Genetics: Published Articles Ahead of Print, published on April 13, 2009 as 10.1534/genetics.108.099614

Evolution of new disease specificity at a simple resistance locus in a crop–weed complex: Reconstitution of the *Lr21* gene in wheat

Li Huang*, Steven Brooks[†], Wanlong Li[‡], John Fellers[§], James C. Nelson[¶], and Bikram Gill[‡1]

[*]Department of Plant Sciences and Plant Pathology, Montana State University, Bozeman, MT 59717-3150. [†]USDA-ARS, Dale Bumpers National Rice Research Center, 2890 Hwy 130 E., Stuttgart, Arkansas 72160. [‡]Wheat Genetic and Genomic Resources Center, Department of Plant Pathology, Kansas State University, Manhattan, KS 66506-5502. [§]USDA-ARS-PSERU, Department of Plant Pathology, 4008 Throckmorton Hall, Kansas State University, Manhattan, KS 66506-5502 [¶]Department of Plant Pathology, 4031 Throckmorton Hall, Kansas State University, Manhattan, KS 66506-5502

[1]*Corresponding author:* Wheat Genetic and Genomic Resources Center, Department of Plant Pathology, Kansas State University, Manhattan, KS 66506-5502.

E-mail: bsgill@ksu.edu; Fax: (785) 532-5692.

Abstract

The wheat leaf-rust resistance gene *Lr21* was first identified in an Iranian accession of goatgrass, *Aegilops tauschii* Coss., the D genome donor of hexaploid bread wheat, and was introgressed into modern wheat cultivars by breeding. In order to elucidate the origin of the gene, we analyzed sequences of *Lr21* and *lr21* alleles from 24 wheat cultivars and 25 accessions of *Ae. tauschii* collected along the Caspian Sea in Iran and Azerbaijan. Three basic nonfunctional *lr21* haplotypes, H1, H2, and H3, were identified. *Lr21* was found to be a chimera of H1 and H2, which were found only in wheat. We attempted to reconstitute a functional *Lr21* allele by crossing cultivars Fielder (H1) and Wichita (H2). Rust inoculation of 5876 F2 progeny revealed a single resistant plant that proved to carry the H1H2 haplotype, a result attributed to intragenic recombination. These findings reflect how plants balance the penalty and necessity of a resistance gene and suggest that plants can reuse "dead" alleles to generate new disease resistance specificity and lead to a "death–recycle" model of plant resistance-gene evolution at simple loci. We suggest that selection pressure in crop–weed complexes contributes to this process.

Introduction

Plants possess large numbers of resistance genes (*R* gene) as a part of an elaborate plant defense system. In different plants, *R* gene structure may consist of a simple (single-copy) or complex loci where genes are in clusters as a result of gene duplication events. This duplication is considered as birth of an *R* gene. *R* genes are necessary for plants to respond to pathogen attacks and survive when the pathogens are around. Mutations, gene

conversion and recombination were found to be the means to create new specificities to various pathogens (for review, Leister, 2004). However, an *R* gene could bring a penalty when the pathogen is absent (Stahl, *et al*., 1999). In such a case, plants have better fitness when they get rid of the *R* gene function. In nature, presence of different pathogens maintains diversity of *R* gene specificities. So far, there is no report on the fates of nonfunctional *R* genes.

In native agricultural ecosystems wild plants often grow as weeds intermixed with or adjacent to their crop relatives. Extensive gene flow occurs between wild and domesticated forms, spawning numerous crop landraces adapted to diverse environments and occasionally new species. Common (hexaploid or bread) wheat (*Triticum aestivum* L., 2n=6x=42, genome formula AABBDD) arose from such a process by hybridization of domesticated tetraploid wheat (*T. turgidum* L. 2n=4x=28, AABB) with goatgrass (*Aegilops tauschii* Coss., 2n=2x=14, DD) growing as a weed in farmers' fields along the Caspian Sea in Iran ca. 8,000 years ago (Kihara, 1944; McFadden and Sears, 1946; Nesbitt and Samuel, 1998).

Because of the pivotal importance of *Ae. tauschii* in wheat evolution and crop improvement, Kihara and colleagues (Kihara *et al*., 1965) made extensive collections from Iran, Afghanistan and adjacent regions. They suggested that Caspian Iran was the center of genetic diversity of *Ae. tauschii*, a proposition later confirmed by molecular-marker analysis (Lubbers *et al*., 1991) as well as of resistance to leaf rust. Nine named and 12 new leaf rust resistance genes have been documented in *Ae. tauschii* and many more remain to be identified (Gill *et al*., 2008). Leaf rust, a scourge of wheat since before Roman times, is caused by the fungus *Puccinia triticina* (Eriks). It attacks mainly the leaf

blade, producing small elliptical orange-red pustules on the upper surface and causing premature defoliation resulting in as much as 40% yield loss (McIntosh *et al*., 1995).

One *Ae. tauschii* accession, TA1599, collected in Caspian Iran, carries a gene named *Lr21* that confers resistance to all known *P. triticina* races. *Lr21*, transferred to wheat in the 1970s (McIntosh *et al*., 1995; Rowland and Kerber, 1974) was recently cloned (Huang *et al*., 2003) and shown to be a simple (single-copy) locus encoding a NBS-LRR protein of 1080 amino acids. Here we report how a simple locus such as *Lr21* evolved novel resistance specificities in a unique "crop–weed" system and how fragments of nonfunctional alleles could be reused in this process.

Materials and methods

Plant materials: Twenty-five accessions of *Ae. tauschii* were used for this study (Table 1). Of these, 12 are the *Lr21* carriers identified from the entire collection of 528 accessions of *Ae. tauschii* collected over a large geographic area representing its genetic diversity and maintained by the Wheat Genetic and Genomic Resources Center (WGGRC) at Manhattan, KS, USA. Additional *Lr* genes are present in two accessions: *Lr39* in TA2450 and *Lr42* in TA2467. The remaining accessions carry no known *Lr* genes and are susceptible to leaf rust in the field. Among the 13 *lr21* accessions, five were sampled at the same collection sites as the *Lr21* accessions, three were collected along the Caspian Sea of Iran within 51 km of the *Lr21* accessions, and the remaining five were from places where no *Lr21* accessions were found.

