University of Zurich[UZH]

Zurich Open Repository and Archive
University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2009

Two different CC-NBS-LRR genes are required for Lr10-mediated leaf rust resistance in tetraploid and hexaploid wheat

Loutre, C ; Wicker, T ; Travella, S ; Galli, P ; Scofield, S ; Fahima, T ; Feuillet, C ; Keller, B

DOI: https://doi.org/10.1111/j.1365-313X.2009.04024.x

Posted at the Zurich Open Repository and Archive, University of Zurich
ZORA URL: https://doi.org/10.5167/uzh-29489
Journal Article
Accepted Version

Originally published at:
Loutre, C; Wicker, T; Travella, S; Galli, P; Scofield, S; Fahima, T; Feuillet, C; Keller, B (2009). Two different CC-NBS-LRR genes are required for Lr10-mediated leaf rust resistance in tetraploid and hexaploid wheat. The Plant Journal, 60(6):1043-1054.
DOI: https://doi.org/10.1111/j.1365-313X.2009.04024.x

Two different CC-NBS-LRR genes are required for *Lr10*-mediated leaf rust resistance in tetraploid and hexaploid wheat

Caroline Loutre[1,6], Thomas Wicker[1,6,7], Silvia Travella[1], Paolo Galli[1,2], Steve Scofield[5], Tzion Fahima[3], Catherine Feuillet[1,4] and Beat Keller[1]

[1]Institute of Plant Biology, University of Zurich, Zollikerstrasse 107, 8008 Zürich, Switzerland
[2]Current address: Institute of Integrative Biology, ETH Zürich, 8092 Zürich
[3]Institute of Evolution, University of Haifa, Mt. Carmel, Haifa 31905, Israel
[4]Current address: UMR INRA-UBP 1095, Amélioration et Santé des Plantes 234, Avenue du Brézet, 63100 Clermont-Ferrand, France
[5]Department of Agronomy, 915 W. State Street, Lilly Hall of Life Sciences, Purdue University, West Lafayette, Indiana 47907-2054, USA
[6]Both authors contributed equally to this work
[7]Corresponding author

Corresponding author's email address: wicker@botinst.uzh.ch

Running title: Leaf rust resistance in wheat

Keywords: resistance gene evolution, virus-induced gene silencing, diversifying selection, resistance gene analog, leucine zipper, tetraploid wheat

Author's email addresses

Caroline Loutre	carolineloutre@yahoo.fr
Thomas Wicker	wicker@botinst.uzh.ch
Silvia Travella	stravella@botinst.uzh.ch
Paolo Galli	paolo.galli@agrl.ethz.ch
Steve Scofield	scofield@purdue.edu
Tzion Fahima	fahima@research.haifa.ac.il
Catherine Feuillet	catherine.feuillet@clermont.inra.fr
Beat Keller	bkeller@botinst.uzh.ch

Summary

Comparative study of disease resistance genes in crop plants and their relatives provides insight on resistance gene function, evolution and diversity. Here, we studied the allelic diversity of the *Lr10* leaf rust resistance gene, a CC-NBS-LRR coding gene originally isolated from hexaploid wheat, in 20 diploid and tetraploid wheat lines. Besides a gene in the tetraploid wheat variety '*Altar*' that is identical to the hexaploid wheat *Lr10*, two additional, functional resistance alleles showing sequence diversity were identified by virus-induced gene silencing in tetraploid wheat lines. In contrast to most described NBS-LRR proteins, the N-terminal CC domain of LR10 was found to be under strong diversifying selection. A second NBS-LRR gene at the *Lr10* locus, *RGA2*, was shown through silencing to be essential for *Lr10* function. Interestingly, *RGA2* showed much less sequence diversity than *Lr10*. These data demonstrate allelic diversity of functional genes at the *Lr10* locus in tetraploid wheat and these new genes can now be analysed for agronomic relevance. *Lr10*-based resistance is highly unusual both in its dependence on two, only distantly, related CC-NBS-LRR proteins as well as in the pattern of diversifying selection in the N-terminal domain. This indicates a new and complex molecular mechanism of pathogen detection and signal transduction.

Introduction

Single dominant or co-dominant resistance (*R)* genes represent a highly effective defense system of plants against pathogens, in particular obligate biotrophic pathogens such as rust fungi. The products of *R* genes act as immune receptors which directly or indirectly sense the presence of a pathogen and trigger strong defense responses that frequently result in cell death (Jones and Dangl, 2006) and completely stop further pathogen growth. When used in elite plant cultivars under agricultural conditions, *R* genes are frequently not effective over longer time periods due to genetic adaptation of the pathogen. Newly evolving pathogen races can "overcome" a specific *R* gene by avoiding detection through *R* gene products. Therefore, there is a constant need for additional, new resistance genes in plant breeding. It was suggested that landraces, germplasm from the primary and secondary gene pools of a particular species as well as wild relatives are ideally suited to identify new functional variants of agronomically relevant genes (Feuillet et al. 2008; Baum et al. 1992).

Typically, *R* genes encode proteins with an N-terminal coiled-coil (CC) or Toll-interleukin receptor type (TIR) domain, a nucleotide binding site (NBS) domain and an extended domain of leucine-rich repeats (LRR, for recent reviews see Bent and Mackay, 2007; McDowell and Simon, 2008). Typically, diversifying selection was observed in the LRR domain, particularly in the xxLxLxx motif of the individual LRR units proposed to form a solvent-exposed β-sheet (Dodds et al., 2001). It was suggested that, in cases of a direct interaction of the R gene product with the pathogen effector, these amino acids are conferring biochemical specificity to the interaction (Jones and Jones, 1997; Dodds et al., 2001). In contrast, the N-terminal domain is usually highly conserved in allelic series of *R* genes. For example, in the MLA or PM3 powdery mildew resistance proteins in barley and wheat, respectively, the CC domain is completely conserved (Bieri et al., 2004; Yahiaoui et al., 2006). A whole genome

study of paralogous NBS-LRR genes from Arabidopsis showed that most of the positively selected sites were in the LRR domain, whereas only few such sites were identified in the N-terminal domain (Mondragon-Palomino et al., 2002).

The N-terminal domain of R proteins has been shown to interact with a variety of different proteins. For example, the N-terminal domain of RPS5 interacts with the host protein PBS1 which is the target of a pathogen effector (Ade et al. 2007) whereas the CC domain of the barley MLA protein interacts with WRKY transcription factors (Shen et al., 2006). Furthermore, there is evidence that the N-terminal domain of the tobacco N resistance protein associates with the pathogen ”effector” molecule p50 of the tobacco mosaic virus (Burch-Smith et al. 2007). Only in one case, the TIR domain of an *R* gene (flax *L* gene*)* was found to be partially under diversifying selection (Ayliffe et al., 1999, Luck et al., 2000). These data suggest that the N-terminal domain of R proteins can be involved in both recognition and signalling processes as well as intramolecular interactions (Rairdan et al., 2008).

NBS-LRR genes frequently occur in clusters and the evolution of new alleles is driven primarily by point mutations which are then reshuffled by illegitimate recombination or gene conversion, either within or between genes (Michelmore and Meyers 1998; Kuang et al. 2004; Wicker et al. 2007). There are also simple genetic loci with single *R* genes such as the *Rps5* gene in Arabidopsis (Tian et al., 2002) and the *L* gene in flax (Luck et al., 2000). The allelic diversity of *Rps5* in the gene pool of Arabidopsis was shown to be very low (Tian et al., 2002) while the *L* allelic series shows functionally different variants of the gene.

The resistance gene *Lr10* against wheat leaf rust *Puccinia triticina* encodes a CC-NBS-LRR protein and occurs together with a second CC-NBS-LRR coding gene called *RGA2* (Feuillet et al*.,* 2003). LR10 and RGA2 have only very distantly related protein sequences and belong to different ancient evolutionary lineages in grass genomes (Wicker et al. 2007). While *Lr10* and *RGA2* are present on chromosome 1AS only in a

subset of hexaploid wheat lines, a homoeologous *RGA2* gene is always found on chromosome 1DS in hexaploid wheat (Scherrer et al., 2002). Interestingly, two major haplotypes were observed at the *Lr10* locus on 1AS in diploid, tetraploid and hexaploid wheat: either both genes were present (haplotype H1) or both genes were absent (haplotype H2) (Scherrer et al., 2002). This demonstrated that the two haplotypes are evolutionary ancient, with a presence *vs*. absence polymorphism indicative of balancing selection. Moreover, large-scale sequencing of the *Lr10* locus showed that the H2 haplotype is a deletion derivative of the H1 haplotype (Isidore et al., 2005). Finally, the H1 haplotype was found in less than 25% of a collection of 62 hexaploid lines investigated, but was predominant in diploid and tetraploid accessions (Scherrer et al., 2002).

Here, we wanted to study the molecular diversity and evolutionary forces that have shaped the *Lr10* resistance locus and the possible role of *RGA2* in *Lr10*-mediated resistance. Allele-mining in the wheat gene pool yielded two new functional alleles of *Lr10*. Additionally, the analysis of a 22 *Lr10* alleles from diploid and tetraploid wheat revealed strong diversifying selection acting on the N-terminal domain, suggesting that it is the site of interaction with pathogen effector molecules. Finally, functional studies demonstrated that the *RGA2* gene is required for *Lr10*-mediated resistance, providing evidence for the need of two different CC-NBS-LRR proteins for leaf rust resistance.

Results

Characterisation of *Lr10* orthologous genes from wild and domesticated diploid and tetraploid wheat species

In this work we analyzed two genes present at the wheat *Lr10* locus: *Lr10* and *RGA2*. We named the genes to include the name of the wheat variety as well as the type of gene (e.g. the two genes from cv. *Langdon* will be referred to as *Langdon_Lr10* and

Langdon_RGA2). To study *Lr10* diversity in the wheat gene pool, the full-length coding sequence of *Lr10* was amplified by PCR from 18 accessions of cultivated tetraploid wheat *T. turgidum ssp. durum*, wild tetraploid wheat *T. turgidum ssp. dicoccoides* and the diploid donor of the A-genome, *T. urartu* (Table 1). The previously published *Lr10* and *RGA2* sequences of *[Thatcher+Lr10]* (a near-isogenic line of hexaploid bread wheat cultivar *Thatcher* with the introgressed *Lr10* gene, McIntosh et al. 1995), *Can3842* (Feuillet et al., 2003), the tetraploid wheat variety *Langdon* (Isidore et *al*., 2005) and the diploid wheat *T. monococcum* (cultivar DV92, Wicker et al., 2001) were also included in the analysis.

The position of the single intron was identified by comparison with the *[Thatcher+Lr10]_Lr10* reference sequence and protein sequences were predicted for all 22 sequences. In multiple alignments of the DNA sequences, *[Thatcher+Lr10]_Lr10* was used as reference sequence with which all others were compared (Figure 1). The total number of base substitutions compared to *[Thatcher+Lr10]_Lr10* ranged from 0 (identical sequence *Altar_Lr10*) to 147 in the *T. monococcum* sequence *DV92_Lr10* (Figure 1). Non-synonymous substitutions outnumbered synonymous substitutions, ranging from 0 in *Altar* to 80 in *T. monococcum* (Figure 1). Interestingly, the numbers of substitutions did not correlate with the species' or subspecies' phlylogenetic origin as the accessions within single species/subspecies showed a wide range of base substitution numbers (except *T. monococcum* where only one sequence was available). For example, the eight sequences from tetraploid *T. turgidum ssp. durum* had 0 to 117 substitutions while those from *ssp. dicoccoides* showed between 62 and 128 substitutions (Figure 1).

