Published July 7, 2016

ORIGINAL RESEARCH

# A Cluster of Nucleotide-Binding Site–Leucine-Rich Repeat Genes Resides in a Barley Powdery Mildew Resistance Quantitative Trait Loci on 7HL

Carlos P. Cantalapiedra, Bruno Contreras-Moreira, Cristina Silvar, Dragan Perovic, Frank Ordon, María Pilar Gracia, Ernesto Igartua, and Ana M. Casas*

## Abstract

Powdery mildew causes severe yield losses in barley production worldwide. Although many resistance genes have been described, only a few have already been cloned. A strong QTL (quantitative trait locus) conferring resistance to a wide array of powdery mildew isolates was identified in a Spanish barley landrace on the long arm of chromosome 7H. Previous studies narrowed down the QTL position, but were unable to identify candidate genes or physically locate the resistance. In this study, the exome of three recombinant lines from a high-resolution mapping population was sequenced and analyzed, narrowing the position of the resistance down to a single physical contig. Closer inspection of the region revealed a cluster of closely related NBS-LRR (nucleotide-binding site–leucine-rich repeat containing protein) genes. Large differences were found between the resistant lines and the reference genome of cultivar Morex, in the form of PAV (presence-absence variation) in the composition of the NBS-LRR cluster. Finally, a template-guided assembly was performed and subsequent expression analysis revealed that one of the new assembled candidate genes is transcribed. In summary, the results suggest that NBS-LRR genes, absent from the reference and the susceptible genotypes, could be functional and responsible for the powdery mildew resistance. The procedure followed is an example of the use of NGS (next-generation sequencing) tools to tackle the challenges of gene cloning when the target gene is absent from the reference genome.

Published in Plant Genome
Volume 9. doi: 10.3835/plantgenome2015.10.0101



Powdery mildew (*Blumeria graminis*) is an obligate biotrophic fungal ectoparasite of grasses. It colonizes the surface of leaves, feeding from the epidermal cells by means of specialized organs called haustoria (Jørgensen, 1988). The forma specialis *hordei* causes powdery mildew in barley (*Hordeum vulgare* L.), which leads to severe losses in yield and grain quality in temperate latitudes worldwide (Ames et al., 2015; Zhang et al., 2005). This results in a significant economic impact since barley is one of the most widely grown crops (Verstegen et al., 2014). Consequently, the interaction of barley and powdery mildew has been extensively studied (for a recent review, see Schweizer, 2014) and many resistance genes known as mildew genes (*Ml* genes) have been described (Friedt and Ordon, 2007).

However, most of them are still molecularly uncharacterized. Among cloned genes, the recessive *mlo* stands out; providing durable resistance (Jørgensen, 1992) which has

C.P. Cantalapiedra, B. Contreras-Moreira, M.P. Gracia, E. Igartua, and A.M. Casas, Dep. of Genetics and Plant Production, Estación Experimental de Aula Dei, EEAD-CSIC, Avda. Montañana 1005, 50059 Zaragoza, Spain; C.P. Cantalapiedra, Plant Biology and Biotechnology PhD Program, Universitat Autònoma de Barcelona, Spain; B. Contreras-Moreira, Fundación ARAID, Spain; C. Silvar, Universidade da Coruna, Spain; F. Ordon, D. Perovic, Julius Kühn-Institute JKI, Quedlinburg, Germany. Received 19 Oct. 2015. Accepted 24 Feb. 2016. *Corresponding author (acasas@eead.csic.es).

Abbreviations: BAC, bacterial artificial chromosome; BES, BAC-End sequence; CAPS, cleaved amplified polymorphic sequence; CPC, Coding Potential Calculator; CNV, copy-number variation; FPC, finger-printed contig; HM, heterozygous mapping; IBGSC, International Barley Genome Sequencing Consortium; LRR, leucine-rich repeat; MTP, minimum tiling path; NBS, nucleotide-binding site; NGS, next-generation sequencing; ORF, open reading frame; PAV, presence-absence variation; PCR, polymerase chain reaction; QTL, quantitative trait locus; RIL, recombinant inbred line; RTq, real-time quantitative; SBCC, Spanish Barley Core Collection; SNP, single-nucleotide polymorphism; UCR, University of California Riverside; WGS, whole genome sequencing.

remained effective for over 30 yr and copes with a broad spectrum of pathogen isolates (Büschges et al., 1997). The other major powdery mildew resistance genes cloned so far are located at the *Mla* locus, which consists of a cluster of genes encoding for related proteins (Wei et al., 1999). Several *Mla* alleles have been cloned (Zhou et al., 2001; Halterman et al., 2001) out of the many resistance specificities described for this locus (Jørgensen and Wolfe, 1994).

Cloning of *mlo* and *Mla* involved long and laborious efforts. Specifically, fine-mapping of these genes consisted in recurrent steps of marker development, polymorphism detection and genotyping, looking for recombinants. This was done to narrow down the respective genetic intervals until an affordable physical size of the region was achieved, and subsequently resolved by chromosome walking or sequencing of subclones developed using yeast or bacterial artificial chromosome (BAC) clones. This cumbersome procedure was most challenging for species like barley due to the lack of genomic resources and its large and highly repetitive genome (Krattinger et al., 2009). However, the recent advent of high-throughput sequencing, by means of NGS technologies, has accelerated the development of synteny resources (Mayer et al., 2011), sequenced enriched physical maps (Ariyadasa et al., 2014; International Barley Genome Sequencing Consortium [IBGSC], 2012; Mascher et al., 2013b; Muñoz-Amatriaín et al., 2015), genotyping (Comadran et al., 2012; Poland et al., 2012), and sequence capture platforms (Mascher et al., 2013a). In consequence, gene cloning now benefits from the easier and faster genotyping of high-resolution mapping populations, high-throughput polymorphism detection in parental lines, and new fine mapping approaches, such as mapping-by-sequencing (Mascher et al., 2014).

Typical disease resistance genes from plant innate immunity encode receptors usually activated through recognition of molecules from the pathogen (Flor, 1971). These receptors are usually subdivided in two classes. Transmembrane pattern-recognition receptors represent the first active line of defense at the plant cell surface (Jones and Dangl, 2006). They enable the recognition of microbe-associated molecular patterns and induce pattern-triggered immunity. In contrast, a second class of resistance proteins induces elicitor-triggered immunity, detecting either the action or the structure of pathogen molecules inside host cells. These receptors are polymorphic, defining a repertoire for the detection of distinct pathogen effectors (Maekawa et al., 2011). Most genes in this second class encode proteins of the NBS-LRR family (McHale et al., 2006).

NBS-LRRs are abundant in plant genomes (Yue et al., 2012) and are encoded by genes often located in clusters of closely related members (Michelmore and Meyers, 1998). These evolve through rapid expansion and contraction of gene families (Meyers et al., 2003; Monosi et al., 2004; Zhou et al., 2004). In barley, an example of an NBS-LRR cluster is that residing in the *Mla* locus (Seeholzer et al., 2010). NBS-LRR genes encode two protein domains. The NBS domain bears a string of motifs largely conserved in plants, both in sequence and in order (Marone et al., 2013). NBS domains are followed by a LRR domain, which is generally more variable, often associated with direct or indirect non-self-recognition (Spoel and Dong, 2012). Besides *Mla* genes, many other disease resistance genes have been associated to NBS-LRR loci in plants (reviewed in Marone et al., 2013). For instance, in barley *Rpg5*/*rpg4* confers resistance to *Puccinia graminis* (Brueggeman et al., 2008), and *Rdg2a* to *Drechslera graminea* (Bulgarelli et al., 2010). Additional NBS-LRR genes have been cloned in wheat and its wild relatives (discussed in Gu et al., 2015).

This study took advantage of the sequencing-based genomic resources available for barley to fine map a powdery mildew resistance QTL. A high-resolution mapping population was developed to narrow down the QTL interval, followed by exome sequencing of recombinant lines with contrasting resistance phenotypes. The results revealed that genes located in the physical region corresponding to the genetic interval where the QTL is placed, formed a cluster of closely related NBS-LRRs, of which the resistant lines have unique haplotypes.

## Materials and Methods

### Plant Material and Mapping Population

A $BC_1F_2$ population was obtained from the cross Plaisant × RIL151. Recombinant inbred line (RIL) 151 derives from the SBCC097 × Plaisant population (SBCC, Spanish Barley Core Collection; Silvar et al., 2010). This line has only one of the two resistance QTL identified in the original donor landrace, on 7HL (Silvar et al., 2012). $BC_1F_2$ seeds were planted in 96-well trays and sampled 10 d after sowing. For each individual $BC_1F_2$ plant, a 0.6 cm leaf disk was cut. DNA extraction and amplification was performed with the Extract-N-Amp Plant polymerase chain reaction (PCR) kit (Sigma, San Antonio, TX). A cleaved amplified polymorphic sequence (CAPS) marker, QBS58, and a microsatellite, EBmac0755, were used as flanking markers to delimit the QTL interval. Restriction digestion of PCR products was performed in a 20 μL volume using 1.5 U of the respective restriction endonuclease (Fermentas). Plants were selected if they showed recombination between both markers. Data from another four markers (QBS52, QBS46, QBS44, and QBS36) were used to perform linkage analysis with JoinMap 4.0 (van Ooijen, 2006), using Kosambi's map function. Selected plants were vernalized for 6 wk at 3 to 8°C, 8 h light, then transplanted to pots and transferred to a growth chamber, where the plants were grown under long-day conditions (16 h light, 250 μmol $m^{-2}$ $s^{-1}$, 20°C, 60% relative humidity; 8h dark, 16°C, 65% relative humidity). Plants were bagged before seed setting.

To select homozygous recombinants in the $BC_1F_3$ generation, 20 progeny plants of each selected $BC_1F_2$ plant were screened as explained above. Additional CAPS and pyrosequencing markers were incorporated at this stage. To verify

the genotype of the $BC_1F_4$ recombinant lines, genomic DNA was isolated from frozen leaves using the NucleoSpin Plant II kit (Macherey-Nagel, Germany). The complete set of markers used can be found in Supplemental File 1.

## Pathogen Isolates and Disease Assessment

Four isolates of *B. graminis* f. sp. *hordei* (R79, R126, R164, and R225) were used to score resistance and susceptibility in the parents and $BC_1F_4$ recombinant lines. These isolates were propagated on plants of the susceptible cultivar Igri. The seedlings were grown under mildew-free conditions, at 20°C with 60 to 70% relative humidity, and a 16 h light/8 h dark photoperiod. Ten days after sowing, when the first leaf was fully expanded, five plants per line were inoculated with the different isolates by brushing them with powdery mildew spores. Inoculated plants were maintained under the same conditions described above. The infection types were recorded on a scale of 0 to 4 (including intertypes) 10 d after inoculation, following the procedure of Torp et al. (1978) and Jensen et al. (1992). Plants with infection scores < 2 were classified as resistant, otherwise were labeled as susceptible. Pictures were also taken 10 d after infection.