Of 24 wheat cultivars (Table 1 and S4) with the *lr21* allele that were tested for polymorphism using the KSUD14-STS marker, a PCR-based molecular marker that

distinguishes between *Lr21* and *lr21* (Huang *et al*., 2003), 'Fielder', a spring wheat, and 'Wichita', a winter wheat, were chosen for this study. WGRC7 is a wheat germplasm developed by WGGRC (http://www.k-state.edu/wgrc/) by direct crossing Wichita with *Ae. tauschii* accession TA1649 and then backcrossing with Wichita twice (Raupp *et al*., 1983).

One dicot species, *Arabidopsis thaliana* and five cereal species, barley (*Hordeum vulgare*), oat (*Avena sativa*), rye (*Secale cereale*), maize (*Zea mays*), and rice (*Oryza sativa*) were chosen based on the evolutionary timeline to assess the approximate age of the *Lr21* locus.

DNA manipulation and sequence analysis: DNA isolation, digestion, blotting and Southern hybridization followed the online protocols at http://wheat.pw.usda.gov/NSF/project/mapping-data.html. Several pairs of primers were designed based on the coding regions and flanking sequences of *Lr21* from *Ae. tauschii* accession TA1649. Primers Sta (TTGTGATGGAGAAACGAGTGGCC) and Tor (CGGACGAGTAGTTCTTTCAGGA) were designed to amplify the entire gene and 397 bp flanking regions (Fig. 1). Each full-length allele was amplified by long-range PCR using Herculase enhanced DNA polymerase (Stratagene, La Jolla, CA, USA) from genomic DNA of each accession. The PCR products were then cloned directly using pGEM-T easy system (Promega, Madison, WI, USA). Vector primers sp6 and T7 were first used to sequence the ends of each clone. Internal primers were designed later based on the sequences obtained from previous primers. At least three clones from each accession were sequenced from both directions. All the sequences were assembled using MacVector 6.5.3 (Oxford Molecular Ltd., Madison, WI, USA). Primers Sta3

(TGGCTAATGCAGTGGGCACGG) and D14-R (GGACATTAGGCGATGCTTTGAATTC) were used to amplify the NBS region of the alleles (Fig. 1). The marker KSUD14-STS was designed based on a 105-bp or an 88-bp insertion/deletion (Indel) in the first intron of the *Lr21* (Fig. 1). A 1.36-kb fragment from this region is a signature of the *Lr21*, while a 1465-bp or a 1448-bp fragment with a 105-bp or an 88-bp insertion is a tag of *lr21*. Other sizes of fragments amplified with KSUD14-STS represent *Lr21* paralogs (Figs. 2A, 2B).

Gene expression study: Expression of the *Lr21* and *lr21* alleles was characterized by modified quantitative RT-PCR (Kashkush *et al*., 2003). The mRNA was isolated from leaf tissues with or without inoculation of the pathogen isolate PRTUS6 using MicroPoly(A) Pure^{TM} (Ambion, Austin, TX, USA). 1st strand cDNAs were synthesized using oligo-dT primer and 2nd strand cDNAs were amplified with gene-specific primers. *Lr21* or *lr21* were amplified with D14-F (CGAGATTGGTCCTATGAGGTGGT) and D14-R (Fig. 1). Actin gene expression was used for normalization for the expression study. Actin-F (GGTATCGTGAGCAACTGGGATG) and Actin-R (GTGAAGGAGTAACCTCTCTCGGTG) were used to amplify a 383-bp fragment. PCR was performed under the conditions: 95°C for 4 minutes, and 12 cycles of each with 95°C for 30 seconds, 60°C for 30 seconds and 72°C for 1.5 minutes. Amplicons from D14 or actin primers were electrophoresed separately on 1% agarose gels, then transferred to Hybond-N^{+} membranes (Amersham Biosciences, Piscataway, NJ, USA) and probed with KSUD14 or the actin gene.

Results

The *Lr21* NBS-LRR family in cereal species: Hybridization of the *Lr21* NBS region to wheat, rye, barley and oat revealed multiple bands (Fig. 3), indicating that *lr21* homologs are present elsewhere in the genome and that the *Lr21* NBS-LRR family is shared by this group of grasses with a basic chromosome number of 7. No signal was detected in maize, rice and *Arabidopsis* (Fig. 3), indicating that *Lr21* homologs are absent in these species. Primers designed based on the *Lr21* NBS region were able to amplify fragments from rye, barley and oat by PCR. Sequencing of the amplicons showed 95–98% similarity to the *Lr21* NBS sequence (S1). BLASTn searches against the finished rice genome sequence revealed eight rice homologs with an E-value above 20 and identity from 57% to 70%, consistent with the failure of hybridization. None of the homologs were detected on rice chromosome 5, which is homoeologous to the wheat group 1 chromosomes including 1D, where the *Lr21* locus resides. Southern hybridization with probe KSUD14, a part of the *Lr21* gene, revealed two *Lr21* paralogs located on the short arm of chromosome 1D in the *Lr21* donor accession of *Ae. tauschii* (Huang *et al.*, 2003). Two cosmid clones carrying the paralogs were identified. Full-length sequencing of the paralogs showed them to be NBS-LRR-like sequences with 80% identity to *Lr21* in the NBS region (S2) and only about 50% identity in the rest of the gene.

Only one *Lr21* allele was identified. Of 528 accessions representing the geographic diversity of *Ae. tauschii* that were screened against a mixture of leaf rust pathogens, 158 were found to be resistant at seedling stage. The resistant accessions were tested with the *Lr21*-specific marker KSUD14-STS and 12 were found to carry the *Lr21* allele. All 12 had been collected along the Caspian Sea in Iran and Azerbaijan within a range of 675 km (Fig. 4 and Table 1). They shared an identical 1.36-kb fragment at the *Lr21* locus but

were polymorphic at paralogous loci, with four patterns revealed by marker KSUD14-STS (Fig.2A). Amplification, cloning, and alignment of the *Lr21* sequences showed that all 12 accessions carry an identical allele at the *Lr21* locus.