The coding sequence from tetraploid wheat *T. turgidum* ssp. *durum Altar_Lr10* was identical to *[Thatcher+Lr10]_Lr10* (Figure 1). The two only differ by a single base pair insertion in the intron sequence (not shown). The sequence of the second hexaploid wheat variety *Can3842* differs in 81 bases from *[Thatcher+Lr10]_Lr10*. Surprisingly, the

most similar gene to *Can3842_Lr10* was found in the diploid *T. urartu* accession *TRI6735* which differs by only 5 base pairs from *Can3842_Lr10*.

The distribution of nucleotide substitutions is very uneven as the different *Lr10* genes are a mosaic of segments with varying degrees of homology to *[Thatcher+Lr10]_Lr10* (Figure 1). These regions explain much of the observed sequence diversity among *Lr10* homologs and are easily identifiable by simple visual inspection (Figure 1). Analysis with DnaSP (see methods) showed a statistically highly significant probability that they are the result of multiple gene conversion events where homologous regions from different haplotypes had been integrated (Figure 1). The NBS and LRR domains both contain one such region with distinct motifs of at least two different haplotypes where only few nucleotides differ between individual sequences of the same haplotype while there are dozens of polymorphisms between haplotypes (Figure 1, Supplementary Figure 2).

The N-terminal domain of *Lr10* is under strong diversifying selection

Highest sequence diversity was found in the N-terminal domain of *Lr10* where the 22 sequences are complex mosaics of at least four different haplotype segments (Figure 1, Figure 2 and Supplementary Figure 2). In fact, the pattern of breakpoints between haplotype sequences was too complex for the sequences to be sorted out into distinct groups (Supplementary Figure 2). We used PAML to calculate the ratio of non-synonymous to synonymous substitution (Ka/Ks) and to identify sites with a high probability for being under positive (i.e. diversifying) selection. Interestingly, most such sites lay in the CC domain within the N-terminal 130 amino acids where 20 of the 26 sites with a posterior probability of more than 95% and 14 of the 18 sites with a higher posterior probability of more than 99% were found (Figure 1, Figure 2). The 4^{th} and 5^{th} loop of a predicted CC motif (Feuillet et al., 2003) are particularly diverse as 9 of 15 amino acids have a high probability to be under diversifying selection (Figure 2). In this

region, there are several codon positions where up to five different amino acids are encoded in the 22 sequences (loops 3, 4 and 5, Figure 2).

Despite the high sequence diversity, the hydrophobic backbone sequence of the leucine zipper is still perfectly conserved. Only one amino acid position in the hydrophobic backbone of the predicted CC motif has a high probability for being under diversifying selection (Figure 2).

The region further downstream (between amino acid positions 100-130) contains several insertion/deletions (indels) with different breakpoints, reflecting a series of independent indel events (Figure 2). Eleven sites with a high likelihood for diversifying selection were also found in a region of the LRR domain which presumably has its diversity from a gene conversion event (see above). The density of positively selected amino acids in that region is, however, much lower than in the CC domain (Figure 1).

We asked the question whether the evidence for diversifying selection is a mere consequence of the high number of polymorphisms introduced by the several gene conversion events detected. This, however, seems not to be the case because not all gene conversion regions contain positively selected amino acid residues. Only one positively selected site was found in the gene conversion region in the NBS domain of *Lr10* and none at all in a gene conversion region in *RGA2* (see below). Finally, the number of polymorphisms between haplotypes was similar in the gene conversion regions of the NBS and LRR of *Lr10* (Supplementary Figure 2), but the number of positively selected sites differed strongly (one in the NBS versus eleven in the LRR domain).

For comparison, we performed the same analysis on 12 publicly available alleles of flax rust resistance gene *L* whose N-terminal domain has been described as possibly being under diversifying selection (Luck et al. 2000). We found by PAML analysis a lower density of amino acids under diversifying selection than in LR10: within the TIR domain

(N-terminal ~180 aa), we found 6 amino acids with a posterior probability of >99% and 10 with p>95%.

The *RGA2* gene in the *Lr10* haplotype shows little sequence variation

Previous haplotype studies had shown that *Lr10* always occurred in combination with a second NBS-LRR coding gene called *RGA2* (Wicker et al., 2001, Scherrer et al. 2002, Isidore et al., 2005). To compare the evolutionary history of the two genes, we sampled the sequence diversity of *RGA2* in a subset of 15 lines that either showed resistance against a leaf rust avirulent on *Lr10* or contained an *Lr10* sequence of particular interest (e.g. *Can3842_Lr10* which differs strongly from *[Thatcher+Lr10]_Lr10*). For all *RGA2* sequences, protein sequences were predicted by comparison with *[Thatcher+Lr10]_RGA2* . At the DNA level, *Lr10* and *RGA2* can not be aligned and their similarity at the protein level is limited to a short stretch each in the N-terminal and LRR domains and the motifs common to NBS domains. *RGA2* and *Lr10* have probably diverged before the monocot/dicot divergence as they both have homologs in Arabidopsis that are more similar to them than they are to each other (data not shown). *RGA2* is longer than *Lr10* as it contains a partially duplicated NBS domain (Figure 3).

Multiple alignments showed that sequence variability in *RGA2* is much lower than in *Lr10* as the 15 *RGA2* sequences differ by a maximum of 36 nucleotide substitutions and one insertion/deletion from the *[Thatcher+Lr10]_RGA2 .* In contrast, the *T. monococcum* A-genome *Lr10*-orthologs differs from *Lr10* in 147 nucleotide positions. Even the D genome copy of *RGA2* in cultivar *[Thatcher+Lr10]* (Feuillet et al. 2003) differs in only 122 nucleotide positions and in 22 amino acid positions from *[Thatcher+Lr10]_RGA2* and is thus more similar to the latter than many *Lr10* genes from tetraploid wheat are to *[Thatcher+Lr10]_Lr10* (not shown). The overall nucleotide

diversity (i.e. the average number of nucleotide differences per site between all pairs of sequences) in *RGA2* is 0.0053 whereas for *Lr10*, this value is 7 times higher (0.0357). The *RGA2* sequences from the tetraploid *T. turgidum ssp. durum* varieties *Altar*, *Langdon* and *Russello* are identical at the DNA level to *[Thatcher+Lr10]_RGA2* . The most similar sequences to *RGA2* of the the hexaploid variety *Can3842* gene were found in *T. urartu* accessions (*TRI17122* and *TRI6735)* which differ in 12 and 13 bases, respectively, from the *Can3842* sequence (Figure 3). The highest sequence variation among *RGA2* genes was found in three sequences from tetraploid *T. turgidum ssp. dicoccoides*. These contain one divergent region of about 500 bp that was probably introgressed through a gene conversion event (Figure 3). PAML analysis failed to identify any amino acid residues which are under diversifying selection. In fact, the overall ω for the *RGA2* genes was 0.315, indicating strong purifying selection.

Functional *Lr10* genes are present in the tetraploid wheat gene pool

As some of the *Lr10* and *RGA2* sequences from tetraploid wheat are highly similar or (in the case of *Altar*) even identical to those in *[Thatcher+Lr10]*, they might be functional resistance genes. To test this hypothesis, we selected the five tetraploid wheat cultivars *Altar*, *Langdon*, *Russello*, *Bufala* and *L382A* which are all resistant to the avirulent leaf rust isolate *BRW 97512-19* for further study. The recently developed BSMV-based virus-induced gene silencing (VIGS) system in wheat (Scofield et al., 2005) was used to evaluate the resistance function of the *Lr10* genes in these lines. In all five lines, the N-terminal and NBS regions of *Lr10* were targeted by VIGS with constructs CL126 and CL128, respectively (Supplementary Figure 1) while *[Thatcher+Lr10]* was used as a positive control. A fragment targeting the CC coding domain of the *Lr1* gene (CL118, Cloutier et al., 2007) was used as a control for the specificity of silencing. Seedlings were infected with the modified virus RNA and, after

development of viral symptoms, infected with leaf rust isolate *BRW 97512-19*. Both CL126 and CL128 resulted in full susceptibility in *[Thatcher+Lr10]*, *Russello, Altar* and *Bufala* (Figure 4b,d). The results for *Altar* were identical to those of [Thatcher+Lr10] (not shown). No change of resistance was observed in *Langdon* and *L382A* (not shown). No difference was observed between the constructs targeting the N-terminal or the NBS domain. In the control experiments, neither the *Lr1*-derived sequence in CL118 nor a wild-type virus interfered with leaf rust resistance in hexaploid as well as tetraploid wheat lines (Figure 4a,c, e and g, Cloutier et al. 2007). Thus, we conclude that the leaf rust resistance observed in tetraploid lines *Russello*, *Altar* and *Bufala* is depending on the activity of an *Lr10* orthologous gene, whereas other genes are responsible for leaf rust resistance in *Langdon* and *L382A*.

***Lr10*-mediated leaf rust resistance depends on the *RGA2* gene**

The strong conservation in the *RGA2* gene sequences in a broad germplasm as well as the fact that *Lr10* and *RGA2* always occur either as combined H1 haplotype or both are deleted in the H2 haplotype (Isidore et al., 2005) lead to the hypothesis that *RGA2* might also be involved in the resistance reaction to leaf rust. The absence of DNA sequence homology between *Lr10* and *RGA2* genes allowed to easily define short gene fragments for specific virus-induced gene silencing of each gene. The *RGA2* genes of the same cultivars as for *Lr10* (*[Thatcher+Lr10]*, *Bufala*, *Russello*, *Altar*, *L382A* and *Langdon*) were silenced with construct CL127 targeting the NBS domain. After silencing of *RGA2*, a conversion of resistance to susceptibility was observed in hexaploid cv. *[Thatcher+Lr10]* and the tetraploid wheat lines *Altar*, *Bufala* and *Russello* (Figure 4f and h, data not shown). In contrast, resistance was not affected in the tetraploid cultivars *Langdon* and *L382A* (not shown). In the control experiment, none of the lines tested for resistance isolate *BRW 97512-19* was affected by specific targeting of the *Lr1* leaf rust resistance gene (Figure 4g, Cloutier et al. 2007). These data

demonstrate that *Lr10*-based leaf rust resistance in tetraploid and hexaploid wheat depends on both the *Lr10* gene earlier identified by mutagenesis as well as on the *RGA2* gene, making it one of the rare examples in which race-specific resistance depends on more than one NBS-LRR protein.

Expression of *Lr10* and *RGA2* genes after virus-induced gene silencing

To determine the efficiency of silencing, expression of the *Lr10* and *RGA2* genes in *[Thatcher+Lr10]* was investigated by quantitative, real-time PCR after virus-induced gene silencing. After infection with the construct CL128 targeting *Lr10*, the expression level of *Lr10* was reduced to about one third of the level observed in plants infected with a control virus (Supplementary Table 1), whereas expression of *RGA2* was mostly unaffected. No expression was found in cultivar *Thatcher*, in agreement with the absence of an *Lr10* allele in this line with haplotype H2. After specific silencing of *RGA2* with construct CL127, its expression was also reduced to about 30% of the control level (Supplementary Table 1). Surprisingly, in these seedlings *Lr10* expression almost doubled (Supplementary Table 1), but it remains unclear if this is a specific response to lower *RGA2* expression. These results demonstrate a significant and specific reduction of *Lr10* or *RGA2* gene expression after VIGS.

