Theor Appl Genet (2007) 115:159–168
DOI 10.1007/s00122-007-0551-z

ORIGINAL PAPER

Physical mapping and identification of a candidate for the leaf rust resistance gene *Lr1* of wheat

Ji-Wen Qiu · Anita Christina Schürch · Nabila Yahiaoui · Ling-Li Dong · Hua-Jie Fan · Zhong-Juan Zhang · Beat Keller · Hong-Qing Ling

Received: 24 November 2006 / Accepted: 10 April 2007 / Published online: 4 May 2007

Abstract *Lr1* is a dominant leaf rust resistance gene located on chromosome 5DL of bread wheat and the wild species *Aegilops tauschii*. In this study, three polymorphic markers (WR001, WR002, and WR003) were developed from resistance gene analogs (RGAs) clustering around the *Lr1* locus. Using these and other markers, *Lr1* was mapped to a genetic interval of 0.79 cM in *Ae. tauschii* and 0.075 cM in wheat. The CAPS marker WR003, derived from *LR1RGA1*, co-segregated with *Lr1* in both mapping populations of wheat and *Ae. tauschii*. For isolation of *Lr1*, two genomic BAC libraries (from *Ae. tauschii* and hexaploid wheat) were screened using the tightly flanking marker PSR567F and a set of nested primers derived from the conserved region of the RGA sequences. Approximately 400 kb BAC contig spanning the *Lr1* locus was constructed. The *LR1RGA1* encoding a CC-NBS-leucine-rich repeat (LRR) type of protein was the only one of the four RGAs at the *Lr1* locus, which co-segregated with leaf rust resistance. Therefore, it represents a very good candidate for *Lr1*. The allelic sequences of *LR1RGA1* from resistant and susceptible lines revealed a divergent DNA sequence block of ~605 bp encoding the LRR repeats 9–15, whereas the rest of the sequences were mostly identical. Within this sequence block, the 48 non-synonymous changes resulted in 44 amino acid differences. This indicates that *LR1RGA1* likely evolved through one or more recombination or gene conversion events with unknown genes.

Communicated by J. W. Snape.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-007-0551-z) contains supplementary material, which is available to authorized users.

J.-W. Qiu · L.-L. Dong · H.-J. Fan · Z.-J. Zhang · H.-Q. Ling (✉)
State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Datun Road, Chaoyang District, Beijing 100101, China
e-mail: hqling@genetics.ac.cn

J.-W. Qiu
Graduate University of the Chinese Academy of Sciences, 19 Yuquan Road, Beijing 100049, China

A. C. Schürch · N. Yahiaoui · B. Keller
Institute of Plant Biology, University of Zurich, Zollikerstrasse 107, 8008 Zürich, Switzerland

Introduction

Leaf rust, caused by *Puccinia recondita* f. sp. *tritici*, is one of the major fungal diseases in wheat. It can lead to severe yield losses and poor seed quality. The use of fungicides results in an increase of farmer input and may cause environmental pollution. Thus, growing resistant cultivars is the most effective, economic, and environmentally friendly approach for controlling this disease. Isolation of resistance genes from wheat and related species as well as the development of highly diagnostic markers is the basis for understanding and improving wheat resistance.

The gene pool of wheat and its relatives contains several hundred major disease resistance (*R*) genes (McIntosh et al. 1995). Isolation of such resistance genes from the wheat genome is challenging due to a very large genome size (1.6×10^{10} bp for bread wheat), high content of repetitive sequences (80%), redundancy of three subgenomes (A, B, and D genomes) as well as absence of an active transposon tagging system. Considering that most of the *R* genes are characterized by their location on the genetic map, map-based cloning is the best option to isolate such genes. Recently, comparative genomic mapping has revealed that

the difficulties of gene isolation in hexaploid wheat can be partially overcome through chromosome walking using the genomes of the diploid wheat progenitors (Feuillet and Keller 2002). Using these strategies, three resistance genes and active alleles have been cloned from hexaploid wheat (Huang et al. 2003; Feuillet et al. 2003; Yahiaoui et al. 2004, 2006; Srichumpa et al. 2005).

The leaf rust resistance gene *Lr1* was first identified in the hexaploid wheat variety Malakoff (Ausemus et al. 1946), and then found in a wide range of wheat varieties in the world (McIntosh et al. 1995). *Lr1* is a dominant gene located on wheat chromosome 5DL (Feuillet et al. 1995). An orthologous gene of wheat *Lr1* was recently identified in *Aegilops tauschii* (Ling et al. 2004). This gene provides resistance to a set of leaf rust races very similar to the one avirulent on the bread wheat *Lr1* and maps at a homoeologous location in *Ae. tauschii* (Ling et al. 2004). *Lr1* was delimited in the region between RFLP markers ABC718 and PSR567 and was tightly linked to PSR567. Nine positive BAC clones were isolated from the BAC library of *Ae. tauschii* accession Aus18913 by RFLP markers ABC718 and PSR567 (Ling et al. 2003). Three contigs were constructed, one of them containing five BAC clones isolated by ABC718, whereas the other two contigs consisted of four BAC clones which were isolated with PSR567. The three contigs did not overlap with each other and, therefore did not span the *Lr1* locus. Thus, additional physical mapping is needed to identify the *Lr1* gene. Ling et al. (2003) identified three resistance gene analogs (RGAs) from BACs screened by PSR567 and designated them as the 567R family. BLAST analysis indicated that the three RGAs have high amino acid sequence identity to the leucine-rich repeat (LRR) domain encoded by the rice bacterial blight resistance gene *Xa1* (Ling et al. 2003). These results demonstrated that the *Lr1* locus contains a family of related RGAs, with *Lr1* possibly being a member of this cluster. Here, we report a high-resolution genetic and physical map of the *Lr1* region and a candidate for the *Lr1* resistance gene.

Materials and methods

Plant materials

Susceptible spring wheat variety Thatcher and its near isogenic resistant line Thatcher*Lr1*, susceptible (TA1704) and resistant (Tr.t213) accessions of *Ae. tauschii* as well as a number of recombinant lines from the crosses of Thatcher*Lr1* × Thatcher (Ling et al. 2003) and Tr.t213 × TA1704 (Ling et al. 2004) were used for high-resolution genetic mapping.

Genomic DNA extraction and RFLP analysis

For RFLP and Southern analysis, genomic DNA was extracted from leaf tissue as described by Graner et al. (1990). Restriction digestion and Southern hybridization were performed as described by Ling et al. (2003).

Primer design

The sequences of *Ae. tauschii* RGAs were aligned to detect divergent and conserved sequences using ClustalW (http://www.ebi.edu/clustalw). Gene specific primers were designed based on the sequences in divergent regions. For the screening of a wheat BAC library, two pairs of nested primers were designed from the conserved sequences of the four RGAs. The oligonucleotide sequences of all primers used in this work are shown in Table 1.