The *lr21* alleles are pseudogenes. To elucidate the mechanisms underlying the origin of *Lr21* function, we characterized nonfunctional *lr21* alleles from 13 *Ae. tauschii* accessions (Fig. 4 and Table 1) and two *Ae. tauschii*-derived *lr21* alleles present in the D genome of two hexaploid wheat cultivars (Table 1). Based on KSUD14-STS genotyping in *Ae. tauschii*, six different polymorphic patterns were observed, but none contained the 1.36-kb fragment (Fig. 2B). Based on sequencing, 11 different *lr21* alleles were identified from the 13 *Ae. tauschii* accessions; TA2467, TA2473 and TA2481 carried the same allele (Fig. 5, and S3). The wheat D genome alleles *lr21-f* (Fielder) and *lr21-w* (Wichita) were not detected in the sampled *Ae. tauschii* accessions. The sequences of the 11 *lr21 Ae. tauschii* and two wheat D genome *lr21* alleles revealed a spectrum of the accumulated variation at this locus (Fig. 5). There were 36 single-nucleotide polymorphisms (SNPs), none introducing a stop codon and only 11 representing nonsynonymous substitutions. There were also 15 insertions/deletions (indels), three of which caused frameshifts introducing premature stop codons. With reference to the *Lr21* sequence, indel 3 (a 2-bp deletion at position +761) was present in all *lr21* alleles except *lr21-f* and *lr21-TA1699*. This deletion introduced an early stop codon resulting in a putative 151-aa peptide. Similarly, indel 10, a 1-bp deletion at position +1773, was identified only in the *lr21-f* allele, which encodes a putative 380-aa peptide. Indel 12 was a 4-bp deletion at +3195 identified in 12 *lr21* alleles including the *lr21-TA1699* allele. Each *lr21* allele carried one or two of these three indels and was thus a nonfunctional pseudogene.

The two distinct *lr21* alleles found in wheat D genome were identified and cloned from chromosome 1D each of Wichita *(lr21-w*) and Fielder (*lr21-f*), respectively (Fig. 1). The *lr21-w* and *lr21-f* alleles were distributed among the other 22 sampled wheat cultivars (S4) without bias as to winter or spring habit. Both encoded truncated proteins, resulting from a 2-bp indel at position +761 in *lr21-w* and a 1-bp indel at +1773 in *lr21-f* (Fig. 5), and thus were pseudogenes.

The *Lr21* allele is a recombined allele of recent origin. Sequences from the NBS and part of 3' regions of the gene from rye, barley and oat revealed that the 2-bp insertion at indel 3 (position +761) (Fig. 6B) and nucleotides G at position +1864, T at position +4161 and C at position +4170 are shared among the three outgroup species. The presence in the *lr21-f* allele of this 2-bp insertion at indel 3, and the ancestral SNPs (Fig. 6B), support its ancestral character. Sequence comparisons suggested that the 13 different *lr21* alleles were derived from three basic haplotypes and were subsequently modified by point mutation and insertion/deletion events (Fig. 5 and S3). We designated allele *lr21-f* as haplotype H1, *lr21-w* as H2, and *lr21-TA2467* as H3. The *Lr21* allele appears to be a chimera derived from intragenic recombination between H1 and H2 (Fig. 5 and S3), followed by the deletion of a 105-bp segment within an intron, and was designated $H1H2_{-105.}$ The putative crossover site lies between positions +762 and +1772. The recombined allele could encode a full 1080-aa protein and would be free of both the 2-bp (in H2) and 1-bp (in H1) deletions. Another chimeric haplotype H2H1 appeared in the *lr21-TA2477* allele, a putative product of reciprocal recombination carrying the 2-bp deletion (Fig. 5 and S3), and the lr21 protein product of this allele is truncated. The remaining *lr21* alleles are suggested to have been derived from the H3 haplotype as

shown in the supplementary data (S3). All of the *lr21* alleles are present as truncated pseudogenes but are transcribed as revealed by RT-PCR (S5), suggesting that their promoters are still functioning.

A functional allele can be created from two dead alleles. To test the hypothesis that *Lr21* could originate from two dead (nonfunctional) alleles, we crossed wheat cultivars Wichita (H2) and Fielder (H1) (Fig. 6A) and screened 5876 F_2 progeny in the greenhouse using leaf rust isolate PRTUS6. One plant, a putative recombinant designated as F/W-R, was identified based on the lower infection type (IT) than both parents. The critical region between +762 and +1772 was PCR-amplified from the F/W-R plant, cloned, and sequenced. Six of nine clones had sequences identical to H2. The other three had sequences identical to H1 from the 5'-end to position +843 and thereafter identical to H2 (Fig. 6B). These data indicated that F/W-R was heterozygous at the *Lr21* locus for H2 and H1H2 haplotypes. The recombinant H1H2 was identical to the *Lr21* of the *Ae. tauschii* gene (*Lr21-a* in Fig. 6B) except for a 105-bp insertion derived from H2 and four substitutions: G to A at positions +1689 and +1862, T to G at position +1864, and A to G at position +2175 (Figs. 5 and 6B). The presence of this insertion and the four SNPs showed that the reconstituted *Lr21-b* functional allele was created in this cross. The 105-bp sequence lies in the first intron of the gene, and the deletion or insertion of this fragment does not change the length of the peptide encoded by the gene. Two of four substitutions are synonymous, while the one at position +1862 changes a methionine to a valine (both neutral and nonpolar) and the second at +2175 changes the acidic polar aspartic acid to the neutral nonpolar glycine. However, the reconstituted allele *Lr21-b* still confers resistance to the same pathogen isolate, indicating that the amino-acid

changes at these positions do not change the function of the protein. The F/W-R plant was selfed. Fifty progenies were tested for the reaction to leaf rust isolate PRTUS6. Progeny testing revealed that resistance to leaf rust isolate was conferred by a single dominant gene (Fig.7). Ten resistance and 10 susceptible plants were selected for genotyping using the critical region of the *Lr21-b*. The result suggested that the resistance was conferred by the *Lr21-b* allele.