The *Lr10/RGA2* locus of the tetraploid *cv. Langdon* is not a functional resistance haplotype

The tetraploid cultivar *Langdon* is resistant to leaf rust race *BRW 97512-19* which is avirulent on *Lr10.* The *Langdon*_LR10 protein differs by only three amino acids (two substitutions and one deletion) from *[Thatcher+Lr10]*_LR10, whereas *[Thatcher+Lr10]_RGA2* and *Langdon*_RGA2 are identical. However, silencing of *Langdon_Lr10* did not result in loss of resistance. This was confirmed by genetic mapping in a recombinant inbred line population derived from a cross of *Langdon* with

the susceptible wild wheat *T. turgidum ssp. dicoccoides* accession *Gitit* G18-16. A subset of 27 individuals from a population of 152 recombinant inbred lines revealed that the resistant phenotype was not linked to the *Langdon_Lr10 gene*. Ten of the analysed individuals showed segregation between resistance and the *Lr10* ortholog (i.e. they contained the gene but were susceptible). Surprisingly, *Russello*_LR10 contains the same three amino acid substitutions as *Langdon*_LR10, but *Russello* was shown to be a functional variant. However, *Russello*_LR10 contains two additional amino acid substitutions compared to the *[Thatcher+Lr10]_Lr10* protein which could be responsible for reverting functionality.

To test whether this lack of function of the *Langdon_Lr10* gene might be rather due to a defect in gene expression than to the three amino acid polymorphisms, we determined the expression level of *Lr10* and *RGA2* in young seedlings by quantitative PCR. The *Lr10* gene was expressed at a similar level in *Langdon* compared to *[Thatcher+Lr10]* (Figure 5). However, *Langdon_RGA2* was expressed at a significantly lower level than in the resistant *Altar* (Figure 5). We conclude that the *Lr10* and *RGA2* genes from *Langdon* are not a functional haplotype mediating resistance to *AvrLr10*, either because of alterations in gene expression of the *RGA2* gene or because of the three amino acids changes in the *Langdon*_LR10 protein.

Discussion

In this study we found that the *Lr10* locus mediating resistance against wheat leaf rust differs from most other studied *R* loci in two aspects: first, a strong diversifying selection was observed in the N-terminal domain of the protein, and second, both the *Lr10* and *RGA2* NBS-LRR coding genes are required for *Lr10*-mediated resistance. The diversity study in the wheat gene pool identified a domesticated, tetraploid wheat cultivar (*Altar*) with an identical *Lr10* allele to the one found in *[Thatcher+Lr10]*, demonstrating that the *Lr10* resistance of modern bread wheat was already present in

tetraploid wheat. This is similar to findings for the *Lr1* wheat leaf rust resistance gene which was identified also in an ancestor of bread wheat, i.e. the D-genome donor *Aegilops tauschii* (Qiu et al. 2007). Furthermore, the identification of additional active genes in *Russello* and *Bufala* demonstrates that the tetraploid gene pool is a valid source for allele-mining at the *Lr10* locus. In contrast, there is good evidence that the alleles of the wheat powdery mildew resistance locus *Pm3* specifically evolved in hexaploid wheat, as none of the resistance alleles found in modern bread wheat could be identified in a large collection of accessions of tetraploid wheat (Yahiaoui et al. 2006; Yahiaoui et al. 2009). Thus, different resistance loci in the cultivated species bread wheat have very different evolutionary histories and ages, probably reflecting the type of biological interactions between R protein and pathogen effectors.

Evolution of the *Lr10* locus in the wheat gene pool and origin of the *Lr10* resistance gene

The sequence diversity analyses suggest the presence of an ancient pool of diverse haplotypes at the *Lr10* locus in wheat. Despite regions of high sequence conservation, several *Lr10* genes differ from each other by the presence of distinct segments containing many more nucleotide substitutions than the surrounding sequences. This suggests sequence exchange by gene conversion or unequal crossing over and is reminiscent of previous findings in several other resistance genes where sequence exchange between neighbouring genes in clusters resulted in a complex mosaic of various conserved blocks (Parniske et al., 1997; Noël et al., 1999; Van der Hoorn et al., 2001; Smith and Hulbert, 2005). The arrangement of paralogs in clusters is therefore seen as an important source of variation as sequence exchange can result in new pathogen specificities (Parniske et al., 1997; Noël et al., 1999; Van der Hoorn et al., 2001; Smith and Hulbert, 2005).

However, *Lr10* is not part of a gene cluster but a single-copy gene and the closely located *RGA2* gene is only distantly related to *Lr10* (Feuillet at al. 2003), excluding sequence exchanges between paralogs. Nevertheless, the explanation for the observed mosaic pattern of sequence diversity might be essentially the same as for high-copy genes: allelic recombination and gene conversion between ancient haplotypes. The hypothesis of ancient haplotypes is further supported by the finding of completely identical *Lr10* and *RGA2* genes in tetraploid cultivar *Altar*. In addition, the levels of homology among *Lr10* genes do not reflect the expected phylogenetic relationships. Indeed, in three of the four species examined, we found *Lr10* genes with a wide range of homology to the one from hexaploid wheat. This suggests that the wheat gene pool contains a large number of old haplotypes at the *Lr10* locus which can recombine and thus provide the source of the observed variation. This is consistent with previous findings showing that the two main haplotypes at the *Lr10* locus are ancient, *i.e.* were already present before the radiation of the wheat A-genomes about 4-5 MYA (Isidore et al., 2005).

Diversifying selection in the N-terminal domain of the LR10 protein

In many plant NBS-LRR proteins, the LRR domain and in particular the solvent-exposed residues of the β-sheet substructure are under diversifying selection (DeYoung and Innes, 2007). Surprisingly, the LR10 protein showed conservation in the NBS and LRR domains, whereas strong diversifying selection was detected in many amino acid positions in the N-terminal domain. Despite intense literature search, we found only three studies that specifically mention diversifying selection outside the LRR domain: Mondragon-Palomino et al. (2002) identified one amino acid site under diversifying selection adjacent to the CC domain each in two RGAs. Additionally, Ayliffe et al. (1999) and Luck et al. (2000) describe diversifying selection in the TIR and NBS

domains of the flax rust resistance gene *L.* Our own PAML analysis of *L* genes confirmed these findings but identified a much lower density of amino acids positions under diversifying selection than was found in *Lr10*. Thus, the observed strong diversifying selection (qualitatively and quantitatively) within the CC domain of LR10 is highly unusual for NBS-LRR proteins.

Residues under diversifying selection were located mostly in a predicted CC or leucine zipper structure (Feuillet et al. 2003) which is characterized by heptads with hydrophobic residues at the a and d positions (Lupas and Gruber, 2005). Leucine zippers are known to form strong secondary and tertiary structures, for example in the transcription factor GCN4 where they promote the formation of homodimers (reviewed by Lupas and Gruber, 2005). In LR10, hydrophobic residues were conserved while the other residues in the heptad repeat are highly polymorphic and showed a high probability of being under diversifying selection. Three of the most diverse positions were all in the presumably solvent-exposed position c of the heptad repeat. This unexpected pattern suggests that the selected amino acid residues are directly involved in host protein-pathogen effector interactions.

There are several possible explanations for the observed sequence diversity: for example, the diversifying sequence could be the target of a protease effector protein that inactivates LR10. Proteolytic cleavage would be reduced by mutations in the target sequences, whereas the overall structure of the hydrophobic heptad repeat would be conserved. In the tomato/*Cladosporium fulvum* pathosystem, the fungal protease inhibitor AVR2 was found to target defense-related plant proteases, thereby exhibiting selection for variant proteins less sensitive to inhibition (Shabab et al. 2008). In this system, a strong diversifying selection was found for amino acids around the substrate binding groove.

An alternative explanation for diversifying selection in the N-terminal domain would be based on a fungal effector protein with a similar leucine zipper as LR10 and the ability

to competitively bind to the LR10 N-terminal domain. One can speculate that this might result in reduced availability of LR10 for downstream signalling due to competition with the fungal effector. Consequently, resistance would be lost, as *Lr10*-mediated resistance is dosage-dependent. This hypothesis was inspired by reports on homotypic interactions between N-terminal TIR domains of vertebrate toll-like receptors with downstream signalling molecules (Xu et al. 2000) or to the CC domain of barley MLA which interacts with WRKY type transcription factors (Shen et al. 2007)

Finally, it is possible that the CC domain of the LR10 protein is directly involved in pathogen recognition. This would be a deviation from most previous findings where this function was generally attributed to the LRR domain (Jones and Jones, 1997; Dodds et al., 2001). However, the CC motif is in some sense analogous to the LRR motif, with a hydrophobic backbone that secures the three-dimensional integrity and many solvent-exposed residues that can evolve largely freely. Thus, one can imagine that in LR10, the CC domain has taken over the function the LRR fulfils in other resistance interactions. Indeed, the N-terminal domain of NBS-LRR proteins apparently interacts with a number of other proteins, for example with plant proteins targeted by pathogen effectors (Mackay et al. 2002; Ade et al. 2007) or directly with a viral elicitor molecule (Burch-Smith et al. 2007). In conclusion, the N-terminal domain of NBS-LRR proteins seems to serve multiple biochemical roles and its further analysis promises to be a rich source for a better understanding of molecular plant-pathogen interactions (Rairdan et al., 2008).

Two different, genetically closely linked NBS-LRR coding genes are required for *Lr10*-based resistance

RGA2 differs from *Lr10* as it is extremely well conserved in all wheat accessions analysed. Indeed, the nucleotide diversity among *RGA2* genes is almost 7 times lower than in *Lr10* genes and between 7 and almost 30 times lower than in a recently

published set of 12 R genes from different plant species (Jiang et al., 2007). Surprisingly, VIGS demonstrated that both *Lr10* and *RGA2* are essential for *Lr10*-mediated resistance in tetraploid and hexaploid wheat. This is reminiscent to several other cases where two NBS-LRR proteins are required for effective resistance (Sinapidou et al. 2004; Peart et al. 2005; Staal et al. 2006; Eitas et al. 2008; Ashikawa et al. 2008; Lee et al. 2009). The situation for LR10 and RGA2 seems unique, however, since both involved NBS-LRR proteins have characteristic features of leucine zippers domains in the N-terminal region.

Two different models could explain the requirement for the two proteins. First, LR10 and RGA2 are both involved in recognition of the pathogen effector, for example as a heterodimer or as part of a larger protein complex. Second, RGA2 could be active downstream of LR10 in the signal cascade resulting in resistance. Currently, there is no evidence for any of these models. We did not find any physical interaction of LR10 and RGA2 protein domains in a yeast-two-hybrid screen (our unpublished data), although it has to be considered that a physical interaction might only occur in the presence of the pathogen effector. The strong conservation of RGA2 seems to be more compatible with a downstream activity, whereas the diversifying selection on LR10 would indicate that this protein is in direct interaction with the pathogen effector.

Lr10 and *RGA2*: which one is the *R* gene?

The original identification of *Lr10* as the effective resistance gene was based on the identification of three independent point mutations located in the *Lr10* gene and transformation (Feuillet et al. 2003). In these transformation experiments, *Lr10* and *RGA2* were co-bombarded in overexpression constructs. Interestingly, the lines which showed good resistance were always transgenic for both *Lr10* and *RGA2* (Feuillet et al. 2003; our unpublished data). The only exception was one plant which showed resistance but its progeny was weak and did not set seed (Feuillet et al. 2003). These

transformation studies provide additional evidence that both genes must be present for effective resistance. Possibly, the fact that in three independent mutants of *Lr10*-mediated resistance the mutations were located in the same gene is just accidental and in additional *Lr10* mutants, mutations might be found in *RGA2*.