## Exome Sequencing

Genomic DNA from three $BC_1F_4$ lines (1476, 1766, and 2085) was extracted from leaf tissue using the NucleoSpin Plant II XL kit from Macherey-Nagel. Exome capture and DNA sequencing was performed at CNAG (Centro Nacional de Análisis Genómico, Barcelona). DNA capture was performed in a single reaction with the Roche Nimblegene SeqCap EZ Developer kit (Mascher et al., 2013a), following the instructions from the manufacturer. DNA was barcoded with TruSeq adapters and pooled before hybridization to the exome probes. DNA fragmentation and size selection was performed to produce 2 × 101 bp paired-end reads with average insert size of 150 bp. Sample preparation followed standard Illumina TruSeq procedures. Sequencing was performed in two separate runs of an Illumina HiSeq2000, each in a single lane.

Reads were aligned to the Morex whole genome sequencing (WGS) assembly (IBGSC, 2012) with BWA MEM (Li and Durbin, 2009) with default parameters. Read duplicates were tagged by means of MarkDuplicates from picard-tools-1.113 (http://broadinstitute.github.io/picard). Variant detection was performed combining SAMtools (Li et al., 2009) and GATK (McKenna et al., 2010) (see Supplemental Materials and Methods). Variants were filtered out, requiring a minimum depth of 10 and a minimum quality of 30 in each genotyped line. Polymorphic variants were obtained comparing the data of the $BC_1F_4$ lines with variants for SBCC097 and Plaisant from another exome capture essay (Cantalapiedra, Contreras-Moreira, Gracia, Igartua, and Casas, unpublished data, 2014).

To look for the recombination points in the sequences of the three $BC_1F_4$ lines, a score was assigned to each variant identified after the exome capture. If a variant was like Plaisant, the score was increased by 1. If the variant was like SBCC097, the score was decreased by 1 instead. If it was different to the parents, the score remained unchanged. Therefore, the variants in which the three lines were Plaisant-like received a score of +3 in that position in the genome. On the contrary, if all three lines were like SBCC097, the score was –3. This was repeated for every variant. The scores of the variants lying on a single Morex WGS contig were averaged to obtain a single contig score.

## Identification and Annotation of the BACs Located within the QTL Region

Contigs of each BAC associated to finger-printed contig (FPC) 591, from IBGSC (2012) and University of California Riverside (UCR BACs, hereafter; Muñoz-Amatriaín et al., 2015), were concatenated to build up BAC pseudoscaffolds. Gene annotations were obtained from IBGSC data, by alignment of the associated Morex WGS contigs to Uniref90 and UniprotKB (blastx, maximum e-value $1e^{-50}$) and by identification and annotation of open reading frames (ORFs) with getorf (Rice et al., 2000; -minsize 90) and the script run_predict.sh from CPC (Coding Potential Calculator, v.0.9-r2; Kong et al., 2007). Searches of NBS and LRR motifs (taken from Table 1 in Jupe et al., 2012) were performed with MAST (MEME suite 4.10.1; Bailey and Gribskov, 1998). Structure of the NBS-LRR genes was obtained after alignment of the predicted proteins to NCBI nr protein database (see Supplemental Materials and Methods). Multiple alignments of the proteins were performed with Clustal Omega (Sievers et al., 2011).

## Finding and Assembling Heterozygous Mapping Regions

Although the lines used for this study should all be homozygous in the QTL region, a number of sites with heterozygous variants were found after aligning exome sequences to the reference. To systematically locate these regions, an analysis of the number of different k-mers mapping to the pseudoscaffolds was performed. Read mappings from exome sequencing were surveyed to quantify each different 50-mer aligning to each position in the reference, considering only those sampled at least four times. The scripts used for k-mer analysis are available in Supplemental File 2. Sets of reads from the segments with more than one kind of k-mer (therefore annotated as heterozygous mappings, HMs) and mapping to disease resistance proteins were assembled with Trinity (Grabherr et al., 2011; parameters located in Supplemental Materials and Methods). The sequence contigs obtained for the different $BC_1F_4$ lines were compared and clustered. A representative sequence was chosen from each cluster and a genotype was assigned to it based on its presence-absence pattern across $BC_1F_4$ lines. Several overlapping contigs, which showed the same PAV in the lines, were assembled together.

## Validation of the Genotypes Found with the Exome Capture by PCR

The genotypes of the parents and the recombinant lines were checked for those Morex WGS contigs which had

polymorphisms associated with the resistance or susceptibility phenotype. These included contigs 1622651, 167712, 211721, and 50573. Amplicons were used to validate the genotypes of the lines corresponding to sequences present in BACs M01 and D03 from FPC 591. In addition, the PAV polymorphism of the lines was assessed for the two largest new assembled sequence contigs (ELOC1 and ELOC2), including cultivar Morex. Primers were designed with Primer 3 (Untergasser et al., 2012) and validated by running isPCR (https://genome.ucsc.edu/cgi-bin/hgPcr, verified 22 Apr. 2016) against the WGS assemblies from IBGSC data. In addition, primers were designed to amplify the unknown fragments between Morex WGS contig 50573 and both ELOC1 and Morex WGS contig 44875, by Long Range PCR. The primers and their respective PCR conditions can be found in Supplemental File 1.

### Characterization of the New Assembled Sequence Contigs

Putative ORFs encompassing the assembled ELOCs were searched with ORF Finder. In addition, CPC was conducted to evaluate their protein-coding potential. The resulting DNA sequences were searched for in the Uniprot Plants and NCBI nr databases. Both sequences were also compared against the IBGSC databases and Haruna Nijo flcDNAs (Matsumoto et al., 2011) with Barleymap (Cantalapiedra et al., 2015). The predicted aminoacid sequences coded by those ORFs were compared to each other with blastp.

### Real-Time PCR of the Assembled Sequence Contigs

For Real-Time quantitative polymerase chain reaction (RTq-PCR) experiments, 7-d-old plants were inoculated with powdery mildew isolate R79 in the greenhouse. Two samples per line were collected at 12, 24, 48, and 72 h after infection. Each sample consisted of the pooled leaf tissue of two plants.

Total RNA was extracted from frozen samples using the Aurum TM Total RNA Mini Kit (BioRad, Hercules, CA) following the manufacturer's instructions. First-strand cDNA was synthesized from 100 ng of total RNA by using the iScript cDNA Synthesis Kit (BioRad). Primers were designed with Primer Express 3.0 (Applied Biosystems, Carlsbad, CA). RTq-PCR was performed in 50 μL of reaction mixture made up of 2.5 μL of cDNA, 1 × iQ SYBR Green Supermix (BioRad) and 0.3 μM of each specific primer. Primers and PCR conditions can be found in Supplemental File 1. The Actin gene was used as a constitutively expressed reference gene to normalize expression as in Trevaskis et al. (2006).

## Results

### Fine Mapping of the Resistance Locus

To fine map the resistance QTL identified on 7HL in the SBCC097 × Plaisant population (Silvar et al., 2010), a RIL containing only this QTL (RIL151, Silvar et al., 2012) was backcrossed to Plaisant. A large $BC_1F_2$ population was obtained, and tested for recombination between markers QBS58 and EBmac0755, flanking the 7HL QTL. Out of 2899 $BC_1F_2$ plants tested, 152 recombinants were identified and grown until maturity. Twenty-five $BC_1F_3$ families were then screened to identify homozygous recombinants, which were further tested with the markers obtained in previous studies, exploiting synteny and physical information (Silvar et al., 2012; 2013b). This procedure identified 15 $BC_1F_4$ plants covering the whole region (Fig. 1). A genetic map of the region was constructed with the information of the entire $BC_1F_2$ generation and allowed narrowing the position of the QTL down to a 0.07 cM interval between markers QBS46 and QBS44. Furthermore, three $BC_1F_4$ lines, one susceptible (1476) and two resistant (1766 and 2085), showed the same genotype flanking the QTL but different phenotype (Fig. 1, Supplemental Fig. S1). Therefore, the gene or genes responsible for the resistance lay within the interval between QBS46 and QBS44.

### Analysis of Exome Sequencing Polymorphisms

Exome sequencing of the parents and the three $BC_1F_4$ lines was performed in order to identify the differences between the resistant and the susceptible plants (Supplemental Table S2). Analysis of the read data from exome sequencing involves a mapping step using a reference, the Morex WGS assembly (IBGSC, 2012) in this case. However, the region associated to the resistance was majorly of interest here. Therefore, the genetic markers from the previous section were located in the POPSEQ map (Mascher et al., 2013b) and the identified positions (Fig. 1) were used to anchor available genomic resources to the region (Supplemental Materials and Methods, Supplemental Fig. S2). This yielded a set of 973 Morex WGS contigs (Supplemental File 3) associated to 17 FPCs, which are contigs with assigned physical positions (Supplemental Table S1). Comparing the variants between the parents, 1037 polymorphisms were identified, corresponding to 120 Morex WGS contigs (out of the 973 just described). The genotypes of the $BC_1F_4$ lines were checked, looking for variants consistent with the phenotypic profile of the lines (1476 like the susceptible parent, Plaisant; the other two like the resistant parent, SBCC097), as those would be the most informative toward finding candidate genes. Only one of the Morex sequences, contig 50573, presented haplotypes fully in agreement with the phenotypic profile of the lines. This contig has a single annotated gene, a "Pentatricopeptide repeat-containing protein" (MLOC_65722 in IBGSC data). A CAPS marker designed for this gene was assayed on all 15 $BC_1F_4$ lines, and its position within the QTL region was confirmed.

