Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Cloning of the broad-spectrum wheat leaf rust resistance gene Lr47 introgressed from Aegilops speltoides

Hongna Li
Peking University Institute of Advanced Agricultural Sciences
Lei Hua
National Key Laboratory of Wheat Improvement, Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agriculture Sciences in Weifang
Shuqing Zhao
State Key Laboratory of North China Crop Improvement and Regulation, College of Plant Protection, Hebei Agricultural University
Ming Hao
Sichuan Agricultural University https://orcid.org/0000-0003-2693-3864
Rui Song
National Key Laboratory of Wheat Improvement, Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agriculture Sciences in Weifang
Shuyong Pang
National Key Laboratory of Wheat Improvement, Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agriculture Sciences in Weifang
Yanna Liu
National Key Laboratory of Wheat Improvement, Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agriculture Sciences in Weifang
Hong Chen
Triticeae Research Institute, Sichuan Agricultural University
Wenjun Zhang
University of California, Davis
Tao Shen
National Key Laboratory of Wheat Improvement, Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agriculture Sciences in Weifang
Jin-Ying Gou
China Agricultural University https://orcid.org/0000-0002-7540-0354
Hailiang Mao
National Key Laboratory of Crop Genetic Improvement
Guiping Wang

National Key Laboratory of Wheat Improvement, Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agriculture Sciences in Weifang

Xiaohua Hao

National Key Laboratory of Wheat Improvement, Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agriculture Sciences in Weifang

Jian Li

National Key Laboratory of Wheat Improvement, Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agriculture Sciences in Weifang

Baoxing Song

National Key Laboratory of Wheat Improvement, Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agriculture Sciences in Weifang

Caixia Lan

National Key Laboratory of Crop Genetic Improvement

Zaifeng Li

Agricultural University of Hebei

Xing Wang Deng

Peking University https://orcid.org/0000-0001-8709-1467

Jorge Dubcovsky

Howard Hughes Medical Institute and University of California, Davis https://orcid.org/0000-0002-7571-4345

Xiaodong Wang

Hebei Agricultural University

Shisheng Chen (shisheng.chen@pku-iaas.edu.cn)

National Key Laboratory of Wheat Improvement, Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agriculture Sciences in Weifang https://orcid.org/0000-0002-8617-4356

Article

Keywords: wheat, leaf rust, resistance genes, CC-NBS-LRR, Aegilops speltoides, introgression

Posted Date: May 25th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2944166/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Version of Record: A version of this preprint was published at Nature Communications on September 28th, 2023. See the published version at https://doi.org/10.1038/s41467-023-41833-2.

Abstract

Leaf rust, caused by *Puccinia triticina*Eriksson (*Pt*), is one of the most severe foliar diseases of wheat. Breeding for leaf rust resistance is a practical and sustainable method to control this devastating disease. Here, we report the identification of *Lr47*, a broad-spectrum leaf rust resistance gene introgressed into wheat from*Aegilops speltoides*. The *Lr47* gene encodes a coiled-coil nucleotide-binding leucine-rich repeat protein that is both necessary and sufficient to confer *Pt* resistance, as demonstrated by loss-of-function mutations and transgenic complementation. New *Lr47* introgression lines with no or reduced linkage dragwere generated using the *ph1b* mutation, and a diagnostic molecular marker for *Lr47* was developed. The CC domain of the Lr47 protein was not able to induce cell death, nor did it have self-protein interaction. The cloning of *Lr47* expands the number of *Pt*-resistance genes that can be incorporated into multigene transgenic cassettes to control this devastating disease.

Introduction

Wheat is one of the leading food crops providing one-fifth of the total food calories and proteins consumed by humans. Reducing yield losses inflicted by fungal pathogens is an effective measure to increase wheat production. *Puccinia triticina* Eriksson (*Pt*), the causal agent of wheat leaf rust (or brown rust), is one of the most devastating fungal pathogens in wheat. This disease occurs in most wheat growing regions [1,2] and can cause significant yield losses in susceptible wheat varieties under favorable climatic conditions [3]. According to recent studies, global annual wheat yield losses due to this pathogen could be approximately 25 million tons [4,5]. Breeding for leaf rust resistance is considered the most practical and sustainable way to control this devastating disease.

To date, over 80 leaf rust resistance (*Lr*) genes have been officially designated in wheat and its wild relatives [6]. However, due to the size and complexity of the wheat genome, only ten *Lr* genes have been cloned to date either by classical map-based cloning (*Lr1*, *Lr10*, *Lr21*, *Lr34*, *Lr42*, and *Lr67*) [7-12] or by rapid gene-cloning approaches including MutRenSeq (*Lr13*), TACCA (*Lr22a*), MutChromSeq (*Lr14a*), and MutIsoSeq (*Lr9*) [13-16]. Among the cloned *Lr* genes, *Lr34* and *Lr67* are known as slow rusting genes encoding a putative ATP-binding cassette transporter and a hexose transporter, respectively [10,11]. *Lr14a* encodes a protein containing 12 ankyrin repeats, and *Lr9* encodes a protein with an N-terminal tandem kinase domain followed by vWA/Vwaint domains at its C-terminus [15,16]. The other six isolated genes have been shown to encode typical coiled-coil nucleotide-binding leucine-rich repeat (NLR) proteins [7-9,12-14].

Cloning additional *Lr* genes is desirable to diversify the combinations of *Pt* resistance genes used in transgenic cassettes or gene pyramiding to achieve durable resistance. The diploid wheat species *Aegilops speltoides* Tausch (2n = 2x = 14, SS), the closest extant donor of the B-genome to bread wheat [17,18], harbors valuable *Lr* genes, such as *Lr28*, *Lr35*, *Lr36*, *Lr47*, *Lr51*, and *Lr66* [18-20]. These genes have been successfully transferred from *Ae. speltoides* to common wheat, but none of them has been

cloned so far due to the limited recombination observed between wheat and *Ae. speltoides* chromosomes.

The resistance gene *Lr47* was transferred from chromosome 7S#1 of *Ae. speltoides* to the short arm of chromosome 7A of hexaploid wheat translocation line T7AS-7S#1S-7AS·7AL using homoeologous recombination in the presence of the *Pairing homeologous1* mutation (*ph1b*) [18], but the ancestral origin of this *Ae. speltoides* segment is not clear [21,22]. Using C-banding and restriction fragment length polymorphism (RFLP) markers, it was determined that the genetic length of the translocated segment 7S#1S was approximately 20–30 centimorgans (cM) long and that it was located 2–10 cM from the centromere [18]. In a recent study, the length of the *Ae. speltoides* segment 7S#1S was estimated to be between 157 and 174 Mb based on a set of simple sequence repeat (SSR) markers [23]. This interstitial translocation segment carrying *Lr47* was subsequently backcrossed into several spring wheat cultivars, such as Pavon, Express, Kern, RSI5, Yecora Rojo, and UC1041 [22,24]. However, the presence of the *Lr47* introgression was found to be associated with several detrimental effects on agronomic and quality traits, including reduced grain yield, decreased flour yield, and increased flour ash [22]. Therefore, new rounds of homoeologous recombination using the *ph1b* mutation are needed to reduce the size of the introgressed segment containing *Lr47* and minimize linkage drag. The linkage between *Lr47* and these negative effects can also be broken by cloning the gene and generating transgenic plants. This is a worthwhile endeavor because *Lr47* is one of a few genes known to confer strong levels of resistance against a wide range of *Pt* isolates [1,18,23,25-30], and may play an important role in the improvement of wheat resistance to leaf rust.

Here, we report the identification of the *Lr47* gene through a combination of map-based cloning, ethyl methanesulfonate (EMS) mutagenesis and transcript assembly (EMTA) approaches. We also validate the function of the cloned candidate gene using independent susceptible EMS mutants, barley stripe mosaic virus (BSMV)-mediated gene editing [31], and stable transgenic complementation. To minimize potential linkage drag, we generated smaller *Ae. speltoides* chromosome segments carrying *Lr47* using the *ph1b* mutation. Finally, we developed a diagnostic molecular marker to accelerate the deployment of *Lr47* in wheat breeding programs.

Results

Lr47 **provides broadly effective resistance to diverse leaf rust pathotypes.** The bread wheat line Kern *Lr47* (PI 603918*7/Kern) is near isogenic to wheat cultivar cv. Kern and carries the introgressed *Ae. speltoides* chromosome segment 7S#1S. In the seedling tests, all Kern *Lr47* plants showed strong resistance to *Pt* race THDB, whereas all Kern plants were susceptible (Fig. 1a). In a subset of 128 F_2 plants from the cross Kern *Lr47* × ZZ5389 evaluated with race THDB, 93 plants were resistant and 35 were susceptible, which fits a 3:1 segregation ratio expected for a single dominant genetic locus ($\chi^2 = 0.38$, $P = 0.54$). To quantify the growth of *Pt* pathogen at 2, 4, 6, and 8 dpi in Kern *Lr47* and its recurrent parent Kern, we measured the average infection areas using fluorescence microscopy and the fluorescent dye WGA-FITC, which stains

the *Pt* pathogen. At all four time points, the average infection areas observed microscopically were significantly smaller ($P < 0.001$) in plants with *Lr47* than in those without the gene (Fig. 1b, c). Fluorescent images of *Pt* growth in Kern showed a diffuse network of fungal hyphae at the edges of the infected sites, indicating that *Pt* appeared to spread unimpeded. In contrast, pathogen growth in Kern *Lr47* was restricted and no significant differences were observed in the fungal infection areas from 4 to 8 dpi (Fig. 1c).

Seedlings of *Lr47* NILs (Express *Lr47*, Yecora Rojo *Lr47*, UC1041 *Lr47*, and RSI5 *Lr47*) and their recurrent parents (Express, Yecora Rojo, UC1041, and RSI5) were challenged with the other 23 *Pt* pathotypes collected in China (Supplementary Table 3). All *Lr47* NILs exhibited strong levels of resistance (ITs = 0;) to the tested races (Fig. 1d, Supplementary Table 3). In contrast, the recurrent parents showed a wide range of responses to different *Pt* races, probably due to other *Lr* resistance genes present in these genetic backgrounds. Nevertheless, inoculation of these recurrent parents revealed that at least one accession was fully susceptible to each *Pt* race (Fig. 1d), confirming that *Lr47* confers resistance to these *Pt* pathotypes.

Characterization of the Ae. speltoides **segment introgressed into common wheat.** To determine the physical location and size of the *Ae. speltoides* chromosome segment introgressed into hexaploid wheat, we compared the SNPs identified in the RNA-seq of *Lr47* NILs (Kern *Lr47*, Yecora Rojo *Lr47*, and UC1041 *Lr47*) with those from another 10 sequenced hexaploid wheat varieties and three *Ae. speltoides* accessions (AE915, AE1590, and PI 554292). We focused on the SNPs that were present in the three *Ae. speltoides* accessions, but absent in the 10 sequenced hexaploid wheat varieties, and that are hereafter referred to as *Ae. speltoides*-specific SNPs. Based on this rule, we identified 3,169 *Ae. speltoides*-specific SNPs (Supplementary Table 5) that were shared with *Lr47* NILs in the proximal region of chromosome arm 7AS starting from 40.4 Mb to 190.5 Mb (CS RefSeq v1.1 coordinates, Supplementary Fig. 1). The translocation breakpoints ranged from 40,050,116 bp to 40,398,277 bp at the terminal end, and from 190,476,279 bp to 191,833,539 bp at the proximal end (Supplementary Table 5). Thus, the *Ae. speltoides* segment in the three *Lr47* NILs ranges from 150.1 Mb to 151.8 Mb.

To better define the translocation breakpoints, we designed 16 7A/7S genome-specific primers across the ~ 150 Mb introgressed *Ae. speltoides* segment (Supplementary Table 4), and used them to genotype the *Lr47* NILs and their recurrent parents. We found that translocation breakpoints in *Lr47* NILs were located between markers *pku0738* and *pku0745* at the terminal end, and between markers *pku2216* and *pku2233* at the proximal end (Supplementary Fig. 2).