However, there is also an alternative explanation: in hexaploid bread wheat the A- and D- genome copies of *RGA2* are very similar and it might be that the D-genome copy can fully substitute for a defect A-genome copy of *RGA2*. This would not be the case in diploid and tetraploid wheat (which do not contain a D genome) and therefore, there would have been selection for maintaining the A genome copy, a situation which is in agreement with the observation of high frequency of the H1 haplotype (where *Lr10* and *RGA2* are both present). Following this line of reasoning, the *Lr10* gene might be deleterious to wheat if its function would not be "controlled" in some way by *RGA2*. In addition, this hypothesis postulates that the ratio of LR10 and RGA2 is critical for correct functioning. Thus, it remains to be investigated how such a control mechanism might work and if RGA2 is a protein which specifically modulates the LR10 output.

Experimental procedures

Plant material, PCR experiments and gene silencing

Accessions or varieties of wheat lines, leaf rust infection type, geographical origins and seed providers are listed in Table 1. Primers used for PCR, sequencing or gene silencing are shown in Supplementary Table 2 and sequence organization and positions of primers are indicated in Supplementary Figure 1. Details on PCR conditions, sequencing and gene silencing are available as Supplementary text.

For virus-induced gene silencing (Holzberg et al., 2002), two constructs targeting the *Lr10* gene in the N-terminal (CL126) and NBS region (CL128) and one construct (CL127) targeting the NBS2 region of *RGA2* were used (Supplementary Figure 1). A

fragment of the CC domain (CL118) of the resistance gene *Lr1* (EF439840) targeting base-pairs 104 to 388 (primers 118-CL and 122-CL) was used as a control (Cloutier et al. 2007). *In vitro* transcription of the viral subgenomes was performed using the kit mMESSAGE mMACHINE® T7 Kit (Ambion, Austin, USA). Leaf rust infections were performed on five day old seedlings as described by Schachermayr et al. (1995) using the isolate *Puccinia triticina BRW 97512-19*.

For quantitative real-time PCR, total RNA was isolated from leaves of *Thatcher*, *[Thatcher+Lr10]* and *cvs. Langdon* and *Altar* from 19 days-old, virus infected and 10 days-old untreated seedlings. Real-time PCR assays were performed with the ABI PRISM 7700 Sequence Detection System (Applied Biosystems) using SYBR Green PCR Master Mix (Applied Biosystems, Rotkreuz, Switzerland) as described by Travella et al. (2006). The *Lr10* and *RGA2* regions targeted by real-time PCR were different from the regions targeted by VIGS. To normalize results, the *glyceraldehyde-3-phosphate dehydrogenase* gene (*GAPDH*, AF251217) was used as an internal standard amplifying a 61 bp fragment (440 to 501 bp) using GAPDH-RT-F and GAPDH-RT-R.

Lr10* mapping in *cv. Langdon

The mapping population consisted of 152 F_6 recombinant inbreed lines (RILs) derived from a cross between *Triticum turgidum ssp. durum* (female) cultivar *Langdon* and a wild emmer *T. turgidum ssp. dicoccoides* (male) accession G18-16. The wild emmer accession was sampled from Gitit (32°06 N, 35°24 E; 288.4 m above sea level, Peleg et al., 2005). F_1 progeny from the initial cross were taken through five generations of selfing *via* single seed decent to produce homozygous RILs. *Lr10*-specific primers were designed for parental lines: primers Lgd_F1 and Lgd_R1 amplified a 805 bp fragment in *Langdon*, while primers Git_F1 and Git_R1 amplified a 530 bp fragment from *Gitit* G18-16. Phenotypes were determined following a standard procedure (Schachermayr

et al., 1995) after infection with leaf rust isolate isolate *BRW 97512-19* avirulent on Langdon.

Sequence analysis

Multiple sequence alignments were done with CLUSTALW (Thompson et al., 1994) using a gap creation penalty of 3.0 and a gap extension penalty of 0.01. Putative gene conversion events were identified with DnaSP (ub.es/dnasp/) using the two-gamete method by choosing two sets of at least three and five sequences, respectively (the minimum required by DnaSP) containing the region of interest. CC structures were predicted with the program COILS (ch.embnet.org/software/COILS_form.html). Visual representation of multiple alignments was done with a home-made Perl script available upon request. PAML (http://abacus.gene.ucl.ac.uk/software/paml.html) was used to determine Ka/Ks ratios. Likelihoods of more complex models to null models M_0 (which assumes neutral codon evolution) were compared. Also Likelihoods of *M1a* (nearly neutral codon evolution with two codon classes allowed to take on values from $0 \le \omega_0 \le 1$ or $\omega_0 = 1$) were compared with those of a more complex model M2a (diversifying selection), which incorporates an additional positively selected sites class ($\omega_2 > 1$), and M3 (which assumes three discrete site classes). Results of analyses using model M7 and M8 were also compared, which both assume a β-distribution for $0 \le \omega \le 1$, with the latter model allowing for an extra class of sites with $\omega > 1$. Posterior probabilities for each codon site to belong to one of the selection classes was calculated. Codons with a posterior probability >0.95 were considered positively selected (i.e. $\omega > 1$).

Accession numbers

Lr10 and *RGA2* DNA sequences are deposited at GenBank under the accession numbers EU675942 - EU675971 (see also Supplementary Table 3).

Acknowledgements

This work was supported by the Swiss National Science Foundation grant 3100-105620.

References

Ade, J., DeYoung, B.J., Golstein, C., and Innes, R.W. (2007) Indirect activation of a plant nucleotide binding site-leucine-rich repeat protein by a bacterial protease. *Proc. Natl. Acad. Sci. USA,* **104**, 2531-2536.

Ashikawa, I., Hayashi, N., Yaman, H., Kanamori, H., Wu, J., Matsumoto, T., Ono, K. and Yano, M. (2008) Two adjacent nucleotide-binding site-leucine-rich repeat class genes are required to confer Pikm-specific rice blast resistance. *Genetics,* **180**, 2267-2276.

Ayliffe, M.A., Frost, D.V., Finnegan, E.J., Lawrence, G.J., Anderson, P.A., and Ellis J.G. (1999) Analysis of alternative transcripts of the flax *L6* rust resistance gene. *Plant J.,* **17**, 287-292.

Baum, M., Lagudah, E.S., and Appels, R. (1992) Wide crosses in cereals. *Annu. Rev. Plant Biol.,* **43**, 117-143.

Bent, A.F., and Mackey, D. (2007) Elicitors, effectors, and R genes: the new paradigm and a lifetime supply of questions. *Annu. Rev. Phytopathol.,* **45**, 399-436.

Bieri, S., Mauch, S., Shen, Q.H., Peart, J., Devoto, A., Casais, C., Ceron, F., Schulze, S., Steinbiß, H.H., Shirasu, K., and Schulze-Lefert P. (2004) *RAR1* positively controls steady state levels of barley MLA resistance proteins and enables sufficient MLA6 accumulation for effective resistance. *Plant Cell,* **16**, 3480-3495.

Burch-Smith, T.M., Schiff, M., Caplan, J.L., Tsao, J., Czymmek, K., and Dinesh-Kumar, S.P. (2007) A novel role for the TIR domain in association with pathogen-derived elicitors. *PLoS Biology,* **5**, e68 doi:10.1371

Cloutier, S., McCallum, B.D., Loutre, C., Banks, T.W., Wicker, T., Feuillet, C., Keller, B., and Jordan, M.C. (2007) Leaf rust resistance gene *Lr1*, isolated from bread wheat (*Triticum aestivum* L.) is a member of the large *psr567* gene family. *Plant Mol Biol.,* **65**, 93-106.

DeYoung, B.J., and Innes, R.W. (2007) The rules of engagement. The XIII International Congress on Molecular Plant-Microbe Interactions, Sorrento, Italy, July 2007. *New Phytol.,* **176**, 506-10.

Dodds, P.N., Lawrence G.J. and Ellis, J.G. (2001) Six amino acid changes confined to the leucine-rich repeat beta-strand/beta-turn motif determine the difference between the *P* and *P2* rust resistance specificities in flax. *Plant Cell,* **13**, 163-178.

Eitas, T.K., Nimchuk, Z.L., and Dangl, J.L. (2008) *Arabidopsis* TAO1 is a TIR-NB-LRR protein that contributes to disease resistance induced by the *Pseudomonas syringae* effector AvrB. *Proc. Natl. Acad. Sci. USA,* **105**, 6475-6480.

Feuillet, C, Travella, S., Stein, N., Albar, L., Nublat, A. and 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.

Feuillet, C., Langridge, P., and Waugh, R. (2008) Cereal breeding takes a walk on the wild side. Trends Genet., **24**, 24-32.

Holzberg, S., Brosio, P., Gross, C., and Pogue, G.P. (2002) Barley stripe mosaic virus-induced gene silencing in a monocot plant. *Plant J.,* **30**, 315-327.

Isidore, E., Scherrer, B. Chalhoub, B., Feuillet, C. and Keller, B. (2005) Ancient haplotypes resulting from extensive molecular rearrangements in the wheat A genome have been maintained in species of three different ploidy levels. *Genome Res.,* **15**, 526-536.

Jiang, H., Wang, C., Ping, L., Tian, D., and Yang, S. (2007) Pattern of LRR nucleotide variation in plant resistance genes. *Plant Science,* **173**, 253-261.

Jones, J.D., and Dangl, J.L. (2006) The plant immune system. *Nature,* **444**, 323-329.

Jones, D. and Jones, J.D. (1997) The role of leucine-rich repeat proteins in plant defenses. *Adv. Bot. Res. Adv. Plant Pathol.,* **24**, 89–167.

Kuang, H., Woo, S.S., Meyers, B.C., Nevo, E., and Michelmore, R.W. (2004) Multiple genetic processes result in heterogeneous rates of evolution within the major cluster Disease resistance genes in lettuce. *Plant Cell,* **16**, 2870-2894.

Lee, S.-K., Song, M.-Y., Seo, Y.-S., Kim, H.-K., Ko, S., Cao, P.-J., Suh, J.-P., Yi, G., Roh, J.-H., Lee, S., An, G., Han, T.-R., Wang, G.-L., Ronald, P. and Jeon, J.-S. (2009) Rice *Pi5*-mediated resistance to *Magnaporthe oryzae* requires the presence of two coiled-coil-nucleotide-binding-leucine-rich repeat genes. *Genetics,* **181**, 1627-1638.

Luck, J.E., Lawrence, G.J., Dodds, P.N., Shepherd, K.W., and Ellis, J.G. (2000) Regions outside of the leucine-rich repeats of flax rust resistance proteins play a role in specificity determination. *Plant Cell,* **12**, 1367-1378.

Lupas, A.N., and Gruber, M. (2005) The structure of alpha-helical coiled coils. *Adv Protein Chem.,* **70**, 37-78.

McDowell, J.M., and Simon, S.A. (2008) Molecular diversity at the plant-pathogen interface. *Dev. Comp. Immunol.,* **32**, 736-44.

Mackey, D., Holt, B.F., Wiig, A., and Dangl, J.L. (2002) RIN4 interacts with *Pseudomonas syringae* type III effector molecules and is required for RPM1-mediated resistance in *Arabidopsis*. *Cell,* **108**,743-54.

McIntosh, R.A., Wellings, C.R. and Park, R.F. (1995) *Wheat Rusts: An Atlas of Resistance Genes*. Dordrecht: Kluwer Academic Publishers.

Michelmore, R.W. and Meyers, B.C. (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Res.,* **8**, 1113-1130.