Discussion

We selected *Lr21* gene for map-based cloning more than ten years ago because a dozen *Ae. tauschii* accessions harbored a leaf rust resistance gene mapped to the *Lr21* locus. Our hypothesis was that *Lr21* was a complex locus and was spawning new specificities as a result of unequal crossing over similar to the *Rp1* locus of maize (Hulbert et al.,). It was surprising when molecular cloning revealed it to be a simple, single copy locus (Huang et al. 2003). After investigating the sequence variation at the *Lr21* locus in the 12 *Lr21*-carrier tauschii accessions and the 15 *lr21* alleles in a sample of tauschii and bread wheat accessions, our results showed unexpected monomorphism at the *Lr21* locus among the 12 *Lr21*-carriers. The sequence data revealed *Lr21* to be a chimeric allele providing possible clues to its recent origin through intragenic recombination. This hypothesis was experimentally verified by the recovery of a functional allele in the progenies of a cross between two susceptible parents. These results have important implications about the age of the *Lr21* locus, its origin, evolution and fixation in the context of crop-weed coevolutionary process as distinct from resistance evolution in wild populations and these aspects of this study are discussed below.

An ancient locus with a young allele of *Lr21* NBS-LRR: The evolutionary timeline indicates that wheat diverged from rice and maize ~65 MYA, from barley ~14 MYA and from rye ~7 MYA (Huang *et al*., 2002). Compared to 80% identity with its paralogs (S2), *Lr21* shared more than 95% identity with barley, oat and rye homologs in the NBS region (S1). Since in general, orthologous genes in different species are more similar in sequence to one another than paralogous copies within a species (Hulbert *et al*., 2001; Michelmore *et al*., 1998), it is plausible that *Lr21* is an ancient locus shared by wheat, barley and oat. However, its restricted geographic distribution and DNA-level monomorphism strongly suggest that the *Lr21* allele of *Ae. tauschii* originated more recently in a single event. In this scenario, it spread most likely by rare cross-pollination among *Ae*. *tauschii* populations, farming activity and commerce of wheat grains contaminated with goatgrass seeds.

Chimeric origin of *Lr21*: *Ae. tauschii* is a self-pollinated species with outcrossing rate lower than 5%. The presence in the wheat D genome of the H1 and H2 haplotypes indicates that these alleles were present in *Ae. tauschii* (donor of wheat D genome) growing alongside domesticated tetraploid wheat. It thus appears that *Ae. tauschii* parents carrying H1 and H2 or similar haplotypes were involved in hybridization events leading to origin of bread wheat 8,000 years ago and that bread wheat originated in at least two independent hybridization events (Talbert *et al*., 1998). As H1 and H2 haplotypes were only detected in bread wheat, our failure to detect *Ae. tauschii* accessions carrying haplotypes H1 and H2 most probably was due to limited sampling.

Alternatively, these haplotypes may be extinct in *Ae. tauschii* and preserved only in wheat.

The molecular mechanism underlying the origin of *Lr21* function may be associated with its location at the most distal point of chromosome 1D short arm, a recombination hot spot (Qi *et al*., 2004; Spielmeyer *et al*., 2000). We previously reported one intragenic recombination event between positions -61 and +1354 involving alleles *Lr21* and *lr21-w* in a sample of 332 F2 plants (Huang *et al*., 2003). It involved a conversion tract of a minimum of 191 bp and maximum of 1415 bp of DNA from the *lr21-w* to *Lr21*, rendering the latter ineffective. We have now experimentally reconstituted *Lr21* through another intragenic recombination event. In addition, one H2H1 haplotype, an obvious product of intragenic recombination, was detected in a small sample of 13 *lr21* alleles. These findings suggest that recombination events have occurred multiple times at the *Lr21* locus and that the evolutionary history of *Lr21* has been shaped by its location in a high-recombination region.

"Birth–recycle" at the *Lr21* locus: The presence of only one functional allele among an assortment of nonfunctional alleles at the *Lr21* locus suggests a cost of carrying the resistance allele in the absence of virulent pathogen strains. Two evolutionary classes of NBS-LRR genes have been characterized. One supports the so-called "arms-race" model represented by the *L* locus of flax (Ellis *et al*., 1999) and the *RPP13* locus of *Arabidopsis* (Rose *et al*., 2004), containing large numbers of different functional alleles and high degree of variation in the regions responsible for specificity. The other class, consistent with a "trench warfare" model, is represented by the *RPM1* (Stahl *et al*., 1999) and *RPS2* (Mauricio *et al*., 2003) loci of *Arabidopsis* where variation is low with no evidence of

diversifying selection between functional and non-functional forms. The functional *RPM1* allele has been shown to impose a penalty in the absence of the pathogen. Complete deletion is one way to remove the deleterious effect of an allele such as *RPM1*. An alternative way is truncation, as seen with the *lr21* alleles.

Our discovery suggests a "death–recycle" model of plant resistance-gene evolution at simple loci. A "birth-and-death" process similar to that of the vertebrate major histocompatibility complex (MHC), T-cell receptor (TCR) and immunoglobulin genes has been proposed to explain the evolution of resistance genes at complex loci (Michelmore *et al*., 1998). At a simple locus, there is no "birth" associated with gene duplication. A functional allele may become ineffective because of mutation or defeat by a new race of the pathogen, and then may be reused in the creation of a new functional allele at that locus. Our results have confirmed that plants can reuse nonfunctional alleles to create new resistance specificity. The recycling of *lr21* hints at the potential usefulness of truncated alleles. New resistances similar to *Lr21* that occurred in nature may also arise in plant breeding programs more often than recognized because of extensive selection pressure for rare new disease specificities in segregating populations subjected to disease epidemics.

The wheat–goatgrass complex and the fixation of *Lr21*: *Ae. tauschii* accessions carrying the functional *Lr21* allele are predominant in regions where agriculture was practiced, to the point where at one location, Ramsar, Iran, all collected accessions carried this allele (Fig. 4 and Table 1). This predominance would be expected if some evolutinary process in the crop–weed agroecosystem led to fixation of a new disease-resistance specificity created by a rare recombination event in *lr21*. Agricultural practice

favors crop monoculture or a single variety planted over large areas for long periods of time. This often leads to much higher rust disease pressure in wheat fields than in wild *Ae. tauschii* populations, which harbor mixtures of different rust resistance genes or even several leaf-rust resistance genes in a single accession (Gill *et al*., 2008). We propose that leaf rust epidemics in a crop monoculture imposed selection pressure on goatgrass populations in or near wheat fields. A plant carrying the *Lr21* allele would have fitness advantage in an environment with high leaf rust inoculum. In this model, the crop–weed complex coevolutinary process was critical to the selection and retention of the *Lr21* gene in *Ae. tauschii* populations.