### Physical Localization of the Resistance Locus

From the previous analysis, only Morex contig 50573 was unambiguously located within the QTL interval. However, although its genetic POPSEQ map position was known, it

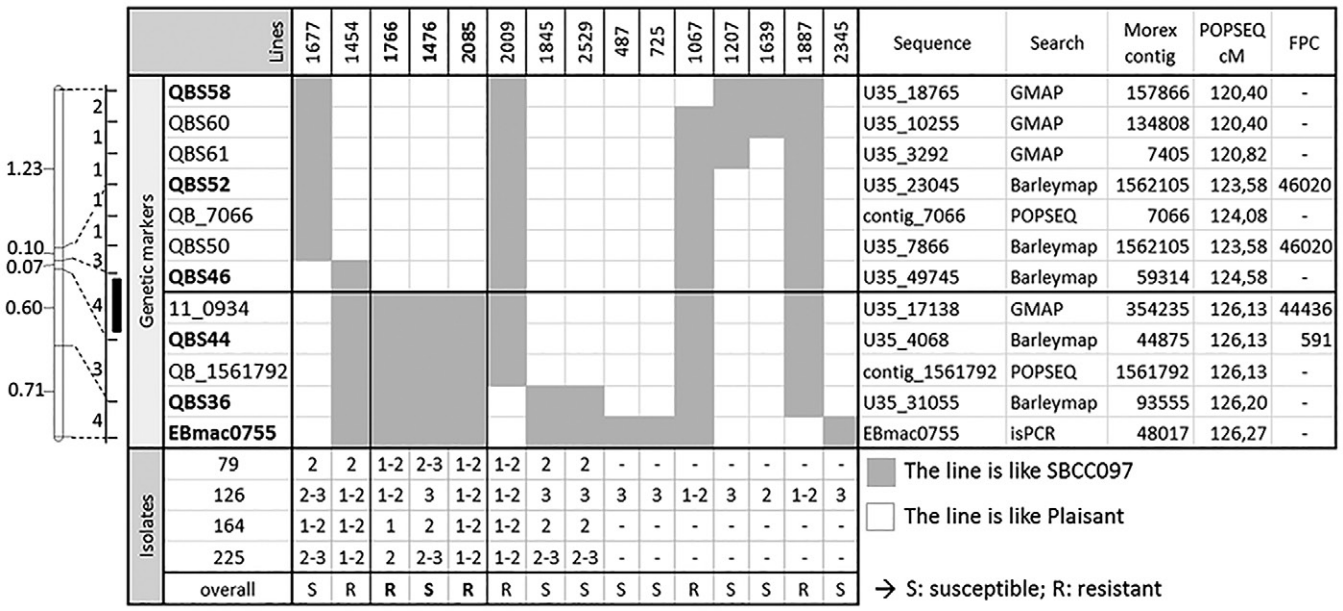


Fig. 1. Fine mapping of the 7HL quantitative trait loci (QTL). Left: Genetic map of $BC_1F_2$ mapping population (distances in cM) showing a schematic distribution of the recombinants found in the $BC_1F_3$ by marker interval. The black vertical bar indicates the position of the QTL. Center: Graphical genotypes of the 15 $BC_1F_4$ lines. Markers assayed in the $BC_1F_2$ are highlighted in bold type. The lines sequenced in this study (1766, 1476, 2085) are separated from the others by thick vertical lines. The thick horizontal line between QBS46 and 11_0934 marks the most likely position of the resistance gene. The bottom table summarizes the evaluation of the lines for resistance to four different powdery mildew isolates. Right: Table showing the sequences used to locate the genetic markers in the barley genome, and the sources (POPSEQ) or search methods used, Barleymap or GMAP (Wu and Watanabe, 2005). The target whole genome sequencing contigs are shown ("Morex contig" column) along with their position in chromosome 7H ("POPSEQ cM" column), as well as the physical contigs ("FPC" column) associated to them.

could not be found in the IBGSC physical map, hindering its direct physical localization. Nonetheless, most of the variants in the remaining Morex WGS contigs were clearly located on either side of the candidate region (i.e., the three lines had the same genotype). Looking at the genotypes of the lines from exome data, the position and order of Morex WGS contigs was not always in agreement with the POPSEQ map (Supplemental Fig. S3). If only Morex WGS contigs with known physical position were considered, the genotypes of the recombinant lines indicated the likely physical location of the recombination breakpoints within FPC 591, more specifically, between contigs 167712 and 44875 (Fig. 2A). The position of yet another Morex WGS contig, 211721, was ambiguous. The genotypes of the lines for these contigs were confirmed by PCR assays.

To further delimit the physical position of the resistance locus, the BACs associated to FPC 591 in the IBGSC physical map were retrieved (Fig. 2B). Among BACs with available sequence data, HVVMRXALLmA0204M01 (M01 hereafter) spans a central segment of FPC 591. Among the Morex WGS contigs aligning to M01 (Supplemental File 4), 167712 and 211721 were identified ~2.5 kb apart. Moreover, Morex contig 44875 was associated to BAC HVVMRXALLEA0187D03 (D03 from now on), both from IBGSC anchoring data and by our homology searches (identity 99.75%, full target coverage, bitscore 1448; to D03 BES MRX2BAD187D03T71). D03 covers the right half of FPC 591, but it has not been fully sequenced yet. No other BACs providing new data within the QTL interval were identified. Candidate genes should thus be placed within the minimum tiling path (MTP) defined by BACs M01 and D03.

During the progress of this work, a new assembly of BACs (UCR BACs) was published. In this assembly, two extra BACs were associated to FPC 591 (Fig. 2B): 0139I11 and 0758B20 (I11 and B20 from now on). BAC I11 (Supplemental File 5) was compared to M01 (Supplemental Fig. S4A). Most of the I11 sequences are already present in M01, but with a different arrangement. In contrast, the comparison of B20 and M01 pseudoscaffolds (Supplemental Fig. S4B) showed that they are mostly different, with only a few related regions. Among the Morex WGS contigs which aligned to B20 (Supplemental File 6), contigs 50573 and 44875 were found, separated by 4234 bases. Note that Morex WGS contig 50573 is the only one with a haplotype in agreement with the phenotypes of the lines, hence supporting the position of the resistance locus within FPC 591.

## Searching for Candidate Genes in the Reference Cultivar Morex

Candidate genes were searched for in the annotated Morex genome. Alignments of Morex WGS contigs, anchored to BAC M01, against IBGSC and Uniref90 sequences, revealed eight gene annotations: five "Disease resistance protein RPM1," two transposon-related and one "Putative disease resistance protein RGA4." In-house annotation of the ORFs identified in the M01 pseudoscaffold (see Materials and Methods) confirmed the presence of the RPM1- and

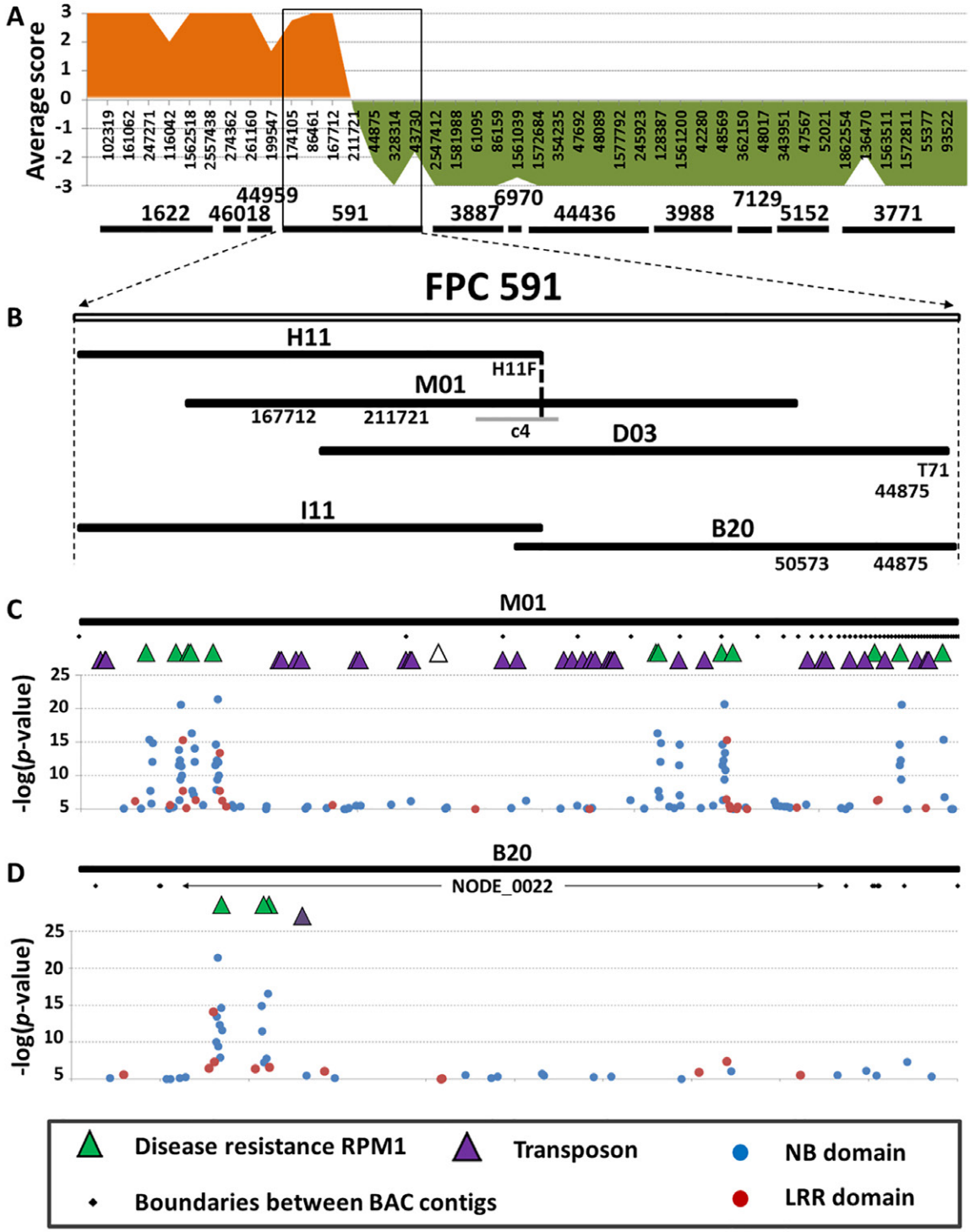


Fig. 2. Analysis of bacterial artificial chromosomes (BACs) in minimum tiling path (MTP) of finger-printed contig (FPC) 591. (A) Average scores of the Morex whole genome sequencing (WGS) contigs considering the genotypes of the $BC_1F_4$ lines in relation to the parents. Orange: positive score, more lines are like Plaisant; green: negative score, more lines are like SBCC097. Contigs are sorted by increasing FPC cM position, and by POPSEQ position to resolve coincidences, from left (120.4 cM) to right (126.6 cM). FPCs are shown as black horizontal bars. (B) IBGSC (H11, M01, and D03) and UCR (I11 and B20) BACs covering FPC 591. Morex WGS contigs 167712 and 211721 are anchored to M01. BAC-End sequence (BES) H11F and BAC contig c4 of M01 match by sequence alignment (vertical dashed line). BES T71 and Morex WGS contig 44875 align to each other. Morex WGS contigs 44875 and 50573 are anchored to B20. (C) Analysis of the pseudoscaffold of BAC M01, represented as a black horizontal bar. Green triangles are ORFs annotated as RPM1 by alignment to Uniref90. A white triangle shows an ORF annotated as RGA4, which seems to be related to transposons. Purple triangles show the position of ORFs annotated as transposons. The scatterplot shows the-$\log_{10}$(*p*-value) of the NBS and LRR motifs identified throughout the pseudoscaffold (blue dots, NBS domains; red dots, LRR domains). (D) Analysis of the pseudoscaffold of BAC B20. Note that NODE_0022 is highlighted as the longest contig in the BAC.