Screening of BAC libraries and characterization of BAC clones

Two genomic BAC libraries were used in this study: the BAC library of *Ae. tauschii* accession AL8/78 kindly provided by Dr. J. Dvorak (UC Davis, Davis, USA), and the BAC library of bread wheat Xiaoyan 54 constructed in our group. The *Ae. tauschii* BAC library was screened by

Table 1 PCR primers defining the molecular markers used in this work

Primer name	Forward sequence (F)	Reverse sequence (R)	T_m (°C)
PSR567	5′ CAGATGCAGGGGTCTACAGTC 3′	5′ TCCAGATACTCTGCAACAGTG 3′	55
PSR567F	5′ CCTGCACATTAATCAGC 3′	5′ TCCAGATACTCTGCAACAGTG 3′	55
WR001	5′ TGGATTAGTTGATGCACCCA 3′	5′ 5′ CGTGGGATGGCACGAGCAT 3′	60
WR002	5′ ATTGGCGTGATCGATGGTGG 3′	5′ GTGCCGAACAGATTGCGAGA 3′	55
WR003	5′ GGGACAGAGACCTTGGTGGA 3′	5′ GACGATGATGATTTGCTGCTGG 3′	55
NestP1	5′ CATAGACTCTGATCACCAAGCC 3′	5′ GAGTGTGGTTGAAGAGCATCA 3′	56
NestP2	5′ TGTTTCTGCACAAGAATGCC 3′	5′ TTGGTTGCCACGTGTTCA 3′	56
Lr1rga1-1	5′ ATGGCGGCGGCTCTCG 3	5′ AAC CTC ATT TCC CCG TCA A 3′	55
Lr1rga1-2	5′ GAGTATTATTTGGATGATCGATGC 3′	5′ TCTGTAGTTGGTCCACCAAGG 3′	55
Lr1rga1-3	5′ GGGACAGAGACCTTGGTGGA 3′	5′ GTTAGCAATACGAGACGGATAAATCTG 3′	55

Southern hybridization with RFLP marker PSR567F, and the hexaploid wheat BAC library was screened by PCR with primers PSR567F, nestP1, and nestP2 (Table 1). The PCR reaction solution (25 µl) contained 0.5–1 U of rTaq DNA polymerase (TaKaRa, Shiga, Japan), 1 × PCR buffer (50 mM KCl, 10 mM Tris–HCl pH 8.3, 1.5 mM $MgCl_2$, and 0.001% gelatin), 200 µM dNTPs, 400 nM primers, and 50 ng of BAC DNA templates. Reactions were performed in a PTC-225 thermocycler (MJ-Research, Watertown, MA, USA) as follows: About 1 cycle of 95°C for 5 min, 36 cycles of 95°C for 1 min, 55–56°C for 1 min, 72°C for 1 min, with a final 5 min extension reaction at 72°C. The PCR products were then fractionated on a 1% agarose gel and visualized by ethidium bromide staining.

For characterization of BAC clones, BAC DNA was purified with the "Qiagen Plasmid Purification System Midi" kit (Qiagen, Hilden, Germany). The purified BAC DNA was digested by *Not*I and *Asc*I, and subsequently separated by pulse field gel electrophoresis (PFGE) on 1% agarose gels. The PFGE was run for 15–17 h at 6 V/cm and 12–14°C with an initial switch of 5 s and a final switch of 15 s. The DNA was visualized by ethidium bromide staining. Southern analysis of BAC clones was performed as described above for genomic DNA.

For fingerprint analysis, the BAC clones were digested by restriction enzymes *Not*I, *Hind*III, *Eco*RI and separated by pulsed field gel electrophoresis or normal electrophoresis. The relationships between the BAC clones were defined by analysis of the restriction enzyme patterns and Southern hybridization with corresponding probes.

Shotgun cloning and low-pass sequencing of BAC clone 68F22

For isolation of low-copy sequences, a shotgun library of BAC clone 68F22 and sequencing of shotgun clones was performed as described by Stein et al. (2000).

PCR isolation of *LR1RGA1* alleles

The allelic sequences of *LR1RGA1* were isolated by PCR amplification from genomic and BAC DNA of diploid and hexaploid species using the three primer pairs Lr1rga1-1, Lr1rga1-2, and Lr1rga1-3 (Table 1), which were designed with 200–300 bp overlap between fragments. Each fragment of each allele was produced by two independent PCR reactions, and the PCR products were cloned into the pGEM-T vector. Two to three clones of each PCR reaction were sequenced. Whole sequence of each allele was assembled using the software Contig Express (Vector NTI Suite 7.0).

Prediction of protein sequence and functional domains

The *Ae. tauschii* RGA sequences of the 567R family and the homologous gene sequences of *LR1RGA1* from wheat and *Ae. tauschii* lines were translated into protein sequences using the online gene prediction program Fgenesh (http://sunl.Softberry.com/cgi-bin/programs/gfind/Fgenesh.pl). Identification of all functional motifs in the LR1RGA predicted protein [coiled coil domain, nucleotide binding site, and LRR (CC-NBS-LRR)] was performed according to Meyers et al. (1999, 2003). The coiled-coil domain prediction was performed using the online program COILS (http://www.ch.embnet.org/software/COILS-form.html).

Construction of a phylogenetic tree

The amino acid sequences of four RGAs from *Ae. tauschii*, three RGAs from barley, three cloned *R* genes (*Lr10*, *Lr21*, and *Pm3b*) from wheat, and *Xa1* from rice were aligned using ClustalX (Thompson et al. 1997) and the program MEGA Version 3.1 (Kumar et al. 2004) was used to produce the phylogenetic tree based on the neighbor–joining (NJ) method.

The K_a/K_s analysis

The sequence alignment of the putative *Lr1* alleles was performed using ClustalX and visualized using the Genedoc software (http://iubio.bio.indiana.edu/soft/molbio/ibmpc/genedoc-readme.html). DnaSP Version 4.0 (Rozas et al. 2003) was used for analysis of polymorphism data. The average frequency of synonymous (K_s) and non-synonymous (K_a) nucleotide substitutions between the alleles was calculated using the program MEGA Version 3.1. A Student's *t*-test was used to test for significance of differences between K_a and K_s ratios.

Results

New marker development and high-resolution genetic mapping of the *Lr1* locus

In earlier studies, we have shown that the first 250 bp of the RFLP probe PSR567 (1,285 bp), which is genetically closely linked to the *Lr1* resistance gene, encodes a protein sequence which is highly similar to the LRR region of the bacterial blight-resistance protein *Xa1* of rice, whereas the rest of the probe was a non-coding sequence (Ling et al. 2003). Here, a new probe with a length of 710 bp was derived from the non-coding sequence of PSR567 and designated as PSR567F. Southern hybridization revealed that

PSR567F was more specific than PSR567, showing less hybridizing bands in *Ae. tauschii* and in hexaploid wheat (data not shown). Additionally, PSR567F was mapped at the same locus as PSR567 and was tightly linked to *Lr1* both in wheat (0.04 cM) and *Ae. tauschii* (0.47 cM) (Fig. 1).