Fine mapping of Lr47 **using 7A/7S homoeologous recombinants induced by the** ph1b **mutant.** No recombination was expected between wheat chromosome 7A and the introgressed 7S chromosome segment in the presence of the *Ph1* gene. Genotyping of 128 F_2 plants from the cross Kern *Lr47* × ZZ5389 using the flanking markers *pku0745* and *pku2216* (Supplementary Table 4) revealed no recombination between these two markers, confirming that recombination between *Ae. speltoides* and *T. aestivum* chromosomes is suppressed. To induce recombination between the 7A/7S homoeologous

chromosomes, the introgression line Kern *Lr47* was crossed with CS*ph1b*. Supplementary Fig. 3 describes the procedure for identifying recombinants between the 7S and 7A chromosomes. A total of 15 F_2 plants heterozygous for *Lr47* and homozygous for *phlb* were obtained and self-pollinated to generate F_3 seeds for the recombination screening. We genotyped 2,654 plants from eight selected segregating F_3 families with the flanking markers *pku0745* and *pku2216*, and identified 542 plants with recombination events.

The recombinants were genotyped using another 13 7A/7S-genome specific markers across the introgressed *Ae. speltoides* chromosome segment (Fig. 2a, b). Among the recombinants, we only selected plants heterozygous for a smaller *Ae. speltoides* chromosome segment for phenotyping. Finally, we identified *Ae. speltoides* segments of seven different lengths, which were designated as L1 to L7 (Fig. 2c). The selected recombinants were evaluated for resistance to race THDB at the three-leaf stage in growth chambers. Plants carrying introgressions L1 and L4 were susceptible, so *Lr47* was mapped to a 3.5-Mb region (CS RefSeq v1.1 coordinates) flanked by markers *pku1104* and *pku1152* (Fig. 2c). The candidate gene region includes 49 annotated high-confidence genes in Chinese Spring (*TraesCS7A01G110400*–*TraesCS7A01G115200*, Supplementary Table 6). The functional annotation of these genes revealed six typical NLR genes (Supplementary Table 6), a gene class frequently associated with disease resistance in plants [32–37].

To refine the mapping of *Lr47* in the Kern *Lr47* × CS*ph1b* population, we crossed the susceptible EMS mutant m118 (in the Kern *Lr47* background) with the non-mutagenized resistant Kern *Lr47*, and generated an F_2 population consisting of 1,141 individuals. We performed whole genome re-sequencing of the parental lines, identified EMS-induced SNPs, and generated sequence-based markers in the *Lr47* candidate region. We used these markers to genotype lines with recombination events in the target region and phenotyped these lines for resistance to *Pt.* Based on these results, we mapped *Lr47* between markers *pkus675* and *pkus175* (Fig. 2d), within the same interval as in the Kern *Lr47* × CS*ph1b* population. The reduced candidate region corresponds to a physical interval of 2.5 Mb in the reference genome of the *Ae. speltoides* accession 'TS01' (Fig. 2e). Within the candidate gene region in the TS01 genome, we found six typical NLR genes (Fig. 2e, Supplementary Table 7).

Identification of the Lr47 **candidate gene using EMS mutagenesis and transcript assembly (EMTA).** Among the progenies of 662 M_2 mutant families screened with *Pt* race THDB, we identified ten independent families with susceptible plants and confirmed their susceptibility to race THDB by evaluting M_3 seeds derived from the susceptible plants (Fig. 3a). Genotyping using six 7A/7S-genome specific markers confirmed that all the susceptible mutant lines carried the complete *Ae. speltopides* chromosome segment.

To identify the candidate gene of *Lr47*, we first generated RNA-seq reads (≥ 69.8 million 150-bp paired-end reads per sample) from *Pt*-inoculated leaves of Kern *Lr47* and ten independent susceptible M_3 mutants (Supplementary Table 8). We used inoculated leaves because we did not know if the *Lr47* gene is induced by the pathogen or not. Assembly of the Kern *Lr47* sequenced reads yielded 146,715 high confidence transcript contigs (≥ 500 bp). We performed BLASTN searches of the Kern *Lr47* transcriptome

database using the sequences of the candidate NLR genes in CS and TS01 as queries, and obtained 45 transcript contigs with a BLAST e-value = 0. We mapped the sequenced reads of ten susceptible mutants against the 45 selected Kern *Lr47* transcript contigs and found one contig named KN638873_g379_i12 with EMS-type (G/C-to-A/T) point mutations in all ten mutants (Fig. 3b and Supplementary Fig. 4a). Using the primer pairs *EMS8054F2R1*, *EMS8054F1R4*, and *EMS8054F7R7* (Supplementary Table 4) developed from this contig, we performed PCRs to amplify the region containing the mutations and confirmed the presence of nucleotide transitions in these susceptible mutants (Supplementary Fig. 4b). In addition, genotyping of the critical recombinants derived from the cross m118 × Kern *Lr47* with marker *EMS8054F1R4* confirmed that the C-to-T SNP in m118 co-segregated with the disease phenotype.

To determine the structure of the *Lr47* candidate gene in contig KN638873_g379_i12, we (i) assembled the whole genome re-sequencing reads to obtain a genomic sequence containing the *Lr47* candidate; (ii) confirmed the transcript and genomic sequences of the *Lr47* candidate by PCR and Sanger sequencing; and (iii) compared the transcript with the corresponding genomic sequence. These analyses revealed that the *Lr47* candidate gene has six exons encoding a typical NLR protein of 928 amino acids, which is hereafter referred to as CNL2 (Fig. 3c; GenBank accession number OQ919262). Using 5’ and 3’ RACE (Supplementary Fig. 5), we determined that the 5’-untranslated region (UTR) of *CNL2* is 1,142 bp long with three introns, and the 3’UTR is 607 bp long with only one intron (Fig. 3c). The predicted 2,787 bp coding sequence of *CNL2* is 95.6% and 91.9% identical to its closest homologs within the candidate regions in *Ae. speltoides* TS01 and CS, respectively. Based on the predicted coding sequence, we found that four of the mutations introduced premature stop codons and the others introduced nonsynonymous amino acid changes (Fig. 3c). AlphaFold prediction of the full length CNL2 protein yielded a structural model with the expected structures for the CC and NB domains as well as an LRR domain containing multiple repeat units forming a typical α/β horseshoe fold (Supplementary Fig. 6). Four amino acid substitutions were located in the CC domain, three in the NB domain, and four in the LRR domain. The amino acid substitutions are associated with large changes in the predicted CNL2 structure (Supplementary Fig. 7).

Validation of CNL2 **using BSMV-sgRNA-based gene editing.** Because the BSMV-sgRNA-based gene editing method requires a wheat line expressing the Cas9 gene, we first crossed the introgression line Kern *Lr47* with the cultivar Bobwhite carrying a highly expressed Cas9 allele [31]. The resulting F_1 plants, carrying both the *Lr47* and Cas9 genes, were inoculated with BSMV-sgRNA constructs targeting the candidate gene *CNL2*. Leaves of the inoculated F_1 plants showed chlorotic spots and white streaks, consistent with BSMV-induced viral symptoms (Supplementary Fig. 8). Next-generation sequencing of amplicons prepared with DNA isolated from the infected leaves of F_1 plants showed a somatic editing efficiency of ~ 16%.

Infected F_1 (M_0) plants were self-pollinated to produce M_1 seeds. Genotyping of 216 M_1 plants identified 24 plants (11.1%) with heterozygous or homozygous mutations. Among them, two plants (mut-1 and mut-2) that were homozygous for “T” or “TT” deletions at position 2,280 downstream of the ATG in the

cDNA (Fig. 4a) were selected for further characterization. These frameshift deletions alter 18% of the protein sequence, resulting in loss of function of the CNL2 protein. Evaluations of the progeny of the two selected plants with race THDB showed that the plants homozygous for the deletions displayed susceptible reactions similar to those of Cas9-transgenic Bobwhite, whereas Kern *Lr47* and its sister line without editing were resistant (Fig. 4b). These results indicate that *CNL2* is required for *Lr47*-mediated resistance to *Pt*.

Wheat plants transformed with CNL2 **were resistant to leaf rust.** To determine whether *CNL2* is sufficient to confer resistance to leaf rust, a 7,234 bp genomic DNA fragment from the introgression line Kern *Lr47*, including the complete coding region, introns, and regulatory sequences (Fig. 4c), was transformed into the spring wheat cultivar Fielder via *A. tumefaciens*-mediated transformation. A total of 80 independent transgenic T_0 plants were generated, and the presence of the transgene was confirmed with the primer pairs *EMS8054F7R7* and *Lr47speF5R5* (Supplementary Table 4), which amplify the transcribed region of *CNL2*. All transgenic plants containing the transgene and Kern *Lr47* (positive control) showed high levels of resistance to *Pt* race PHQS, whereas the untransformed control Fielder was completely susceptible (Supplementary Figs. 9a and 10). Some of the resistant transgenic plants showed lower or even higher levels of resistance than Kern *Lr47* (Supplementary Figs. 9a and 10), suggesting that these transgenic plants may have different numbers of *CNL2* insertions or expression levels.

To test this hypothesis, the qRT-PCR primer pair *Lr47qPCRF2R3* were used to assess the transcript levels of *CNL2* in 11 randomly selected transgenic T_0 plants (T_0CS401-1 to T_0CS401-11) using *ACTIN* as an endogenous control. *CNL2* transcript levels were significantly higher in all selected transgenic T_0 plants than in the Fielder control ($P < 0.01$), and eight of them showed higher *CNL2* transcript levels than Kern *Lr47* ($P < 0.05$; Supplementary Fig. 9b). Approximately 25 T_1 plants from each selected transgenic family were genotyped with the primer pair *Lr47speF5R5*, and all except three showed significant departures from the expected segregation ratio of 3:1 (transgenic/non-transgenic), with an excess of transgenic plants (Supplementary Table 9). The number of *CNL2* insertions was determined by TaqMan copy number assays (Supplementary Table 9). Overall, plants derived from three of the transgenic events (T_1CS401-2, T_1CS401-7, and T_1CS401-8) were estimated to have only a single copy of the transgene, whereas the other eight transgenic families were estimated to have between two and five *CNL2* copies (Supplementary Table 9). Estimated *CNL2* copy number based on TaqMan assays correlated significantly with the transcript levels in the selected transgenic plants ($R = 0.90$, $P < 0.001$).

Transgenic T_1 plants from the 11 selected transgenic events were challenged with races PHQS and FHJL, which are virulent on Fielder. All plants from the T_1 transgenic families T_1CS401-1 and T_1CS401-10, which were fixed for the transgene, exhibited a high level of resistance, and resistance in the transgenic families T_1CS401-2 and T_1CS401-7 perfectly co-segregated with the presence of the transgene (Fig. 4d and Supplementary Fig. 11). These results demonstrated that *CNL2* is sufficient to confer resistance to *Pt* races PHQS and FHJL. Inoculation of transgenic plants and the Fielder control with race THDB did not

provide useful information because Fielder was highly resistant (ITs = 1-) to this *Pt* race (Supplementary Fig. 12).

Taken together, the high-resolution genetic map, loss-of-function EMS and BSMV-Cas9-induced mutations, and the transgenic results demonstrated that the candidate *CNL2* is *Lr47*.

The Lr47 **translocation segment originated from** Ae. speltoides **var.** speltoides. A search of the released reference genomes identified the closest homologs of *Lr47* on chromosomes 7A, 7B, and 7D of diploid, tetraploid, and hexaploid wheat, as well as on chromosome 7 of *Sitopsis* species (including *Ae. speltoides*, *Ae. longissimi*, *Ae. sharonensis*, and *Ae. searsii*), which had 83.29–92.78% protein similarity to Lr47. The phylogenetic tree in Supplementary Fig. 13 shows that Lr47 and its homologs from *Sitopsis* species are in the same phylogenetic clade. Multiple alignment of the protein sequences revealed that most of the variation was present in the LRR domain (Supplementary Fig. 13).

A dominant marker, *Lr47mas*, was designed based on the DNA polymorphism C1059G (resulting in amino acid change D353E), which distinguishes *Lr47* from all other homologs found so far (Supplementary Fig. 14). The forward primer was designed in the first intron and the reverse primer contained the critical nucleotide 1059C at its terminal 3' end. PCR amplification with the marker *Lr47mas* (Supplementary Table 4) at an annealing temperature of 54 °C yielded an amplicon of 488 bp only when *Lr47* was present (Supplementary Fig. 15). Using this marker, we examined a large collection of wheat genotypes, including 144 accessions of *T. aestivum*, 78 of *T. turgidum*, 24 of *T. monococcum*, and 118 of *Ae. speltoides* (Supplementary Table 10). PCR amplicons of the predicted size were present in only three of the *Ae. speltoides* accessions but were absent in all diploid, tetraploid, and hexaploid wheat accessions (except for six *Lr47* introgression lines) tested in this study (Supplementary Table 10). To confirm the presence of *Lr47* in the three *Ae. speltoides* lines (T2140002, Y162, and Y397), we designed two pairs of gene-specific primers, *Lr47SHF3/R7* and *Lr47SHF2/R1* (Supplementary Table 4), that amplify the complete coding region of *Lr47*. Sanger sequencing confirmed that these three *Ae. speltoides* accessions carry a functional *Lr47* allele identical to that of Kern *Lr47*.