Mondragón-Palomino, M., Meyers, B.C., Michelmore, R.W. and Gaut, B.S. (2002) Patterns of positive selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*. *Genome Res.,* **12**, 1305-1315.

Noël, L., Moores, T.L., van Der Biezen, E.A., Parniske, M., Daniels, M.J., Parker, J.E. and Jones, J.D. (1999) Pronounced intraspecific haplotype divergence at the *RPP5* complex disease resistance locus of *Arabidopsis. Plant Cell,* **11**, 2099-2112.

Parniske, M., Hammond-Kosack, K.E., Golstein, C., Thomas, C.M., Jones, D.A., Harrison, K., Wulff, B.B. and Jones, J.D. (1997) Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the *Cf-4/9* locus of tomato. *Cell,* **91**, 821-832.

Peart, J.R., Mestre, J., P., Lu R., Malcuit I. and Baulcombe, D. (2005) NRG1, a CC-NB-LRR protein, together with N, a TIR-NB-LRR protein, mediates resistance against tobacco mosaic virus. *Curr. Biol.,* **15**, 968-973.

Peleg, Z., Fahima, T., Abbo, S., Krugman, T., Nevo, E., Yakir, D. and Saranga, Y. (2005) Genetic diversity for drought resistance in wild emmer wheat and its ecogeographical associations. *Plant, Cell Environ.,* **28**, 176-191.

Qiu, J.W., Schürch, A.C., Yahiaoui, N., Dong, L.L., Fan, H.J., Zhang, Z.J., Keller, B., and Ling, H.Q. (2007) Physical mapping and identification of a candidate for the leaf rust resistance gene *Lr1* of wheat. *Theor. Appl. Genet.,* **115**, 159-68.

Rairdan, G.J., Collier, S.M., Sacco, M.A., Baldwin, T.T., Boettrich, T. and Moffett, P. (2008). The coiled-coil and nucleotide binding domains of the potato Rx disease resistance protein function in pathogen recognition and signaling. *Plant Cell,* **20**, 739-741.

Schachermayr G.M., Messmer, M.M., Feuillet, C., Winzeler, H., Winzeler, M., and Keller, B. (1995) Identification of molecular markers linked to the *Agropyron elongatum*-derived leaf rust resistance gene *Lr24* in wheat. *Theor. Appl. Genet.,* **90**, 982-990.

Scherrer, B., Keller, B., and Feuillet, C. (2002) Two haplotypes of resistance gene analogs have been conserved during evolution at the leaf rust resistance locus *Lr10* in wild and cultivated wheat. *Funct. Int. Genomics,* **2**, 40-50.

Scofield, S.R., Huang, L., Brandt, A.S., and Gill, B.S. (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** , 2165-73.

Shabab, M., Shindo, T., Gu, C., Kaschani, F., Pansuriya, T., Chintha, R., Harzen, A., Colby, T., Kamoun, S., and van der Hoorn, R.A. (2008) Fungal effector protein AVR2 targets diversifying defense-related cys proteases of tomato. *Plant Cell,* **20**, 1169-83.

Shen, J., Araki, H., Chen, L., Chen, J.Q., and Tian, D. (2006) Unique evolutionary mechanism in *R*-genes under the presence/absence polymorphism in *Arabidopsis thaliana*. *Genetics,* **172**, 1243-1250.

Shen, Q.-H., Saijo, Y., Mauch, S., Biskup, C., Bieri, S., Keller, B., Seki, H., Uelker, B., Somssich, I.E. and Schulze-Lefert, P. (2007) Nuclear activity of MLA immune

receptors links isolate-specific and basal disease-resistance responses. *Science,* **315**, 1098-1103.

Sinapidou, E., Williams, K., Nott, L., Bahkt, S., Tör, M., Crute, I., Bittner-Eddy, P., and Beynon, J. (2004) Two TIR:NB:LRR genes are required to specify resistance to *Peronospora parasitica* isolate Cala2 in *Arabidopsis.* Plant J., **38**, 898–909.

Smith, S.M. and Hulbert, S.H. (2005) Recombination events generating a novel *Rp1* race specificity. *Mol. Plant Microbe Interact.,* **18**, 220-228.

Staal, J., Kaliff, M., Bohman, S., and Dixelius, C. (2006) Transgressive segregation reveals two *Arabidopsis* TIR-NB-LRR resistance genes effective against *Leptosphaeria maculans*, causal agent of blackleg disease. *Plant J.,* **46**, 218-30.

Tian, D., Araki, H., Stahl, E., Bergelson, J., and Kreitman, M. (2002) Signature of balancing selection in *Arabidopsis. Proc. Natl. Acad. Sci. USA,* **99**, 11525-11530.

Van der Hoorn, R.A., Kruijt, M., Roth, R., Brandwagt, B.F., Joosten, M.H. and De Wit, P.J. (2001) Intragenic recombination generated two distinct *Cf* genes that mediate *AVR9* recognition in the natural population of *Lycopersicon pimpinellifolium*. *Proc. Natl. Acad. Sci. USA,* **98**, 10493-10498.

Wicker, T., Stein, N., Albar, L., Feuillet, C., Schlagenhauf E. and Keller B. (2001) Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. *Plant J.,* **26**, 307-316.

Wicker, T., Yahiaoui, N., and Keller, B. (2007) Illegitimate recombination is a major evolutionary mechanism for initiating size variation in plant resistance genes. *Plant J.,* **51**, 631-641.

Xu, Y., Tao, X., Shen, B., Horng, T., Medzhitov, R., Manley, J.L., and Tong, L. (2000) Structural basis for signal transduction by the Toll/interleukin-1 receptor domains. *Nature,* **408**, 111-115.

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

Yahiaoui, N., Kaur, N., and Keller, B. (2009) Independent evolution of functional *Pm3* resistance genes in wild tetraploid wheat and domesticated bread wheat. *Plant J.,* **57**, 846- 856 .

Legends for Supplementary Material

Supplementary Table 1. Expression of *Lr10* and *RGA2* after VIGS

Supplementary Table 2. PCR Primers

Supplementary Table 3. GenBank accession numbers and sequence names

Supplementary Figure 1. *Lr10* and *RGA2* constructs

Supplementary Figure 2. *Lr10* gene conversion events

Supplementary Text for Experimental Procedures

Table 1. Wheat accessions used in this study. *T. urartu* accessions are diploid, *T. turgidum* accessions are tetraploid.

Variety	Species	Origin	Infection type[a]
DV92[f]	*T. monococcum*	?	0
TRI 6735[b]	*T.urartu*	Armenia	3
TRI 17122[b]	*T.urartu*	Turkey	4
TRI 17 159[b]	*T.urartu var.urartu*	Turkey	4
TRI 17119[b]	*T.urartu var. pubescentiurartu*	Turkey	3
Gitit[c]	*T.turgidum ssp. dicoccoides*	Israel	4
IG 46352[d]	*T.turgidum ssp. dicoccoides*	Jordan	1
L382A[d]	*T.turgidum ssp. dicoccoides*	Iran	1
TRI 11504[b]	*T.turgidum ssp.dicoccoides var. namuricum*	Lebanon	3
TRI 14095[b]	*T.turgidum ssp.dicoccoides var. aaronsohnii*	Italy	3
TRI 14196[b]	*T.turgidum ssp.dicoccoides var. kotschyi*	Italy	3+
TRI 9865[b]	*T.turgidum ssp.dicoccoides var.macraaronsohnii*	?	4
Altar 84[e]	*T.turgidum ssp.durum*	?	1
Russello[d]	*T.turgidum ssp.durum*	Italy	1
Bufala[d]	*T.turgidum ssp.durum*	Italy	1
Waha[d]	*T.turgidum ssp.durum*	Italy	1
Roqueno[d]	*T.turgidum ssp.durum*	Spain	1
Simeto[d]	*T.turgidum ssp.durum*	Italy	1
Razzak[d]	*T.turgidum ssp.durum*	Tunisia	1
Langdon[f]	*T.turgidum ssp.durum*	USA	1
[Thatcher+Lr10][f]	*T.aestivum*	Canada	1
Can3842[f]	*T.aestivum*	Canada	4

[a]Infection type of the leaf rust isolate *BRW 97512-19* (AvrLr10) determined according to McIntosh et al. (1995): 0: immune 1: Small uredia with necrosis 3: medium-size uredia 4: susceptible, large uredia.

[b]Seeds provided by the Institute of Plant Genetics and Crop Plant Research Gatersleben, Germany.

[c]Seeds provided by Dr. Tzion Fahima, Institute for Evolution, University of Haifa, Israel.

[d]Seeds provided Prof. Salamini, Max-Planck-Institute, Cologne, Germany.

[e]Seeds provided by CIMMYT, Mexico.

[f]Sequence previously published

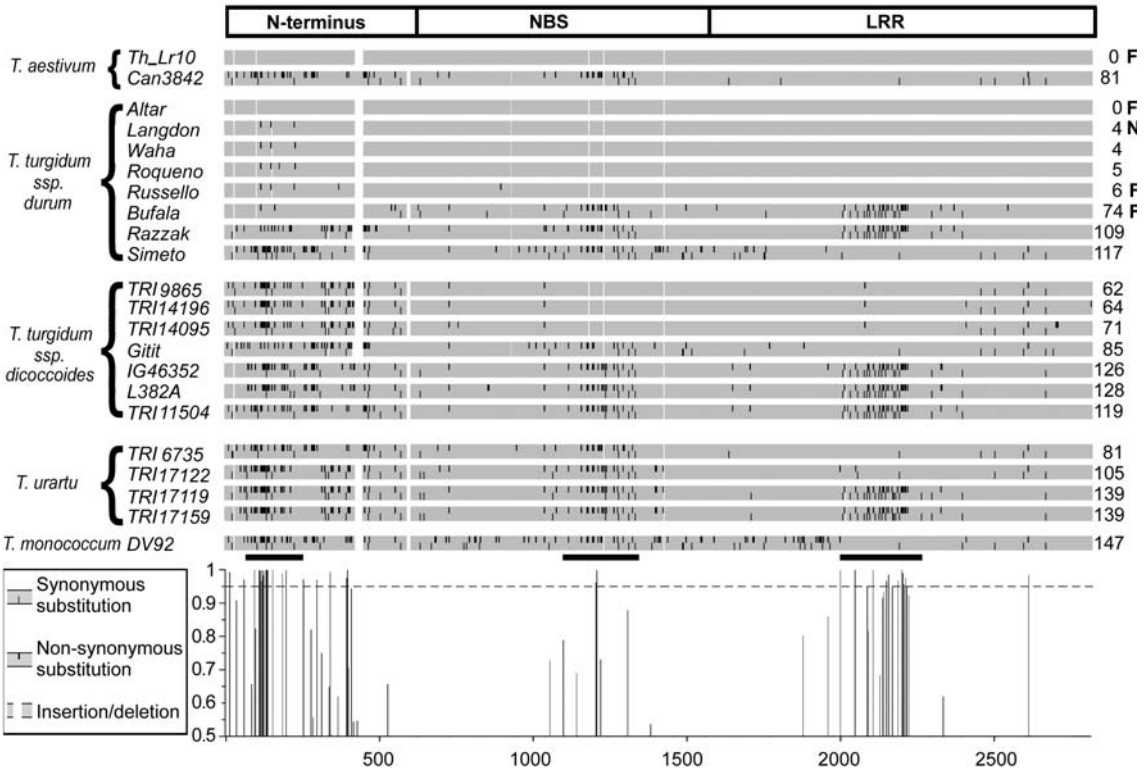


Figure 1. Comparison of *Lr10* coding sequences from different wheat species. The structure of the gene is indicated at the top. The origin of the sequences from different wheat (sub-) species is indicated at the left. The *Lr10* gene from hexaploid wheat variety *[Thatcher+Lr10]* (*Th_Lr10*) was used as a reference with which all others were compared. Synonymous and non-synonymous substitutions are indicated as vertical bars, while interruptions indicate gaps in the alignment. The number of substitutions in each sequence compared to *[Thatcher+Lr10]_Lr10* is indicated on the right. Genes that were functionally tested by VIGS for resistance against leaf rust isolate *BRW 97512-19* are indicated at the right with "F" for functional and "N" for non-functional. Black bars underneath the alignment indicate regions which are statistically likely to have resulted from gene conversion events (determined with DnaSP). The graph at the bottom indicates positions of codons which encode amino acids under diversifying selection (Ka/Ks > 1). The dashed line indicates a posterior probability of 0.95.