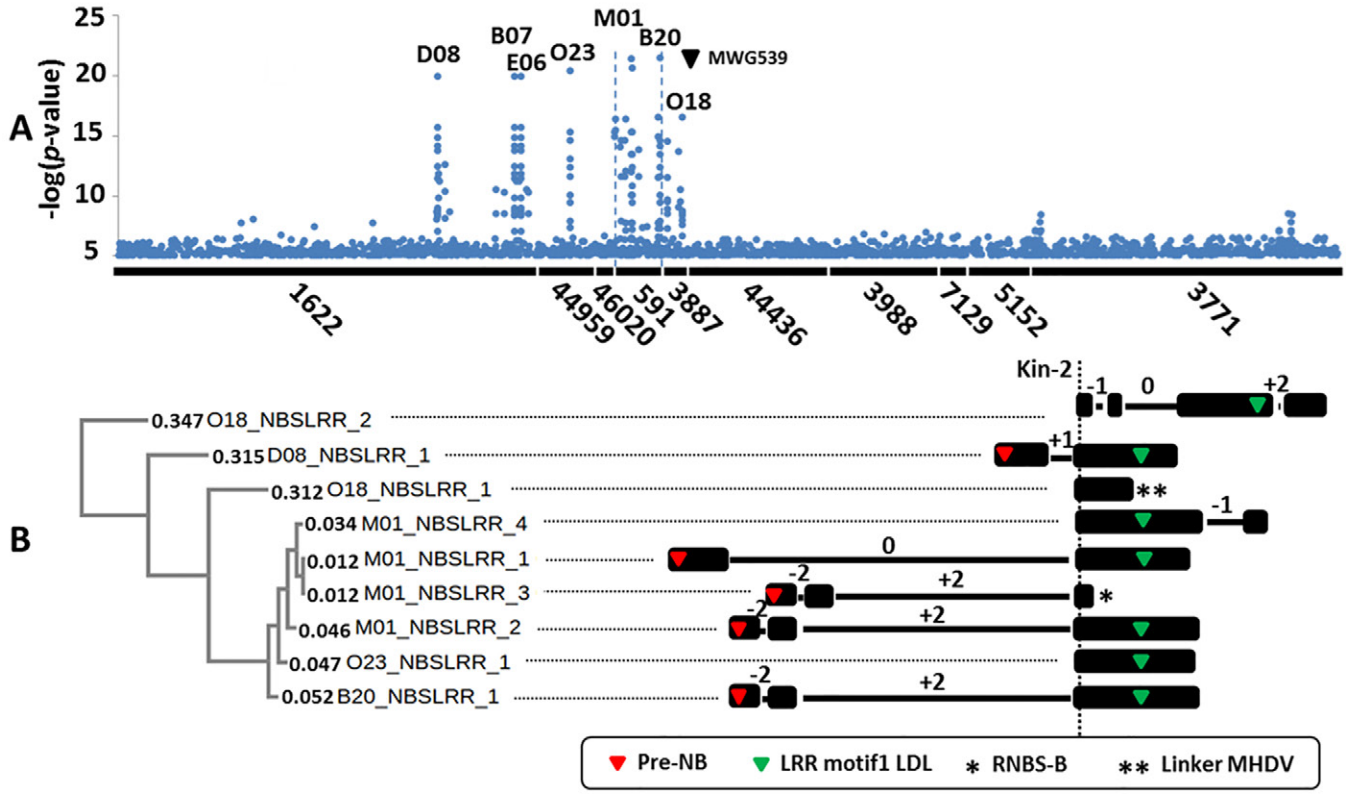


Fig. 3. Nucleotide-binding site (NBS) and leucine-rich repeat (LRR) motifs found in the region of FPC 591. (A) Significance of the motifs found in the whole region (of about 5.6 Mb). Vertical dashed blue lines demarcate the motifs found within FPC 591. A black triangle indicates the physical position of RFLP marker MWG539, close to the *Mlf* locus (Schönfeld et al., 1996). (B) UPGMA clustering of the predicted proteins containing NBS-LRR motifs. Protein names are prefixed with their respective BAC codes. Distances obtained from the multiple alignment are shown to the left of each protein name. Inferred gene structures are shown to the right (black boxes: exons; black horizontal lines: introns). The number on each intron shows the frame change from one exon to the next. Motifs shown on gene structures are named after Table 1 in Jupe et al. (2012). A vertical dashed line shows the position of the Kinase-2 motif, to which the structures of genes have been aligned. Asterisks indicate the presence of a specific motif at the end of the available sequence of the corresponding gene.

transposon-related sequences, including loci not associated to Morex WGS contigs and, therefore, lacking exome capture probes. When the whole pseudoscaffold was self-aligned, the ORFs annotated as RPM1 proteins appeared to be related to each other (Supplemental Fig. S5). Since RPM1 belongs to the NBS-LRR family of resistance-genes, motifs which are known to be conserved in domains of NBS-LRR genes (Jupe et al., 2012) were searched for in the region using the software MAST. Most RPM1-related loci were also confirmed by the MAST scan (Fig. 2C). Overall, nine segments were identified with highly significant motifs from the N-terminal, NBS and linker domains; three of them with LRR motifs (Supplemental File 4). The same analysis was applied to BAC I11, which showed almost the same features as M01, as expected (Supplemental File 5).

On the other hand, IBGSC annotation of the Morex WGS contigs associated to UCR BAC B20 showed up 2 genes: a "Pentatricopeptide repeat-containing protein" in contig 50573, mentioned earlier, and a "WD-repeat protein 57 IPR015943" in contig 44875. Both results were confirmed with alignments to Uniref90. In addition, another 3 Uniref90 hits to the left of contig 50573 were obtained; all labeled as "Disease resistance protein RPM1," both using raw Morex WGS contigs and in silico identified ORFs as queries. Again MAST scans of NBS-LRR motifs confirmed these results (Fig. 2D) and, as with M01, several hits related to transposons were obtained close to them (Supplemental File 6).

Analysis of NBS-LRR motifs in a wide physical region around FPC 591 (55 UCR BACs, spanning 5.6 Mb) revealed that the cluster is mostly circumscribed to the resistance locus (Fig. 3A). A few other NBS-LRR genes were detected outside the locus, but these were unrelated both in terms of sequence and gene structure (Fig. 3B, Supplemental Files 7 and 8).

Therefore, besides a Pentatricopeptide repeat-containing protein and a WD-repeat protein, the MTP spanning the resistance locus in Morex is rich in transposons and contains a cluster of closely related NBS-LRR genes.

## Analysis of Heterozygous Mappings in Morex

As shown above, only Morex WGS contig 50573 had a haplotype consistent with being within the resistance locus. However, there were other Morex WGS contigs for which some variants were consistent, but others were not. Many of the variants in those contigs were apparently heterozygous. This was highly unlikely, as the

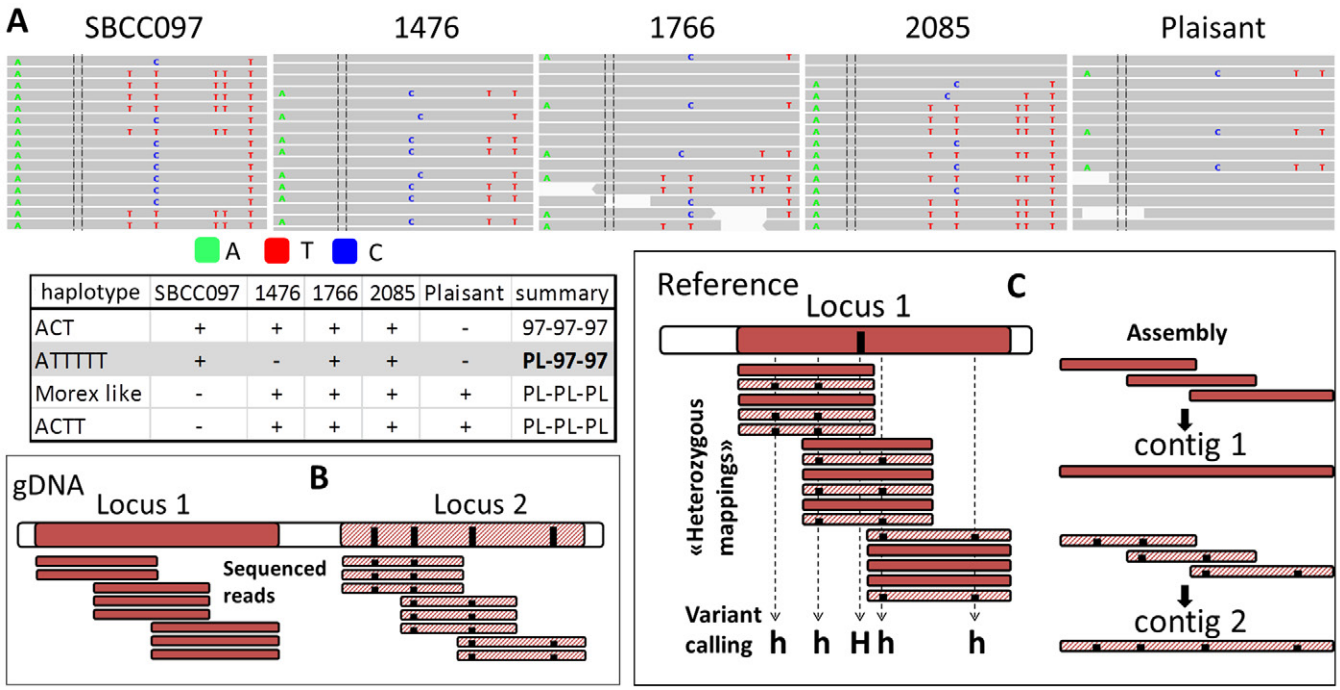


Fig. 4. Heterozygous mappings (HMs). (A) Images captured from Integrative Genomics Viewer (Integrative Genomics Viewer [IGV]; Robinson et al., 2011), showing reads (gray horizontal bars) mapping to a specific interval of Morex whole genome sequencing contig 1622651. Colored characters show the variants detected for each genotype in relation to the Morex reference. The table summarizes the haplotypes identified, along with their presence-absence type (+ or –) in the lines. Genotypes of the three $BC_1F_4$ recombinant lines relative to the parents are shown in the "summary" column. One group of variants (ATTTTT, light gray background) is consistent with the phenotypic resistance profile of the lines (PL-97-97 or susceptible-resistant-resistant). (B) schematic representation of the reads that would be obtained after sequencing two closely related loci. The two loci are represented by horizontal bars (red background; plain for Locus 1, striped for Locus 2), with a few hypothetical differences (black vertical bars). (C) Reads from (B) are mapped back to the reference. In the example shown, the reference lacks one locus (Locus 2), and all sequenced reads hit the existing one (Locus 1), producing apparent HMs. As a result, variant calling yields heterozygous calls (h) and homozygous calls (H) intermixed. A new assembly could solve this region, yielding independent contigs resembling the original loci, due to the presence of the four genotypic variants between the two loci.

