1	5.8%	*Avr 2,5,6,7,10*
SNP-1:A-2:2–3:3–47:4–6:1-J59:1-10A:2-10B:2–11:1	4.6%	*Avr 4,5,6,7,10*
SNP-1:P-2:1–3:1–47:1–6:1-J59:1-10A:1-10B:1–11:1	3.9%	*Avr 1,2,4,5,6,7,10*
SNP-1:A-2:1–3:1–47:1–6:1-J59:1-10A:1-10B:1–11:3	2.0%	*Avr 2,4,5,6,7,10*
SNP-1:A-2:1–3:3–47:4–6:1-J59:1-10A:1-10B:1–11:2	2.0%	*Avr 2,4,5,6,7,10*
SNP-1:A-2:2–3:1–47:1–6:1-J59:1-10A:2-10B:2–11:2	2.0%	*Avr 4,5,6,7,10*

* *AvrLm11* protein function was undetermined due to the unavailability of a differential line carrying *Rlm11*.

** The naming convention for SNP haplotype groups is described in 'Materials and Methods' (Construction of haplotypes). Briefly, a haplotype groups is a concatenated haplotype of individual avirulence genes with separator '_'. For example, 'SNP-1:A_2:1_3:2_47:9_6:1_J59:1_10A:1_10B:1_11:1' indicates a haplotype group comprising of the following avirulence gene haplotypes, 1:A, 2:1, 3:2, 47:9 and so on, which stand for certain haplotypes detailed in Table 3.

Since a cDNA sequence determines that of its protein amino acids, protein sequence polymorphism was then examined, and a unique set of amino acid changes corresponding to an Avr gene SNP haplotype was referred to as a protein haplotype. For most of the Avr genes, a SNP haplotype translated into a unique protein haplotype, but for both *AvrLm10A* and *AvrLm10B*, the two SNPs did not cause any amino acid change and they corresponded to only one protein haplotype (Table 3). Similar to SNP haplotype group, protein haplotype of each Avr gene in an isolate were combined to form a protein haplotype group, and a total of 44 protein haplotype groups were identified (Table S6).

Avirulence gene frequency and race structure assessment

Since Avr gene SNP haplotypes potentially dictates protein function, we were tempted to utilize the haplotypes to assess the frequency of Avr genes in a *L. maculans* population. For most of the Avr genes, the link between SNP haplotypes and avirulence was determined by comparing SNP haplotypes and phenotypic ratings. Because of genetic code redundancy, different SNP haplotypes and protein haplotypes might point to the same functional protein. For instance, AvrLm2SNP_Haplo_2 and AvrLm2SNP_Haplo_3 both encoded virulent protein avr2 due to two amino acid changes, either $G^{133}H$ $D^{146}Q$ or G^{133}**N** $D^{146}Q$, both disrupted the protein function. Despite the fact that all three SNP haplotypes of *AvrLm3* were able to encode functional proteins, they were considered virulent unless *AvrLm7* was inactivated [62]. As for *AvrLm4-7*, all isolates carrying pre-mature STOP codons were observed to be double virulent (*avr4avr7*). *AvrLm10* was reported to require two genes, namely, *AvrLm10A* and *AvrLm10B* [63] to interact with the *Rlm10* in host plant. All isolates carried *AvrLm10* and shared the same coding sequences of *AvrLm10A* and *AvrLm10B* as isolate v23.1.3 (JN3). *AvrLm11* was not included in the assessment for the lack of host genetic sources to decide if the two amino acid changes, namely $D^{31}T$ and $K^{34}D$, affected its protein function. Based on SNP haplotype groups, 10 *L. maculans* races in the isolates were differentiated with incidence significantly different from each other (Fig. 4A). Two most prevalent races, *AvrLm2,3,5,6,9,10* and *AvrLm2,4,5,6,7,10*, collectively accounted for about 61% of the isolates. While *AvrLm5*, *AvrLm6* and *AvrLm10* were carried by all isolates, the other avirulence genes were found present in some isolates with frequency ranging from 17.5% (*AvrLm1*) to 64.3% (*AvrLm7*) (Fig. 4B).

Discussion

Monitoring changes in the race structure of *L. maculans* populations is important for selecting effective R genes for blackleg resistance breeding and management [3]. In Canada, this has been conducted in the prairies for almost 30 years, first based on pathogenicity group and later on the race structure according to *L. maculans* Avr profile [3, 32, 64]. These isolates provided the unique resources for this study to investigate the relationship between DNA variants and the function of Avr genes. It would be unnecessary and uneconomic to perform DNA re-sequencing on all the isolates phenotyped, only representative isolates were selected (Fig. 1, Table S2) to capture the genetic variation in Avr genes and explore their applications in pathogen race profiling.

What first caught our attention is that the SNP numbers considerably varied with avirulence genes, led by *AvrLm4-7* with 38 SNPs, followed by *AvrLm3* with 9 SNPs, but other avirulence genes only had 1 ~ 3 SNPs. It was believed that avirulence gene mutates in response to selection pressure imposed by resistant genes, so intuitively, *Rlm4* and *Rlm7*, the cognate resistant genes corresponding to *AvrLm4* and *AvrLm7* were supposed to be the most dominant resistant gene. It was found, however, that *Rlm1*, *Rlm2* and *Rlm3* were the most frequently used R genes in Canadian *B. napus* varieties [12]. *AvrLm1* was deleted in about 80% of isolates, but all isolates still carry functional *AvrLm3* genes despite of SNPs. Synonymous polymorphisms in human could affect messenger RNA splicing, stability, and structure as well as protein folding [65], therefore, the SNPs in *AvrLm3*, either synonymous or non-synonymous, might change *AvrLm3* protein transformation to avoid host recognition. On the other hand, varieties carrying *Rlm4* and *Rlm7* represented only about 2%, but *AvrLm4-7* mutates more frequently than *AvrLm2* and *AvrLm3*. Now that *AvrLm4-7* interacts with *AvrLm3* [25], it is reasonable to ask if the interaction serves as an advanced mechanism for *L. maculans* to overcome host defense system.

Although some SNPs were chosen for *L. maculans* avirulence/virulence diagnostic test [20, 38], there is still an understanding gap between SNP patterns and *L. maculans* pathogenicity. Investigation of highly dense DNA variants in the pathogen isolates offers a panoramic view of the DNA variant profile, which may help better discern the pathogen's strategy to cope with biological and environmental changes at the DNA level. In the current study, about 21,000 SNP sites were identified among 158 *L. maculans* isolates. We first examined variant composition and distribution among genomic regions, including AT-blocks, GC-blocks, and coding and non-coding regions. Ts/Tv is considered as an important parameter in evolutionary genetics [66]. Theoretically transversions are much less common than transition mutations, because the generation of transversions during replication requires much greater distortion of the double DNA helix than transition mutations [67], for this reason, nucleotide transitions are favored several fold over transversions, which was suggested to be a result of selection [65]. Relatively lower Ts/Tv ratios were reported in the ranges of 1.21 ~ 2.46 and 0.75 ~ 1.83 for some bacteria and unicellular eukaryotes respectively [68], but higher for fungi, for example, a ratio of 5 for rice blast fungus, *Magnaporthe oryzae* [69]; however, all these ratios were calculated at whole genome level. With great interest, we computed and compared the ratios across different genetic regions. The Ts/Tv of the whole *L. maculans* genome was found to be 3.1 for *L. maculans*, less than 4 as previously reported [23]. In addition, Ts/Tv changed dramatically between AT block and GC blocks. Ts/Tv in AT-block was 3 times higher than that in GC blocks and genes, suggesting that transversion in GC blocks and genes occurs more frequent than AT block. The results suggested that GC blocks and gene regions contribute more to the pathogen evolution, because transversion could impose greater impact on functional regulatory element activity [70].