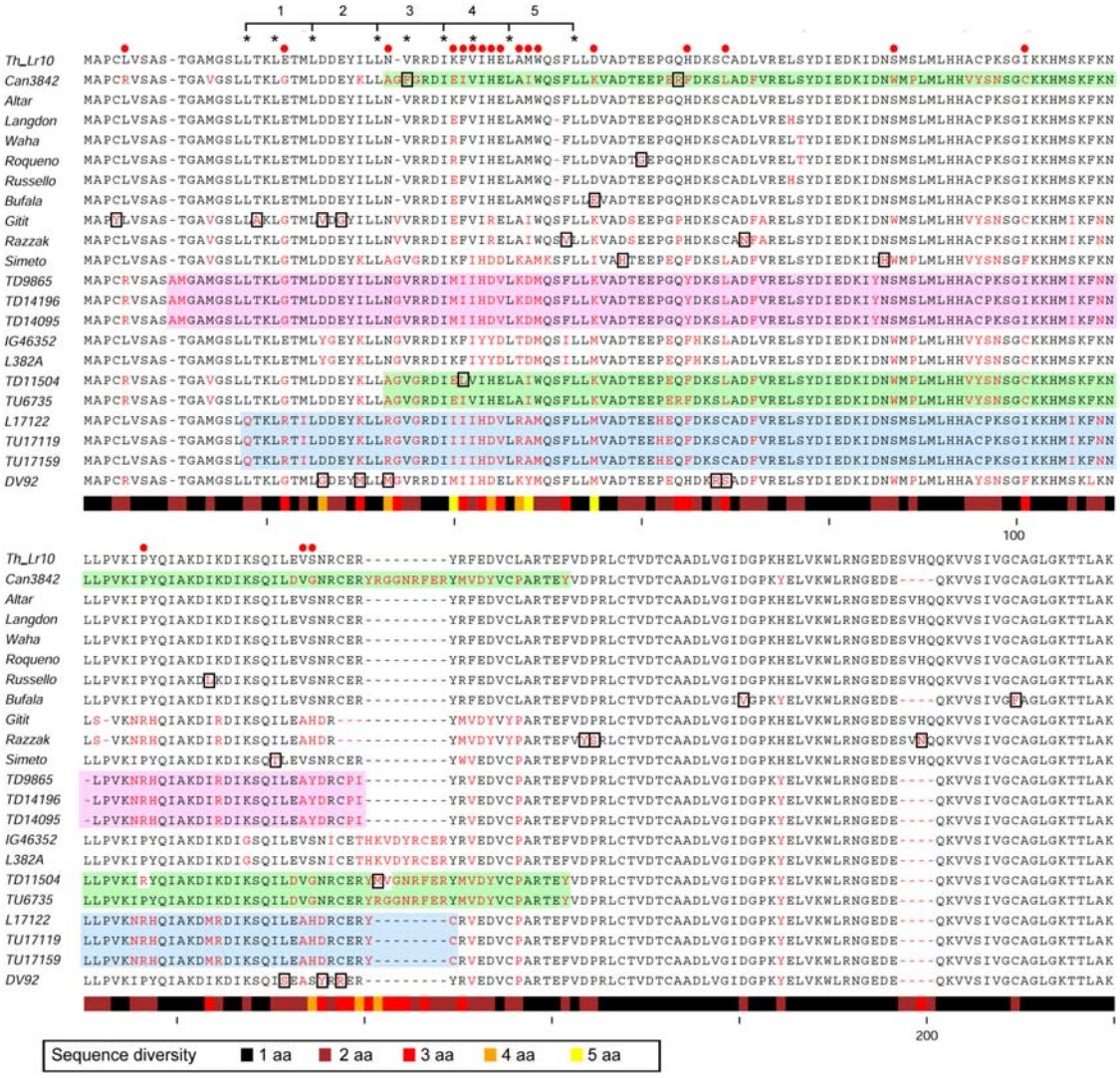


Figure 2. Alignment of the first 220 amino acids of the N-terminal domain of *Lr10* sequences. The areas shaded in pink, light blue and light green indicate likely gene conversion events. Asterisks indicate positions which correspond to the backbone of a predicted leucine zipper motif. The 5 loops of this motif are numbered. Note that deletions in some sequences are compensated by additional leucine residues at the 4th and the last leucine of the motif. Amino acid residues under diversifying selection are indicated with red dots. The "heat map" underneath the alignment indicates the sequence diversity for each position. Gaps are counted as amino acids. Amino acid substitutions that are exclusive to one single sequence are marked with black boxes (*Th_Lr10*:*[Thatcher+Lr10]_Lr10*).

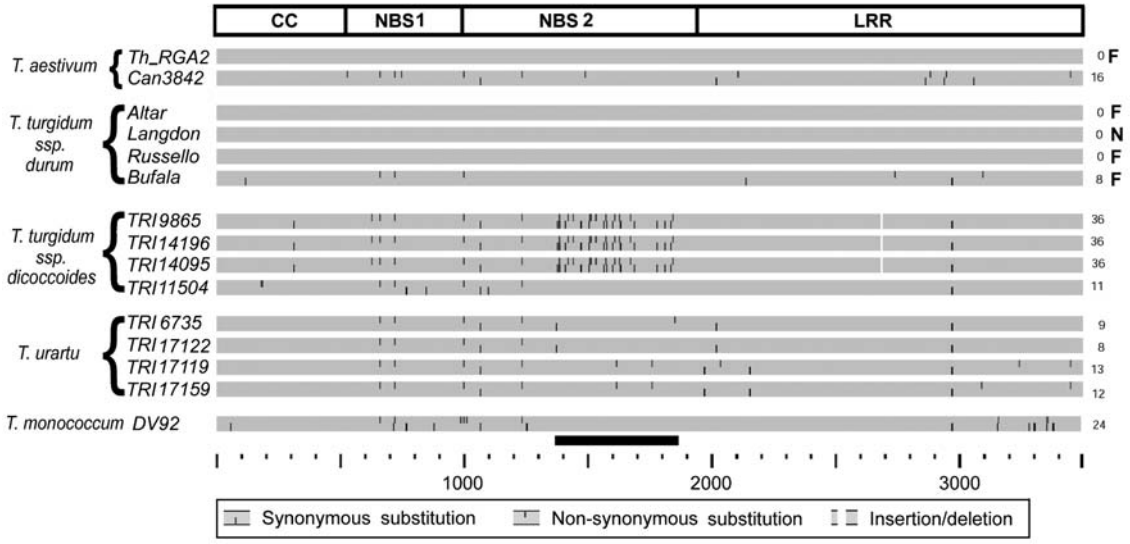


Figure 3. Comparison of *RGA2* coding sequences from different wheat species. *RGA2* from the hexaploid wheat variety *[Thatcher+Lr10]* (*Th_RGA2*) was used as a reference with which all others were compared. Symbols and labels are analogous to those in Figure 1.

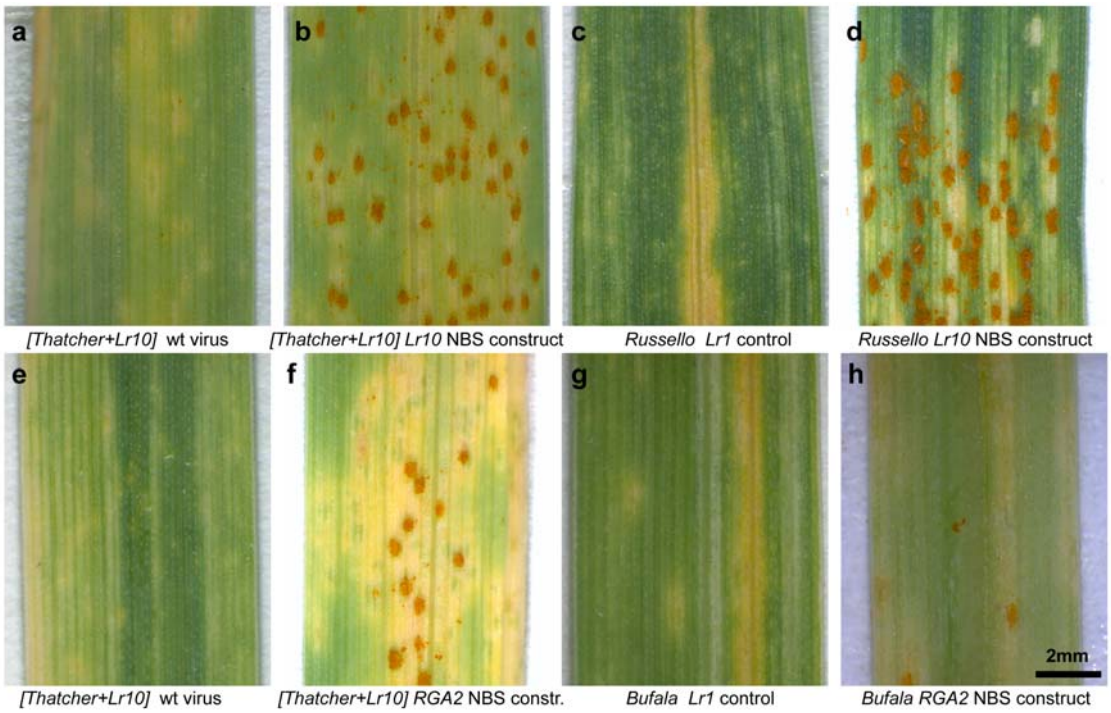


Figure 4. Virus-induced gene silencing with *Lr10* and *RGA2* and *Lr1* as control. **a.** Leaves of *[Thatcher+Lr10]* infected with a wild type virus and (**b.**) with construct CL128 targeting the NBS domain of *Lr10.* **c.** Tetraploid cv. *Russello* infected with CL118 of the rust gene *Lr1* as control and (**d.**) with CL126 targeting the N-terminal domain of *Lr10.* **e.** and **f.**: Functionality test of the *RGA2* construct. **e.** Infection with wild type virus on *[Thatcher+Lr10].* **f.** Infection with the *RGA2*-construct CL127 on *[Thatcher+Lr10].* **g.** Tetraploid *Bufala* infected with CL118 of the rust gene *Lr1* as control and (**h.**) with CL127 of *RGA2.*

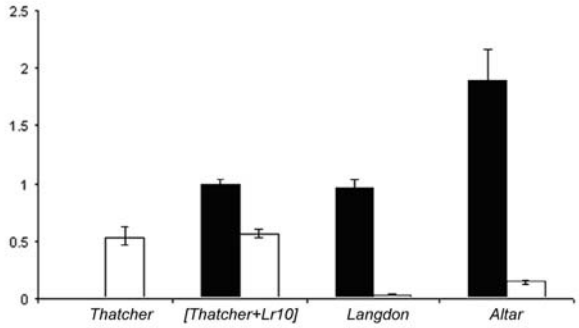


Figure 5. Gene expression of the *Lr10* (black bars) and *RGA2* (white bars) genes in 10 days old wheat seedlings of four wheat varieties. Relative gene expression was standardised to the expression of *Lr10* in *[Thatcher+Lr10]* (expression =1).