Given the higher specificity of PSR567F compared to PSR567, the genomic BAC library of *Ae. tauschii* accession AL8/78 was screened by hybridization with PSR567F. Three positive BAC clones 63N23, 68F22, and 88K16 were identified and isolated. PFGE of BAC DNAs digested with restriction enzymes *Not*I and *Asc*I showed that the approximate sizes of 63N23, 68F22, and 88K16 were 100, 200, and 120 kb, respectively. Fingerprint analysis revealed that the three BAC clones overlapped, spanning ~200 kb at the PSR567 locus and that BAC clone 68F22 covered the other two BAC clones almost completely (Fig. 2). Thus, 68F22 was selected for further characterization.

The DNA of 68F22 was sheared and fragments of ~2 kb were cloned to generate two low-pass shotgun libraries. Ninety-eight clones from each shotgun library were sequenced. The sequences of the 196 clones were analyzed using BLASTN and BLASTX programs (Altschul et al. 1997) against the public database. About 21 clones represented contaminating genomic sequences of *Escherichia coli*, whereas 25 of the remaining 175 clones were confirmed to be low-copy sequences by Southern hybridization of shotgun clones with wheat genomic DNA. Two (F22–98 and F22–139) of the 25 low-copy sequences gave a BLASTX hit to the bacterial blight resistance protein *Xa1* of rice. They overlapped and were assembled into a

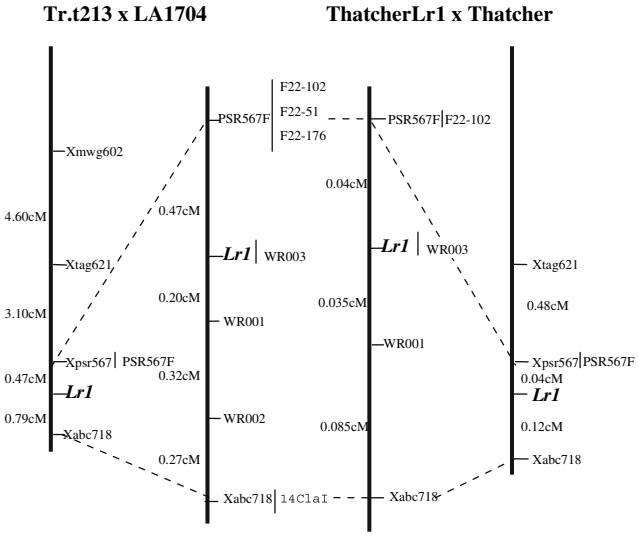

Fig. 1 Genetic mapping of new markers developed from low-pass sequences of BAC clones 68F22 and RGAs identified around the *Lr1* locus in the mapping populations of hexaploid Thatcher*Lr1* × Thatcher and *Ae. tauschii* Tr.t 213 × LA1704

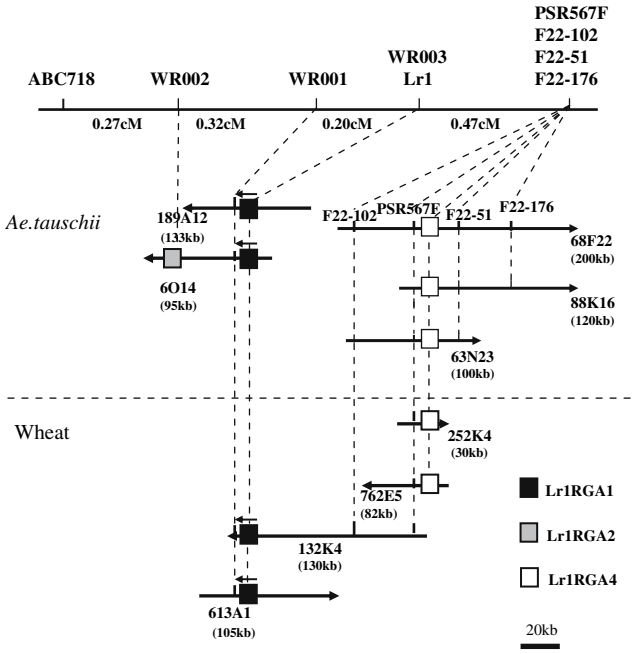

Fig. 2 BAC contig encompassing the *Lr1* locus. BAC clones 88K16, 63N23, and 68F22 were isolated from the BAC library of *Ae. tauschii* accession AL8/78, 6O14, and 189A12 were two BACs isolated previously from the BAC library of *Ae. tauschii* accession Aus18913 by Ling et al. (2003), and the other BACs were from hexaploid wheat. The direction of *arrows* indicates BAC direction from forward (*left*) to reverse (*right*) end. The *arrows* above the *black boxes* indicate the transcriptional orientation of the gene

3,410 bp sequence. Its full coding sequence (4,529 bp) was then obtained using primer walking on BAC 68F22. Sequence analysis revealed that the RGA from BAC 68F22 had a high sequence identity at the DNA level to 567A (91%), 567B (88%), and 567C (86%) (hereafter called *LR1RGA1*, *LR1RGA2*, and *LR1RGA3*, respectively), described by Ling et al. (2003). The high homology to *LR1RGA1*, *LR1RGA2*, and *LR1RGA3* suggests that it represents a new member of the 567R family and it was therefore designated *LR1RGA4*.

To develop new RFLP markers from shotgun sequences, polymorphisms of the low-copy sequences were analyzed with parental lines of the mapping populations. Five of twenty-five sequences (F22–102, F22–51, F22–176, F22–133, and F22–99) displayed polymorphism between Thatcher and Thatcher*Lr1* and/or between *Ae. tauschii* accessions TA1704 and Tr.t213. The five sequences were derived from different regions of the 68F22 BAC clone (Fig. 2). Even though they spanned more than 100 kb physical distance, they all mapped at the same locus as PSR567F and no recombination event was detected in the two large mapping populations.