As mentioned above, monitoring the changes in *L. maculans* race structure plays a pivotal role in blackleg management. The conventional approach involves multiple steps including isolate collection, bioassays on differential hosts and disease severity rating. The approach has served its purpose for many years, but obviously with room for improvement on efficiency, accuracy and clarity. Use of SNP genotyping can be more cost-effective, with a high-throughput process amenable to automation. However, SNP genotyping approach has its limitations. Firstly, it relies on a single SNP to determine Avr gene functionality, which can be true if the SNP chosen is the sole nucleotide associated with the compatible/incompatible interaction. This might not be always the case because such a SNP was often selected from a limited number of isolates or populations, and some undetected SNPs or variants in other populations might compromise the avirulence gene function. Secondly, pathogens evolve because of host-pathogen interactions, and mutation in the pathogen occurs frequently for survival under selection pressure. While a SNP can be selected empirically to detect loss-of-function of an Avr gene, it cannot be ruled out that other SNPs within the gene region that may disrupt the Avr gene despite the presence of the SNP normally would predict the avirulence. For instance, none of SNPs we detected in *AvrLm3* interferes with the gene function, but the functionality of *AvrLm3* is actually dictated by SNPs in *AvrLm4-7* [62]. Because of the masking effect of *AvrLm7* on *AvrLm3* [25], *AvrLm7* always disrupts *AvrLm3* regardless of *AvrLm3* gene sequence. Thirdly, SNP genotyping assumes that the site of interest is biallelic, so commercial SNP genotyping chemistry is designed generally to interrogate biallelic SNPs. While biallelic SNPs have been reported to be the majority of polymorphic sites, triallelic SNPs exist in *L. maculans*, and more importantly, a biallelic site will probably turn out to be multiallelic when more populations or individuals are tested. We found approximately 2% of total SNPs in *L. maculans* were triallelic in this study. For *AvrLm4-7*, the SNPs at SC12-1374707 are C/G/T (Fig. 5B). Empirically and coincidentally, this SNP site was selected as a biallelic marker (C/G) to determine whether *AvrLm4* is dysfunctional [20, 26, 32]. Consequently, the SNP assay based on the biallelic assumption was not able to interrogate the third SNP allele T, which was present in some isolates like MT07-35 (Fig. 3B), leading to missing or even false calls. Similarly, the issue will arise for other triallelic sites, in *AvrLmJ1-5-9* (SNP A/C/G at the 164th nucleotide) [27], and *AvrLm2* (SNP A/C/G at the 397th nucleotide)[24]. Fourthly, pre-mature STOP codons raise an issue for SNP genotyping. Six pre-mature STOP codons were identified in *AvrLm4-7* in this study. Pre-mature STOP codons are usually associated with gene loss-of-function (Fig. 4). Despite the observation that these nonsense SNPs in *AvrLm4-7* were linked with the SNP T at SC12_1374707, they may co-exist with the SNP C or G at the same site, causing double virulence of *avr4avr7*, defying the SNP genotyping results of *Avr4Avr7* or avr4*Avr7*. Therefore, without prior information, a diagnostic method based only on the site SC12_1374707 could be inaccurate or erroneous. Taken all together, a single SNP appears inadequate for genotyping *L. maculans* Avr genes.

Some DNA mutations in Avr genes are shared by *L. maculans* isolates from different continents. For example, *AvrLm1* deletion was found in France [71], Australia [72] and this study, and SNP G/C at SC12_1374707 discovered in this study was previously identified in France [26]. Some mutations, however, only found in one continent. For example, a $K^{55}T$ and a pre-mature STOP codon ($R^{29}X$) were previously detected in *AvrLmJ1-5-9* in Australia [27], but not reported in Canadian isolates. It is also noteworthy that the deletion of *AvrLm4-7* was detected in 516 of 845 European isolates [62], but in none of the Canadian isolates examined in this study. Partial or complete deletion of *AvrLm4-7* was also suggested to be responsible for the double virulence phenotype of *avr4avr7* [26]. An attempt to investigate the InDel location and size in *AvrLm4* failed to find the forward primer (TATCGCATACCAAACATTAGGC) in either masked or unmasked assembly of scaffords available in the *L. maculans* genome. We used the forward primer sequence as a query to nucleotide-BLAST the *L. maculans* genome deposited in the NCBI gene bank, and retrieved the genomic region of accession GenBank: AM998635.1 surrounding the Avr gene *AvrLm4-7* as a hit. The forward primer sequences were located in SC12 between the 1375938th and 1375960th bases, about 1,350 base downstream of the start codon of *AvrLm4-7*, which happened to be in a gap identified between the 1,375,400th and 1,376,400th in all isolates examined (data not shown), however the whole DNA sequence for *AvrLm4-7* remained complete and intact. Because of the above-mentioned gap between the 1,375,400th and 1,376,400th bases, if the two

primers were used to amplify *AvrLm4-7* in Canadian prairie isolates, wrong genotyping result of deletion could be expected for some of isolates. This indicates that *L. maculans* isolates of different geographical regions mutate in different ways, suggestive of different evolutionary pathways.

With reduced cost of NGS and broader use of target amplicon sequencing for studying microbe genomes [73], it may be possible to use sequencing and haplotypes as a more reliable metric to characterize Avr profile and race structure of *L. maculans*. Statistical analysis based on haplotypes may often be more efficient than analyses of individual markers through an empirical process [74] or simulation studies [75], because haplotype analyses take tightly linked markers into account, providing much more information than individual markers do [59]. SNP haplotype has been applied mostly in identifying genomic polymorphism and other genetic studies, such as the work on the honeybee pathogen *Nosema ceranae* [76]. It is well-documented that the methods for SNP haplotype inference require family and population information [77] either for unrelated [54] or related individuals based on exact-likelihood [78], approximate-likelihood [79] computations, or rule-based strategies [80]. Although the SNP haplotype construction algorithms are intended to identify co-segregating SNPs and then establish reliable genotype-phenotype connection, they are essentially a family-based analysis and the haplotypes generated from one pedigree might not be extrapolated to other populations. These methods were developed for diploid or polyploid organisms, and none of these methods was considered immaculate without limitations. Furthermore, this study dealt with haploid *L. maculans*, and aimed at establishing most accurate association of avirulence/virulence with DNA mutations in Avr genes. Therefore, in this study all discovered SNPs were taken into account, even synonymous SNPs that could compromise protein functions to some extent [65]. This simplified SNP construction strategy can accommodate any SNPs, already-identified or newly-emerged in *L. maculans* isolates collected from any fields or populations, and alleviates concerns about the accuracy or inadvertent ignorance of SNPs resulting from any complicated SNP construction algorithm.