Supplementary Table 1. Quantitative expression of *Lr10* and *RGA2* genes after virus-induced gene silencing. The expression level was normalized based on expression levels in *[Thatcher+Lr10]* after infection with a control virus targeting the *Lr1* leaf rust resistance gene. Note that hexaploid wheat (*[Thatcher+Lr10]* and *Thatcher*) contains a homoeologous *RGA2* gene on the D genome.

Cultivar	Gene silenced	*Lr10* expression	*RGA2* expression
[Thatcher+Lr10]	*Lr1*	1.00	1.00
[Thatcher+Lr10]	*Lr10*	0.31 ± 0.15	0.63 ± 0.23
[Thatcher+Lr10]	*RGA2*.	1.82 ±0.22	0.30 ± 0.06
Thatcher	*Lr1*	0.001	0.43 ± 0.07

Supplementary Table 2. Primers used for PCR, sequencing, genetic mapping and virus-induced gene silencing.

Primer	Sequence
ThLr10-V[a]	CGGAACTATGGAGAGTGAAC
ThLr10-U[a]	GGGAAATGTAGACAGGTACAT
PibTm1-4[a]	CCCATTCTCTCGTGGATGAT
PibTm10A[a]	GAAGCCGGATTATAGTGTC
PibTm10E[a]	TTCATAGCTCATTGCATC
PibTm10H	AGAAACGCCCAACTGACG
PibTm10U[a]	AGGCAAGTTTGAGAAGTG
RGA23B5[a]	CAATTGTGATGAACTCCTCA
CL1[b]	ATATTAATTAAGCCGCGACATCAAGTTCGTCATA
CL2[b]	TATGCGGCCGCTACTGACCTCCAAGATTTGGCTC
CL3[b]	ATATTAATTAAGTCACGAAAAGGGATTTCCCCCT
CL4[b]	TATGCGGCCGCAAGATCTTCCCCATCTTCAT
127-CL[b]	GCAGCTTAATTAAGAGATGGCTGCCAATTTGAG
130-CL[b]	GCAGAGCGGCCGCGGCTTGATGTTTCCAATGC
118-CL[b]	AGCAGTTAATTAACAACTCCCAGCAGATCAG
122-CL[b]	AGTATGCGGCCGCATCTTGTGCAGAGGAACAAG
RT-ST1[c]	GAGGAACCTGGACAGCATGAC
RT-ST2[c]	CAGGCGTGATGGAGCATAAGT
RT-ST3[c]	GCCACAGAACGTAGCACCAGA
RT-ST4[c]	CCTCCGTCGAATCTCCATCTT
GAPDH-RT-F[c]	TTAGACTTGCGAAGCCAGCA
GAPDH-RT-R[c]	AAATGCCCTTGAGGTTTCCC
Lgd_F1[d]	CATCCTGCTCAATGTTCG
Lgd_R1[d]	GAGTAGTACTACCAATCGCA
Git_F1[d]	GCAATCATTTCTCCTCAAGG
Git_R1[d]	CCATATCAGGGCTTCGTG

[a]Used for amplification and sequencing of genes.

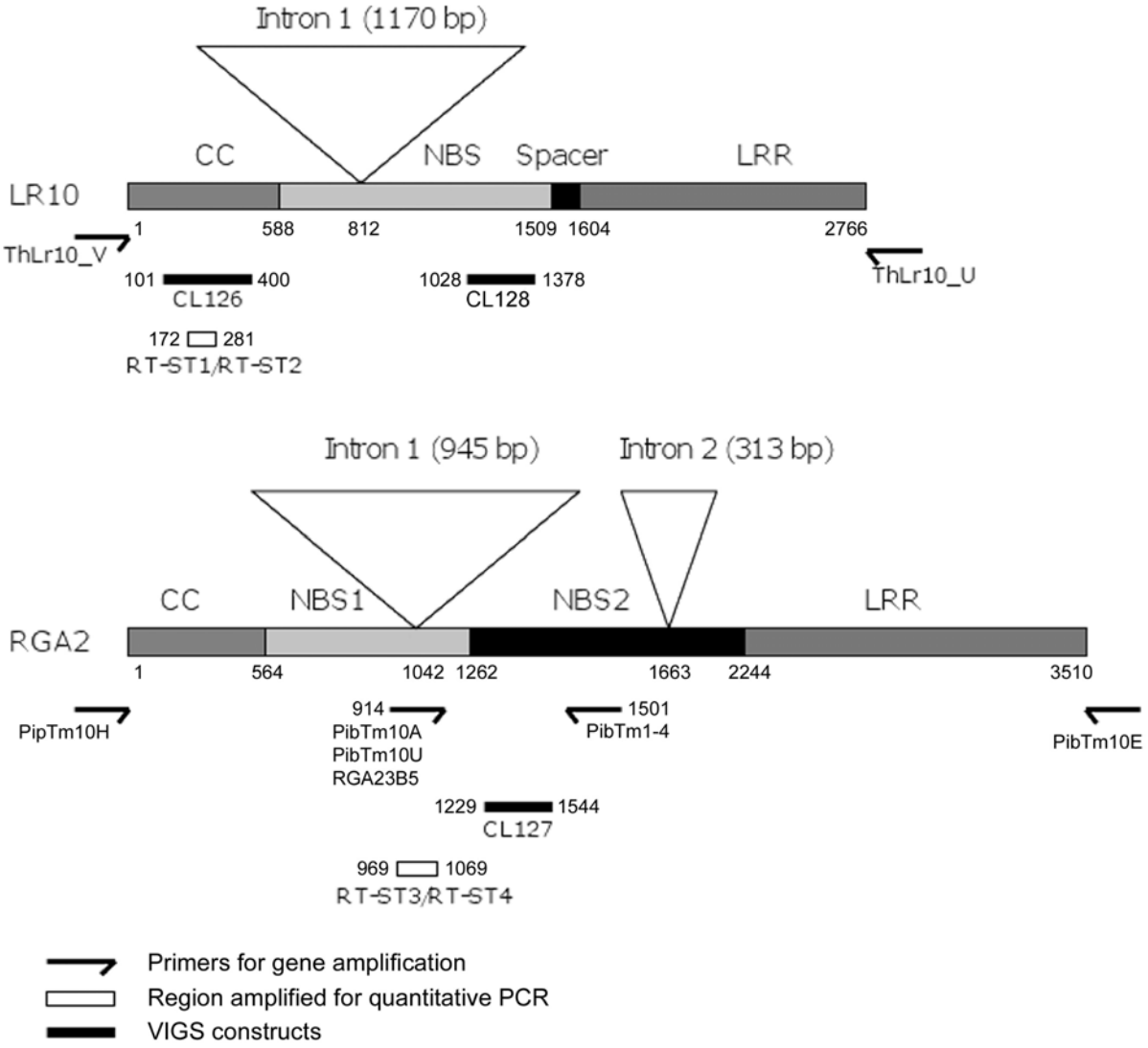
[b]Used for virus induced gene silencing.

[c]Used for quantitative real-time PCR

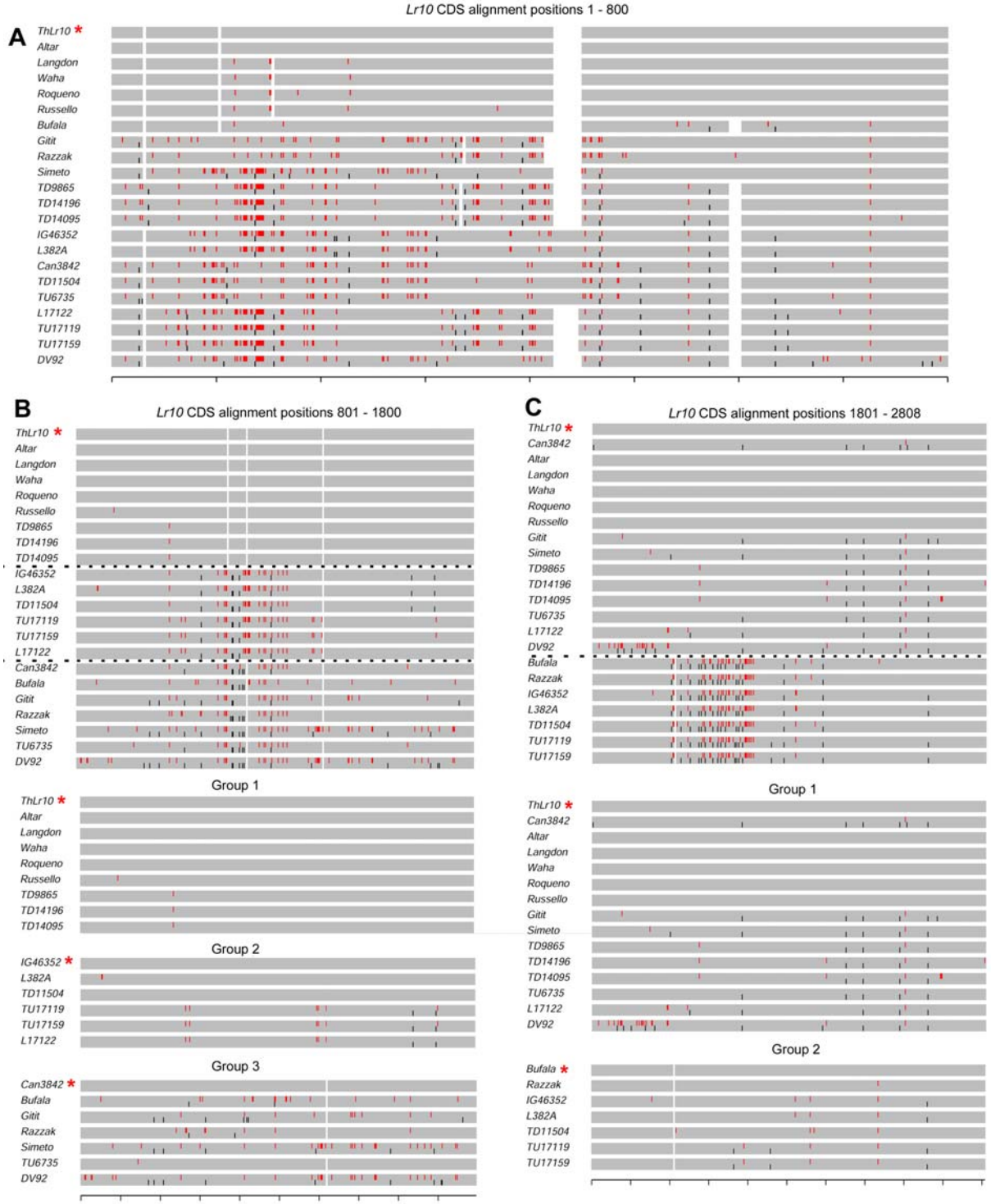
[d]Used for genetic mapping

Supplementary Table 3. GenBank accession numbers and their corresponding sequence names.