Thus, here and in earlier work (Ling et al. 2003) a total of four RGAs (LR1RGA1, LR1RGA2, LR1RGA3, and LR1RGA4) were detected from *Ae. tauschii* BACs identified

by RFLP markers PSR567 and PRS567F. Development of gene-specific markers from the four RGAs was expected to enable the identification of a candidate gene for *Lr1*. Therefore, the genomic sequences of the four RGAs were amplified from susceptible (Thatcher and TA1704) and resistant lines (Thatcher*Lr1* and Tr.t213) by PCR using gene specific primers (Table 1). The PCR products were cloned and sequenced. Sequence differences between genes in the parental lines of the mapping populations were detected by ClustalW alignment. Based on sequence polymorphisms, three new markers could be developed from *LR1RGA1* and *LR1RGA2*. The first one was the SSR marker WR001 derived from the region 1,482 bp downstream of the predicted stop codon of *LR1RGA1*. This microsatellite showed a simple sequence repeat difference of 9 bp between Thatcher and Thatcher*Lr1* and 6 bp between TA1704 and Tr.t213 (Fig. 3a). WR001 was mapped 0.035 cM from *Lr1* in wheat and 0.20 cM from *Lr1* in *Ae. tauschii* mapping populations, on the same side as RFLP marker ABC718 (Fig. 1). The second marker (WR002) is based on an insertions and deletions (Indel) and was derived from the 5′-end of *LR1RGA2* sequence (Fig. 3b). WR002 displayed polymorphism only in the population of Tr.t213 × TA1704. It was localized between WR001 and ABC 718, 0.32 cM from WR001 and 0.27 cM from ABC718 (Fig. 1). The third marker was a CAPS marker (WR003) derived from a 760 bp PCR product of *LR1RGA1*, which showed polymorphism between parental lines of the *Ae. tauschii* and wheat mapping populations when digested with *Bgl*II (Fig. 3c). Genetic linkage analysis showed that WR003 co-segregated with *Lr1* both in wheat and *Ae. tauschii* populations (Fig. 1).

Construction of physical map of the *Lr1* locus

To establish a complete physical map of the *Lr1* locus, probe PSR567F was used to screen the BAC library of bread wheat cv. Xiaoyan 54 by PCR. Five BAC clones

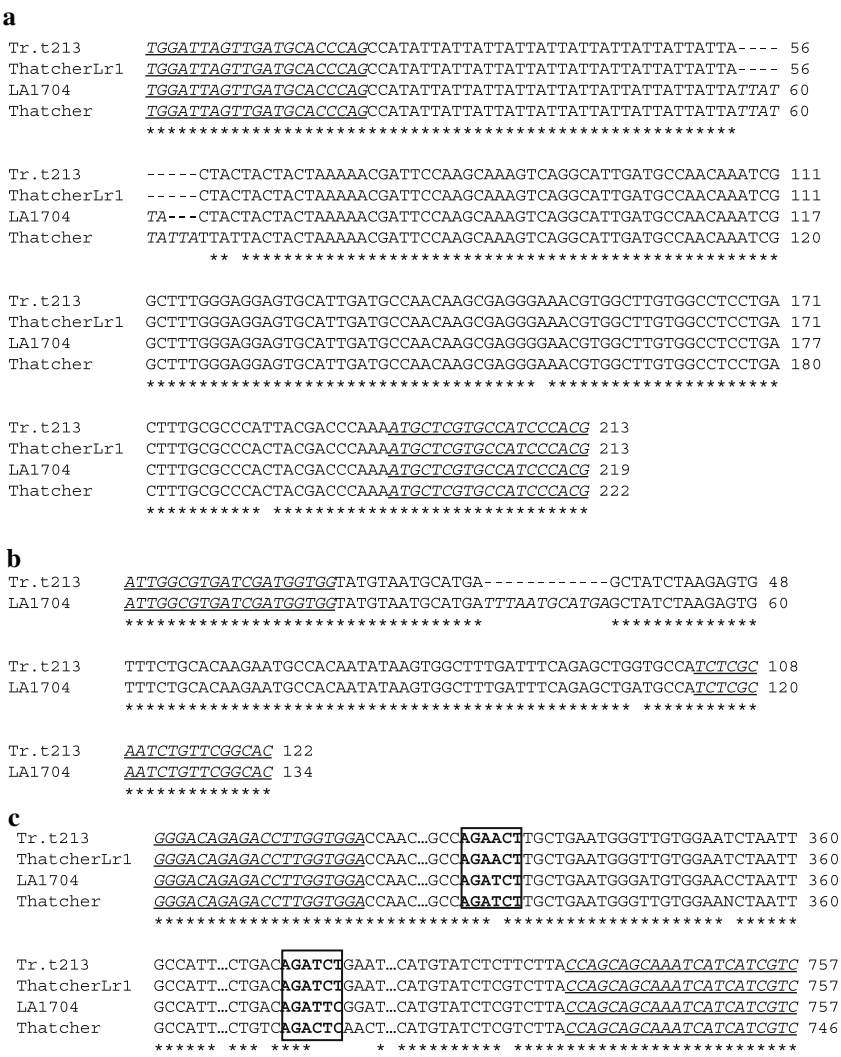

Fig. 3 Sequence comparisons of PCR-based markers. Alignment of DNA sequences from the parental lines of the mapping populations at the WR001 locus (**a**), the WR002 locus (**b**), and the WR003 locus (**c**). *Italic letters* indicate sequence differences, *underlined italic letters* indicate forward and reverse primers for the PCR reactions, *bold letters* in *boxes* indicate polymorphic recognition sites of restriction endonuclease *Bgl*II

(132K4, 823J3, 657D4, 762E5, and 444I5) were isolated and assigned to the same genomic region based on DNA fingerprinting patterns with the restriction endonucleases *Asc*I, *Not*I, *Hind*III, and *Eco*RI. The BAC clones 132K4 (130 kb) and 762E5 (82 kb) were the two largest clones among the five BACs and overlapped by ~50 kb (Fig. 2). The contig of 132K4 and 762E5 spanned about 160 kb and completely covered the other three BACs of the contig (data not shown). Two fragments 132-567-1 (1,007 bp) and 132-567-2 (607 bp) were amplified by PCR with PSR567 primers (Table 1) in the BAC clone 132K4, whereas only one band (132-567-1) was produced in BAC 762E5. Further sequence analysis revealed that 132-567-1 has 89% and 132-567-2 90% sequence identity to PSR567 at the DNA level. These results indicate that there are two homologues of PSR567 present in this contig.

To establish a physical contig containing markers flanking *Lr1* from bread wheat, two pairs of nested primers (Table 1), designed from the conserved region of the four RGAs described above, were also used to screen the wheat BAC library of Xiaoyan 54. Twenty-three BAC clones were isolated and 46 BAC ends were obtained by direct sequencing of BACs. Sequence analysis showed that most of them were repetitive DNA, which could not be mapped. Fingerprint analysis and Southern hybridization of BAC ends resulted in two additional new BAC clones (613A1 and 252K4) extending the contig obtained with the PSR567F probe (132K4 and 762E5; Fig. 2). Using the genetic markers in the linkage map as anchor markers, the BAC clones were positioned around the *Lr1* locus by PCR analysis and Southern hybridization. BAC 613A1 gave a positive signal with markers WR001 and WR003, and 132K4 gave a positive result with WR001, WR003, and F22–102, indicating that BACs 613A1 and 132K4 are localized at the *Lr1* locus, and the whole contig spans ~220 kb around *Lr1*. To confirm the physical map, full-length sequences of the *LR1RGA1* coding region from BAC clones 613A1 and 132K4, and of the *LR1RGA4* from 252K4 and 762E5 were isolated and sequenced. The *LR1RGA1* and *LR1RGA4* sequences of Xiaoyan54 BAC clones shared 98% sequence identity to the corresponding *LR1RGA1* and *LR1RGA4* sequences of *Ae. tauschii*, demonstrating that the 220 kb large BAC contig from bread wheat derives from the *Lr1* region and contains the RGAs *LR1RGA1* and *LR1RGA4*. To identify whether there exists other related RGA(s) in the contig, the BACs were analyzed by Southern hybridization with the conserved sequence of the four RGAs. No additional hybridization signal was detected, indicating that there are only two related RGAs existing in the bread wheat BAC contig. However, we cannot exclude the possibility that there may be other RGAs in this interval, which are not related at the DNA level. Furthermore, sequence analysis revealed that the forward end of BAC clone 1135F4 contained a partial 639 bp match to LR1RGA3. Due to lack of polymorphism, this BAC could not be localized at its exact location around the *Lr1* locus. The wheat BAC contig overlapped with *Ae. tauschii* BAC 189A12 (Ling et al. 2003) in a stretch of about 100 kb, and with 68F22 in 90–100 kb. In summary, a BAC contig of ~400 kb spanning the *Lr1* locus was constructed (Fig. 2).