In conclusion, there are at least three advantages with genotyping-by-haplotyping when compared to other methods of analysis for *L. maculans* Avr profile or race structure, including SNP genotyping. Firstly, it considers all DNA variants in an Avr gene. Secondly, it can be readily translated into protein isoforms. Thirdly, it is able to capture the emergence of new SNPs in any pathogen populations. Therefore, we propose genotyping-by-haplotyping as a new method for large-scale *L. maculans* Avr profile analysis (Fig. 5). All existing SNP haplotypes with connection with Avr gene functions can be categorized, indexed, and stored in a database for inquiries. Any new haplotype will be added to the database once its relationship with the avirulence/virulence has been determined through the conventional phenotyping process. To this end, we are in the process of developing a new strategy for target sequencing of *L. maculans* Avr genes to improve the reliability of *L. maculans* Avr profiling while reducing the cost of genotyping-by-haplotyping procedure.

Conclusion

In this study, identification of genome-wide DNA variants and SNP haplotypes associated with avirulence genes of *Leptosphaeria maculans* were performed with 158 isolates selected from 1590 isolates originating from western Canada. There were 21,016 polymorphic variants identified in the isolates. Forty eight SNP haplotype groups were discovered and linked with different avirulence gene functionality. Being more informative and accurate than SNP genotyping, SNP haplotyping was hence proposed to be a more reliable and informative strategy for large-scale survey of *L. maculans* race structure.

Materials And Methods

Selection of *L. maculans* isolates

To monitor changes in race structure of blackleg pathogen in western Canada, 800 *L. maculans* isolates were collected from trap plots of the susceptible cultivar 'Westar' at nine sites: Brandon, Carberry and Plum Coulee in Manitoba, Scott, Melfort and Indian Head in Saskatchewan, Vermilion, Camrose and Vegreville in Alberta during 2007–2008. Additional 790 isolates were from commercial canola field collected between 2012 and 2014. A set of 14 cultivars or breeding lines of *Brassica* spp. (Table S1) were used as differential hosts to determine the presence/absence of Avr genes following established methodology [3]. Phenotypic similarity of the 1590 isolates was determined using Euclidean distance of cluster analysis in the R package [681], and 162 isolates were selected from major branches of the resulting dendrogram for further analyses.

Extraction of fungal DNA

The selected isolates were cultured in 50 ml sterile centrifuge tubes containing 10 ml liquid V8 medium (800 ml of distilled water, 200 ml of V8 juice, 0.7 g of calcium carbonate, and 100 mg of streptomycin sulfate) in an incubator (Lab Line Orbit Environment shaker) at 20^{o}C and 300 rpm for 7 days when a mycelial ball of *L. maculans* had formed in each tube. The mycelial balls were collected and dried in a Labconco Benchtop Freeze Dryer (Labconco, Kansas City, MO), and then employed for DNA extraction using the EZNA Fungal DNA mini kit (Omega Bio-Tech, Norcross, GA) following the manufacturer's instructions.

Preparation of DNA libraries and DNA sequencing

DNA samples of each isolate were quantified using the Quant-iT™ PicoGreen® dsDNA Assay kit (Thermo Fisher Scientific, Waltham, MA). Libraries were generated using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England BioLabs, Whitby, ON) following the manufacturer's instructions. Illumina adapters containing barcodes were purchased from PerkinElmer (Waltham, MA). Size selection of libraries at 350–450 bp was performed using SPRIselect beads (Beckman Coulter, Brea, CA) prior to PCR amplification (7 cycles). Libraries were quantified using the Kapa Illumina GA with Revised Primers-SYBR Fast Universal kit (Kapa Biosystems, Wilmington, MA). The average size of the fragments was determined using a LabChip GX (PerkinElmer, Waltham, MA) instrument. The libraries were normalized, denatured in 0.05 N NaOH and then diluted to 200 pM, and neutralized using HT1 buffer. Clustering was performed on an Illumina cBot system and the flow cell was run on a HiSeq 4000 for 2 × 100 cycles (paired-end mode) following the manufacturer's instructions. A phiX library was mixed with libraries at 1% as control. The Illumina control software was HCS HD 3.4.0.38, the real-time analysis program was RTA v. 2.7.7. The program bcl2fastq v2.18 was used to demultiplex samples and generate fastq reads.

Sequence alignment and SNP discovery

The NGS data was analyzed using the software package Lasergene Genomics Suite 13 including SeqMan NGen, SeqMan Pro and QSeq (DNASTAR, Madison, WI). Short reads were first assembled with the haploid option to the genome sequence of *L. maculans* 'brassicae' isolate v23.1.3 [15] using SeqMan NGen. The sequence of wa74_scaffold00338 (GenBank: FO906748.1) containing the gene *AvrLm3*, which was previously reported to be absent from the aforementioned reference genome [25], was included as the reference template to detect variants in *AvrLm3* and its flanking regions. Variants (SNPs and InDels) relative to the *L. maculans* reference isolate were determined using QSEq. Variant tables of isolates were exported, filtered (MAF ≥ 5, depth ≥ 10), and combined for further SNP data mining. The SNP distribution in specific genomic regions and variant composition was investigated. To explore variant density in different blocks, AT- and GC- block segments in supercontigs (SC) were first determined by Excel VBA programming with sliding windows of 120 nucleotides. In this study, any continuous nucleotide stretch meeting the criteria of GC content < 33% [15] and length > 2000 base pairs was considered an AT-block. All “N” masked genome regions were excluded from the AT-block assessment.

The confirmation of a triallelic SNP of SC12_1374707 was confirmed by Sanger sequencing. Briefly, *AvrLm4-7* whole gene fragments were amplified using a forward primer CTCACCTCCGTATCTTTAGTCGCA and a reverse primer CAGTTAACAACATGCCACTATCCCT, and cloned into the vector PCR2.1-TOPO using the Invitrogen TOPO™ TA cloning kit. The inserts were Sanger-sequenced using regular M13 forward and reverse primers. Sequence profiles were imported into BioEdit for quality examination and sequence alignment among isolates.

Construction of haplotypes

SNP haplotype could be defined as SNP groups inherited together because of genetic linkage or their haploid nature. Pycnidiospores and mycelia of *L. maculans* are haploid fungal propagules so SNPs within a gene of an isolate are always inherited together. Considering that synonymous SNPs might impact protein function by altering protein structure [82], in this study all SNPs in each avirulence gene, both synonymous and non-synonymous were concatenated in the order of their position on supercontigs to form haplotypes. Each SNP haplotype was given a name starting with the gene name (*AvrLm1*, *AvrLm2*, etc.) followed by “SNP_Haplo” plus a serial number; amino acid changes as the result of haplotypes, were referred to as protein haplotypes in a similar fashion [65], but with “SNP” replaced by “PRO”. For ease of description, each SNP haplotype for an individual Avr gene was given a unique code, following the naming convention of ‘gene serial number:haplotype serial number’. A gene serial number refers to an Avr gene, and a haplotype serial number to a SNP haplotype based on their phenotype (virulent or avirulent) and prevalence. For example, the three haplotypes of *AvrLm3* were assigned the codes 3:1, 3:2 and 3:3, whereas SNP or protein haplotype groups were subsequently constructed by concatenating SNP or protein haplotype codes. The functionality of a protein with a specific amino sequence translated from a SNP-incorporated cDNA were assessed based mainly on our phenotypic data along with the information in literatures [25, 31].