Ae. speltoides includes two subspecies: *Ae. speltoides* var. *speltoides* and *Ae. speltoides* var. *ligustica*. Since the taxonomic subspecies and origin of the *Lr47* introgressed segment are not clear [21,22], we compared the SNPs identified from Kern *Lr47* RNA-seq data with those of eight *Ae. speltoides* var. *speltoides* accessions, six *Ae. speltoides* var. *ligustica* lines, and two *Ae. speltoides* lines for which the subspecies is unknown but carry *Lr47* (T2140002 and Y162). We focused only on the polymorphisms that were within the ~ 150 Mb introgressed segment 7S#1S and were polymorphic among the 16 *Ae. speltoides* accessions described above, but were absent in the recurrent parent Kern. Based on this approach, we obtained 199 polymorphic sites (Supplementary Table 11). A Neighbor-Joining tree based on these polymorphisms showed that Kern *Lr47* is in a branch encompassing multiple *Ae. speltoides* var. *speltoides* accessions (Supplementary Fig. 16), suggesting that the translocated segment originated from this subspecies. The two accessions that most closely resemble to Kern *Lr47* are T2140002 and Y162, which carry *Lr47*. The geographical origins of T2140002 and Y397 (carrying *Lr47*) are unknown, but the

most closely related accession Y162 was collected in Iraq, suggesting that Iraq and its neighboring countries are a likely region where *Lr47* originated.

Reducing the length of the introgressed Ae. speltoides **chromosome segment carrying** Lr47. Because the large introgressed *Ae. speltoides* segment 7S#1S was associated with several undesirable traits, we reduced the size of the introgressed segment carrying *Lr47* using the *ph1b* mutation. Supplementary Fig. 17 describes the crosses that led to the shortening of the alien chromosome segment 7S#1S. We started from recombinant L2 (Fig. 2c), which has a smaller introgression segment of approximately 40 Mb with a recombination event between 76.1 and 80.1 Mb (Fig. 2c and Supplementary Fig. 18).

Subsequently, the L2 plant was self-pollinated and its progeny was used for a second round of recombination with the *ph1b* mutant. Using the markers *pku0745* and *pku1176*, we identified 20 recombinants from 313 segregating individuals. These recombinants were genotyped with the 7A/7S genome-specific markers from this region (Supplementary Table 4). Among them, we identified one critical recombinant, which we hereafter refer to as F284-128 (L8). The crossover breakpoint in L8 was between markers *pku0957* (58.8 Mb) and *pku1026* (63.0 Mb), so the 7S chromosome segment in this resistant line is reduced to between 13.1 Mb and 21.3 Mb (Supplementary Fig. 18d).

Recombinant L8 with the truncated 7S segment was crossed and backcrossed once with the Chinese common wheat variety Yangmai21 (YM21) (Supplementary Fig. 17), which is susceptible to multiple *Pt* races, including THDB, PHQS, FHJL, HCJR, PHRT, and PHTT. Four PCR markers *pku1026*, *Lr47speF5R5*, *Lr47mas*, and *pku1176* (Supplementary Table 4), were used to confirm the presence of the truncated 7S segment in the selected BC_1F_2 plants (hereafter referred to as YM21 + *Lr47*-2). Homozygous BC_1F_2 plants from the introgression line YM21 + *Lr47*-2 challenged with six different Chinese *Pt* races showed high levels of resistance (ITs = 0 to 1), whereas the recurrent parent YM21 and the sister line lacking the smaller alien chromosome segment exhibited susceptible infection types (ITs = 3 to 4) in response to the same races (Fig. 5a).

We then performed GISH experiments to validate the results of the molecular marker analysis described above. GISH analyses confirmed the presence of the interstitial *Ae. speltoides* segment in Kern *Lr47* on the recombinant chromosome arm 7A/7S (Fig. 5b), whereas recombinants L2 and L8 had significantly smaller alien chromatin segments (Fig. 5c, d).

FISH-based karyotype analysis revealed strong Oligo-pSc119.2 and Oligo-pTa713 signals on the introgressed *Ae. speltoides* chromosome segment 7S#1S in Kern *Lr47* (Fig. 5e), whereas only the pSc119.2 signal was detected in L2 (Fig. 5f), suggesting that the proximal region of the introgressed segment 7S#1S was eliminated by the recombination event in L2 (Fig. 2c). In L8, neither Oligo-pSc119.2 nor Oligo-pTa713 signal was detected (Fig. 5g), suggesting that the distal region of the alien chromatin was also replaced by wheat chromatin after the second round of homoeologous recombination. These results are consistent with the analysis with PCR markers (Supplementary Fig. 18) and with the GISH results (Fig. 5b-d).

Agronomic and quality evaluation of the wheat line carrying the shortened alien chromosome segment. To evaluate the effects of the truncated alien chromosome segment containing *Lr47* (35.9–40.0 Mb, Supplementary Fig. 18c) on agronomic and quality traits, recombinant L2 was crossed and backcrossed three times with the hexaploid wheat variety YM21 and self-pollinated to generate the BC_3F_3 seeds (Supplementary Fig. 17). This line was selected because it was available earlier than L8. We planted BC_3F_3 sister lines homozygous for the presence (YM21 + *Lr47*-1) or absence of the alien chromosome segment in both the greenhouse and growth chamber (Supplementary Fig. 19a, b) and measured the phenotypic changes. Under disease-free conditions, no significant differences were observed between the new introgression line YM21 + *Lr47*-1 and its sister line for 15 morphological and quality traits examined (Fig. 6). The only exception was the spike length (SL), which was significantly shorter in YM21 + *Lr47*-1 than in its sister line ($P < 0.05$; Fig. 6c). The results were consistent between the growth chamber and the greenhouse experiments (Supplementary Fig. 19).

In a field experiment in Shandong Province, China, the introgression line YM21 + *Lr47*-1 was highly resistant to *Pt*, whereas the sister control line was fully susceptible (Fig. 6q, r). Analysis of morphological traits showed that SL was also significantly shorter in YM21 + *Lr47*-1 than in the isogenic sister line without the *Lr47* introgression (Fig. 6s, t). There was no significant difference in SNS, PH, or TN between YM21 + *Lr47*-1 and the sister control line (Fig. 6u and Supplementary Fig. 20).

Characterization of Lr47, **which encodes a typical NLR immune receptor protein.** Transcript levels of *Lr47* relative to *ACTIN* were measured in Kern *Lr47* by qRT-PCR. We observed no significant transcriptional differences between plants inoculated with *Pt* race PHQS and mock inoculated with water at 1, 2, 4, and 6 dpi (Fig. 7a), indicating that *Lr47* is not induced by the presence of the *Pt* pathogen.

The predicted Lr47 protein includes an N-terminal CC domain, a central NB site, and a C-terminal LRR region, with two predicted monopartite nuclear localization signals determined by the cNLS Mapper program (amino acids 22–54 and 533–539; Fig. 7b). To determine the subcellular localization of Lr47, a GFP tag was used to visualize Lr47 in tobacco (*N. benthamiana*) leaves and wheat protoplasts. Both cytoplasmic and nuclear fluorescence were observed in tobacco leaves for all constructs, namely GFP-Lr47_CDS, Lr47_CDS-GFP, GFP-Lr47_CC, and Lr47_CC-GFP constructs (Fig. 7c). Anti-GFP immunoblots confirmed that the proteins were expressed at the expected sizes for all constructs (Fig. 7d). In wheat protoplasts, cytoplasmic and nuclear signals were also detected in GFP-Lr47_CDS, Lr47_CC-GFP, and GFP-Lr47_CC constructs (Supplementary Fig. 21).

To test whether the full length Lr47 and/or the CC domain alone are capable of triggering cell death in *N. benthamiana*, we investigated signaling after *Agrobacterium*-mediated transient expression in leaves of *N. benthamiana*. We did not observe cell death or obvious yellowing in leaf regions transiently overexpressing Lr47 and its protein domains individually. In contrast, robust cell death was observed in leaf regions expressing BAX and the CC domain of Sr13 (Fig. 7e and Supplementary Fig. 22). Moreover, the self-interaction ability of Lr47 was initially assessed using yeast two-hybrid assays. We did not detect

any direct interaction between the Lr47 protein itself and its protein domains CC, NB, and LRR *in vitro* (Supplementary Fig. 23).

Discussion

Wild relatives of wheat are valuable sources of disease resistance genes, and they have been used previously for introgressing *Lr* genes into common wheat varieties [6]. However, cloning resistance genes in wheat relatives is often hampered by the suppressed recombination in the regions carrying alien introgressions. Several newly developed technologies, such as MutRenSeq and MutChromSeq, have facilitated the isolation of resistance (*R*) genes in wheat and its wild relatives [38]. Here, we report the cloning of the broadly effective leaf rust resistance gene *Lr47* using a combination of *ph1b*-induced recombination and the EMTA method. Using this approach, the *CNL2* candidate gene for *Lr47* was successfully identified. The susceptibility to *Pt* races of the 12 independent EMS or BSMV-based editing mutants and the resistance of the *CNL2* transgenic Fielder lines (Figs. 3 and 4) confirmed that this candidate gene is both necessary and sufficient to confer resistance to leaf rust, and therefore that *CNL2* is *Lr47*.

The EMTA method developed in this study (Fig. 3b) does not require the generation of a physical map/reference sequence across the map interval. This feature is useful and may spark the interest of researchers in identifying target genes in recombination-sparse regions, especially in plant species with large genomes. However, the effectiveness of the EMTA method relies on several factors: (1) the delimitation of the target gene within a physical interval; (2) the isolation of multiple independent mutants carrying loss-of-function mutations; (3) the presence of orthologs or homologs or paralogs in the colinear regions in sequenced reference genomes; and (4) the comparison of the selected mutants with wild type using RNA-seq data. It is worth noting that RNA-seq might be biased by tissue or sampling time, and therefore other sequencing strategies, such as whole genome re-sequencing or exome capture sequencing, could be good options. In this study, the wild-type Kern *Lr47* was sequenced at very high sequencing depth (> 250 million reads, Supplementary Table 8) on the Illumina NovaSeq 6000 platform to ensure a high-quality transcriptome assembly. *Lr47* was found within a complex locus comprising a cluster of NLR genes that exhibit both copy number and structural variations (Supplementary Tables 6 and 7). Since NLR genes are the most common class of genes associated with disease resistance in wheat and other plants [32,34,35,39–41], we hypothesized that one of the NLR genes might be *Lr47* and therefore prioritized NLR genes within the candidate region (Fig. 2d, e) for further analysis. This significantly reduced the amount of sequencing data to be analyzed.

Recently, virus-based sgRNA delivery systems have been developed and tested in wheat and other plant species [31,42]. Compared with the biolistic-based or *Agrobacterium*-mediated gene editing systems, the virus-based delivery systems do not require conventional plant genetic transformation and regeneration procedures [31]. In our study, the BSMV-based sgRNA delivery system was able to generate both somatic and heritable mutations in wheat via infection of previously obtained Cas9-transgenic wheat plants [31,43].

We demonstrated that the Cas9-transgenic wheat plants can be crossed with a line carrying the targeted alien chromosome (Kern *Lr47*) to generate edits in the F_1 plants (Fig. 3b). The results suggest that the BSMV-based sgRNA delivery system could reduce genotype dependency for targeted gene editing by transferring the high-expressing Cas9 locus into different wheat cultivars. Considering the simplicity of the BSMV-sgRNA infection procedure, this genome editing tool has great potential for applications in wheat.

Sequence alignments of the Lr47 protein and representative homologs revealed critical amino acid polymorphisms, which we used to develop a diagnostic marker for the presence of Lr47 (Supplementary Figs. 14 and 15). We validated this diagnostic marker in a larger number of accessions of *T. monococcum*, *T. turgidum*, *T. aestivum*, and *Ae. speltoides* (Supplementary Table 10). We detected *Lr47* in only 2.5% of the *Ae. speltoides* genotypes and in none of the diploid, tetraploid, and hexaploid wheat accessions tested, except for six *Lr47* NILs. This result indicates that the incorporation of *Lr47* has the potential to improve leaf rust resistance in a wide range of modern wheat varieties and highlights the importance of mining new *R* genes from wild progenitors.