Characterization of *LR1RGA1* allelic sequences

As described above, of the four identified RGAs, *LR1RGA1* was located in the BAC contig spanning the *Lr1* locus and was the only known RGA in the contig, which co-segregated with *Lr1*. The *LR1RGA1* gene is 4,036 bp long and encodes a predicted protein of 1,344 amino acids. The analysis of *LR1RGA1* revealed that it encodes a CC-NBS-LRR type of protein containing eight motifs characteristic of the NBS region of plant NBS-LRR proteins (Martin et al. 2003; Meyers et al. 1999, 2003) and 27 imperfect LRRs in the C-terminal domain (Fig. 4). These results suggest that *LR1RGA1* is a very good candidate for *Lr1*. Therefore, the sequences of *LR1RGA1* were isolated from resistant (ThatcherLr1 and Tr.t213) and susceptible (Thatcher, Xiaoyan 54, and TA1704) lines by PCR and they are described hereafter as $LR1RGA1^{ThLr1}$, $LR1RGA1^{213}$, $LR1RGA1^{Th}$, $LR1RGA1^{XY54}$, and $LR1RGA1^{1704}$, respectively. Sequence comparison showed that the coding sequences of *LR1RGA1* alleles from *Ae. tauschii* and wheat had a very high percentage of sequence identity (ranging from 95 to 99%) both at DNA and amino acid levels (see supplementary Table 1). The sequence alignment of the *LR1RGA1* alleles showed a clustering of sequence polymorphisms in the LRR coding region. Seventeen synonymous and 48 non-synonymous changes were detected in the LRR region. Except for two synonymous mutations in the susceptible allelic sequences $LR1RGA1^{XY54}$ and $LR1RGA1^{1704}$, all these polymorphisms were located in a 605 bp sequence block (between 2,479 and 3,084 bp, corresponding to amino acids 827–1,028), encoding the 9th to 15th LRR repeats of *LR1RGA1*. It represents the only difference between resistance alleles ($LR1RGA1^{ThLr1}$ and $LR1RGA1^{213}$) and the susceptible alleles ($LR1RGA1^{XY54}$, $LR1RGA1^{Th}$, and $LR1RGA1^{1704}$) in the LRR region. Within the polymorphic sequence block, the 48 non-synonymous changes resulted in 44 amino acid changes (Fig. 5). Thirteen of them affected solvent exposed residues of the LRR. In addition, the fact that synonymous changes between *LR1RGA1* alleles (15 in total) were clustered in this region suggested that this polymorphic sequence block has an ancient origin. If one assumes *LR1RGA1* is *Lr1*, the pattern of sequence polymorphism restricted to sequence blocks embedded into identical conserved sequences indicates that the *Lr1* resistance evolved

Fig. 4 Amino acid sequence encoded by the *LR1RGA1* from wheat variety Thatcher*Lr1*, representing the candidate gene for *Lr1*. The CC, NBS, spacer, and LRR domains are indicated. The *underlined sequence* represents the coiled-coil motif in the N-terminal domain. Amino acids belonging to characteristic motifs in the NBS domain are indicated in *underlined italics*, and arranged in the following order: P-loop, RNBS-A, kinase2, RNBS-B, RNBS-C, GLPL, RNBS-D, and MHDV. In the LRR domain, the conserved residues in the LRR consensus motif (xxLxxLxxLxLxx(C/N/T)x(x)Lx, Jones and Jones 1997) where L is an aliphatic or hydrophobic residue and x any other residue are indicated in *bold italics*

MAAALGSAATLLGKVFTMLSAAPVAAYVDSLELGHNSQQIRAKLAHTRGLLHNAQAQVSDVGHNPGLQE
LLPALSRNADEAEDLLDELHYFQIHDRLHGTNYAATQANFLRHARNALRHTATSSWAACFSCSSAQDDS
DSTSGDDELRFHRVIFSRKFKSVLQDMQTHCDSVSDLLGTIPTSSMPVAVHRPQ

IGSTIIQDTLYGRRHTFEETVNRIFSCKHPVSVLPIVGP***GGIGKTT***FAQHLYNDARTEEH***FQVRVWVCV***
SSDFTVLKLTREILACIPATEEGGSSSVANETTNLDHLQRSIVRRLKSKRF***LIVLDD***IWKCDSQDQWKT
LLAPFTKGETK***GSMLLVTTRF***PKLAQMMET***IDPLELLGLESNDFFTFFEACIF***GEDNKPEHFEDELAGI
AQKIADKLK***GSPLAAKTVGRL***LHKDLSQKHWNGVLEKHQWLKQQNNDDIMPSLKISYDCLPFDLKK***CFS***
YCGLFPEDHWFTSSEINHFWVAVGIIDSDHQADRNYLEELVDNGFLMKKKEYYLDDRCKQKEFDCYV***MH***
DLMHELSKSVSAQECLN