It appears that native agricultural ecosystems, located in Vavilovian world centers of crop plant origin, are virtual outdoor laboratories for the creation of genetic variation. Other "new" genes spawned by such ecosystems may well be of the same worldwide economic significance as *Lr21.* The example presented here argues for the careful preservation of the native agricultural ecosystems in the face of modern agricultural practices as the success of modern plant breeding hinges on the extensive use of genetic variation present in land races and wild relatives of crop plants

Acknowledgements: This research was supported by the Kansas Wheat Commission and a USDA-CSREES special grant to the WGGRC. We are grateful to Frank White and Eduard Akhunov for critical review of the manuscript. This is contribution 07-305-J from the Kansas Agricultural Experiment Station, Manhattan, Kansas, USA.

End notes:

[All DNA sequences have been deposited in GenBank under accession numbers_____.]

References

Ellis, J. G., G. J. Lawrence, J. E. Luck, and P. N. Dodds, 1999 Identification of regions in alleles of the flax rust resistance gene *L* that determine differences in gene-for-gene specificity. *Plant Cell* **11,** 495-506.

Gill, B. S., L. Huang, V. Kuraparthy, W. J. Raupp, D. L. Wilson, and B. Friebe, 2008 Alien genetic resources for wheat leaf rust resistance, cytogenetic transfer, and molecular analysis. *Austral. J. Agr. Res.* **59,** 197-205.

Huang, S., A. Sirikhachornkit, X. Su, J. Faris, B. S. Gill, R. Haselkorn and P. Gornicki, 2002 Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. *Proc. Natl. Acad. Sci. USA* **99,** 8133-38.

Huang, L., S. A. Brooks, W. Li, J. P. Fellers, H. N. Trick and B. S. Gill, 2003 Map-based cloning of leaf rust resistance gene *Lr21* from the large and polyploid genome of bread wheat. *Genetics* **164,** 655-664.

Hulbert, S. H., C. A. Webb, S. M. Smith and Q. Sun, 2001 Resistance gene complexes: evolution and utilization. *Annu. Rev. Phytopathol.* **39,** 285-312.

Kashkush, K. M. Feldman and A. A. Levy, 2003 Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nature Genet.* **33,** 102-106.

Kihara, H. 1944 Discovery of the DD-analyser, one of the ancestors of *Triticum vulgare* (in Japanese). *Agric. Hortic.* **19,** 13-14.

Kihara, H., K. Yamashita, and M. Tanaka, 1965 *Results of the Kyoto University Scientific Expedition to the Karakoram and Hindukush. 1-118* (Kyoto University.

Leister, D, 2004 Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. *Trends in Genet.* **20**: 116-122.

Lubbers, E. L., K. S. Gill T. S. Cox and B. S. Gill, 1991 Variation of molecular markers among geographically diverse accessions of *Triticum tauschii*. *Genome* **34,** 354-361.

Mauricio, R., E. A. Stahl, T. Korves, D. Tian, M. Kreitman and J. Bergelson, 2003 Natural selection for polymorphism in the disease resistance gene *Rps2* of Arabidopsis thaliana. *Genetics* **163,** 735-46.

McFadden, E. S., and E. R. Sears, 1946 The origin of *Triticum spelta* and its free-threshing hexaploid relatives. *J. Hered.* **37,** 81-89, 107-116.

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

Michelmore, R. W., and B. C. Meyers, 1998 Clusters of resistance genes in plants

evolve by divergent selection and a birth-and-death process. *Genome Res.* **8,** 1113-30.
Nesbitt, M. and D. Samuel, 1998 Wheat domestication: archaeobotanical evidence. *Science* **279,** 1431.
Qi, L. L., B. Echalier, S. Chao, G. R. Lazo, G. E. Butler and O. D. Anderson, 2004 A chromosome bin map of 16,000 expressed sequence tag loci and distribution of genes among the three genomes of polyploid wheat. *Genetics* **168,** 701-712.
Raupp, W. J., L. E. Browder and B. S. Gill, 1983 Leaf rust resistance in *Aegilops squarrosa*: its transfer and expression in common wheat (*Triticum aestivum* L.). *Phytopath*. **73: 818**.
Rose, L. E., P. D. Bittner Eddy, C. H. Langley, E. B. Holub, R. W. Michelmore, and J. L. Beynon, 2004 The maintenance of extreme amino acid diversity at the disease resistance gene, *RPP13*, in *Arabidopsis thaliana. Genetics* **166,** 1517-1527.
Rowland, G. G. and E. R. Kerber, 1974 Telocentric mapping in hexaploid wheat of genes for leaf rust resistance and other characters derived from *Aegilops squarrosa*. *Can. J. Genet. Cytol.* **16,** 137-144.
Spielmeyer, W., O. Moullet, A. Laroche and E. S. Lagudah, 2000 Highly recombinogenic regions at seed storage protein loci on chromosome 1DS of *Aegilops tauschii*, the D-genome donor of wheat. *Genetics* **155,** 361-367.
Stahl, E. A., G. Dwyer, R. Mauricio, M. Kreitman and J. Bergelson, 1999 Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. *Nature* **400**, 667-71.
Talbert, L.E., L.Y. Smith and N.K. Blake, 1998 More than one origin of hexaploid wheat is indicated by sequence comparison of low-copy DNA. *Genome* **41**, 402-407.

Figure legends:

Figure 1. Gene structure and primer location of the *Lr21* and *lr21* alleles. The enlarged regions show the major differences between the *lr21*-Wichita and the *lr21*-Fielder. A-terminus (amino terminus), NBS (nucleotide-binding site), LRR (leucine-rich repeats), and C-terminus (carboxy terminus), 5'UTR (5' untranslated region).

Figure 2. Polymorphism survey of *Ae. tauschii* based on KSUD14-STS marker.