parents were homozygous, the $BC_1F_4$ plants were selected to be homozygous for the interval of interest and the possibility of having double recombinants within such a small region was negligible. In fact, visual inspection of the mappings producing those variants revealed different populations of reads stacking to the same locus (Fig. 4A), in contrast with the mappings from contig 50573, which produced unambiguous homozygous single-nucleotide polymorphisms (SNPs; Supplemental Fig. S6). The apparent heterozygous genotypes were confirmed through PCR amplification of CAPS markers (Supplemental Fig. S7). Note that these variants were abundant and linked in recurrent groups, as independent haplotypes, instead of being spread out randomly among the reads. Thus, it is unlikely that they are the result of sequencing errors. Instead, these mappings could have been produced by piling up closely related sequences (repeats, paralogous genes) which were captured by the exome baits (Mascher et al., 2013a; Jupe et al., 2013), but for which the original locus would not be present in the reference. Since they affect variant calling, producing apparent heterozygous variants, from now on this kind of mappings will be referred to as HMs (Fig. 4B and 4C). Almost all Morex WGS contigs with HMs, whose variants had genotypes in agreement with the phenotypic profile of the lines, could be annotated as homologs to "Disease resistance protein RPM1" or "Disease resistance protein RPP13" (Supplemental Table S3), after alignment to the Uniprot Plants database (http://www.uniprot.org/blast/, verified 22 Apr. 2016). Some of those contigs are the ones located within or close to FPC 591 (Supplemental Fig. S8–S11). Taken together, these results suggest that there are sequences related to disease resistance proteins, which are not present in the Morex reference, but are likely within the resistance locus in the genomes of SBCC097 or Plaisant.

In this study, the distribution and abundance of HMs in the resistance locus region was analyzed in more detail to (i) assess whether the differences between the recombinant lines were likely to be related with the disease resistance, (ii) verify whether the presence of HMs was a feature exclusive of the sequences related to NBS-LRR genes in the region of interest, and to (iii) identify and demarcate the segments of the reference in which they occur. This last objective would allow obtaining the reads which produce the HMs and assembling them into sequence contigs (Fig. 4C).

Therefore, we analyzed the number of different 50-mers, fragments of reads of 50 bases, mapping to each position of Morex WGS contigs anchored to BACs M01 and B20 in the three $BC_1F_4$ lines. Note that the reads from our sequencing

data are 101-mers, but to be able to capture diversity in a given position a smaller k-mer size had to be chosen, since mapping duplicates were removed in a previous step. Wherever several 50-mers mapped to the same position, HMs would be likely found; each 50-mer being possibly derived from a different genomic locus. Notably, we found different 50-mers mapping to most of the loci related with NBS-LRR genes, although not all the mapped loci belonged to that class (Supplemental Fig. S12). Out of the covered positions, 74.4 and 89.5% had a single 50-mer in M01 and in B20, respectively. Interestingly, differences among the lines seemed to be associated mostly to disease resistance loci. First, the resistant lines had a larger percentage of positions with several 50-mers (i.e., with HMs) in M01 (Supplemental Fig. S13A), although not in B20. Furthermore, taking into account only the reference positions within annotated NBS-LRR genes, the difference between the resistant lines and the susceptible one increased in both BACs (Supplemental Fig. S13B). Therefore, the differences between the two BACs can to a large extent be explained by the greater abundance of NBS-LRR related sequences in M01 and B20 (49.6 and 11.7% of the mapped bases, respectively).

### De Novo assembly of Exome Sequence Reads Spanning the Resistance Locus

Analysis of HMs pointed toward the presence of NBS-LRR related sequences within the resistance locus, absent from the Morex reference. In light of this, a template-guided assembly of reads producing HMs was performed. First, Morex WGS contig fragments located within FPC 591, related to disease resistance genes and producing HMs were chosen (11 loci). Second, six further Morex WGS contig fragments with HMs and variants in agreement with the phenotypes of the lines were selected. Finally, Morex WGS contig 50573, harboring the "Pentatricopeptide repeat-containing protein," was included as a control. Read subsets mapping to the 18 selected segments were retrieved, and an independent assembly for each genotype was performed (for both parents and the three $BC_1F_4$ lines, Supplemental File 9). These operations yielded 203 sequence contigs, with an average of almost 41 contigs per line. These new contigs were clustered, and a representative sequence per cluster was selected (see Supplemental materials and methods and Supplemental Fig. S14), yielding 31 representative sequences. Based on the presence or absence of those sequences, PAV genotypes for each cluster were assigned to each line. Representative sequences showing the same PAV genotypic profiles were then compared with each other, leading to the assembly of five of them into a contig of 981 nucleotides (ELOC1), and another four into a contig of 787 bases (ELOC2). Therefore, the final set comprised 24 sequence contigs, for which the lines had different PAV genotypes (Supplemental File 9). ELOC1 and ELOC2 were the largest assembled contigs. ELOC1 was absent in Plaisant and 1476, while ELOC2 was only present in SBCC097 and 1766. The absence of ELOC2 from the resistant line 2085 was in agreement with the fewer number of 50-mers identified in this line in comparison with 1766, and it suggested that 2085 and 1476 contained the smallest interval flanking the resistance locus.

### Validation and Characterization of the New Assembled Sequence Contigs

We designed primers to perform PCR amplification of ELOC1 and ELOC2. The PCRs confirmed the PAV genotypes of the 15 $BC_1F_4$ lines and the parents (Fig. 5). In addition, the absence of both sequences in cultivar Morex was verified (data not shown). To check whether this result was a consequence of polymorphism on the primers, the reads from the exome capture of SBCC097, Plaisant, Morex (from the same exome capture experiment), and lines 1476, 1766, and 2085 were realigned to the new contigs. This confirmed the PAV variation found on them. Moreover, the products of amplification of the lines SBCC097 and 1766 were Sanger-sequenced and further validated.

In silico ORF calling was performed with both ELOCs, obtaining two partial ORFs of 322 and 252 amino acids for ELOC1 and ELOC2, respectively. In addition, their protein-coding potential was checked, with log-odds scores of 82.73 and 57.46 for ELOC1 and ELOC2, respectively. The percentage of identity between the two amino acid sequences was 92%, and their alignment covered most of ELOC2. Looking for similar proteins in Uniprot Plants and NCBI nr databases, results were found (Supplemental File 10) within the range of identities obtained when comparing the NBS-LRR proteins in the QTL region in Morex (Supplemental File 8), and comparable with paralogous genes found in other NBS-LRR clusters (Bulgarelli et al., 2010; Kuang et al., 2004; Wei et al., 1999). Moreover, the ELOCs were aligned against the Morex NBS-LRR predicted proteins of the region. The best hits had almost full coverage and 87.9 and 91.6% identity, for ELOC1 and ELOC2, respectively. Alignment of DNA sequences of the ELOCs to the IBGSC databases produced similar results. Also, these alignments revealed that the contigs contained only the LRR domain, lacking the NBS one.

RTq-PCR was used to check the expression of both new contigs. No specific amplicon was obtained for ELOC2 and, therefore, it could either be a pseudogene (Kuang et al., 2004) or be expressed in another tissue or developmental stage (Tan et al., 2007). Nonetheless, amplification was positive for ELOC1, confirming its transcription in leaves of SBCC097 and the two resistant $BC_1F_4$ lines, although this is not a definitive evidence of the gene being functional (Monosi et al., 2004; Wei et al., 2002). The RTq-PCR was performed for SBCC097 at different time points, spanning 72 h after infection. Apparently, there was no change in ELOC1 expression in response to the infection, although this is not irreconcilable with being involved in the resistance or even being regulated at another stage than transcription (Tan et al., 2007).

## Discussion

Barley research has been accelerated by the availability of abundant genomic resources published over the last years. In some cases, this has led to faster gene cloning, like