Abbreviations

Avr: avirulence;InDel: insertion and deletion; MAF: minimum allele frequency; MC: median coverage; NGS: Next Generation Sequencing ; PCR: Polymerase chain reaction;RAPD: random amplified polymorphic DNAs (RAPD); RFLP: restriction fragment length polymorphisms; SC: Supercontig; SNP: Single-nucleotide polymorphism; SSRs: Simple sequence repeats ; TC: template coverage; Ts: transition; Tv: transversion

Declarations

Ethics approval and consent to participate

Not applicable

Consent to publish

Not applicable

Availability of data and materials

The datasets used and/or analyzed during the current study available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interest.

Funding

This work was funded by a competitive grant from Agriculture and Agri-Food Canada. The funding body played no role in the design of the study and collection, analysis, and

Author contributions

FY conceived of and designed the study; QC conducted the experiments, analyzed data and drafted the manuscript; GP and RK provided important information and materials. All authors reviewed the manuscript and approved the final draft.

Acknowledgement

We would like to thank Agriculture and Agri-Food Canada for supporting this research.

References

1. Fitt BD, Brun H, Barbetti M, Rimmer S: **World-wide importance of phoma stem canker** (Leptosphaeria maculans **and** L. biglobosa) **on oilseed rape (**Brassica napus**). In**: **Sustainable strategies for managing** Brassica napus (**oilseed rape**) **resistance to** Leptosphaeria maculans (**phoma stem canker**). Springer; 2006: 3–15.
2. Smith E, Kutcher H, Brandt S, Ulrich D, Malhi S. Johnston: **The profitability of short-duration canola and pea rotations in western Canada**. Can J Plant Sci. 2013;93:933–40.
3. Kutcher H, Balesdent M, Rimmer S, Rouxel T, Chevre A, Delourme R, Brun HJ. **Frequency of avirulence genes in** *Leptosphaeria maculans* **in western Canada**. Can J Plant Path. 2010;32:77–85.
4. Kutcher H, Fernando W, Turkington T, McLaren D. Best management practices for blackleg disease of canola. Prairie Soils Crops. 2011;4:122–34.
5. Fernando W: **Managing Blackleg Resistance Breakdown and Trade Barriers through Blackleg Resistance Stewardship in Canola**. In: MB Agronomists Conference, December, University of Manitoba: 2010.
6. Peng G, Liu X, Mclaren DL, Mcgregor L, Yu FJ. **Seed treatment with the fungicide fluopyram limits cotyledon infection by** *Leptosphaeria maculans* **and reduces blackleg of canola**. *Canadian Journal of Plant Pathology* 2020(just-accepted).
7. Gout L, Eckert M, Rouxel T, Balesdent M-H. **Genetic variability and distribution of mating type alleles in field populations of** *Leptosphaeria maculans* **from France**. Appl Environ Microbiol. 2006;72:185–91.
8. Daverdin G, Rouxel T, Gout L, Aubertot J-N, Fudal I, Meyer M, Parlange F, Carpezat J, Balesdent M-H. Genome structure and reproductive behaviour influence the evolutionary potential of a fungal phytopathogen. PLoS Pathog. 2012;8:e1003020.
9. McDonald BA, Linde C. The population genetics of plant pathogens and breeding strategies for durable resistance. Euphytica. 2002;124:163–80.
10. Rouxel T, Balesdent M. **The stem canker (blackleg) fungus**, *Leptosphaeria maculans*, **enters the genomic era**. Molecular Plant Pathology. 2005;6:225–210.
11. Li H, Sivasithamparam K, Barbetti M. **Breakdown of a** *Brassica rapa* **subsp**. *sylvestris* **single dominant blackleg resistance gene in** *B. napus* **rapeseed by** *Leptosphaeria maculans* **field isolates in Australia**. Plant Dis. 2003;87:752.
12. Zhang DD, Wang XY, Chen JY, Kong ZQ, Gui YJ, Li NY, Bao YM, Dai XF. **Identification and characterization of a pathogenicity-related gene VdCYP1 from** *Verticillium dahliae*. Scientific reports. 2016;6:1–2.
13. Yu F, Gugel RK, Kutcher HR, Peng G, Rimmer SR: **Identification and mapping of a novel blackleg resistance locus LepR4 in the progenies from** Brassica napus × B. rapa **subsp**. sylvestris. *Theoretical and Applied Genetics*. 2013, 126:307–315.
14. Yu F, Lydiate DJ, Gugel R, Sharpe A, Rimmer S. **Introgression of** *Brassica rapa* **subsp**. *sylvestris* **blackleg resistance into** *B. napus*. Molecular breeding. 2012;30:1495–506.
15. Rouxel T, Grandaubert J, Hane JK, Hoede C, Van de Wouw AP, Couloux A, Dominguez V, Anthouard V, Bally P, Bourras S. **Effector diversification within compartments of the** *Leptosphaeria maculans* **genome affected by Repeat-Induced Point mutations**. Nature communications. 2011;2:1–10.
16. Raffaele S, Kamoun S. Genome evolution in filamentous plant pathogens: why bigger can be better. Nat Rev Microbiol. 2012;10:417–30.
17. Soyer JL, El Ghalid M, Glaser N, Ollivier B, Linglin J, Grandaubert J, Balesdent M-H, Connolly LR, Freitag M, Rouxel T. **Epigenetic control of effector gene expression in the plant pathogenic fungus** *Leptosphaeria maculans*. PLoS Genet. 2014;10:e1004227.
18. Sánchez-Vallet A, Fouché S, Fudal I, Hartmann FE, Soyer JL, Tellier A, Croll D. The Genome Biology of Effector Gene Evolution in Filamentous Plant Pathogens. Annu Rev Phytopathol. 2018;56:21–40.
19. Purwantara A, Barrins JM, Cozijnsen AJ, Ades PK, Howle BJ. **Genetic diversity of isolates of the** *Leptosphaeria maculans* **species complex from Australia, Europe and North America using amplified fragment length polymorphism analysis**. Mycol Res. 2000;104:772–81.
20. Zou Z, Liu F, Fernando WD. **Rapid detection of** *Leptosphaeria maculans* **avirulence gene** *AvrLm4-7* **conferring the avirulence/virulence specificity on** *Brassica napus* **using a tetra-primer ARMS-PCR**. Eur J Plant Pathol. 2018;152:515–20.
21. Upson JL, Zess EK, Białas A, Wu C-H, Kamoun S. The coming of age of EvoMPMI: evolutionary molecular plant–microbe interactions across multiple timescales. Curr Opin Plant Biol. 2018;44:108–16.
22. Bousset L, Sprague SJ, Thrall PH, Barrett LG. **Spatio-temporal connectivity and host resistance influence evolutionary and epidemiological dynamics of the canola pathogen** *Leptosphaeria maculans*. Evol Appl. 2018;11:1354–70.
23. Zander M, Patel DA, Van de Wouw A, Lai K, Lorenc MT, Campbell E, Hayward A, Edwards D, Raman H, Batley J. **Identifying genetic diversity of avirulence genes in** *Leptosphaeria maculans* **using whole genome sequencing**. Funct Integr Genom. 2013;13:295–308.
24. Ghanbarnia K, Fudal I, Larkan NJ, Links MG, Balesdent MH, Profotova B, Fernando WD, Rouxel T, Borhan MH: **Rapid identification of the** Leptosphaeria maculans **avirulence gene** AvrLm2 **using an intraspecific comparative genomics approach**. *Molecular plant pathology* 2015, 16:699–709.