One limitation for the incorporation of alien segments by homologous recombination is that the introgressed chromosome segments are usually large and can carry linked genes with negative agronomic or quality effects. Examples of linkage drag of alien introgressions carrying resistance genes include *Lr19* from *Ag. elongatum*, *Pch1* from *Ae. ventricosa*, and *Pm16* from *Ae. speltoides*[44,45]. Similarly, the introgression of the *Ae. speltoides* segment 7S#1S carrying *Lr47* was reported to be associated with several negative effects on agronomic and quality traits [24]. To enhance the utility of *Lr47* in wheat breeding, we attempted to eliminate the deleterious linkage drag by two rounds of *ph1b*-induced homoeologous recombination. Using a combination of marker analysis, cytology, and *Lr* phenotypic screening (Fig. 5, Supplementary Fig. 18), two recombinants with much smaller *Ae. speltoides* chromosome segments (L2: ~36 Mb and L8: ~13 Mb) were identified.

A preliminary characterization of the introgression line YM21 + *Lr47*-1 in greenhouse and chamber experiments showed no significant adverse effects on morphological and quality traits in the absence of disease. The only exception was SL, which was significantly shorter in YM21 + *Lr47*-1 than in its isogenic sister line (Fig. 6c). Shorter spikes were also observed in the field experiment (Fig. 6s, t). Despite reducing SL, the presence of the *Lr47* introgression was not associated with reduction in spike grain yield (Fig. 6 and Supplementary Fig. 19). We are currently introgressing the shorter ~ 13 Mb *Ae. speltoides* chromosome segment into the Chinese common wheat cultivar YM21 to determine whether the shorter spikes observed in YM21 + *Lr47*-1 is due to linkage with other alien genes or pleiotropic effects of the *Lr47* gene. The latter is unlikely because no significant differences in SL were observed between transgenic plants homozygous for the transgene (1 copy) and the untransformed control Fielder under disease-free conditions (Supplementary Fig. 24). If necessary, the flanking markers and the diagnostic marker for *Lr47* (Supplementary Table 4) can be used to develop wheat lines with even smaller introgressed segments carrying the *Lr47* gene.

For NLR genes in plants, the CC domain is thought to play a crucial role in downstream signaling. Previous studies reported that the CC domains of numerous NLR proteins (such as MLA10, Pm21, Sr33, and Sr50) are sufficient to cause cell death after transient expression in *N. benthamiana* leaves [46]. However, other NLR proteins, including RPM1, RPS5, Rx, and Sr35, have CC domains that do not trigger cell death in leaves of *N. benthamiana*[47]. In this study, we found that the CC domain of the Lr47 protein cannot directly induce cell death (Fig. 7e and Supplementary Fig. 22), nor does it have self-protein interaction (Supplementary Fig. 23), indicating that an unknown activation mechanism is essential for Lr47 function. Although NLR proteins have been shown to mediate hypersensitive response by forming pentameric disease resistosomes in *Arabidopsis* (ZAR1) [48] and wheat (Sr35) [49], the activation and downstream pathways of resistance for Lr47 require further investigation.

Lr47 provides strong resistance to virulent *Pt* pathotypes in a wide range of regions, including the United States [23], Mexico [1], Russia [50], South Africa [50], the Indian sub-continent [51], and China (the current study), indicating broad-spectrum resistance to leaf rust. However, elucidation of the typical NLR structure of *Lr47* indicates that its resistance may not be durable in the field because of rapid changes in leaf rust pathogen populations. The rust pathogen may evolve through the mutation or deletion of the corresponding *Avr* genes, leading to the breakdown of resistance conferred by NLR genes [52,53]. For instance, the Ug99 race group of the wheat stem rust pathogen (*Puccinia graminis* f. sp. *tritici*) has successfully broken the resistance of wheat varieties carrying the resistance genes *Sr24*, *Sr31*, *Sr36*, *Sr38*, and *SrTmp*[39,54–56]. Therefore, a combination of *Lr47* with other slow-rusting multi-pathogen resistance genes, such as *Lr34*/*Sr57*[10] and *Lr67*/*Sr55*[11], is a preferred strategy for developing wheat cultivars with durable resistance against this devastating rust pathogen.

In conclusion, the identification of *Lr47*, the new introgression lines with eliminated or reduced deleterious linkage drag, and the available diagnostic marker developed in this study provide useful tools to diversify the deployed *Lr* genes and accelerate the use of *Lr47* in modern wheat breeding programs.

Methods

Plant materials and mapping populations. The six pairs of hard red spring wheat near-isogenic lines (NILs) with and without the leaf rust resistance gene *Lr47*[18,22] used in this study are listed in Supplementary Table 1. A total of 7,590 recombinant gametes from two mapping populations were used to construct a high-density genetic map of *Lr47*. The first one included 2,654 plants from selected F_3 families homozygous for the *ph1b* mutation and segregating for the *Ae. speltoides* chromosome segment (population Kern *Lr47* × CS*ph1b*). The second population, included 1,141 F_2 individuals, was generated from a cross between a Kern *Lr47* EMS mutant susceptible to *Pt* designated m118 and the wild type (Kern *Lr47*). Since both parents (m118 and Kern *Lr47*) carry the same *Ae. speltoides* chromosome, recombination is normal in the presence of the *Ph1b* wild-type allele. Moreover, Kern *Lr47* was crossed with the susceptible wheat line ZhengZhou 5389 (ZZ5389) to characterize leaf rust resistance in the presence of the *Ph1* gene. A collection of 118 accessions of *Ae. speltoides* (Supplementary Table 2), 24

of *Triticum monococcum*, 78 of *T. turgidum*, and 144 of *T. aestivum* (including six *T. aestivum* genetic stocks carrying the *Lr47 Ae. speltoides* segment 7S#1S) was used to determine the value of the diagnostic marker developed in this study for marker assisted selection.

Leaf rust assays and pathogen growth in infected leaves. Leaf rust seedling evaluations were performed at the Peking University Institute of Advanced Agricultural Sciences, Weifang, China and Hebei Agricultural University, Baoding, China. A total of 24 *Pt* isolates and their avirulence/virulence profiles are presented in Supplementary Table 3. Three-leaf stage seedlings were inoculated with fresh *Pt* urediniospores mixed with talcum powder at a ratio of 1:20 using the shaking off method [57]. The inoculated plants were placed in a dark dew chamber set at 22 ℃ for approximately 24 h and then maintained at 22–24 ℃ with a 16-h photoperiod. Infection types (ITs) of plants were scored at ~ 12 days post inoculation (dpi) using a 0–4 scale as described previously [58].

To estimate the growth of *Pt* pathogen in NILs, leaf segments (~ 6 cm long) from Kern *Lr47* and its recurrent parent Kern inoculated with race THDB were sampled at 2, 4, 6, and 8 dpi. The collected leaves were autoclaved in 1 M KOH, stained with wheat germ agglutinin labeled with fluorescein isothiocyanate (WGA-FITC; Cat No. L4895-10MG, Sigma-Aldrich, USA) and visualized by a Zeiss Discovery V20 fluorescence dissecting microscope (Zeiss, Jena, Germany) according to previously described methods [32,59].

EMS mutant screening. Seeds of wheat line Kern *Lr47* were treated with 250 mL of 0.8% EMS (Cat No. M0880-25G, Sigma-Aldrich, USA) and incubated for 18 h at 25 ℃ on a shaker at 150 rpm. Next, the treated seeds were washed five times with 300 mL tap water and then placed under running water for 3 h. All surviving M_1 plants were grown in a greenhouse, and a total of 4,756 independent M_2 families were obtained. Approximately 25 M_2 seeds per family were challenged with *Pt* race THDB in growth chambers at Peking University Institute of Advanced Agricultural Sciences, Weifang, China. The M_3 seeds derived from susceptible M_2 plants were re-evaluated with races PHQS and THDB. Finally, all the identified mutants were genotyped with 7S-specific primers to rule out the possibility of seed contamination.

Sequences, markers, and bioinformatics analysis. RNA sequencing (RNA-seq) of *Lr47* NILs, *Ae. speltoides* accessions (AE915, AE1590, and PI 554292), and the susceptible EMS mutant lines (accession number CRA011051) were performed at Beijing Novogene Bioinformatics Technology Co., Ltd. (Beijing, China). The released reference genomes of hexaploid wheat Chinese Spring (CS; RefSeq v1.1) and other 10 wheat genotypes from the Wheat Pan Genome Project were used in our analyses [60,61]. RNA-seq data of Yecora Rojo *Lr47*, UC1037 *Lr47*, and White Yecora *Lr47* were from a previous study [26]. Raw RNA-seq reads were trimmed using Trimmomatic software version 0.32 to remove low-quality reads and adaptors [62]. Trimmed reads were aligned to the CS reference sequence using STAR software v2.7.10a [63]. Variant calls were made using Freebayes v1.3.6 [64] and filtered using BCFtools v1.14 (https://github.com/samtools/bcftools).

EMS mutagenesis and transcript assembly (EMTA) was performed as follows: RNA-seq reads from Kern *Lr47* were assembled *de novo* using Spades version 3.14.1 [65]. We performed BLASTN searches of the transcriptome database of Kern *Lr47* using sequences of typical NLR genes within the *Lr47* candidate regions from the reference genomes of CS and TS01 [66] as queries, and obtained transcript contigs with similarity values greater than 85% to the homologs or paralogs within the interval that *Lr47* was mapped to. Next, the RNA-seq reads of the ten susceptible EMS mutants were mapped to these transcript contigs as reported previously [67].

To develop markers in the m118 × Kern *Lr47* population, we performed whole genome re-sequencing for both parents (accession number CRA011051). We aligned the m118 and Kern *Lr47* sequences of the genes in the candidate region and identified the polymorphic sites. Genome-specific primers were designed using the Primer 3 online software (v0.4.0, https://bioinfo.ut.ee/primer3-0.4.0/primer3/) and used to amplify regions carrying putative *Ae. speltoides*-specific single nucleotide polymorphisms (SNPs). The procedures for developing Insertion-Deletion (InDel) and cleaved amplified polymorphic sequence (CAPS) markers were reported previously [68,69]. The other published reference genome of *Ae. speltoides* AEG-9674-1 [70] was used to support marker development.

Rapid amplification of cDNA ends (RACE). 5'- and 3'-RACE reactions were carried out using total RNA extracted from leaves of the introgression line Kern *Lr47*. RACE reactions were performed using the Invitrogen FirstChoice RLM-RACE Kit (Cat. no. AM1700, ThermoFisher Scientific, MA) according to the manufacturer's instructions. PCR products from RACE reactions were cloned into the pMD18-T vector (TaKaRa Biotechnology, Kyoto, Japan) using the TA cloning method. The selected positive clones were sequenced using the Sanger method.

BSMV-sgRNA-based gene editing. The BSMV-based guide RNA delivery system [31] was used to validate the *Lr47* candidate gene. The methods and procedures for plasmid construction and virus infection were reported previously [31]. The wheat transgenic line exogenously expressing Cas9 gene in genetic background of Bobwhite (Cas9-transgenic Bobwhite) was crossed with the introgression line Kern *Lr47*, and the resulting F_1 plants were infected with BSMV vectors targeting the *Lr47* candidate gene. The infected M_0 plants were self-pollinated to produce M_1 seeds. We identified mutations in the M_1 plants using the primer pair *Lr47BSMVF2R1* (Supplementary Table 4). Finally, the edited plants and their progenies were challenged with *Pt* race THDB.