ISGFDFRADAIPQSVRHLSINIEDRYDANFEEEMSKLREKIDIANVRTLMIFREYEEERTAKILKDS

FKE***I***NS***L***RV***L******F******I***VVKSAQSFPDM
FSK***L***IH***L***QY***L***K***I***SSPHIDGEMRLPST
LSR***F***YH***L***KF***L******D******L***DDWRGSSDLPED
FSH***L***EN***L***HD***F******R******A***ESKLHSNIRNVGK
MKH***L***QR***L***EE***F***H***V***KKESMGFE
LSE***L***GP***L***TE***L***E***G***GLTVRGLEHVATKEEAT
AAK***L***MLKRN***L***KQ***L******E******L***LWDRDLGGPTTDADI
LDA***L***QPHSN***L***RV***L******A******I***VNHGGTVGPSW
LCLDIW***L***TS***L***ET***L***T***L***AGVCWSTLPPFAK
LPN***L***KG***L******K******L***MRISG***M***HQFGSLCGGTPGKC
FMR***L***KT***V******E******F***YEMPE***L***AEWVVESNCHS
FPS***L***EE***I***R***C***RN***C***PN***L***RVMPFSEVS
FTN***L***RT***L******F***V***S***R***C***PK***M***SLPSMPHTST
LTD***L***N***V***GIGDSEGLHYDGKKLIVIGYGGALASHN
LDT***V***ED***M******I***V***E***R***C***DG***L***FPEDLDGSFV
FRS***V***KN***L******T******L***HVSRLTSSKSSSSKV
LNC***F***P***A******L***S***V******L******V******I***VGYEECVMQFPSSSS
LQK***L******T******F***SG***C***R***G******L***VLVPEEKENGGGIQEDNSL
LQS***L******T******I***VG***C***G***K******L***FSRWPMGMGESETICPFPAS
LKK***L***D***V***FQEPS***M***KSMALLSN
LTS***L***TT***L***QL***N***Y***C***SN***L***TVDGFNPLI
AVN***L***IE***L***Q***V***HR***C***N***T******L***AADMLSEAASHSQRAKLLPAGY
ISR***L***EK***L***N***V***DN***N***CG***L***LVAPICNLL
APA***L***HT***L******V******F***WIDET***M***ESLTEEQEKA
LQL***L***TS***L***QN***L******T******F***F***R******C***RG***L***QSLPQG
LHR***L***SS***L***KE***L******C******V***RG***C***LK***I***QSLPKEGLPLS
LRR***L******K******M***NWRSAEINEQIEKIKRSNPDLSVSYC

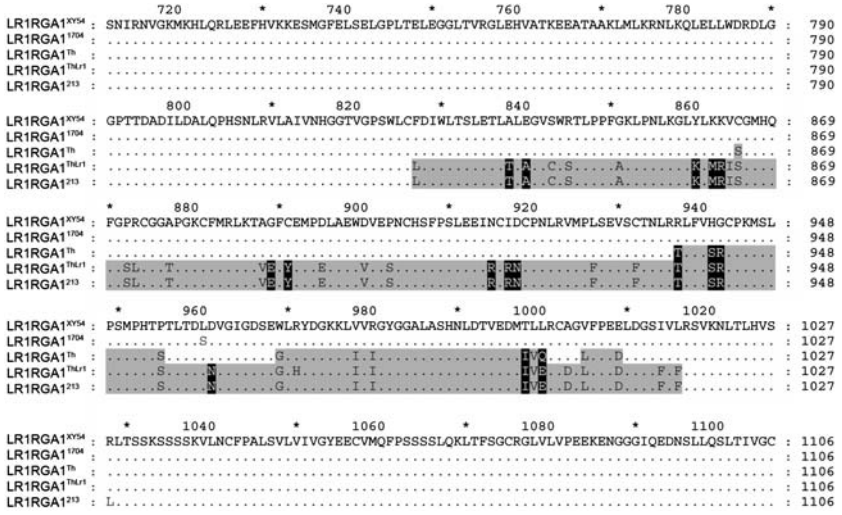

Fig. 5 Partial alignment of the LRR domain encoded by susceptible (*LR1RGA*^{XY54}, *LR1RGA1*^{1704}, and *LR1RGA1*Th) and resistant (*LR1RGA1*^{ThLr1} and *LR1RGA1*^{213}) alleles (the complete alignment of LR1RGA1 protein sequences is available as supplementary figure 1). The comparison of amino acid sequences encoded by susceptible and resistant *LR1RGA1* alleles identifies polymorphic sequence blocks in the LRR region (amino acids 827–1,016) that are possibly resulting from recombination/gene conversion events. *Dots* represent identical amino acids. *Grey shaded areas* represent the blocks of sequences indicative of recombination/gene conversion events. *Black shaded* amino acids within these sequence blocks are polymorphic solvent-exposed amino acids (x residues in the LxxLxLxx motif in LRR repeats)

through one or more recombination or gene conversion events between unknown *Lr1*-related genes. In this region, the sequence of Thatcher (susceptible cultivar) shows a mosaic pattern of conservation with either the resistance alleles or the susceptible alleles suggesting that several small recombination/gene conversion events occurred in this region between *Lr1* alleles (Fig. 5).

Phylogenetic analysis of 567R family

Partial sequences of *LR1RGA1* (567A, 4.7 kb), *LR1RGA2* (567B, 3.2 kb), and *LR1RGA3* (567C, 3 kb) were previously described by Ling et al. (2003). In this study, using primer walking, inverse PCR as well as plasmid rescue from corresponding BAC clones, the sequences of the three RGAs were further extended. In total, 6,052 bp for *LR1RGA1*, 3,913 bp for *LR1RGA2*, and 5,103 bp for *LR1RGA3* were obtained from *Ae. tauschii* BAC clones 189A2, 6O14, and 134C15, respectively (Ling et al. 2003). The sequence of *LR1RGA4*, as described above, originated from two overlapping shotgun clones of *Ae. tauschii* accession AL8/78 BAC 68F22 and subsequent primer walking on BAC 68F22. The sequence alignment revealed that the four RGAs are highly similar in encoded protein sequences (LR1RGA1 has 78% amino acid identity to LR1RGA2, 75% to LR1RGA3, and 81% to LR1RGA4). The four RGAs also show homology to the disease resistance gene analogs of barley (*Hordeum vulgare*) designated as S-T21, S197-4, and S-372 (Madsen et al. 2003). All these seven predicted *R* genes are homologous to the *Xa1* resistance gene of rice (Yoshimura et al. 1998). The phylogenetic tree was constructed based on the amino acid sequences of the seven RGAs, three cloned wheat disease resistance genes (*Lr21*, *Lr10*, and *Pm3b*), and the rice bacterial blight resistance gene *Xa1* (Fig. 6). It shows that LR1RGA1 and LR1RGA4 are most closely related and LR1RGA3 forms a distinct branch. These results suggested that the 567R family might derive from duplication events of an ancient gene during the evolution. *LR1RGA4* is probably a pseudogene, with two frame shift mutations in its coding sequence (one in the spacer region and the other after the 10th LRR) causing stop codons in both cases. The three RGAs from barley (S-T21, S197-4, and S-372) built another branch of this group. *LR1RGA1*, the possible candidate of *Lr1*, has very low similarity to the two other cloned wheat leaf rust resistance genes *Lr10* (13%) and *Lr21* (18%) at the protein level.