25. Plissonneau C, Blaise F, Ollivier B, Leflon M, Carpezat J, Rouxel T, Balesdent MH. **Unusual evolutionary mechanisms to escape effector-triggered immunity in the fungal phytopathogen** *Leptosphaeria maculans*. Molecular ecology. 2017;26:2183–98.
26. Parlange F, Daverdin G, Fudal I, Kuhn ML, Balesdent MH, Blaise F, Grezes-Besset B, Rouxel T. *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. 2009;71:851–63.
27. Van de Wouw AP, Cozijnsen AJ, Hane JK, Brunner PC, McDonald BA, Oliver RP, Howlett BJ. **Evolution of linked avirulence effectors in** *Leptosphaeria maculans* **is affected by genomic environment and exposure to resistance genes in host plants**. PLoS pathogens. 2010;6:e1001180.
28. Plissonneau C, Blaise F, Ollivier B, Leflon M, Carpezat J, Rouxel T, Balesdent MH. **Unusual evolutionary mechanisms to escape effector-triggered immunity in the fungal phytopathogen** *Leptosphaeria maculans*. Molecular ecology. 2017;26:2183–98.
29. Fudal I, Ross S, Gout L, Blaise F, Kuhn M, Eckert M, Cattolico L, Bernard-Samain S, Balesdent M, Rouxel T. **Heterochromatin-like regions as ecological niches for avirulence genes in the** *Leptosphaeria maculans* **genome: map-based cloning of** *AvrLm6*. Mol Plant Microbe Interact. 2007;20:459–70.
30. Petit-Houdenot Y, Degrave A, Meyer M, Blaise F, Ollivier B, Marais CL, Jauneau A, Audran C, Rivas S, Veneault-Fourrey CJ. **A two genes–for–one gene interaction between** *Leptosphaeria maculans* **and** *Brassica napus*. New Phytol. 2019;223:397–411.
31. Balesdent MH, Fudal I, Ollivier B, Bally P, Grandaubert J, Eber F, Chèvre AM, Leflon M, Rouxel T. **The dispensable chromosome of** *Leptosphaeria maculans* **shelters an effector gene conferring avirulence towards** *Brassica rapa*. New Phytol. 2013;198(3):887–98.
32. Liban S, Cross D, Kutcher H, Peng G, Fernando W. **Race structure and frequency of avirulence genes in the western Canadian** *Leptosphaeria maculans* **pathogen population, the causal agent of blackleg in brassica species**. Plant Pathol. 2016;65:1161–9.
33. Flor HH. Current status of the gene-for-gene concept. Annual review of phytopathology. 1971;9:275–96.
34. Vanderplank JE: **Genetic and molecular basis of plant pathogenesis**. Springer Science & Business Media. 2012, Dec 6.
35. Goodwin P, Annis S. **Rapid identification of genetic variation and pathotype of** *Leptosphaeria maculans* **by random amplified polymorphic DNA assay**. Appl Environ Microbiol. 1991;57:2482–6.
36. Schleier S, Voigt K, Wostemeyer J. **RAPD-based molecular diagnosis of mixed fungal infections on oilseed rape (***Brassica napus***): evidence for genus-and species-specific sequences in the fungal genomes**. J Phytopathol. 1997;145:81–7.
37. Voigt K, Schleier S, Wostemeyer J. **RAPD-based molecular probes for the blackleg fungus** *Leptosphaeria maculans* **(phoma lingam): evidence for pathogenicity group-specific sequences in the fungal genomes**. J Phytopathol. 1998;146:567–76.
38. Kaczmarek J, Latunde-Dada AO, Irzykowski W, Cools HJ, Stonard JF, Brachaczek A, Jedryczka M. **Molecular screening for avirulence alleles** *AvrLm1* **and** *AvrLm6* **in airborne inoculum of** *Leptosphaeria maculans* **and winter oilseed rape (Brassica napus) plants from Poland and the UK**. J Appl Genet. 2014;55:529–39.
39. Brown JJ. The choice of molecular marker methods for population genetic studies of plant pathogens. New Phytol. 1996;133:183–95.
40. Chambers K, Lowe RG, Howlett BJ, Zander M, Batley J, Van de Wouw AP, Elliott CE. **Next-generation genome sequencing can be used to rapidly characterise sequences flanking T-DNA insertions in random insertional mutants of** *Leptosphaeria maculans*. Fungal biology biotechnology. 2014;1:10.
41. Golicz AA, Martinez PA, Zander M, Patel DA, Van De Wouw AP, Visendi P, Fitzgerald TL, Edwards D, Batley J. **Gene loss in the fungal canola pathogen** *Leptosphaeria maculans*. Funct Integr Genom. 2015;15:189–96.
42. Raman R, Taylor B, Lindbeck K, Coombes N, Barbulescu D, Salisbury P, Raman HJC, Science P. **Molecular mapping and validation of** *Rlm1* **gene for resistance to** *Leptosphaeria maculans* **in canola (***Brassica napus* **L.)**. 2013, 63:1007–1017.
43. Zander M, Patel DA, Van de Wouw A, Lai K, Lorenc MT, Campbell E, Hayward A, Edwards D, Raman H, Batley J. **Identifying genetic diversity of avirulence genes in** *Leptosphaeria maculans* **using whole genome sequencing**. Funct Integr Genom. 2013;13:295–308.
44. Qian L, Hickey LT, Stahl A, Werner CR, Hayes B, Snowdon RJ, Voss-Fels KP. Exploring and harnessing haplotype diversity to improve yield stability in crops. Frontiers in plant science. 2017;8:1534. doi:10.3389/fpls.2017.01534.
45. Ching AD, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet. 2002. doi.org/10.1186/1471-2156-3-19.
46. Judson R, Stephens JC, Windemuth A. The predictive power of haplotypes in clinical response. Pharmacogenomics. 2000;1:15–26.
47. Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, Kautzer CR, Lee DH, Marjoribanks C, McDonough DP. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science. 2001;294:1719–23.
48. Bernardo R. **Breeding for quantitative traits in plants**. Woodbury, MN: Stemma press. 2002. p. 369.
49. Cavalli-Sforza L, Bodmer WJSF. Human population genetics. San Francisco: Free-man; 1971.
50. Kwok P-Y, Gu Z. **Single nucleotide polymorphism libraries: why and how are we building them**? Molecular medicine today. 1999;5:538–43.
51. Hartl DL, Clark AG. Principles of Population Genetics. 3rd ed. Sunderland: Sinauer Associates Inc.; 1997.
52. Clark AG. Inference of haplotypes from PCR-amplified samples of diploid populations. Mol Biol Evol. 1990;7:111–22.
53. Excoffier L, Slatkin M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Mol Biol Evol. 1995;12:921–7.
54. Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, Stanley SE, Jiang R, Messer CJ, Chew A, Han J-H. Haplotype variation and linkage disequilibrium in 313 human genes. Science. 2001;293:489–93.
55. De Bakker PI, McVean G, Sabeti PC, Miretti MM, Green T, Marchini J, Ke X, Monsuur AJ, Whittaker P, Delgado MJ. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. Nat Genet. 2006;38:1166–72.