Wheat transformation and copy number assays. A 7,234-bp genomic DNA fragment consisting of the entire coding region and introns (3,132 bp) of *Lr47*, 2,097 bp upstream of the start codon, and 2,005 bp downstream from the stop codon was amplified from the introgression line Kern *Lr47* by PCR using the PrimeStar Max DNA Polymerase (TaKaRa, Kyoto, Japan). The fragment was inserted into the linearized binary vector pCAMBIA1300 using an In-Fusion® HD Cloning Kit (Clontech, CA, USA). The resulting plasmid pCAMBIA1300-*Lr47* was introduced into the susceptible hexaploid wheat line Fielder via *Agrobacterium tumefaciens* (strain EHA105)-mediated transformation. Primer pairs *EMS8054F7R7* and

Lr47speF5R5 (Supplementary Table 4) were used to confirm the presence of the transgene, and the qRT-PCR primer pair *Lr47qPCRF2R3* (Supplementary Table 4) was used to estimate the transcript levels of the transgene in selected transgenic T_0 plants. The T_0 and T_1 transgenic plants were challenged with three Chinese *Pt* races PHQS, FHJL, and THDB. The number of integrated *Lr47* transgenes in each transgenic line was estimated based on the segregation ratio of T_1 transgenic plants and the results of a TaqMan copy number assay [32,71].

qRT-PCR analysis. At the two-leaf stage, plants from the introgression line Kern *Lr47* were mock or *Pt* inoculated in two independent growth chambers under the same temperature and photoperiod (24°C day/22°C night and 16 h light/8 h darkness). Leaf samples from different plants were collected immediately before inoculation (0 h) and at 1, 2, 4, and 6 dpi. Total RNAs were isolated using the Spectrum Plant Total RNA Kit (MilliporeSigma, MA, USA) and purified using the Direct-zol RNA MiniPrep Plus Kit (Zymo Research, CA, USA). qRT-PCR reactions were carried out on an ABI QuantStudio 5 Real-Time PCR System (Applied Biosystems, CA, USA) using Fast SYBR GREEN Master Mix. qRT-PCR primer pair *Lr47qPCRF2R3* (Supplementary Table 4) was used to evaluate the transcript levels of *Lr47*. Transcript levels were determined in seven biological replicates and presented as fold-*ACTIN* levels as described previously [72].

Phylogenetic analysis. Sequences of Lr47 protein homologs were obtained from the National Center for Biotechnology Information (NCBI) database (https://www.ncbi.nlm.nih.gov) and the published reference genomes of wheat and its wild relatives (https://wheat.pw.usda.gov/blast/). To explore the origin of the introgressed segment containing *Lr47*, we performed RNA-seq and variant calling for seven *Ae. speltoides* genotypes, including T2140002, Y162, PI 542258, PI 542251, PI 560751, PI 542260, and PI 542270 (accession number CRA011051). Moreover, SNPs of four *Ae. speltoides* Tausch var. *speltoides* accessions (KU-2208A, KU-14601, KU-14605, and KU-12963a) and three *Ae. speltoides* Tausch var. *ligustica* accessions (KU-2236, KU-7716, and KU-7848) were obtained from RNA-seq data [17]. Mulitple sequence alignment was conducted using MUSCLE method implemented in MEGA software v7.0. The phylogenetic tree was generated using the pairwise deletion method (bootstrap values based on 1000 iterations) and visualized using the Interactive Tree Of Life (iTOL) v5.0 (https://itol.embl.de/).

Reduction of Ae. speltoides **chromosome segment.** The introgression line Kern *Lr47* was crossed with the CS *ph1b* mutant (CS*ph1b*) [73] to induce homoeologous recombination and reduce the size of the *Lr47* introgressed chromosome segment 7S#1S. Markers *Xwgc2111* and *Xwgc2049* were used to validate the presence of the *ph1b* mutation [74]. The resulting F_1 plants from the cross Kern *Lr47* × CS*ph1b* were self-pollinated. The derived F_2 plants which were heterozygous for the introgressed segment 7S#1S and homozygous for *ph1b* were self-crossed to generate F_3 families. Using 7A/7S genome-specific markers, we identified recombination events within the introgressed alien chromosome segment and tested these plants for leaf rust resistance. The obtained recombinant plants heterozygous for *Lr47* and homozygous for *ph1b* (residual heterozygous lines) were self-crossed to generate F_4 progeny for another round of screening of new recombinants. The identified critical recombinants carrying *Lr47* were crossed and

backcrossed to the Chinese hexaploid wheat variety Yangmai21 (YM21). Finally, plants homozygous for the shortened alien segments were selected using molecular markers and evaluated for resistance to multiple *Pt* races.

Cytogenetic assays. Genomic *in situ* hybridization (GISH) and fluorescence *in situ* hybridization (FISH) were conducted using the methods described previously [75,76]. The FISH probes oligo-pSc119.2, oligo-pTa535, and oligo-pTa713 were suitable for the identification of chromosomes of common wheat and *Aegilops* species [77]. Oligo-pSc119.2 preferentially paints tandem repeats on B- (or S-) genome chromosomes, and oligo-pTa535 hybridized well to wheat D-genome chromosomes [78]. A combination of oligo-pSc119.2 and oligo-pTa535 was used to distinguish the 42 wheat chromosomes [79]. The specific hybridization pattern of probe oligo-pTa713 makes it easy to differentiate chromosomes [75]. The probes were labeled with either FAM or TAMRA and were synthesized by TsingKe Biological Technology Co., Ltd. (Chengdu, China).

Evaluation of agronomic and quality traits. Homozygous BC_3F_3 plants from the introgression line YM21-*Lr47*-1 and its sister line lacking the reduced alien chromosome introgression were grown in a greenhouse (22–30 ℃ day and 16–24 ℃ night with a 16 h photoperiod) and in a controlled walk-in growth chamber (24 ℃ day/22 ℃ night with a 16 h photoperiod). For the greenhouse and growth chamber experiments, a single plant was grown per pot (~ 3.8 L) and agronomic traits were measured for each plant. The number of productive tillers (TN) was counted at the time of physiological maturity. Plant height (PH) was calculated by measuring the height of the main tiller of each plant from the ground level to the tip of spike excluding awns. Spike length (SL), spikelet number per spike (SNS), and grain number per spike (GNS) were measured as the mean of the main spikes (2–3 main spikes per plant). Grain yield per plant (GYP), thousand-seed weight (TSW), grain length (GL), and grain width (GW) were automatically calculated in the laboratory using a crop scanning test system (Wanshen SC-G, Hangzhou, China) [80]. At the maturity stage, the penultimate internodes were cut into segments of equal length (~ 10 cm), dried at 32 ℃ for 10 days, and then used for measurement. The shearing force (SF) of stems was measured using a universal testing machine (Instron 5848 Microtester, Instron, USA) as described previously [81]. Grain moisture content (GMC), grain protein content (GPC), flour water absorption (FWA), and flour ash (FA) were estimated using near-infrared reflectance spectroscopy (Model SpectraStar 2600 XT-R, Unity Scientific, USA). Grain hardness (GH) was determined in a TA-XT-plus Texture Analyzer (Stable Micro Systems, United Kingdom). Flour yield (FY) was expressed as grams of flour per 100 g of grain (conditioned to 15% moisture content, AACC 26 − 10). The field experiment was conducted during the 2022–2023 growing season at Peking University Institute of Advanced Agricultural Sciences, Weifang, China (36°26'04.0"N, 119°26'42.6"E). The introgression line YM21-*Lr47*-1 and its sister line were planted in 10 rows, each one meter long. The significance of differences in agronomic and quality traits between YM21-*Lr47*-1 and its sister line was estimated using Student's *t*-test.

Yeast two-hybrid assay. The coding regions of the full-length Lr47 protein (CDS, amino acids 1–928), coiled-coil (CC) domain (amino acids 1–172), nucleotide-binding site (NB) domain (amino acids 173–

608), and leucine-rich repeat (LRR) domain (amino acids 609–928) were cloned into the yeast two-hybrid system vectors pGBKT7 (bait) and pGADT7 (prey). The recombinant vectors were co-transformed into yeast strain AH109 using the lithium acetate and polyethylene glycol 3350 method. Co-transformed yeast was spotted onto synthetic dropout medium lacking leucine and tryptophan (SD-Leu-Trp) for selection of transformed colonies and then leucine, tryptophan, histidine, and adenine (SD-Leu-Trp-His-Ade) to detect interactions.

In planta **expression of GFP-fused Lr47 protein.** The coding regions of the full-length Lr47 protein (amino acids 1–928) and CC domain (amino acids 1–172aa) were cloned into vectors pJIM19-GFP and pJIT163-Ubi-GFP [82,83]. The recombinant constructs were expressed in *Nicotiana benthamiana* leaves using transformed *A. tumefaciens* strain GV3101. Wheat protoplasts were isolated from the susceptible common wheat variety Fielder and were transformed as previously described [84]. Fluorescence was checked using a confocal microscope (A1 HD25 Nikon, Tokyo, Japan). Total proteins were extracted from *A. tumefaciens*-transformed tobacco leaves. Protein samples were separated by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and then transferred onto polyvinylidene difluoride (PVDF) membranes. Immunoblots were performed by using anti-GFP antibody (Solarbio, Beijing, China).

For the cell death induction assay, the coding regions of the CC domain of a previously identified stem rust resistance protein Sr13 [32] and a mammalian cell death inducer BAX were cloned into pJIM19-GFP as positive controls. Highly concentrated (OD = 1.0) *A. tumefaciens* carrying Lr47_CDS-GFP, GFP-Lr47_CDS, Lr47_CC-GFP, GFP-Lr47_CC, Lr47_NB-GFP, GFP-Lr47_NB, Lr47_LRR-GFP, GFP-Lr47_LRR, Sr13_CC-GFP, BAX, and GFP were infiltrated into *N. benthamiana* leaves. Necrosis induced by BAX was observed at 48 h post inoculation (hpi), while yellowing caused by Sr13_CC-GFP was detected at 120 hpi.

Declarations

Acknowledgements

The authors thank Dr. Jan Dvorak (UC Davis) for the initial *Aegilops speltoides* recombinant line; Dr. Bao Liu and Zhibin Zhang (Northeast Normal University, China) for genomic sequence of *Ae. speltoides* TS01 before release; and Dr. Xiuying Kong (Chinese Academy of Agricultural Sciences, China) and Dr. Lianquan Zhang (Sichuan Agricultural University, China) for *Ae. speltoides* accessions. Work at SC laboratory was supported by the Young Taishan Scholars Program of Shandong Province, the National Key Research and Development Program of China (2022YFD1201300). Work at XW laboratory was supported by Provincial Natural Science Foundation of Hebei (C2022204010 and C2021204008). Work at JD laboratory was supported by the Howard Hughes Medical Institute and by the Agriculture and Food Research Initiative Competitive Grant 2022-68013-36439 (WheatCAP) from the USDA National Institute of Food and Agriculture (NIFA).

Author contributions

H.L., L.H., and S.Z. performed most of the experimental work; M.H. and H.C. conducted cytogenetic assays; R.S., Y.L., and X.H. contributed subcellular localization; S.P., T.S., and G.W. contributed primers development; W.Z. validated the diagnostic marker in *Ae. speltoides*; J.Y.G. contributed EMS population; H.M. and C.L. evaluated quality traits; B.S. contributed AlphaFold prediction; Z.L. contributed phenotyping; X.W.D. provided scientific support; S.C., X.W., and H.L. analyzed the data and wrote the first version of the manuscript. S.C., X.W., and J.D. conceived and supervised the project. All authors revised the manuscript and provided suggestions.

Competing interests

S.C., H.L., L.H., X.W., R.S., and Y.L. are inventors on a Chinese provisional patent application (China patent filing No.202211247709.2) relating to the use of the *Lr47* gene in wheat breeding programs. The remaining authors declare no competing interests. The authors declare that they have no conflict of interests.

Data availability

Data supporting the findings of this work are available within the paper and its supplementary information files. All the raw sequencing data and genome assembly generated for this project are archived at the National Genomics Data Center under BioProject accession number CRA011051 that are publicly accessible at https://ngdc.cncb.ac.cn/gsa. The sequence of *Lr47* is available in the GenBank database as accession number OQ919262.

References

1. Huerta-Espino, J. *et al.* Global status of wheat leaf rust caused by *Puccinia triticina*. Euphytica 179, 143–160 (2011).
2. Prasad, P., Savadi, S., Bhardwaj, S. & Gupta, P. The progress of leaf rust research in wheat. Fungal Biol 124, 537–550 (2020).
3. Knott, D. R. The wheat rusts-breeding for resistance. 7 (In: Monograph on Theoretical and Applied Genetics. Chapter 12. Springer-Verlag Berlin Heidelberg. p. 7, 1989).
4. Savary, S. *et al.* The global burden of pathogens and pests on major food crops. Nat Ecol Evol 3, 430–439 (2019).
5. Kolmer, J. Leaf rust of wheat: pathogen biology, variation and host resistance. Forests 4, 70–84 (2013).
6. Ren, X. *et al.* Genetics of resistance to leaf rust in wheat: an overview in a genome-wide level. Sustainability 15, 3247 (2023).
7. Feuillet, C. *et al.* 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 (2003).