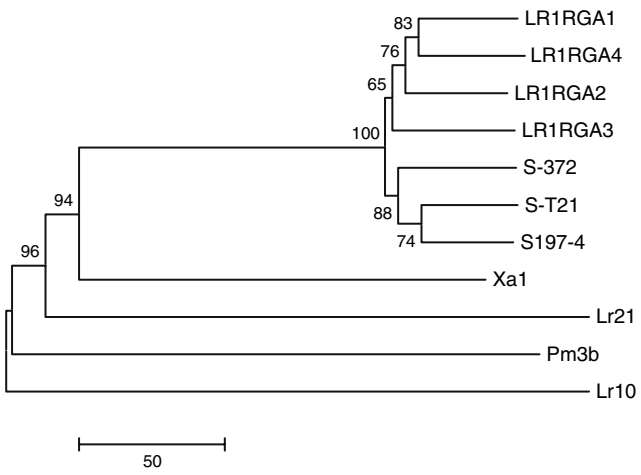

Fig. 6 Phylogenetic tree of the 567R gene family with other NBS-LRR resistance genes from wheat, barley, and rice at the protein level

Discussion

In the present study, we have mapped *Lr1* in a genetic interval of 0.79 cM in *Ae. tauschii* and 0.075 cM in wheat using new markers derived from the RGAs at the *Lr1* locus. A BAC contig of more than 200 kb, encompassing the whole region of *Lr1* was successfully constructed with a hexaploid wheat BAC library. Two (*LR1RGA1* and *LR1RGA4*) of four RGAs identified around *Lr1* were localized in this BAC contig. *LR1RGA1* was the only RGA, which co-segregated with *Lr1* both in the wheat and *Ae. tauschii* mapping populations, whereas *LR1RGA4* was mapped two recombinants away from *Lr1* (Fig. 3). Further sequence analysis revealed that *LR1RGA4* was a pseudogene due to two frame shifts in its coding sequence. Comparison of the coding sequences of *LR1RGA1* alleles isolated from resistant and susceptible lines revealed a divergent sequence block in the region of LRR domain, which includes non-synonymous differences. This polymorphic sequence block clearly distinguishes alleles from resistant and susceptible lines. All these data strongly suggest that *LR1RGA1* is a very good candidate for *Lr1*. The polymorphic sequence block in the LRR domain as well as few point mutations in the NBS of *LR1RGA1* of the susceptible cultivars Thatcher ($LR1RGA1^{Th}$), Xiaoyan54 ($LR1RGA1^{XY54}$), and *Ae. tauschii* accession TA1704 ($LR1RGA1^{1704}$) may cause these genes to be susceptible alleles. Phenotypic complementation via introduction of the candidate gene in a susceptible wheat cultivar or silencing the candidate gene by VIGS [virus induced gene silencing (Scofield et al. 2005)] in a resistant line will be necessary to definitively prove the identity of *Lr1*.

In large size genomes such as wheat, BAC ends can usually not be mapped due to their repetitive nature. Even the non-repetitive BAC ends have several copies in most cases. In addition, most of the plant *R* genes isolated to date are located in a cluster of RGAs (Jones et al. 1993; Song et al. 1997; Hulbert et al. 2001; Jones 2001; Sandhu and Gill 2002; Martin et al. 2003; Qu et al. 2006). The individual members of an RGA gene family normally have high

homology both at the DNA and the protein level and are sometimes separated by repetitive DNA sequences. This makes it difficult to approach the target gene by chromosome walking with BAC ends as RFLP markers. However, sequence differences, such as SNPs and/or small Indel, are present between the allelic RGAs from susceptible and resistant lines. These sequence differences can be converted to polymorphic PCR or CAPS markers as was the case for our markers WR001 and WR003. Since these markers are based on single or a few nucleotide differences, they are more specific than RFLP markers. Therefore, the markers developed from RGAs will be useful for genetic mapping and chromosome walking in large genomes. Furthermore, BACs containing RGAs can be identified by screening BAC libraries using the conserved sequence of RGAs, and the isolated BAC clones can easily be anchored in the target region by sequence analysis of their encoded RGAs. In this work, we used this strategy and successfully developed gene-specific markers based on sequence differences of allelic RGAs from susceptible and resistant parental lines. In summary, the development of gene-specific markers from RGAs is a useful alternative strategy for overcoming the obstacles of genetic mapping and chromosome walking for cloning of genes, especially disease resistance genes, in wheat.

An excess of non-synonymous changes in the solvent exposed residues of the LRR domain (x residues in the LxxLxLxx motif) was shown in several *R* gene families (Michelmore and Meyers 1998). This was indicative of diversifying selection acting on these residues to generate new resistance specificities. We have estimated the average rate of non-synonymous (K_a) and synonymous changes (K_s) for different regions of the *LR1RGA1* resistant and susceptible allelic sequences. An excess of K_a compared to K_s (K_a/K_s ratio higher than 1) was found only for the solvent exposed residues in the LRR coding region of the *LR1RGA1* gene ($K_a/K_s = 1.45$). This seems to indicate a tendency for diversifying selection on these residues although this difference between K_a and K_s was not found to be statistically significant. Since only five sequences were considered here, this relatively low statistical significance may be due to lack of sufficient data. Comparing the resistant and susceptible alleles, we found that the $K_a/K_s > 1$ in solvent exposed residues of the LRR is only significant between resistance versus susceptible alleles but not in the comparison between diploid alleles versus hexaploid alleles. These results indicate that *LR1RGA1* in wheat and *Ae. tauschii* are derived from a common ancient gene. The alleles from two resistant lines share a very similar LRR sequence, which is different from the LRR sequences of the susceptible alleles. If it is assumed that LR1RGA1 is Lr1, this suggests that the resistance phenotype is conferred by the LRR domain. However, we cannot exclude the possibility that some amino acid changes in the CC-NBS domain of the susceptible alleles result in loss of resistance function. The pattern of sequence conservation and divergence between *LR1RGA1* alleles could be explained by gene conversion events introducing clearly delimited polymorphic sequences in the LRR region. Gene conversion was previously identified as an important mechanism of plant resistance gene evolution (Huang et al. 2003; Kuang et al. 2004; Mondragon-Palomino and Gaut 2005). Recently, gene conversion was also shown to be at the origin of polymorphic resistance alleles of the *Pm3* powdery mildew locus in wheat (Yahiaoui et al. 2006). Here, the Thatcher *LR1RGA1* sequence pattern indicated that sequence exchange does occur between *LR1RGA1* alleles. However, the original events that resulted in the polymorphic sequence block between *LR1RGA1* from resistant and susceptible lines might have involved sequence exchange with a paralog of *LR1RGA1*.

A group of RGAs, which are highly homologous to the 567R family at protein level (identities ranged from 71 to 79%), was identified in barley. Among them, RGA S-T21 was mapped to the distal region on chromosome 5HL near MWG602 (Madsen et al. 2003), which maps at 8.2 cM proximal to the *Lr1* gene in *Ae. tauschii* (Ling et al. 2004). Moreover, a copy of PSR567 mapped 1.4 cM distal to MWG602 in the same region on chromosome 5HL (Gallego et al. 1998). Considering that the barley genome is highly collinear with that of *Ae. tauschii* (Dubcovsky et al. 1996), we predict that there may exist an orthologous 567R family in barley.