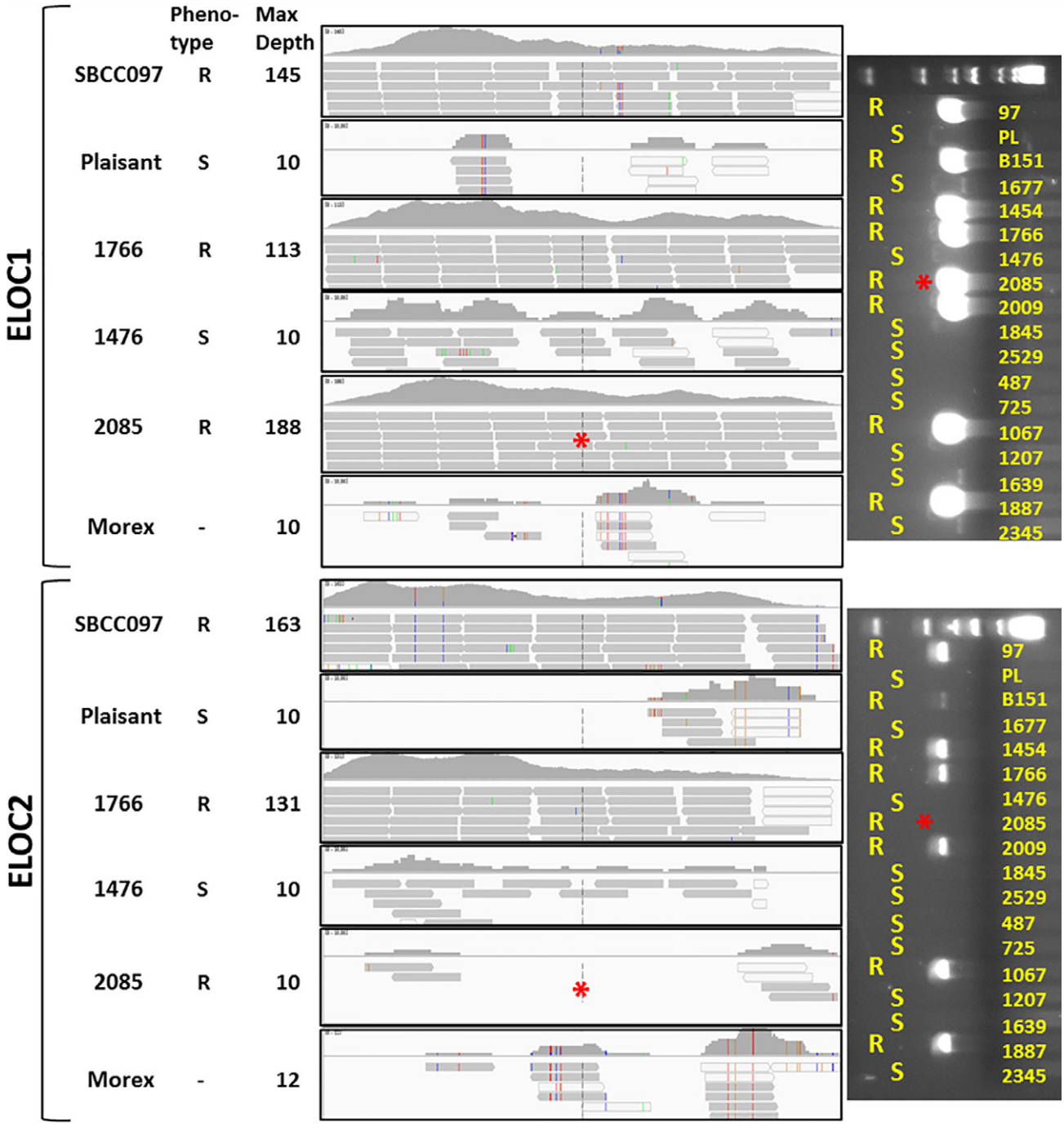


Fig. 5. Presence-absence genotypes for ELOC1 (top) and ELOC2 (bottom). Left: Phenotypes of the two parents, the three sequenced lines, and Morex, along with the maximum depth of coverage (Max Depth) obtained after mapping the exome sequencing reads to ELOC1 and ELOC2 (the two new assembled contigs). Center: Images captured from Integrative Genomics Viewer (IGV), showing the profile of depth of coverage throughout the contigs (top) and individual reads mapped (bottom). Resistant lines have large depths of coverage and similar profiles, covering the whole contigs, with the exception of 2085 in ELOC2 (red asterisk). Susceptible lines have low depth of coverage and irregular, incomplete mapping profiles. Right: Gel electrophoresis of polymerase chain reaction (PCR) amplicons of ELOC1 and ELOC2 for the two parents, the resistant line RIL151 and the 15 $BC_1F_4$ lines, along with their phenotypes. Resistant lines have presence genotypes whereas susceptible lines have absence genotypes, with the exception of 2085 in ELOC2 (red asterisk). R, resistant; S, susceptible.

cloning of *HvCEN* by Comadran et al. (2012). However, other barley genes have not been cloned yet despite their known phenotypic effect and genetic localization, partly due to the lack of such resources until recently. The continuous improvement of barley physical resources (Ariyadasa et al., 2014; IBGSC, 2012; Mascher et al., 2013b; Muñoz-Amatriaín et al., 2015) allows the adoption of more efficient methodologies for genetic studies involving high-throughput genotyping, marker development, gene discovery, expression analysis, synteny and genome comparative studies. The exome capture probe set developed by Mascher et al. (2013a) for barley is already being used

for gene cloning purposes. Mascher et al. (2014) used it to identify *HvMND*, a gene that regulates the rate of leaf initiation, and Pankin et al. (2014) to identify a candidate for *HvPHYC*. In both cases, exome capture was performed on bulked plants with extreme phenotypes from $BC_1F_2$ populations between mutants and the wild-type.

In this work, the same exome capture probe set was used to sequence three recombinant lines for a powdery mildew resistance QTL. The resistance allele was contributed by a Spanish landrace, showing a wide resistance profile (resistance to 23 out of 27 isolates tested) after a thorough disease survey (Silvar et al., 2011) with the accessions from the SBCC (Igartua et al., 1998). Such line had two QTL conferring race-specific resistances on chromosome 7H (Silvar et al., 2010). The mechanism of resistance of this line was classified as consistent with "intermediate-acting" genes, governing resistance mainly at the postpenetration stage (Silvar et al., 2013a). Genomic approaches allowed the development of new markers to narrow down the QTL intervals (Silvar et al., 2012, 2013b), but were insufficient to definitely locate a manageable physical location or a set of candidate genes for the stronger QTL on 7HL, which is the subject of this work.

From that point, a large $F_2$ population was created and screened with markers from those previous studies, aiming to identify recombinant lines to further narrow down the QTL interval. The final interval, just 0.07 cM wide, was apparently small enough to land on potential candidates, as this size is comparable with other intervals used in successful gene cloning attempts in barley (reviewed in Krattinger et al., 2009). Again, the analysis of available genomic resources was insufficient to locate candidate genes or to delimit the resistance to a single physical contig. Although the markers were found in the Morex WGS assembly and a POPSEQ map position could be assigned to them, many other Morex WGS contigs with positions within the QTL interval were identified, leading to a large list of annotated genes. Moreover, since the current barley maps are incomplete, additional contigs could have gone unnoticed. Finally, since not all the contigs to which the markers hit were anchored to physical contigs, the physical localization of the QTL remained unknown. An additional challenge was the search of genetic markers from previous studies in the reference. Several of the markers were only found through the analysis of chimeras from GMAP alignments, likely due to the fragmented nature of the Morex WGS assembly.

Exome sequencing of the parents and three recombinant lines allowed the identification of abundant polymorphic variants. This is a faster and more powerful alternative to the search of markers by in silico comparison of genomic resources from different genotypes or by extrapolation of markers from other populations, since many of these are not necessarily polymorphic between the parental lines of the population under study. However, in this work, most of the homozygous SNPs were located outside the QTL. Only a single Pentatricopeptide-repeat containing protein was easily identified within the QTL region, and its corresponding Morex WGS contig lacked physical anchoring. Despite that, the analysis of the profile of variants along the physical contigs in the region was enough to point toward a single FPC which could contain entirely the QTL. This highlights the usefulness of exome sequencing for fine mapping purposes. However, this work demonstrates the technical challenges encountered. Some positions of Morex WGS contigs were not in agreement with the genotypes of our lines. Differences in collinearity between several genetic maps and the POPSEQ reference have been already described (Cantalapiedra et al., 2015; Silvar et al., 2015). These incongruences are important for fine mapping purposes. A single physical contig holding the resistance locus was identified only after removing the Morex WGS contigs not associated to physical positions and using a score to average together the genotypes of the variants within each Morex WGS contig.

Despite the scarcity of homozygous SNPs found within the QTL region, we observed abundant heterozygous SNPs which were polymorphic between the parents as PAV. Although the work with SNPs and small indels is rather straightforward, working with other kinds of variation such as copy-number variation (CNV) or PAV requires using alternative approaches, for example analyzing mapping depth (Mascher et al., 2014). In this work, HMs are defined as those producing heterozygous variants probably due to the collapse of reads from paralogous genes absent in the reference genome. This phenomenon has been recently described among homeologous genes in an exome sequencing experiment in wheat (King et al., 2015). In studies focused on variant discovery, HMs can confound the discrimination of true variants at a given locus. However, this study used HMs to identify the regions with polymorphic HMs, through k-mer analysis, to further assemble different paralogous genes and assess their expression. Though this approach aimed to locate regions with HMs, k-mer abundance could be directly used for genotyping purposes. As with CNV, analysis of HMs is related to the number of copies of a given sequence. However, the analysis of CNV through mapping depth should cope with the different efficiencies in the hybridization and PCR amplification steps during exome sequencing when the sequences are different. In contrast, the analysis of k-mer abundance has the drawback of being unable to differentiate the copies when they are identical to each other. In addition, analysis of HMs could provide insights into the loci and gene families for which the reference genome is incomplete or shows larger variation between different genotypes. Finally, we genotyped the HMs as PAV polymorphisms by means of template-guided assembly and clustering of the resulting sequence contigs. An alternative approach would be to directly compare the presence or absence of the individual k-mers mapping to a given position in the genotypes, although this would not provide assembled contigs. In both cases, the main difficulty resides in differentiating between orthologous and paralogous genes, allelic

variants and isoforms (Kuang et al., 2004; Seeholzer et al., 2010), either when clustering the contigs from the assembly or when considering that all orthologous k-mers from the different genotypes are mapping to the same reference locus, and not to another closely related one. In any case, the methods used in this study were implemented from standard tools which were combined to accomplish our specific goals, and thus could be further developed and optimized to cope with peculiarities of HMs.

Both the analysis of the sequenced BACs and the genotyping of HMs pointed toward a cluster of related NBS-LRR genes in the resistance locus. These are good candidates for a resistance gene, although we have to be aware that the sequences captured are limited by the baits used and it cannot be ruled out that the actual resistance gene is absent from the capture reactions and/or from the reference genome. NBS-LRR genes are abundant in many plant genomes and are often organized in clusters of one or more groups of related paralogous genes (Michelmore and Meyers, 1998), which makes their assembly difficult. This problem was evident in this study as revealed by the huge difference in size, number and composition of contigs in equivalent sequenced BACs from independent assemblies (e.g., M01 from IBGSC and I11 from UCR). In addition, a common trend observed in NBS-LRR genes in grasses is the rapid expansion and loss of members from those groups (Li et al., 2010; Yang et al., 2013), leading to PAV and CNV between genotypes. Genes found in that region in Morex were poorly annotated and most of them were split into different WGS contigs. Therefore, the exact number and structure of the genes in this cluster remains unknown both in cultivar Morex and in the resistant line SBCC097. In our assembly, the NBS-LRR genes were incomplete, lacking the NBS domains. We do not know whether these genes are actually incomplete or the NBS domains do exist but were not captured. Lack of exome capture reads covering the genes completely, for instance due to the presence of large introns in them, could lead to incomplete assemblies. Nonetheless, the NBS domains are usually more conserved than the LRR ones (Meyers et al., 1999; Pan et al., 2000; Seeholzer et al., 2010), and this could hinder the independent assembly of the different paralogous genes.