Russello_RGA2	EU675942
Altar_RGA2	EU675943
Bufala_RGA2	EU675944
TRI11504_RGA2	EU675945
TRI15095_RGA2	EU675946
TRI15196_RGA2	EU675947
TRI9865_RGA2	EU675948
TRI11498_RGA2	EU675949
TRI17119_RGA2	EU675950
TRI17122_RGA2	EU675951
TRI17159_RGA2	EU675952
TRI6735_RGA2	EU675953
382A_Lr10	EU675954
Simeto_Lr10	EU675955
Russello_Lr10	EU675956
G46352_Lr10	EU675957
Roqueno_Lr10	EU675958
Gitit_Lr10	EU675959
Razzak_Lr10	EU675960
Waha_Lr10	EU675961
Bufala_Lr10	EU675962
Altar_Lr10	EU675963
TRI11504_Lr10	EU675964
TRI14095_Lr10	EU675965
TRI14196_lr10	EU675966
TRI9865_Lr10	EU675967
TRI17119_Lr10	EU675968
TRI17122_Lr10	EU675969
TRI17159_Lr10	EU675970
TRI6735_Lr10	EU675971



Supplementary Figure 1. Schematic representation of the genetic structure of *Lr10* and *RGA2* genes. CC, NBS, spacer and LRR domains are represented in different shades of grey and black. Regions used in PCR, qRT-PCR and for silencing experiments are indicated. The 5' sequence of *RGA2* sequences was amplified using the primer pair PibTm10-H/PibTm1-4 for. For the 3' sequence, primer pair PibTm10A/PibTm10E was used in lines TRI6735 and TRI17122, PibTm10U/PibTm10E in lines TRI11498, TRI14196, TRI9865 and TRI14095, and primer pair RGA23B5/PibTm10E in all other accessions.



Supplementary Figure 2. Graphical display of alignments of the 22 *Lr10* coding sequences. Because of several gene conversion events, the alignment was split into three parts to allow for a better separate analysis of the regions containing gene conversion events. Note that the scale in (**A**) is different to allow the depiction of more details because of the complexity of the gene conversion events in that region. The asterisk indicates the sequence that was used as reference to which all others were compared. In **B** and **C**, an alignment of all sequences with *[Thatcher+Lr10]* (*ThLr10*) as reference is given at the top as well as separate alignments for the individual groups to illustrate levels of polymorphisms between as well as within groups.

Supplementary Text

PCR amplification of *Lr10* and *RGA2* from genomic DNA

Five units of proofreading polymerase (PfuUltra^{TM}High-Fidelity DNA Polymerase, Stratagene, California) was used in a PCR mix supplemented with PfuUltraPolymerase buffer, 1 mM dNTPs, 200 ng of genomic DNA, 2 mM primers. Amplification was for 2 min at 95°C, followed by 31 cycles of 25 s at 95°C, 30 s at the appropriate annealing temperature, 15 min at 68°C followed by a final 68°C-extension for 20 min. Two clones per independent PCR reaction were used to generate a sequence for each amplified gene using the Gap4 Program (Staden Package). Sequencing was done on an ABI3730 automated sequencer (Applied Biosystems) and sequence reads were assembled using Clone Manager DNA for Windows (Sci-Ed Software, http://www.scied.com/index.htm).

Virus-induced gene silencing

γ-capsids of the viral barley stripe mosaic virus were engineered to integrate fragments of genes of interest which were used to infect wheat seedlings (Holzberg et al., 2002). Primer sequences included a *Pac*I restriction site at the 5' end and a *Not*I restriction site at the 3' end of the gene fragment to allow antisense cloning in the viral capsid. Two constructs targeting the *Lr10* gene in the N-terminal (bp positions 101-400 of the CDS, construct CL126) and NBS region (positions 1,028-1378, construct CL128) were amplified with primer pairs CL1/CL2 and CL3/CL4, respectively. Construct CL127 targeting the NBS2 region of *RGA2* (positions 1,229-1,544) was amplified with primer combination 127-CL/130-CL. A fragment of the CC domain (CL118) of the resistance gene *Lr1* (EF439840) targeting base-pairs 104 to 388 (primers 118-CL and 122-CL) was used as a control (Cloutier et al. 2007). Primers and positions of constructs are summarized in Supplementary Table 1 and Supplementary Figure 1.

In vitro transcription of each of the linearised viral subunits (α, β, and modified γ) was performed using the kit mMESSAGE mMACHINE® T7 Kit (Ambion, Austin, USA). Wheat seeds were grown in a climate chamber under a 16h light-regimen (600 µE, 20°C). Five days after germination, seedlings were infected with the viral RNA (Scofield et al. 2005) and viral infection was stimulated by increasing daylight temperature to 25°C. Leaf rust infections were performed as described by Schachermayr et al. (1995) using the isolate *Puccinia triticina BRW 97512-19*.

Quantitative real-time PCR

Total RNA was isolated from leaves of *Thatcher*, *[Thatcher+Lr10]* and *cvs. Langdon* and *Altar* using TRIzol reagent (Invitrogen Life Technologies, Basel, Switzerland) on 19 days-old seedlings when viral treatment had been applied or on 10 days-old untreated seedlings. For reverse transcription, 5 µg of total RNA was denatured at 70°C for 5 min in the presence of 0.07 µg of oligo$(dT)_{20}$ primers and reverse-transcribed with 7 units of reverse transcriptase (Invitrogen Life Technologies, Basel, Switzerland), 1x buffer, 0.7 mM of each dNTPs, 10mM dTT and 1.5 units of RNase OUT (Invitrogen Life Technologies, Basel, Switzerland) in a total volume of 30 µl at 42°C for 90 min. Real-time PCR assays were performed with the ABI PRISM 7700 Sequence Detection System (Applied Biosystems) using SYBR Green PCR Master Mix (Applied Biosystems, Rotkreuz, Switzerland) as described by Travella et al. (2006). Primer pair RT-ST1/RT-ST2 targeted the *Lr10*-N-terminal domain (bp positions 172 – 281). Primer pair RT-ST3/RT-ST4 targeted the *RGA2*-NBS1 domain (positions 969 – 1,069, Supplementary Figure 1). The *Lr10* and *RGA2* regions targeted by real-time PCR were different from the regions targeted by VIGS. To normalize results, the *glyceraldehyde-3-phosphate dehydrogenase* gene (*GAPDH*, AF251217) was used as an internal standard amplifying a 61 bp fragment (440 to 501 bp) using GAPDH-RT-F and GAPDH-RT-R.

Mapping of the leaf rust resistance gene in *cv.* Langdon

The mapping population consisted of 152 F_6 recombinant inbreed lines (RILs) derived from a cross between a *Triticum turgidum ssp. durum* wheat (female) cultivar *Langdon* and a wild emmer wheat, *T. turgidum ssp. dicoccoides* (male) accession G18-16. The wild emmer wheat accession was sampled from Gitit (32°06 N, 35°24 E; 288.4 m above sea level, Peleg et al., 2005). F_1 progeny from the initial cross were taken through five generations of selfing *via* single seed decent to produce homozygous RILs. *Lr10*-specific primers were designed for the parental lines: primers Lgd_F1 and Lgd_R1 amplified a 805 bp fragment in *Langdon*, while primers Git_F1 and Git_R1 amplified a 530 bp fragment from *Gitit* G18-16. Standard PCR reactions were carried out at 52 °C for 1 min in a 30-cycle reaction. The phenotypes were determined following a standard procedure (Schachermayr et al., 1995) after infection with the leaf rust isolate isolate *BRW 97512-19* avirulent on Langdon.

Sequence analysis

Database searches were done with the stand alone Blast software from NCBI (ncbi.nih.gov). Pairwise sequence alignments were done with the EMBOSS program WATER (http://emboss.sourceforge.net/) and DOTTER (Sonnhammer and Durbin, 1995). Multiple alignments were done with CLUSTALW (Thompson et al., 1994) using a gap creation penalty of 3.0 and a gap extension penalty of 0.01.

Homologs of *Lr10* and *RGA2* were identified by BLASTN and/or BLASTX against the whole genome sequences of Arabidopsis, *Brachypodium distachyon*, rice and sorghum. The genome sequences were retrieved from the following resources: Arabidopsis version 6 and rice version 5 from tigr.org, *Brachypodium* from brachypodium.org and sorghum from phytozome.net/sorghum. Coding sequences of all

homologs were determined by comparison with *Lr10* and *RGA2* and annotated by hand.

Putative gene conversion events were identified with DnaSP (ub.es/dnasp/) using the two-gamete method. Regions with apparent gene conversions were tested separately by choosing two sequence sets of at least three and five sequences, respectively (the minimum required by DnaSP), which contained the region of interest. CC structures were predicted with the program COILS (ch.embnet.org/software/COILS_form.html). The visual representation of multiple alignments was done with a home-made Perl script which is available upon request.

Phylogenetic analyses

Phylogenetic analysis was performed with the PHYLIP package (http://evolution.genetics.washington.edu/phylip/) using the protein sequence parsimony method (PROTPARS) on 100 bootstrap replicates with jumbling the order of sequences 3 times for each replicate.

PAML (http://abacus.gene.ucl.ac.uk/software/paml.html) was used to determine non-synonymous-synonymous substitution (Ka/Ks) ratios. Likelihoods of more complex models to null models M_0 (which assumes neutral codon evolution) were compared. Also Likelihoods of *M1a* (nearly neutral codon evolution with two codon classes allowed to take on values from $0 \le \omega_0 \le 1$ or $\omega_0 = 1$) were compared with those of a more complex model M2a (diversifying selection), which incorporates an additional positively selected sites class ($\omega_2 > 1$), and M3 (which assumes three discrete site classes). Results of analyses using model M7 and M8 were also compared, which both assume a β-distribution for $0 \le \omega \le 1$, with the latter model allowing for an extra class of sites with $\omega > 1$. Posterior probabilities (p.p.) fore each codon site to belong to one of

the selection classes was calculated. Codons with a posterior probability >0.95 were considered positively selected (i.e. $\omega > 1$).

References

Cloutier, S., McCallum, B.D., Loutre, C., Banks, T.W., Wicker, T., Feuillet, C., Keller, B., and Jordan, M.C. (2007) Leaf rust resistance gene *Lr1*, isolated from bread wheat (*Triticum aestivum* L.) is a member of the large *psr567* gene family. Plant Mol Biol. **65:** 93-106.

Holzberg, S., Brosio, P., Gross, C., and Pogue, G.P. (2002) Barley stripe mosaic virus-induced gene silencing in a monocot plant. Plant J. **30:** 315-327.

Peleg, Z., Fahima, T., Abbo, S., Krugman, T., Nevo, E., Yakir, D. and Saranga, Y. (2005) Genetic diversity for drought resistance in wild emmer wheat and its ecogeographical associations. Plant, Cell Environ. **28:** 176-191.

Schachermayr G.M., Messmer, M.M., Feuillet, C., Winzeler, H., Winzeler, M., and Keller, B. (1995) Identification of molecular markers linked to the *Agropyron elongatum*-derived leaf rust resistance gene *Lr24* in wheat. Theor. Appl. Genet. **90:** 982-990.

Scofield, S.R., Huang, L., Brandt, A.S., and Gill, B.S. (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 :** 2165-73.

Sonnhammer, E.L., and Durbin, R. (1995) A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene **167:** 1-10.

Travella, S., Klimm, T.E., and Keller B. (2006) RNA interference-based gene silencing as an efficient tool for functional genomics in hexaploid bread wheat. Plant Physiol. **142:** 6-20.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W, improving the sensitivity of progressive multiple sequence alignment through sequence weighting position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673-4680.