8. Huang, L. *et al.* Map-based cloning of leaf rust resistance gene *Lr21* from the large and polyploid genome of bread wheat. Genetics 164, 655–664 (2003).
9. Cloutier, S. *et al.* 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 (2007).
10. Krattinger, S. G. *et al.* A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. Science 323, 1360–1363 (2009).
11. Moore, J. W. *et al.* A recently evolved hexose transporter variant confers resistance to multiple pathogens in wheat. Nat Genet 47, 1494–1498 (2015).
12. Lin, G. *et al.* Cloning of the broadly effective wheat leaf rust resistance gene *Lr42* transferred from *Aegilops tauschii*. Nat Commun 13, 1–12 (2022).
13. Hewitt, T. *et al.* Wheat leaf rust resistance gene *Lr13* is a specific *Ne2* allele for hybrid necrosis. Mol Plant 14, 1025–1028 (2021).
14. Thind, A. K. *et al.* Rapid cloning of genes in hexaploid wheat using cultivar-specific long-range chromosome assembly. Nat Biotechnol 35, 793–796 (2017).
15. Kolodziej, M. C. *et al.* A membrane-bound ankyrin repeat protein confers race-specific leaf rust disease resistance in wheat. Nat Commun 12, 1–12 (2021).
16. Wang, Y. *et al.* An unusual tandem kinase fusion protein confers leaf rust resistance in wheat. available at Research Square https://doi.org/10.21203/rs.3.rs-1807889/v1 (2022).
17. Miki, Y. *et al.* Origin of wheat B-genome chromosomes inferred from RNA sequencing analysis of leaf transcripts from section *Sitopsis* species of *Aegilops*. DNA Res 26, 171–182 (2019).
18. Dubcovsky, J., Lukaszewski, A., Echaide, M., Antonelli, E. & Porter, D. Molecular characterization of two *Triticum speltoides* interstitial translocations carrying leaf rust and greenbug resistance genes. Crop Sci 38, 1655–1660 (1998).
19. Schneider, A., Molnár, I. & Molnár-Láng, M. Utilisation of *Aegilops* (goatgrass) species to widen the genetic diversity of cultivated wheat. Euphytica 163, 1–19 (2008).
20. Marais, G. *et al.* Attempts to remove gametocidal genes co-transferred to common wheat with rust resistance from *Aegilops speltoides*. Euphytica 171, 71–85 (2010).
21. Wells, D. G., Kota, R., Sandhu, H., Gardner, W. & Finney, K. Registration of one disomic substitution line and five translocation Lines of winter wheat germplasm resistant to wheat streak mosaic virus 1 (Reg. No. GP 199 to GP 204). Crop Sci 22, 1277–1278 (1982).
22. Brevis, J. C. *et al.* Agronomic and quality evaluation of common wheat near-isogenic lines carrying the leaf rust resistance gene *Lr47*. Crop Sci 48, 1441–1451 (2008).
23. Xu, X. *et al.* Characterization of an incomplete leaf rust resistance gene on chromosome 1RS and development of KASP markers for *Lr47* in wheat. Phytopathology 111, 649–658 (2021).
24. Helguera, M., Khan, I. & Dubcovsky, J. Development of PCR markers for the wheat leaf rust resistance gene *Lr47*. Theor Appl Genet 100, 1137–1143 (2000).

25. Yuan, L., Gebrewahid, T. W., Zhang, P., Li, Z. & Liu, D. Identification of leaf rust resistance genes in common wheat varieties from China and foreign countries. J Integr Agr 20, 1302–1313 (2021).
26. Wu, J. *et al.* Genome-wide expression profiling of genes associated with the *Lr47*-mediated wheat resistance to leaf rust (*Puccinia triticina*). Int J Mol Sci 20, 4498 (2019).
27. Figlan, S. *et al.* Breeding wheat for durable leaf rust resistance in Southern Africa: variability, distribution, current control strategies, challenges and future prospects. Front Plant Sci 11, 549 (2020).
28. Omara, R. & Abdelaal, K. Molecular and genetic analysis of leaf rust resistance genes in two new Egyptian wheat cultivars. Egypt J Phytopathol 45, 33–52 (2017).
29. Gultyaeva, E., Shaydayuk, E. & Gannibal, P. Leaf rust resistance genes in wheat cultivars registered in Russia and their influence on adaptation processes in pathogen populations. Agriculture 11, 319 (2021).
30. Bhardwaj, S. *et al.* Physiologic specialization and genetic differentiation of *Puccinia triticina* causing leaf rust of wheat on the Indian subcontinent during 2016 to 2019. Plant Dis 105, 1992–2000 (2021).
31. Li, T. *et al.* Highly efficient heritable genome editing in wheat using an RNA virus and bypassing tissue culture. Mol Plant 14, 1787–1798 (2021).
32. Zhang, W. *et al.* Identification and characterization of *Sr13*, a tetraploid wheat gene that confers resistance to the Ug99 stem rust race group. Proc Natl Acad Sci USA 114, E9483–9492 (2017).
33. Zhang, J. *et al.* A single amino acid change can alter the specificity of the multi-allelic wheat stem rust resistance locus *SR9*. available at Research Square https://doi.org/10.21203/rs.21203.rs-2208386/v2208381 (2022).
34. Luo, J. *et al.* Identification and characterization of *Sr22b*, a new allele of the wheat stem rust resistance gene *Sr22* effective against the Ug99 race group. Plant Biotechnol J 20, 554–563 (2022).
35. Li, H. *et al.* Mapping and characterization of a wheat stem rust resistance gene in durum wheat “Kronos”. Front Plant Sci 12 (2021).
36. Saintenac, C. *et al.* Identifcation of wheat gene *Sr35* that confers resistance to Ug99 stem rust race group. Science 341, 783–786 (2013).
37. Wang, S. *et al. RppM*, encoding a typical CC-NBS-LRR protein, confers resistance to southern corn rust in maize. *Front Plant Sci* 13, 951318 (2022).
38. Hafeez, A. N. *et al.* Creation and judicious application of a wheat resistance gene atlas. Mol Plant 14, 1053–1070 (2021).
39. Chen, S., Zhang, W., Bolus, S., Rouse, M. N. & Dubcovsky, J. Identification and characterization of wheat stem rust resistance gene *Sr21* effective against the Ug99 race group at high temperature. PLoS Genet 14, e1007287 (2018).
40. Saintenac, C. *et al.* Identification of wheat gene *Sr35* that confers resistance to Ug99 stem rust race group. Science 341, 783–786 (2013).

41. Gassmann, W., Hinsch, M. E. & Staskawicz, B. J. The *Arabidopsis RPS4* bacterial-resistance gene is a member of the TIR-NBS-LRR family of disease-resistance genes. Plant J 20, 265–277 (1999).
42. Hu, J. *et al.* A barley stripe mosaic virus-based guide RNA delivery system for targeted mutagenesis in wheat and maize. Mol Plant Pathol 20, 1463–1474 (2019).
43. Wang, W. *et al.* Multiplexed promoter and gene editing in wheat using a virus-based guide RNA delivery system. Plant Biotechnol J 20, 2332–2341 (2022).
44. Knott, D. Mutation of a gene for yellow pigment linked to *Lr19* in wheat. Can J Genet Cytol 22, 651–654 (1980).
45. Bohra, A. *et al.* Reap the crop wild relatives for breeding future crops. Trends Biotechnol 40, 412–431 (2022).
46. Gao, A. *et al.* Pm21 CC domain activity modulated by intramolecular interactions is implicated in cell death and disease resistance. Mol Plant Pathol 21, 975–984 (2020).
47. Bolus, S., Akhunov, E., Coaker, G. & Dubcovsky, J. Dissection of cell death induction by wheat stem rust resistance protein Sr35 and its matching effector AvrSr35. Mol Plant Microbe In 33, 308–319 (2020).
48. Wang, J. *et al.* Reconstitution and structure of a plant NLR resistosome conferring immunity. Science 364, eaav5870 (2019).
49. Förderer, A. *et al.* A wheat resistosome defines common principles of immune receptor channels. Nature 610, 532–539 (2022).
50. Volkova, G., Kudinova, O., Vaganova, O. & Agapova, V. Effectiveness of leaf rust resistance genes in the adult and juvenile stages in southern Russia in 2011–2020. Plants 11, 793 (2022).
51. Kumar, S. *et al.* Characterization of five new pathotypes of *Puccinia triticina* identified from Northeast India, Nepal, and Bangladesh. Australas Plant Path 51, 315–325 (2022).
52. Jones, J. D. G. & Dang, J. L. The plant immune system. Nature 444, 323–329, doi:10.1038/nature05286 (2006).
53. Chen, G. *et al.* Cloning southern corn rust resistant gene *RppK* and its cognate gene *AvrRppK* from *Puccinia polysora*. Nat Commun 13, 4392 (2022).
54. Jin, Y. *et al.* Detection of virulence to resistance gene *Sr24* within race TTKS of *Puccinia graminis* f. sp *tritici*. Plant Dis 92, 923–926 (2008).
55. Jin, Y. *et al.* Detection of virulence to resistance gene *Sr36* within the TTKS race lineage of *Puccinia graminis* f. sp *tritici*. Plant Dis 93, 367–370 (2009).
56. Patpour, M. *et al.* First report of the Ug99 race group of wheat stem rust, *Puccinia graminis* f. sp. *tritici*, in Egypt in 2014. Plant Dis 100, 863–863 (2016).
57. Chen, S. *et al.* Stripe rust resistance gene *Yr34* (synonym *Yr48*) is located within a distal translocation of *Triticum monococcum* chromosome $5A^{m}L$ into common wheat. *Theor Appl Genet* 134, 2197–2211 (2021).

58. Stakman, E. C., Stewart, D. M. & Loegering, W. Q. Identification of physiologic races of *Puccinia graminis* var. *tritici*., (Washington DC, 1962).
59. Dawson, A. M. *et al.* The development of quick, robust, quantitative phenotypic assays for describing the host–nonhost landscape to stripe rust. Front Plant Sci 6, 876 (2015).
60. The International Wheat Genome Sequencing Consortium. Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 361, eaar7191 (2018).
61. Walkowiak, S. *et al.* Multiple wheat genomes reveal global variation in modern breeding. Nature 588, 277–283 (2020).
62. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics (Oxford, England) 30, 2114–2120 (2014).
63. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. Bioinformatics (Oxford, England) 29, 15–21 (2013).
64. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. Preprint at http://arxiv.org/abs/1207.3907 (2012).
65. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19, 455–477 (2012).
66. Li, L.-F. *et al.* Genome sequences of five *Sitopsis* species of *Aegilops* and the origin of polyploid wheat B subgenome. Mol Plant 15, 488–503 (2022).
67. Yu, G. *et al. Aegilops sharonensis* genome-assisted identification of stem rust resistance gene *Sr62*. Nat Commun 13, 1607 (2022).
68. Konieczny, A. & Ausubel, F. M. A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. Plant J 4, 403–410 (1993).
69. Bhattramakki, D. *et al.* Insertion-deletion polymorphisms in 3' regions of maize genes occur frequently and can be used as highly informative genetic markers. Plant Mol Biol 48, 539–547 (2002).
70. Avni, R. *et al.* Genome sequences of three *Aegilops* species of the section Sitopsis reveal phylogenetic relationships and provide resources for wheat improvement. Plant J 110, 179–192 (2022).
71. Diaz, A., Zikhali, M., Turner, A. S., Isaac, P. & Laurie, D. A. Copy number variation affecting the *Photoperiod-B1* and *Vernalization-A1* genes is associated with altered flowering time in wheat (*Triticum aestivum*). PloS One 7, e33234 (2012).
72. Pearce, S., Vanzetti, L. S. & Dubcovsky, J. Exogenous gibberellins induce wheat spike development under short days only in the presence of *VERNALIZATION 1*. Plant Physiol 163, 1433–1445 (2013).
73. Sears, E. R. An induced mutant with homoeologous pairing in common wheat. Can J Genet Cytol 19, 585–593 (1977).
74. Gyawali, Y., Zhang, W., Chao, S., Xu, S. & Cai, X. Delimitation of wheat *ph1b* deletion and development of *ph1b*-specific DNA markers. Theor Appl Genet 132, 195–204 (2019).