This study made extensive use of state-of-the-art genomic resources available for barley. Several aspects which could be considered when working with these resources arise from our analysis. We have already mentioned some of them, like the lack of position of many Morex WGS contigs or the incomplete annotation of genes in the region. Regarding contig positions, we describe the combined use of both POPSEQ map of Morex WGS contigs and their anchoring to BACs to obtain as many sequences as possible close to our resistance locus. Additional information from the recent publication of sequenced BACs from UCR, a different assembly to that of IBGSC, allowed to complete the MTP of the region and confirmed the features identified using IBGSC data. Furthermore, it highlighted the discrepancies between assemblies, even when corresponding to the same barley genotype, at least in regions with repetitive sequences like the clustered NBS-LRR genes and transposons found in our region.

Finally, identification of the full sequence at these loci would require obtaining BAC libraries and the use of long-read sequencing technologies. Sequencing the whole region could reveal candidate genes which have gone unnoticed, and it could contribute to the understanding of structure and diversification of NBS-LRR genes. Furthermore, sequencing the region, which is rich in resistance genes in barley, could help identifying other resistances. For example, *Mlf* (Schönfeld et al., 1996), which has been associated to this region previously (Backes et al., 2003), given the close physical location of its linked RFLP probe to our QTL. Although BAC libraries are available for cultivar Morex and a few more accessions, this is still not the case for most barley genotypes. Until those resources are available, the exploitation of exome capture to assemble reads from HMs was used in this study to identify candidates not present in the reference or in the exome capture target space, through similarity with closely related genes.

## Accession Numbers

The NGS data for both parents and recombinant lines are accessible at European Nucleotide Archive under project no. PRJEB11739.

## Supplemental Information Available

Supplemental information is included with this article. Supplement Part 1: Supplemental Materials and Methods, Supplemental Tables S1 to S3, and Supplemental Figures S1 to S13. Supplement Part 2: Spreadsheets, multiple alignments, scripts.


### Acknowledgments

The authors would like to thank Dr. Nils Stein, Dr. Robbie Waugh, and IBGSC members for allowing early use of barley exome capture platform. We are also grateful to Dr. Nils Stein and Dr. Uwe Scholz for allowing the use of computational infrastructure at IPK, Dr. María Muñoz-Amatriaín for her hints to work with UCR BACs, and Drs. Kerstin Flath and Doris Kopahnke for providing and multiplying mildew isolates. This work was funded by Spanish projects AGL2010-21929, AGL2013-48756, RFP2012-00015-00-00, and EUI2009-04075 (national code for Plant-KBBE project ExpResBar). CPC was funded by Spanish MINECO (Grant no. BES-2011-045905 linked to project AGL2010-21929).


### References


Ames, N., A. Dreiseitl, B.J. Steffenson, and G.J. Muehlbauer. 2015. Mining wild barley for powdery mildew resistance. Plant Pathol. 64(6):1396–1406. doi:10.1111/ppa.12384

Ariyadasa, R., M. Mascher, T. Nussbaumer, D. Schulte, Z. Frenkel, N. Poursarebani, et al. 2014. A sequence-ready physical map of barley anchored genetically by two million single-nucleotide polymorphisms. Plant Physiol. 164:412–423. doi:10.1104/pp.113.228213

Backes, G., L.H. Madsen, H. Jaiser, J. Stougaard, M. Herz, V. Mohler, et al. 2003. Localisation of genes for resistance against *Blumeria graminis* f.sp. *hordei* and *Puccinia graminis* in a cross between a barley cultivar

and a wild barley (*Hordeum vulgare* ssp. *spontaneum*) line. Theor. Appl. Genet. 106:353–362. doi:10.1007/s00122-002-1148-1
Bailey, T.L., and M. Gribskov. 1998. Combining evidence using p-values: Application to sequence homology searches. Bioinformatics 14:48–54. doi:10.1093/bioinformatics/14.1.48
Brueggeman, R., A. Druka, J. Nirmala, T. Cavileer, T. Drader, N. Rostoks, et al. 2008. The stem rust resistance gene *Rpg5* encodes a protein with nucleotide-binding-site, leucine-rich, and protein kinase domains. Proc. Natl. Acad. Sci. USA 105:14,970–14,975. doi:10.1073/pnas.0807270105
Bulgarelli, D., C. Biselli, N.C. Collins, G. Consonni, A.M. Stanca, P. Schulze-Lefert, et al. 2010. The CC-NB-LRR-Type *Rdg2a* resistance gene confers immunity to the seed-borne barley leaf stripe pathogen in the absence of hypersensitive cell death. PLoS ONE 5(9):e12599. doi:10.1371/journal.pone.0012599
Büschges, R., K. Hollrichner, R. Panstruga, G. Simons, M. Wolter, et al. 1997. The barley *Mlo* gene: A novel control element of plant pathogen resistance. Cell 88:695–705. doi:10.1016/S0092-8674(00)81912-1
Cantalapiedra, C.P., R. Boudiar, A.M. Casas, E. Igartua, and B. Contreras-Moreira. 2015. BARLEYMAP: Physical and genetic mapping of nucleotide sequences and annotation of surrounding loci in barley. Mol. Breed. 35:13. doi:10.1007/s11032-015-0253-1
Comadran, J., B. Kilian, J. Russell, L. Ramsay, N. Stein, M. Ganal, et al. 2012. Natural variation in a homolog of *Antirrhinum CENTRORADIALIS* contributed to spring growth habit and environmental adaptation in cultivated barley. Nat. Genet. 44:1388–1392. doi:10.1038/ng.2447
Flor, H.H. 1971. Current status of the gene-for-gene concept. Annu. Rev. Phytopathol. 9:275–296. doi:10.1146/annurev.py.09.090171.001423
Friedt, W., and F. Ordon. 2007. Molecular markers for gene pyramiding and disease resistance breeding in barley. In: R.K. Varshney and R. Tuberosa, editors, Genomics assisted crop improvement: 2. Genomics applications in crops. Springer, Dordrecht, the Netherlands. p. 81–101.
Grabherr, M., B.J. Hass, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29:644–652. doi:10.1038/nbt.1883
Gu, L., W. Si, L. Zhao, S. Yang, and X. Zhang. 2015. Dynamic evolution of NBS-LRR genes in bread wheat and its progenitors. Mol. Genet. Genomics 290:727–738. doi:10.1007/s00438-014-0948-8
Halterman, D., F.S. Zhou, F.S. Wei, R.P. Wise, and P. Schulze-Lefert. 2001. The MLA6 coiled-coil, NBS-LRR protein confers *AvrMla6*-dependent resistance specificity to *Blumeria graminis* f. sp. *hordei* in barley and wheat. Plant J. 25:335–348. doi:10.1046/j.1365-313x.2001.00982.x
Igartua, E., M.P. Gracia, J.M. Lasa, B. Medina, J.L. Molina-Cano, J.L. Montoya, et al. 1998. The Spanish barley core collection. Genet. Resour. Crop Evol. 45:475–481. doi:10.1023/A:1008662515059
International Barley Genome Sequencing Consortium (IBGSC). 2012. A physical, genetic and functional sequence assembly of the barley genome. Nature 491:711–716. doi:10.1038/nature11543
Jensen, H.P., E. Christensen, and J.H. Jørgensen. 1992. Powdery mildew resistance genes in 127 Northwest European spring barley varieties. Plant Breed. 108:210–228. doi:10.1111/j.1439-0523.1992.tb00122.x
Jones, D.G.J., and J.L. Dangl. 2006. The plant immune system. Nature 444:323–329. doi:10.1038/nature05286
Jørgensen, J.H. 1988. *Erysiphe graminis*, powdery mildew of cereals and grasses. In: G.S. Sidhu, editor, Genetics of plant pathogenic fungi. Adv. Plant Pathol. 6:137–157.
Jørgensen, J.H. 1992. Discovery, characterization and exploitation of *Mlo* powdery mildew resistance in barley. Euphytica 63:141–152. doi:10.1007/BF00023919
Jørgensen, J.H., and M. Wolfe. 1994. Genetics of powdery mildew resistance in barley. Crit. Rev. Plant Sci. 13:97–119. doi:10.1080/07352689409701910
Jupe, F., L. Pritchard, G.J. Etherington, K. MacKenzie, P.J. Cock, F. Wright, et al. 2012. Identification and localisation of the NB-LRR gene family within the potato genome. BMC Genomics 13:75. doi:10.1186/1471-2164-13-75
Jupe, F., K. Witek, W. Verweij, J. Sliwka, L. Pritchard, G.J. Etherington, et al. 2013. Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. Plant J. 76:530–544. doi:10.1111/tpj.12307
King, R., N. Bird, R. Ramirez-Gonzalez, J.A. Coghill, A. Patil, K. Hassani-Pak, et al. 2015. Mutation scanning in wheat by exon capture and next-generation sequencing. PLoS ONE 10(9):e0137549. doi:10.1371/journal.pone.0137549
Kong, L., Y. Zhang, Z.Q. Ye, X.Q. Liu, S.Q. Zhao, L. Wei, et al. 2007. CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. 35:W345–W349. doi:10.1093/nar/gkm391
Krattinger, S., T. Wicker, and B. Keller. 2009. Map-based cloning of genes in Triticeae (wheat and barley). Chapter 12. In: C. Feuillet and G.J. Muehlbauer, editors, Genetics and genomics of the Triticeae. Springer, Dordrecht, the Netherlands. p. 337–357. doi:10.1007/978-0-387-77489-3_12
Kuang, H., S. Woo, B.C. Meyers, E. Nevo, and R.W. Michelmore. 2004. Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. Plant Cell 16:2870–2894. doi:10.1105/tpc.104.025502
Li, J., J. Ding, W. Zhang, Y. Zhang, P. Tang, J. Chen, et al. 2010. Unique evolutionary pattern of numbers of gramineous NBS-LRR genes. Mol. Genet. Genomics 283:427–438. doi:10.1007/s00438-010-0527-6
Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics 25:1754–1760. doi:10.1093/bioinformatics/btp324
Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, et al. 2009. The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25:2078–2079. doi:10.1093/bioinformatics/btp352
Maekawa, T., T.A. Kufer, and P. Schulze-Lefert. 2011. NLR functions in plant and animal immune systems: So far and yet so close. Nat. Immunol. 12:817–826. doi:10.1038/ni.2083
Marone, D., M. Russo, G. Laidò, A. De Leonardis, and A. Mastrangelo. 2013. Plant nucleotide binding site-leucine-rich repeat (NBS-LRR) genes: Active guardians in host defense responses. Int. J. Mol. Sci. 14:7302–7326. doi:10.3390/ijms14047302
Mascher, M., M. Jost, J.E. Kuon, A. Himmelbach, A. Aßfalg, S. Beier, et al. 2014. Mapping-by-sequencing accelerates forward genetics in barley. Genome Biol. 15:R78. doi:10.1186/gb-2014-15-6-r78
Mascher, M., G.J. Muehlbauer, D.S. Rokhsar, J. Chapman, J. Schmutz, K. Barry, et al. 2013b. Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). Plant J. 76:718–727. doi:10.1111/tpj.12319
Mascher, M., T.A. Richmond, D.J. Gerhardt, A. Himmelbach, L. Clissold, D. Sampath, et al. 2013a. Barley whole exome capture: A tool for genomic research in the genus *Hordeum* and beyond. Plant J. 76:494–505. doi:10.1111/tpj.12294
Matsumoto, T., T. Tanaka, H. Sakai, N. Amano, H. Kanamori, K. Kurita, et al. 2011. Comprehensive sequence analysis of 24,783 barley full-length cDNAs derived from 12 clone libraries. Plant Physiol. 156:20–28. doi:10.1104/pp.110.171579
Mayer, K.F.X., M. Martis, P.E. Hedley, H. Simková, H. Liu, J.A. Morris, et al. 2011. Unlocking the barley genome by chromosomal and comparative genomics. Plant Cell 23:1249–1263. doi:10.1105/tpc.110.082537
McHale, L., X. Tan, P. Koehl, and R. Michelmore. 2006. Plant NBS-LRR proteins: Adaptable guards. Genome Biol. 7:212. doi:10.1186/gb-2006-7-4-212
McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, et al. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–1303. doi:10.1101/gr.107524.110
Meyers, B.C., A.W. Dickerman, R.W. Michelmore, S. Sivaramakrishnan, B.W. Sobral, and N.D. Young. 1999. Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Plant J. 20:317–332. doi:10.1046/j.1365-313X.1999.t01-1-00606.x
Meyers, B.C., A. Kozik, A. Griego, H. Kuang, and R.W. Michelmore. 2003. Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. Plant Cell 15:809–834. doi:10.1105/tpc.009308
Michelmore, R.W., and B.C. Meyers. 1998. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res. 8:1113–1130. doi:10.1101/gr.8.11.1113
Monosi, B., R.J. Wisser, L. Pennill, and S.H. Hulbert. 2004. Full-genome analysis of resistance gene homologues in rice. Theor. Appl. Genet. 109:1434–1447. doi:10.1007/s00122-004-1758-x
Muñoz-Amatriaín, M., S. Lonardi, M.C. Luo, K. Madishetty, J.T. Svensson, M.J. Moscou, et al. 2015. Sequencing of 15,622 gene-bearing