2A: Four patterns (1~4) were revealed among 12 *Lr21* accessions. The 1.36-kb fragment (indicated by an arrow) is a tag of *Lr21* and other size fragments are

Lr21 paralogs. 2B: Six patterns (5~10) were identified among the 13 *lr21* accessions, of which none has the 1.36-kb fragment.

Figure 3. Southern hybridization of genomic DNAs digested with restriction enzyme *Xba*I and probed with KSUD14 (NBS region of *Lr21*). The hybridization stringency amounted to 80% homology.

Figure 4. The sampling sites and geographic distribution of the 25 *Ae. tauschii* accessions used in this study. Twelve accessions (in ***bold italics***) are the *Lr21* carriers.

Figure 5. Sequence variation among tested alleles of *Lr21* and *lr21*. Position numbers are based on the *Lr21* sequence and are shown as offsets from the transcription start site. *lr21-f* is the Fielder allele and *lr21-w* is the Wichita allele. SNP: single-nucleotide polymorphism; indel: insertion/deletion; asterisk: identity to *Lr21*; hyphen: deletion. Positions shaded in grey are those where the nucleotide change resulted in an amino-acid change. The Xs represent deletions in *Lr21*. Four functional domains: carboxy terminus, NBS (nucleotide-binding site), LRR (leucine-rich repeats), and amino terminus are indicated above the sequence.

Figure 6. (A) A crossing scheme for reconstituting a functional *Lr21* allele from two nonfunctional alleles. (B) Shows the indel and SNPs shared among the three outgroup species rye, barley and oat; and the region in which intragenic recombination between H1 and H2 resulted in a chimerical allele *Lr21-b. Lr21-a* is the allele identified from *Ae. tauschii*. Position numbers are based on the *Lr21* sequence. In both panels black shading corresponds to haplotype H1 representing allele *lr21-f* from cultivar Fielder, and gray to H2 representing *lr21-w* from Wichita. Hatching indicates SNPs distinguishing *Lr21-b* from *Lr21-a*.

Figure 7. Infection types of WGRC7 (*Lr21-a/Lr21-a*), progenies of the F/W recombinant (*Lr21-b/lr21-w*), Fielder (*lr21-f/lr21-f*) and Wichita (*lr21-w/lr21-w*) nine days after inoculation with the leaf rust isolate PRTUS6.

Supporting information S1. Sequence alignment in the NBS region of *Lr21* among barley, oat, rye and *Ae. tauschii*. The nucleotides highlighted in yellow are shared between *Lr21, lr21-f* and three outgroup.

Supporting information S2. Sequence alignment in the NBS region of *Lr21* and the two paralogs on the chromosome 1D of *Ae. tauschii*.

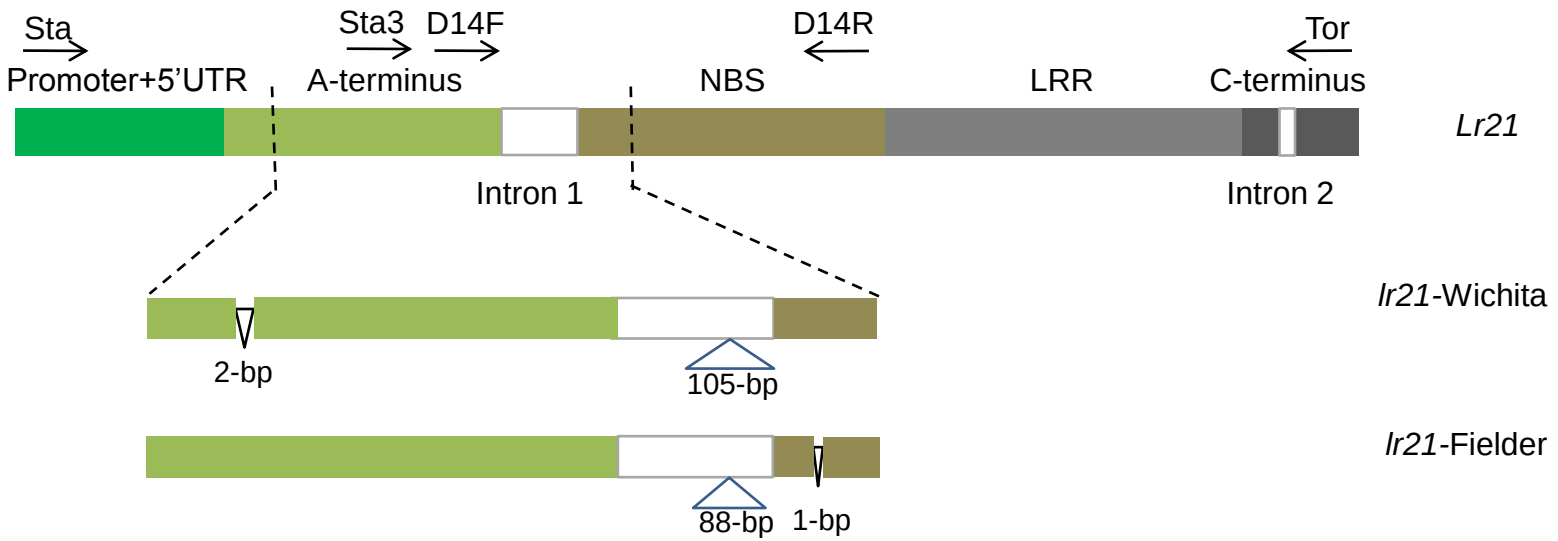
Supporting information S3. A reconstruction of the haplotype structure at the *Lr21* locus based on the nucleotide variation corresponding to figure 4. The figure shows that three haplotypes H1, H2 and H3 (triangles) account for all the derived alleles but does not suggest their evolutionary order. A black bar indicates a SNP and a white bar an indel. Implied haplotypes X1, X2 and X3 were not observed in this study. Both *Lr21* and *lr21-TA2477* are recombinants of H1 and H2.