75. Komuro, S., Endo, R., Shikata, K. & Kato, A. Genomic and chromosomal distribution patterns of various repeated DNA sequences in wheat revealed by a fluorescence *in situ* hybridization procedure. Genome 56, 131–137 (2013).
76. Hao, M. *et al.* Comparison of homoeologous chromosome pairing between hybrids of wheat genotypes Chinese Spring *ph1b* and Kaixian-luohanmai with rye. Genome 54, 959–964 (2011).
77. Song, Z. *et al.* Analysis of structural genomic diversity in *Aegilops umbellulata*, *Ae. markgrafii*, *Ae. comosa*, and *Ae. uniaristata* by fluorescence *in situ* hybridization karyotyping. Front Plant Sci 11, 710 (2020).
78. Tang, Z. *et al.* New types of wheat chromosomal structural variations in derivatives of wheat-rye hybrids. PloS One 9, e110282 (2014).
79. Zhang, A. *et al.* Molecular cytogenetics identification of a wheat-*Leymus mollis* double disomic addition line with stripe rust resistance. Genome 60, 375–383 (2017).
80. Liang, Y. *et al.* Drought-responsive genes, late embryogenesis abundant group3 (*LEA 3*) and vicinal oxygen chelate, function in lipid accumulation in *Brassica napus* and *Arabidopsis* mainly via enhancing photosynthetic efficiency and reducing ROS. Plant Biotechnol J 17, 2123–2142 (2019).
81. Li, Y. *et al. BRITTLE CULM1*, which encodes a COBRA-like protein, affects the mechanical properties of rice plants. The Plant Cell 15, 2020–2031 (2003).
82. Wang, J. *et al.* SAUR17 and SAUR50 differentially regulate PP2C-D1 during apical hook development and cotyledon opening in Arabidopsis. Plant Cell 32, 3792–3811 (2020).
83. Li, C. *et al.* Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. Genome Biol 19, 1–9 (2018).
84. Shan, Q., Wang, Y., Li, J. & Gao, C. Genome editing in rice and wheat using the CRISPR/Cas system. Nat Protoc 9, 2395–2410 (2014).

Figures

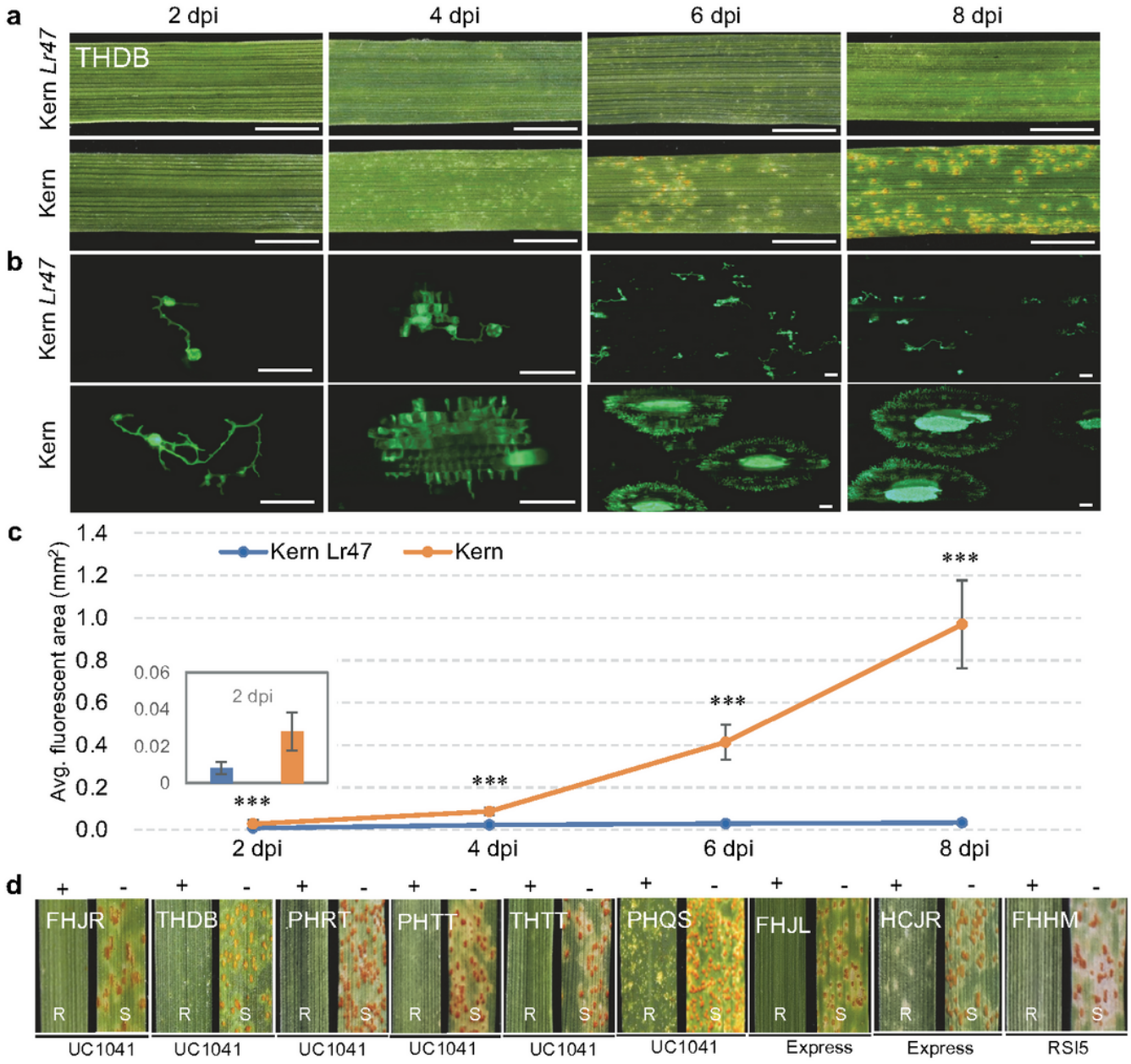


Figure 1

Resistance responses of leaf rust resistance gene *Lr47*. a Infection types in the introgression line Kern*Lr47* and its recurrent parent Kern in response to *Pt* race THDB at 2, 4, 6, and 8 dpi. Scale bars represent 0.5 cm. **b** *Pt* infection areas visualized by WGA-FITC staining. Leaves were collected at 2, 4, 6, and 8 dpi, cleared with KOH, and stained with WGA-FITC. Scale bars represent 100 μm. **c** Interaction graphs showing the average size of individual fungal infection areas of Kern *Lr47* (blue) and Kern (orange) estimated by fluorescence microscopy at 2, 4, 6, and 8 dpi (n = 40). ***, $P < 0.001$. Error bars are standard errors of the mean. **d** Infection types of *Lr47* NILs and their recurrent parents in response to

selected *Pt* races FHJL, PHQS, FHJR, THDB, PHRT, PHTT, THTT, HCJR, and FHHM. +, resistant *Lr47* allele present; -, no resistant *Lr47* allele. R, Resistant; S, Susceptible.

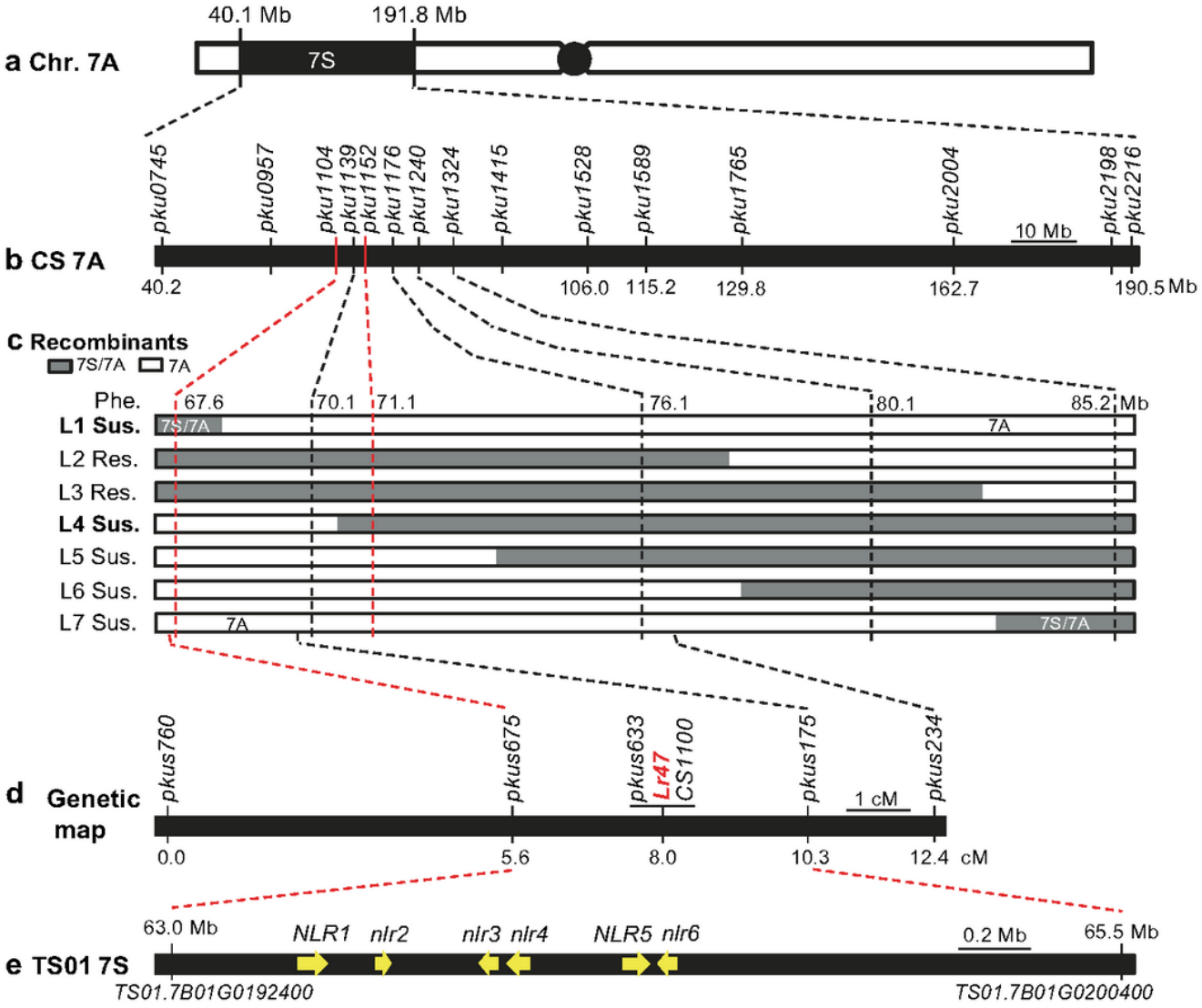


Figure 2

Fine mapping of the leaf rust resistance gene *Lr47*. a Schematic representation of chromosome 7A in Kern *Lr47* carrying the introgressed *Ae. speltoides* segment 7S#1S (indicated by a black rectangle). **b** Fine mapping of *Lr47* using 7A/7S homoeologous recombinants induced by the *ph1b* mutant. The genome-specific markers (Supplementary Table 4) distributed along the introgressed 7S#1S chromosome segment and their physical locations based on the Chinese Spring reference genome Refseq v1.1. **c** Recombinant haplotypes (L1–L7) with *Ae. speltoides* segments of different lengths determined by marker analysis. White rectangles represent chromosome 7A; gray rectangles indicate heterozygous for

7S/7A; Sus., susceptible; Res., resistant. **d** Genetic map of *Lr47* based on 1,141 F_2 plants from cross m118 × Kern *Lr47*. **e** Physical map of *Lr47* in the sequenced reference genome of *Ae. speltoides* TS01. Yellow arrows represent NLR genes.