56. Lorenz AJ, Hamblin MT, Jannink J-LJPo. **Performance of single nucleotide polymorphisms versus haplotypes for genome-wide association analysis in barley**. 2010 5:e14079. doi:10.1371/journal.pone.0014079.
57. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart MJ. The structure of haplotype blocks in the human genome. Science. 2002;296:2225–9.
58. Lu Y, Shah T, Hao Z, Taba S, Zhang S, Gao S, Liu J, Cao M, Wang J, Prakash AB. Comparative SNP and haplotype analysis reveals a higher genetic diversity and rapider LD decay in tropical than temperate germplasm in maize. PloS one. 2011;6:e24861. doi.org/10.1371/journal.pone.0024861.
59. Simko I. One potato, two potato: haplotype association mapping in autotetraploids. Trends in plant science. 2004;9:441–8.
60. Fang DD, Xiao J, Canci PC, Cantrell RG. **A new SNP haplotype associated with blue disease resistance gene in cotton (***Gossypium hirsutum L.***)**. Theor Appl Genet. 2010;120:943–53.
61. Bassi FM, Bentley AR, Charmet G, Ortiz R, Crossa J. **Breeding schemes for the implementation of genomic selection in wheat (***Triticum* **spp.)**. Plant Sci. 2016;242:23–36.
62. Plissonneau C, Daverdin G, Ollivier B, Blaise F, Degrave A, Fudal I, Rouxel T, Balesdent MH. **A game of hide and seek between avirulence genes** *AvrLm4***-7 and** *AvrLm3* **in** *Leptosphaeria maculans*. New Phytol. 2016;209:1613–24.
63. Petit-Houdenot Y, Degrave A, Meyer M, Blaise F, Ollivier B, Marais CL, Jauneau A, Audran C, Rivas S, Veneault-Fourrey C, Brun H. **A two genes–for–one gene interaction between** *Leptosphaeria maculans* **and** *Brassica napus*. New Phytol. 2019;223:397–411.
64. Kutcher H, Van Den Berg C, Rimmer SJ. **Variation in pathogenicity of** *Leptosphaeria maculans* **on** *Brassica* **spp. based on cotyledon and stem reactions**. Can J Plant Path. 1993;15:253–8.
65. Spooner W, McLaren W, Slidel T, Finch DK, Butler R, Campbell J, Eghobamien L, Rider D, Kiefer CM, Robinson MJ, Hardman C. Haplosaurus computes protein haplotypes for use in precision drug design. Nature communications. 2018;9:1–12.
66. Stoltzfus A, Norris RW. On the causes of evolutionary transition: transversion bias. Mol Biol Evol. 2016;33:595–602.
67. Beamish C. **Transversion Mutation**. 2001. doi.org/10.1006/rwgn.2001.1335.
68. Long H, Behringer MG, Williams E, Te R, Lynch M. Similar mutation rates but highly diverse mutation spectra in ascomycete and basidiomycete yeasts. Genome Biology Evolution. 2016;8:3815–21.
69. 10.1371/journal.pone.0065416
Jeon J, Choi J, Lee G-W, Dean RA, Lee Y-H: **Experimental evolution reveals genome-wide spectrum and dynamics of mutations in the rice blast fungus**, Magnaporthe oryzae. *PloS one* 2013, 8: e65416. doi.org/10.1371/journal.pone.0065416
70. Guo C, McDowell IC, Nodzenski M, Scholtens DM, Allen AS, Lowe WL, Reddy TE. **Transversions have larger regulatory effects than transitions**. *BMC Genomics* 2017 18:394. doi.org/10.1186/s12864-017-3785-4.
71. Gout L, Fudal I, Kuhn ML, Blaise F, Eckert M, Cattolico L, Balesdent MH, Rouxel T. **Lost in the middle of nowhere: the** *AvrLm1* **avirulence gene of the Dothideomycete** *Leptosphaeria maculans*. Mol Microbiol. 2006;60:67–80.
72. Van de Wouw AP, Lowe RG, Elliott CE, Dubois DJ, Howlett BJ. **An avirulence gene, AvrLmJ1, from the blackleg fungus**, *Leptosphaeria maculans*, **confers avirulence to** *Brassica juncea* **cultivars**. Mol Plant Pathol. 2014;15:523–30.
73. Armanhi JSL, De Souza RSC, De Araújo LM, Okura VK, Mieczkowski P, Imperial J, Arruda PJSr. Multiplex amplicon sequencing for microbe identification in community-based culture collections. Scientific reports. 2016;6:29543. doi.org/10.1016/j.pbi.2017.04.018.
74. Drysdale CM, McGraw DW, Stack CB, Stephens JC, Judson RS, Nandabalan K, Arnold K, Ruano G, Liggett SB: **Complex promoter and coding region β2-adrenergic receptor haplotypes alter receptor expression and predict in vivo responsiveness**. *Proceedings of the National Academy of Sciences* 2000, 97:10483–10488.
75. Zhang F, Boerwinkle E, Xiong M. Epistasis analysis for quantitative traits by functional regression model. Genome research. 2014;2:989–98.
76. Peters MJ, Suwannapong G, Pelin A, Corradi N. **Genetic and Genome Analyses Reveal Genetically Distinct Populations of the Bee Pathogen** *Nosema ceranae* **from Thailand**. Microb Ecol. 2019;77:877–89.
77. Graça A, Lynce I, Marques-Silva J, Oliveira A. **Haplotype inference combining pedigrees and unrelated individuals**. In: *Workshop on Constraint Based Methods for Bioinformatics (WCB 2009)*: 2009: 27–36.
78. Lander ES, Green P: **Construction of multilocus genetic linkage maps in humans**. *Proceedings of the National Academy of Sciences*. 1987, 84:2363–2367.
79. Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet. 1996;58:1347–63.
80. Li J, Jiang T: **Efficient rule-based haplotyping algorithms for pedigree data**. In: *Proceedings of The Seventh Annual International Conference On Research In Computational Molecular Biology*: 2003, 197–206.
81. Maechler M. **"Finding Groups in Data": Cluster Analysis Extended Rousseeuw et." R Packag**e. version 2.0 6 (2019). https://cran.r-project.org/web/packages/cluster/index.html.
82. Hunt R, Sauna ZE, Ambudkar SV, Gottesman MM, Kimchi-Sarfaty C: **Silent** (**synonymous**) **SNPs**: **should we care about them? In**: **Single Nucleotide Polymorphisms**. Springer; 2009: 23–39.
83. Delourme R, Pilet-Nayel ML, Archipiano M, Horvais R, Tanguy X, Rouxel T, Brun H, Renard M, Balesdent MH. **A cluster of major specific resistance genes to** *Leptosphaeria maculans* **in** *Brassica napus*. Phytopathology. 2004;94:578–83.
84. Dilmaghani A, Balesdent MH, Didier JP, Wu C, Davey J, Barbetti MJ, Li H, Moreno-Rico O, Phillips D, Despeghel JP, Vincenot L. **The** *Leptosphaeria maculans– Leptosphaeria biglobosa* **species complex in the American continent**. Plant Pathol. 2009;58:1044–58.