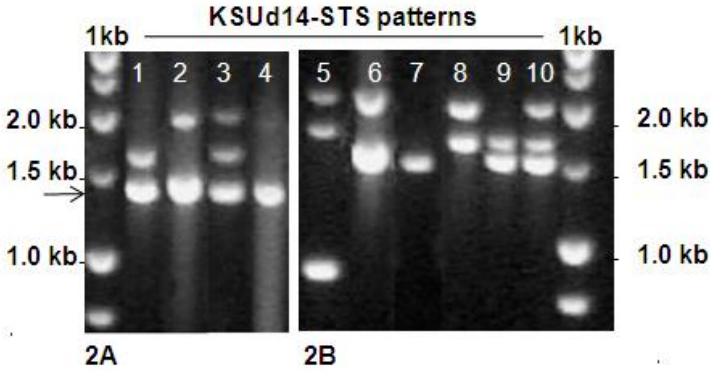
Supporting information S4. Wheat cultivars tested for the *lr21* locus.

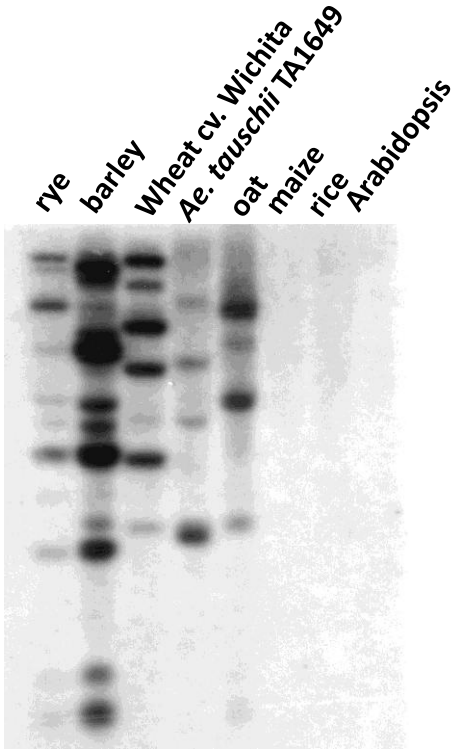
Supporting information S5. Expression patterns of *Lr21* and *lr21* before and 24 hours after pathogen inoculation, as revealed by modified quantitative RT-PCR.

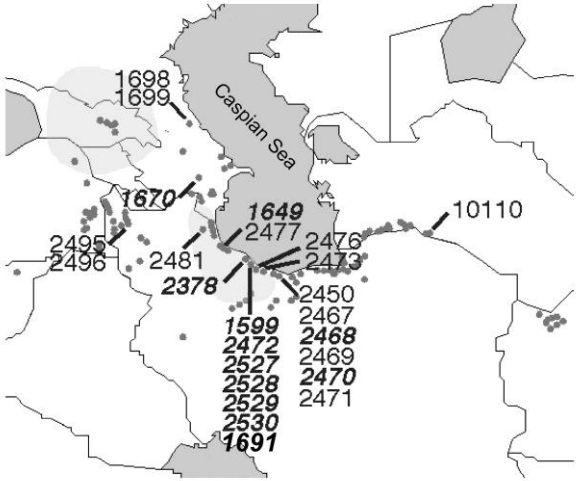
Table legends:

Table 1: *Triticum* and *Aegilops* accessions used for sequencing of *Lr21* and *lr21* alleles. Polymorphism patterns are based on KSUD14-STS marker analysis (Figs. 2A, 2B).


Sta
Promoter+5'UTR
Sta3
D14F
A-terminus
D14R
NBS
LRR
Tor
C-terminus
Lr21
Intron 1
Intron 2
lr21-Wichita
2-bp
105-bp
lr21-Fielder
88-bp
1-bp


KSUd14-STS patterns
1kb
1kb
1 2 3 4
5 6 7 8 9 10
2.0 kb
1.5 kb
1.0 kb
2.0 kb
1.5 kb
1.0 kb
2A
2B

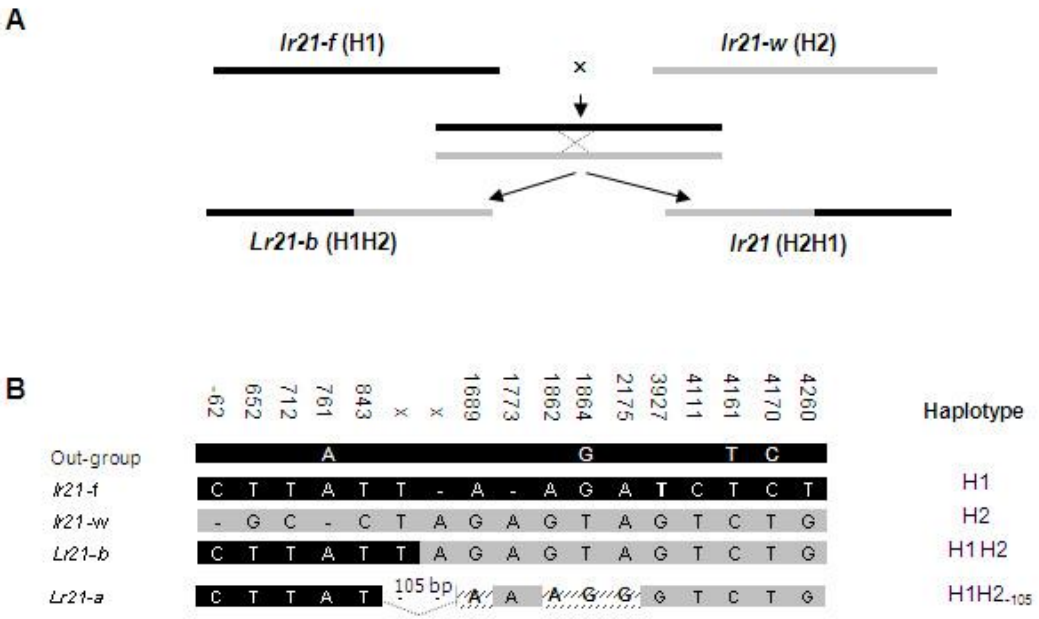

rye
barley
Wheat cv. Wichita
Ae. tauschii TA1649
oat
maize
rice
Arabidopsis


Caspian Sea
1698
1699
1670
1649
2477
2476
2473
2495
2496
2481
2378
2450
2467
1599
2472
2527
2528
2529
2530
1691
2468
2469
2470
2471
10110