In conclusion, our genetic and physical mapping results strongly suggest that *LR1RGA1* is the *Lr1* gene, encoding a protein with a CC-NBS-LRR structure similar to many resistance gene products. A highly polymorphic sequence segment clearly distinguishes the alleles from the resistant and susceptible lines of wheat and *Ae. tauschii*, and may cause the functional difference.

Acknowledgments This work was supported by National Natural Science Foundation of China grant 30671130 to HQL and by Swiss National Science Foundation grant 3100A0-105620 to BK.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Ausemus ER, Harrington JB, Reitz LP, Worzella WW (1946) A summary of genetic studies in hexaploid and tetraploid wheats. J Am Soc Agron 38:1083–1099

Dubcovsky J, Luo MC, Zhong GY, Bransteitter R, Desai A, Kilian A, Kleinhofs A, Dvorak J (1996) Genetic map of diploid wheat, *Triticum monococcum* L. and its comparison with maps of *Hordeum vulgare* L. Genetics 143:983–999

Feuillet C, Messmer M, Schachermayr G, Keller B (1995) Genetic and physical characterization of the *Lr1* leaf rust resistance locus in wheat (*Triticum aestivum L.*). Mol Gen Genet 248:553–562

Feuillet C, Keller B (2002) Comparative genomics in the grass family: molecular characterization of grass genome structure and evolution. Ann Bot 89:3–10

Feuillet C, Travella S, Stein N, Albar L, Nublat A, Keller B (2003) Map-based isolation of the leaf rust disease resistance gene *Lr10* from the hexaploid wheat (*Triticum aestivum L.*) genome. Proc Natl Acad Sci USA 100:15253–15258

Gallego F, Feuillet C, Messmer M, Penger A, Graner A, Yano M, Sasaki T, Keller B (1998) Comparative mapping of the two wheat leaf rust resistance loci *Lr1* and *Lr10* in rice and barley. Genome 41:328–336

Graner A, Siedler H, Jahoor A, Hermann RG, Wenzel G (1990) Assessment of the degree and type of polymorphism in barley (*Hordeum vulgare*). Theor Appl Genet 80:826–832

Huang L, Brooks SA, Li W, Fellers JP, Trick HN, Gill BS (2003) Map-based cloning of leaf rust resistance gene *Lr21* from the large and polyploid genome of bread wheat. Genetics 164:655–664

Hulbert SH, Webb CA, Smith SM, Sun Q (2001) Resistance gene complexes: evolution and utilization. Annu Rev Phytopathol 39:285–312

Jones DA, Dickinson MJ, Balint-Kurti, Dixon MS, Jones JDG (1993) Two complex resistance loci revealed in tomato by classical and RFLP mapping of the *cf-2*, *cf-4*, *cf-5*, and *cf-9* genes for resistance to *Cladosporium fulvum*. Mol Plant Microbe Interact 6:348–357

Jones DA, Jones JDG (1997) The role of leucine-rich repeat proteins in plant defences. Adv Bot Res Incorporating Adv Plant Pathol 24:89–167

Jones JDG (2001) Putting knowledge of plant disease resistance genes to work. Curr Opin Plant Biol 4:281–287

Kuang H, Woo SS, Meyers BC, Nevo E, Michelmore RW (2004) Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. Plant Cell 16:2870–2894

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 5:150–163

Ling HQ, Zhu YX, Keller B (2003) High-resolution mapping of the leaf rust disease resistance gene *Lr1* in wheat and characterization of BAC clones from the *Lr1* locus. Theor Appl Genet 106:875–882

Ling HQ, Qiu JW, Singh RP, Keller B (2004) Identification and genetic characterization of an *Aegilops tauschii* ortholog of the wheat leaf rust disease resistance gene *Lr1*. Theor Appl Genet 109:1133–1138

Martin GB, Bogdanove AJ, Sessa G (2003) Understanding the functions of plant disease resistance proteins. Annu Rev Plant Biol 54:23–61

Madsen LH, Collins NC, Rakwalska M, Bacges G, Sandal N, Krusell L, Jensen J, Waterman EH, Jahoor A, Ayliffe M, Pryor AJ, Langridge P, Schulze-Lefert P, Stoufaard J (2003) Barley disease resistance gene analogs of the NBS-LRR class: identification and mapping. Mol Genet Genomics 269:150–161

McIntosh RA, Wellings CR, Park RF (1995) Wheat rusts: an atlas of resistance genes. Kluwer, Dordrecht, pp 1–200

Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Plant J 20:317–332

Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW (2003) Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. Plant Cell 15:809–834

Michelmore RW, Meyers BC (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res 8:1113–1130

Mondragon-Palomino M, Gaut B (2005) Gene conversion and the evolution of three leucine-rich repeat gene families in *Arabidopsis thaliana*. Mol Biol Evol 22:2444–2456

Qu S, Liu G, Zhou B, Bellizzi M, Zeng L, Dai L, Han B, Wang GL (2006) The broad-spectrum blast resistance gene *Pi9* encodes a nucleotide-binding site-leucine-rich repeat protein and is a member of a multigene family in rice. Genetics 172:1901–1914

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497

Sandhu D, Gill KS (2002) Gene- containing regions of wheat and the other grass genomes. Plant Physiol 128:803–811

Scofield SR, Huang L, Brandt AS, Gill BS (2005) Development of a virus-induced gene-silencing system for hexaploid wheat and its use in functional analysis of the *Lr21*-mediated leaf rust resistance pathway. Plant Physiol 138:2165–2173

Song WY, Pi LY, Wang GL, Gardner J, Holsten T, Ronald P (1997) Evolution of the rice *Xa21* disease resistance gene family. Plant Cell 9:1279–1287

Srichumpa P, Brunner S, Keller B, Yahiaoui N (2005) Allelic series of four powdery mildew resistance genes at the *Pm3* locus in hexaploid bread wheat. Plant Physiol 139:885–895

Stein N, Feuillet C, Wicker T, Schlagenhauf E, Keller B (2000) Subgenome chromosome walking in wheat: a 450-kb physical contig in *Triticum monococcum* L. spans the *Lr10* resistance locus in hexaploid wheat. Proc Natl Acad Sci USA 97:13436–13441

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Yahiaoui N, Srichumpa P, Dudler R, Keller B (2004) Genome analysis at different ploidy levels allows cloning of the powdery mildew resistance gene *pm3b* from hexaploid wheat. Plant J 37:528–538

Yahiaoui N, Brunner S, Keller B (2006) Rapid generation of new powdery mildew resistance genes after wheat domestication. Plant J 47:85–98

Yoshimura S, Yamanouchi U, Katayose Y, Toki S, Wand ZX, Kono I, Kurata N, Yano M, Nobuo I, Sasaki T (1998) Expression of *Xa1*, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. Proc Natl Acad Sci USA 95:1663–1668