85. Rouxel T, Willner E, Coudard L, Balesdent MH. Screening and identification of resistance to Leptosphaeria maculans (stem canker) in Brassica napus accessions. Euphytica. 2003;133:219–31.
86. Delourme R, Brun H, Ermel M, Lucas MO, Vallee P, Domin C, Walton G, Li H, Sivasithamparam K, Barbetti MJ. **Expression of resistance to** *Leptosphaeria maculans* **in Brassica napus double haploid lines in France and Australia is influenced by location**. Ann Appl Biol. 2008;153:259–69.
87. Rashid MH, Zou Z, Fernando WD. **Development of molecular markers linked to the Leptosphaeria maculans resistance gene Rlm6 and inheritance of SCAR and CAPS markers in** *Brassica napus* × *Brassica juncea* **interspecific hybrids**. Plant Breeding. 2018;37:402–11.
88. Kutcher HR, Keri M, McLaren DL, Rimmer SR. **Pathogenic variability of** *Leptosphaeria maculans* **in western Canada**. Can J Plant Path. 2007;29:388–93.
89. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. Nat Biotechnol. 2011;29:24–6. https://doi.org/10.1038/nbt.1754.

Figures

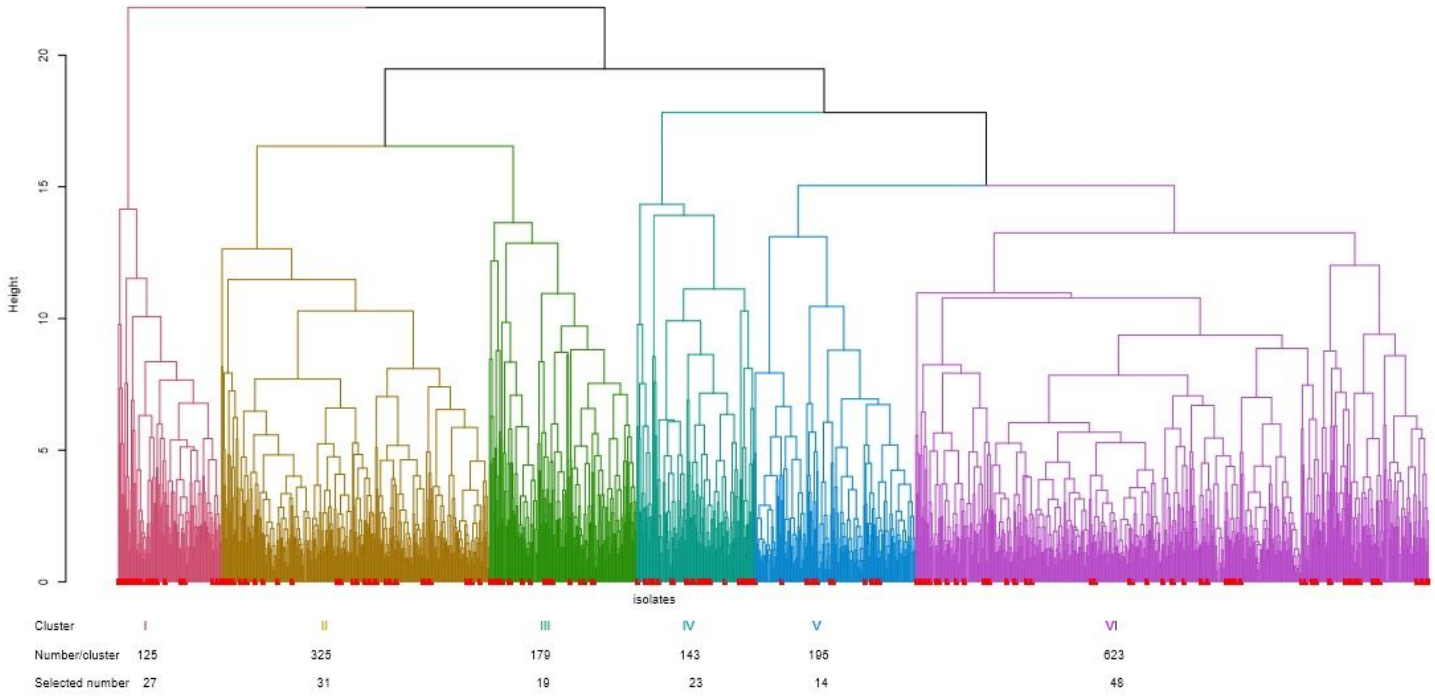


Figure 1

Cluster analysis of 1590 L. maculans isolates collected from western Canada between 2007 and 2013. Selected isolates for re-sequencing were labeled by red leaves.

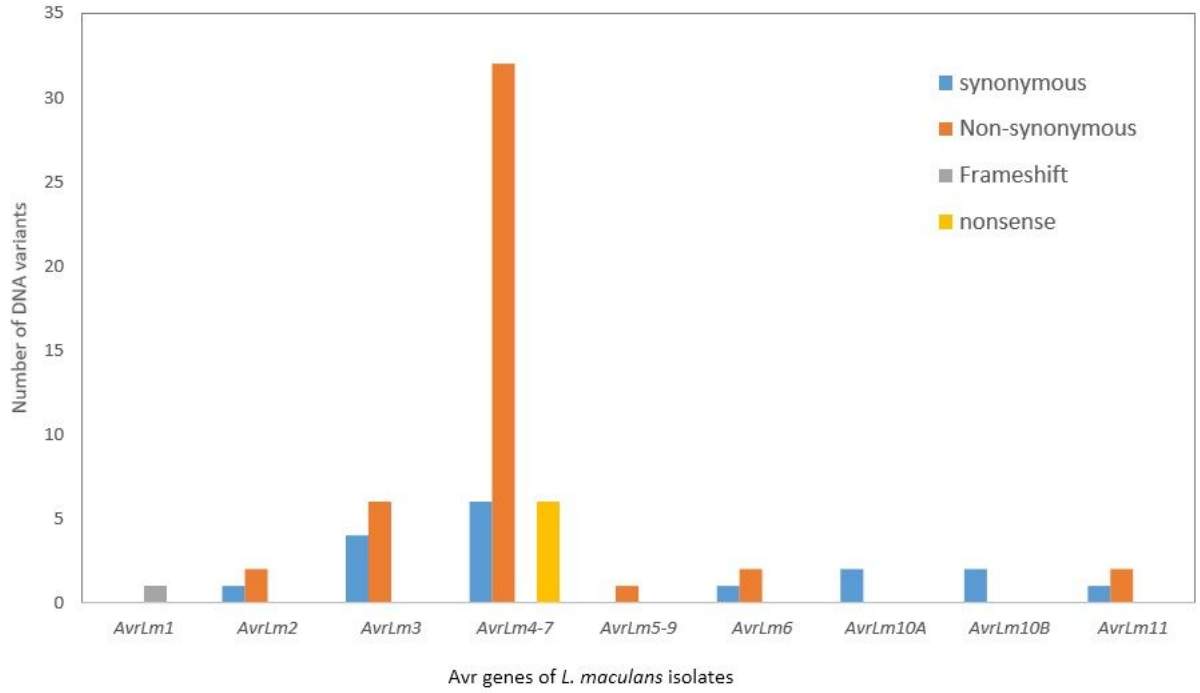
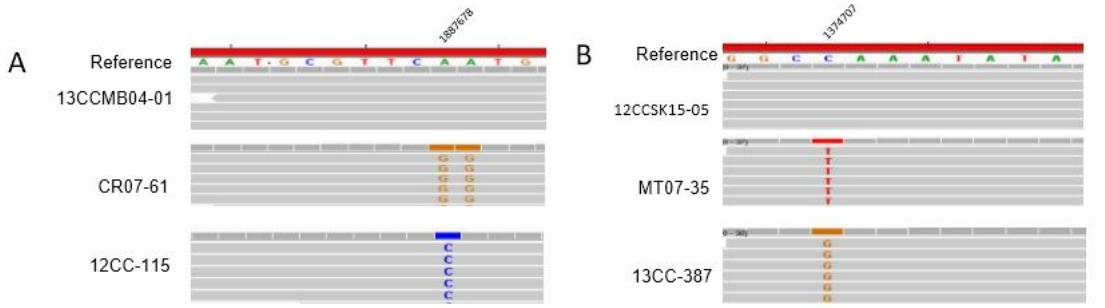


Figure 2

Number of DNA variants identified in different Avr genes of L. maculans.