---------- Amino terminus ----------| ---------- NBS ----------+--------- LRR ----------| ------Carboxy terminus ------

SNPs	1		2		3	4	5		6		7					8				9	10	11	12	13	14	15
Indels		1		2				3		4		5	6	7	8		9	10	11							
Position	-104	-62	487	489	515	652	712	761	843	x	1276	x	x	x	x	1689	x	1773	1861	1862	1864	2015	2095	2175	2271	2564
Lr21	C	C	A	G	T	T	T	A	T	-	T	-	-	-	-	A	-	A	A	A	G	C	G	G	G	G
lr21 -f	*	*	*	*	*	*	*	*	*	*	*	T	*	G	*	*	*	-	*	*	*	*	*	A	*	*
lr21 -w	*	-	*	*	*	G	C	-	C	*	*	T	A	G	*	G	*	*	*	G	T	*	*	A	*	*
lr21 -TA10110	*	-	*	*	*	G	C	-	C	A	C	T	*	G	*	-	*	-	-	-	-	-	-	-	*	*
lr21 -TA2467	*	-	*	*	*	G	C	-	C	*	*	T	A	G	*	*	*	*	*	*	T	*	*	A	*	*
lr21 -TA2469	*	-	*	*	*	G	C	-	C	*	*	T	A	G	*	*	*	*	*	*	T	*	*	A	*	*
lr21 -TA2450	*	-	*	*	*	G	C	-	C	*	*	T	A	G	*	*	*	*	-	-	-	*	T	A	*	*
lr21 -TA2476	*	-	*	*	*	G	C	-	C	*	*	T	A	G	*	*	*	*	*	*	T	*	*	A	*	*
lr21 -TA2473	*	-	*	*	*	G	C	-	C	*	*	T	A	G	*	*	*	*	*	*	T	*	*	A	*	*
lr21 -TA2481	*	-	*	*	*	G	C	-	C	*	*	T	A	G	*	*	*	*	*	*	T	*	*	A	*	*
lr21 -TA2471	*	-	*	*	*	G	C	-	C	*	*	T	A	G	*	*	*	*	*	*	T	*	*	A	*	*
lr21 -TA1698	*	-	T	-	C	G	C	-	C	*	*	T	A	G	*	*	*	*	-	-	-	A	*	A	*	A
lr21 -TA1699	*	-	T	-	C	*	C	*	C	*	*	T	A	G	*	*	*	*	-	-	-	A	*	A	*	A
lr21 -TA2495	*	-	*	*	*	G	C	-	C	*	*	T	A	G	*	*	A	*	*	*	T	*	*	A	*	*
lr21 -TA2496	*	-	*	*	*	G	C	-	C	*	*	T	A	G	C	*	*	*	*	*	T	*	*	A	*	*
lr21 -TA2477	A	-	*	*	*	G	C	-	C	*	*	T	A	G	*	*	*	*	*	*	T	*	*	*	C	*

SNPs	16	17	18	19	20				21	22	23	24	25	26	27	28	29	30	31	32	33		34	35	36
Indels						12	13	14														15			
Position	2797	2924	2991	3022	3193	3195	3243	3244	3361	3691	3827	3844	3902	3927	3972	4032	4111	4161	4167	4170	4201	4220	4255	4260	4282
Lr21	T	T	A	T	A	A	T	G	G	A	G	A	T	G	C	C	T	C	C	T	C	T	G	G	C
lr21 -f	*	*	*	*	*	*	*	*	*	*	*	*	*	T	*	*	C	T	*	C	*	*	*	T	*
lr21 -w	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
lr21 -TA10110	*	*	*	*	C	-	-	-	*	G	*	*	C	T	*	*	C	T	*	C	*	*	*	T	*
lr21 -TA2467	G	*	*	*	C	-	-	-	*	G	A	*	C	T	*	*	C	T	*	C	*	*	*	T	*
lr21 -TA2469	G	C	*	*	C	-	-	-	*	G	A	*	C	T	T	*	C	T	*	C	*	*	*	T	*
lr21 -TA2450	G	*	*	*	C	-	-	-	*	G	*	G	C	T	*	T	C	T	*	C	A	*	*	T	*
lr21 -TA2476	G	*	*	*	C	-	-	-	*	G	A	*	C	T	*	*	C	T	T	C	*	*	A	T	*
lr21 -TA2473	G	*	*	*	C	-	-	-	*	G	A	*	C	T	*	*	C	T	T	C	*	*	A	T	*
lr21 -TA2481	G	*	*	*	C	-	-	-	*	G	A	*	C	T	*	*	C	T	T	C	*	*	A	T	*
lr21 -TA2471	G	*	*	*	C	-	-	-	*	G	A	*	C	T	*	*	C	T	T	C	*	*	*	T	*
lr21 -TA1698	G	*	*	*	C	-	-	-	A	G	*	*	C	T	*	*	C	T	*	C	*	*	*	T	*
lr21 -TA1699	*	*	*	*	C	-	-	-	A	G	*	*	C	T	*	*	C	T	*	C	*	*	*	T	*
lr21 -TA2495	G	*	*	C	C	-	-	-	*	G	*	*	C	*	*	*	C	T	*	C	*	*	*	T	*
lr21 -TA2496	G	*	*	*	C	-	-	-	*	G	*	*	C	T	*	*	C	T	*	C	*	*	*	T	*
lr21 -TA2477	*	*	G	*	*	*	C	*	*	*	*	*	*	T	*	*	C	T	*	C	*	-	*	T	T


A
Ir21-f (H1)
Ir21-w (H2)
×
Lr21-b (H1H2)
Ir21 (H2H1)
B
-62
652
712
761
843
x
x
1689
1773
1862
1864
2175
3927
4111
4161
4170
4260
Haplotype
Out-group
A
G
T
C
Ir21-f
C T T A T T - A - A G A T C T C T
H1
Ir21-w
- G C - C T A G A G T A G T C T G
H2
Lr21-b
C T T A T T A G A G T A G T C T G
H1 H2
Lr21-a
C T T A T
105 bp
A A A G G G T C T G
H1H2-105

Infection types