BACs clarifies the gene-dense regions of the barley genome. Plant J. 84:216–227. doi:10.1111/tpj.12959
Pan, Q., J. Wendel, and R. Fluhr. 2000. Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. J. Mol. Evol. 50:203–213. doi:10.1007/s002399910023.
Pankin, A., C. Campoli, X. Dong, B. Kilian, R. Sharma, A. Himmelbach, et al. 2014. Mapping-by-sequencing identifies *HvPHYTOCHROME C* as a candidate gene for the early maturity 5 locus modulating the circadian clock and photoperiodic flowering in barley. Genetics 198:383–396. doi:10.1534/genetics.114.165613
Poland, J.A., P.J. Brown, M.E. Sorrells, and J.L. Jannink. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotype-by-sequencing approach. PLoS ONE 7(2):E32253. doi:10.1371/journal.pone.0032253
Rice, P., I. Longden, and A. Bleasby. 2000. EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet. 16:276–277. doi:10.1016/S0168-9525(00)02024-2
Robinson, J.T., H. Thorvaldsdóttir, W. Winckler, M. Guttman, E.S. Lander, G. Getz et al. 2011. Integrative genomics viewer. Nat. Biotech. 29:24–26. doi:10.1038/nbt.1754
Schönfeld, M., A. Ragni, G. Fischbeck, and A. Jahoor. 1996. RFLP mapping of three new loci for resistance genes to powdery mildew (*Erysiphe graminis* f. sp. *hordei*) in barley. Theor. Appl. Genet. 93:48–56. doi:10.1007/BF00225726
Schweizer, P. 2014. Host and nonhost response to attack by fungal pathogens. Chapter 11. In: J. Kumlehn and N. Stein, editors, Biotechnological approaches to barley improvement. Springer, Berlin, Germany. p. 197–235. doi:10.1007/978-3-662-44406-1_11
Seeholzer, S., T. Tsuchimatsu, T. Jordan, S. Bieri, S. Pajonk, W. Yang, et al. 2010. Diversity at the *Mla* powdery mildew resistance locus from cultivated barley reveals sites of positive selection. Mol. Plant Microbe Interact. 23:497–509. doi:10.1094/MPMI-23-4-0497
Sievers, F., A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7:539. doi:10.1038/msb.2011.75
Silvar, C., H. Dhif, E. Igartua, D. Kopahnke, M.P. Gracia, J.M. Lasa, et al. 2010. Identification of quantitative trait loci for resistance to powdery mildew in a Spanish barley landrace. Mol. Breed. 25:581–592. doi:10.1007/s11032-009-9354-z
Silvar, C., K. Flath, D. Kopahnke, M.P. Gracia, J.M. Lasa, A.M. Casas, et al. 2011. Analysis of powdery mildew resistance in the Spanish barley core collection. Plant Breed. 130:195–202. doi:10.1111/j.1439-0523.2010.01843.x
Silvar, C., D. Kopahnke, K. Flath, A. Serfling, D. Perovic, A.M. Casas, et al. 2013a. Resistance to powdery mildew in one Spanish barley landrace hardly resembles other previously identified wild barley resistances. Eur. J. Plant Pathol. 136:459–468. doi:10.1007/s10658-013-0178-7
Silvar, C., M.M. Martis, T. Nussbaumer, N. Haag, R. Rauser, J. Keilwagen, et al. 2015. Assessing the barley genome zipper and the genomic resources for breeding purposes. Plant Gen. 8(3) doi:10.3835/plantgenome2015.06.0045
Silvar, C., D. Perovic, T. Nussbaumer, M. Spannagl, B. Usadel, A. Casas, et al. 2013b. Towards positional isolation of three quantitative trait loci conferring resistance to powdery mildew in two Spanish barley landraces. PLoS ONE 8(6):e67336. doi:10.1371/journal.pone.0067336
Silvar, C., D. Perovic, U. Scholz, A.M. Casas, E. Igartua, and F. Ordon. 2012. Fine mapping and comparative genomics integration of two quantitative trait loci controlling resistance to powdery mildew in a Spanish barley landrace. Theor. Appl. Genet. 124:49–62. doi:10.1007/s00122-011-1686-5
Spoel, S.H., and X. Dong. 2012. How do plants achieve immunity? Defence without specialized immune cells. Nat. Rev. Immunol. 12:89–100. doi:10.1038/nri3141
Tan, X., B.C. Meyers, A. Kozik, M.A. West, M. Morgante, et al. 2007. Global expression analysis of nucleotide binding site-leucine rich repeat-encoding and related genes in Arabidopsis. BMC Plant Biol. 7:56. doi:10.1186/1471-2229-7-56
Torp, J., H.P. Jensen, and H.J. Jorgensen. 1978. Powdery mildew resistance genes in 106 northwest European spring barley varieties. In: Yearbook. Royal Veterinary and Agricultural University, Copenhagen, Denmark. p. 75–102.
Trevaskis, B., M.N. Hemming, W.J. Peacock, and E.S. Dennis. 2006. *HvVRN2* responds to daylength, whereas *HvVRN1* is regulated by vernalization and developmental status. Plant Physiol. 140:1397–1405. doi:10.1104/pp.105.073486
Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B.C. Faircloth, M. Remm, et al. 2012. Primer3—New capabilities and interfaces. Nucleic Acids Res. 40(15):e115. doi:10.1093/nar/gks596
van Ooijen, J.W. 2006. JoinMap®4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen, the Netherlands.
Verstegen, H., O. Köneke, V. Korzun, and R. von Broock. 2014. The world importance of barley and challenges to further improvements. In: Kumlehn, J., and N. Stein, editors, Biotechnological approaches to barley improvement. Springer, Berlin, Germany. p. 3–19. doi:10.1007/978-3-662-44406-1
Wei, F., K. Gobelman-Werner, S.M. Morroll, J. Kurth, L. Mao, R. Wing, et al. 1999. The *Mla* (powdery mildew) resistance cluster is associated with three NBS-LRR gene families and suppressed recombination within a 240-kb DNA interval on chromosome 5S (1HS) of barley. Genetics 153:1929–1948.
Wei, F., R.A. Wing, and R.P. Wise. 2002. Genome dynamics and evolution of the *Mla* (powdery mildew) resistance locus in barley. Plant Cell 14:1903–1917. doi:10.1105/tpc.002238
Wu, T.D., and C.K. Watanabe. 2005. GMAP: A genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21:1859–1875. doi:10.1093/bioinformatics/bti310
Yang, S., J. Li, X. Zhang, Q. Zhang, J. Huang, J.Q. Chen, et al. 2013. Rapidly evolving R genes in diverse grass species confer resistance to rice blast disease. Proc. Natl. Acad. Sci. USA 110:18572–18577. doi:10.1073/pnas.1318211110
Yue, J., B.C. Meyers, J. Chen, D. Tian, and S. Yang. 2012. Tracing the origin and evolutionary history of plant nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes. New Phytol. 193:1049–1063. doi:10.1111/j.1469-8137.2011.04006.x
Zhang, Z., C. Henderson, E. Perfect, T.L.W. Carver, B.J. Thomas, P. Skamnioti, et al. 2005. Of genes and genomes, needles and haystacks: *Blumeria graminis* and functionality. Mol. Plant Pathol. 6:561–575. doi:10.1111/j.1364-3703.2005.00303.x
Zhou, F.S., J.C. Kurth, F.S. Wei, C. Elliott, G. Vale, N. Yahiaoui, et al. 2001. Cell-autonomous expression of barley *Mla1* confers race-specific resistance to the powdery mildew fungus via a *Rar1*-independent signaling pathway. Plant Cell 13:337–350. doi:10.1105/tpc.13.2.337
Zhou, T., Y. Wang, J. Chen, H. Araki, Z. Jing, K. Jiang, et al. 2004. Genome-wide identification of NBS genes in *japonica* rice reveals significant expansion of divergent non-TIR NBS-LRR genes. Mol. Genet. Genomics 271:402–415. doi:10.1007/s00438-004-0990-z