C

ATG CCA CTA TCC CTC GAG ATA ATC TTA ACG CTA CTC GCT CTC TCT ATC CCT ACA ATT ACA GCT TGT AGA GAG GCC
TCA ATA TCT GGA GAA ATT CGC TAT CCT (35)^{C}TAA GGC ACT TGT CCC ACA AAG ACT GAA GCT TTG AAT GAT TGT AAC
AAA GTA ACG AAG GGC TTA ATT GAC TTT AGT (60)^{C}TAA TCG CAT (63)^{C}TAA CGT GCC (66)T^{G}AG GGT ATA GAT ATG ACG
GCG AAA GTC CAA TGT GCG CCC TGG ATA ACT ACC GAC CCT (85) T^{G}AG GAT GTA GTT CTT TGC ACT TGC AAG ATC ACG
GCG CAT AGA TAT CGA GAA TTC GTT CCC AAA ATT CCC TAT AGC AGC TTT AGC TCA GCA CCT GGA GTT ATA TTT GGC
CAG GAG ACT GGT TTA GAC CAT GAC CCT GAA (131)T^{G}AG GTT GTT AAC ATG AAA GCA AGG ACT CGT GGT TGC GAC TAA

Figure 3

Triallelicand nonsense mutations identified in the isolates of L. maculans. A. The triallelic SNP site for AvrLm2 is at the position 1,887,678 in supercontig 6. B.AvrLm4-7 at 1,374,707 in supercontig supercontig12. The snapshots show short reads assembly in the Integrative Genome Viewer [89]. For simplicity, only 6 tracks were taken from original assembly alignment; grey color represents identical bases to the reference sequence. C. Premature STOP codons were identified in the isolates collected in western Canada.Six SNPs leading to premature STOP codons identified in AvrLm4-7, the reference bases are left-superscripted.

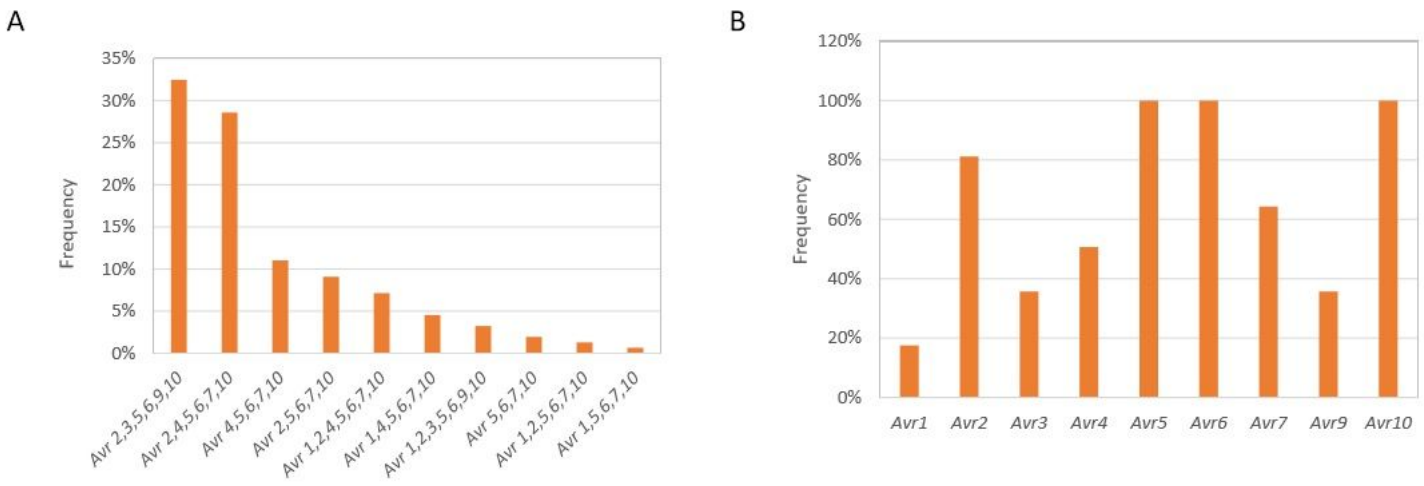


Figure 4

Race structure assessment of L. maculansisolates by SNP haplotyping. A. Race structure; B.Avirulence gene frequencies.

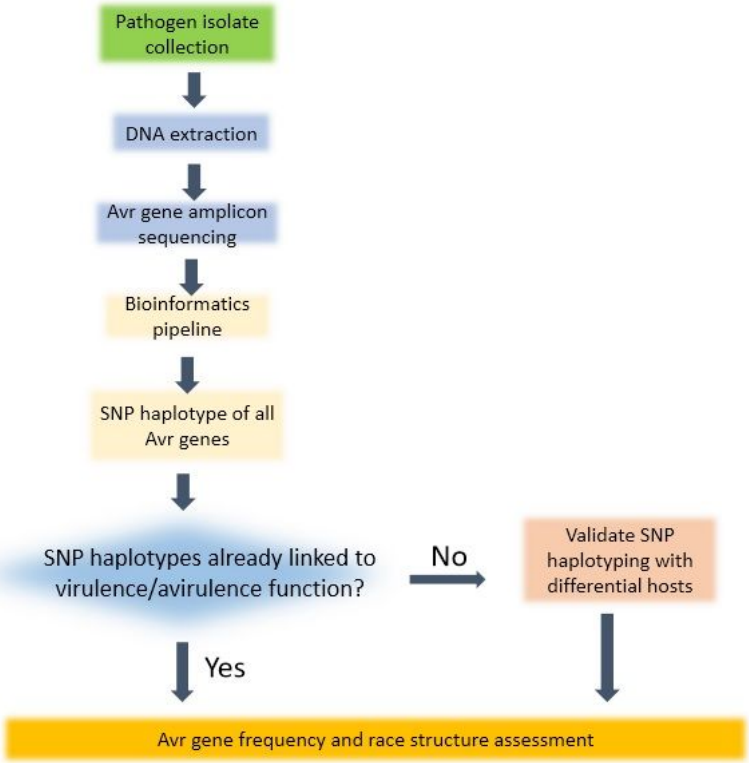


Figure 5

Proposed approach for L. maculansAvr profiling using haplotyping

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalTable1.xlsx
- AdditionalTable6.xlsx
- AdditioanlFigure1.pdf
- AdditionalTable5.xlsx

- AdditionalFigure3.pdf
- AdditionalTable3.xlsx
- AdditionalTable4.xlsx
- AdditionalTable2.xlsx
- AdditionalFigure2.pdf