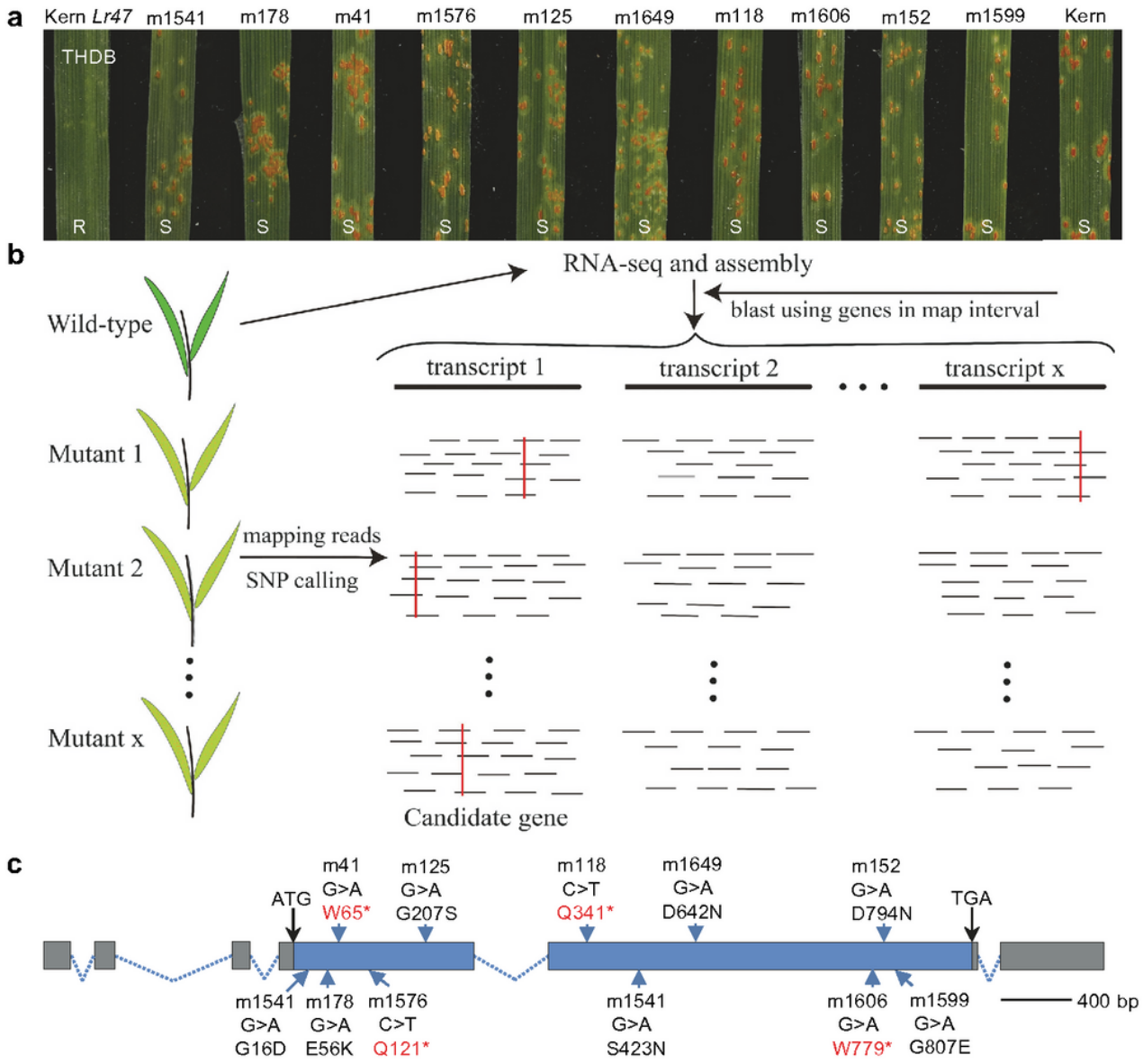


Figure 3

Identification of *Lr47* using the EMS mutagenesis and transcript assembly (EMTA) method. a Susceptible EMS mutants used to isolate the*Lr47* candidate gene. Infection types for Kern *Lr47* (positive control), ten independent EMS mutants, and Kern (negative control) inoculated with *Pt* race THDB. R, resistant; S, susceptible. **b** Schematic representation of the EMTA method. RNA-seq reads from the resistant parent

(wild type) are *de novo* assembled. A search of the transcriptome database of the resistant parent using the sequences of genes within the mapping interval in sequenced reference genomes yields transcripts with high similarity. Next, RNA-seq readsfrom susceptible EMS mutants are mapped to the obtained transcripts. Transcript 1 in this example, which has a preponderance of single-nucleotide variants (red lines) across the susceptible mutants, is considered a good candidate. **c** Gene structure of the *Lr47*candidate gene. The positions of the EMS-derived loss-of-function mutations are indicated by blue arrows. Gray boxes indicate untranslated regions, blue boxes represent coding exons, and dotted lines represent introns.

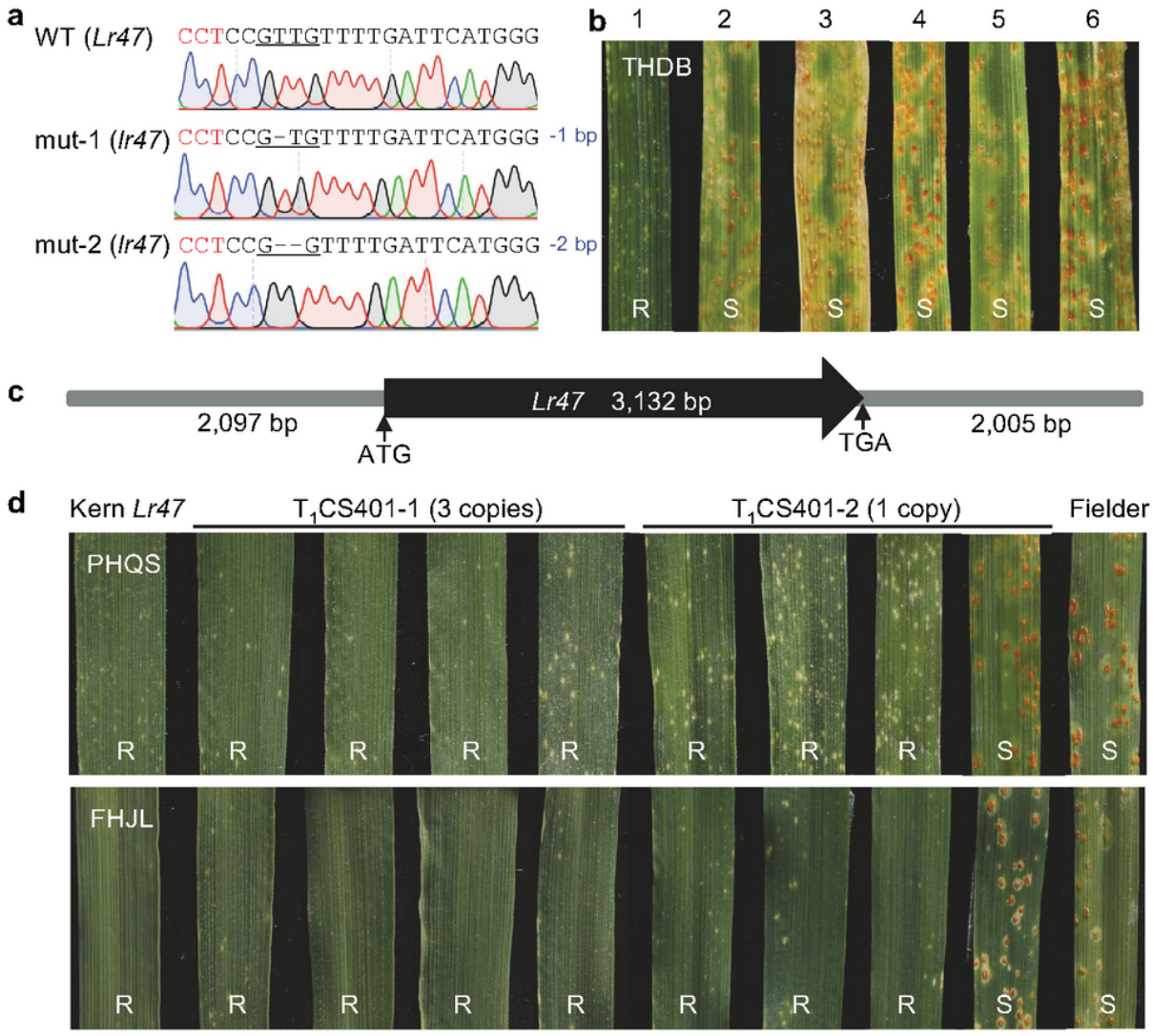


Figure 4

Functional validation of *Lr47* by BSMV-mediated gene editing and transgenic complementation. a Sequencing chromatogram showing the induced polymorphisms between WT and the selected BSMV-

Cas9-induced editing mutants. The mutated nucleotides are underlined and the CCT PAM sequence is highlighted in red. **b** Reactions to *Pt* race THDB. 1) Kern *Lr47*; 2-3) mutant line mut-1; 4-5) mutant line mut-2; and 6) Cas9-transgenic Bobwhite. **c** The 7,234-bp genomic DNA fragment including *Lr47* that was used for transformation. The black arrow indicates the *Lr47* gene (from ATG to TGA). **d** Seedling infection types of Kern *Lr47*, two transgenic families T_1CS401-1 and T_1CS401-2, and Fielder control in response to *Pt* races PHQS and FHJL. Copy number of transgenes was estimated using TaqMan assays (Supplementary Table 9). R, resistant; S, susceptible.

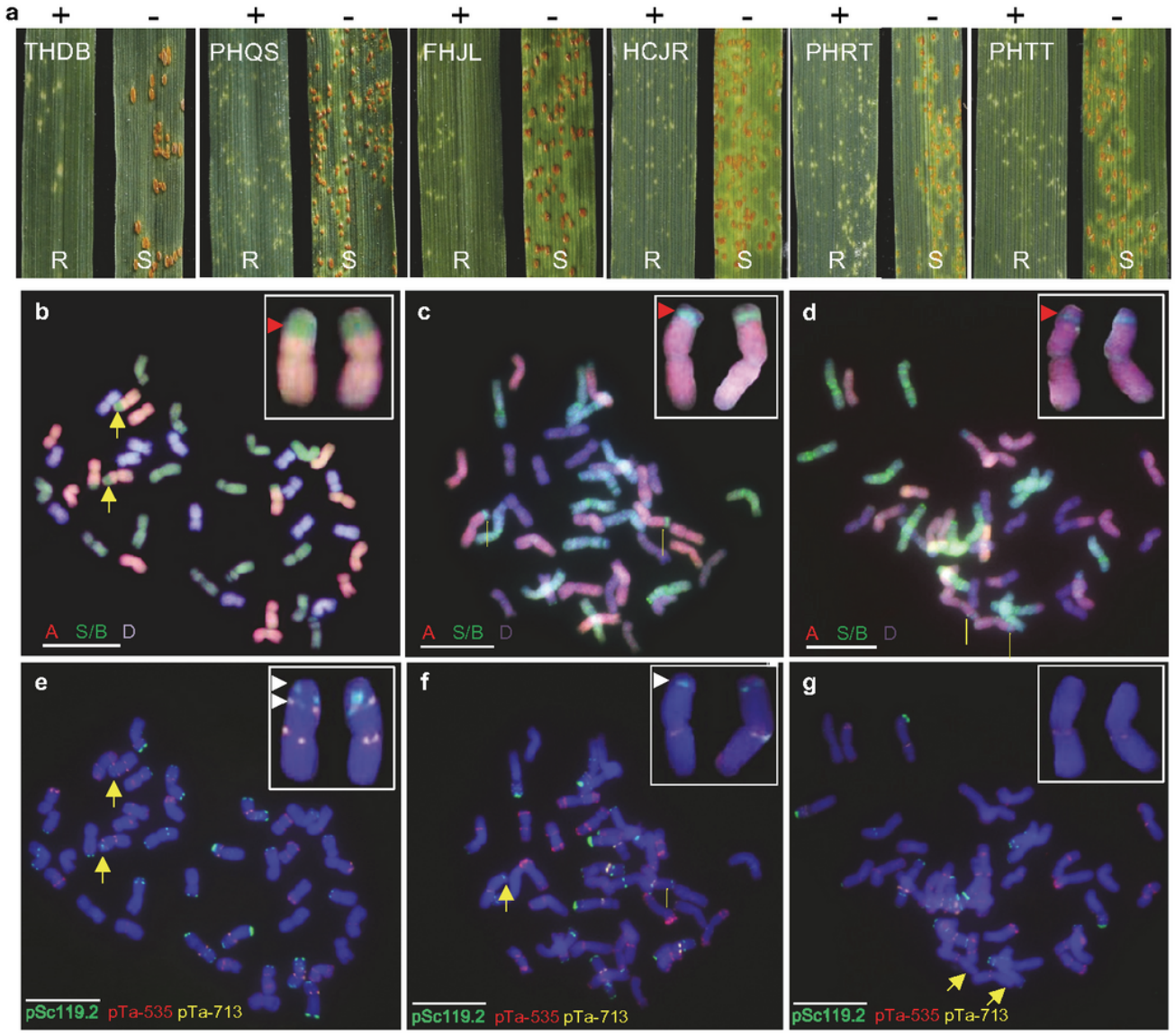


Figure 5

Reduction in the length of the introgressed *Ae. speltoides* chromosome segment carrying *Lr47*. a Infection types of homozygous BC_1F_2 plants from the introgression line YM21+*Lr47*-2 (+) and the plants lacking

Lr47 (-). Six *Pt* races (THDB, PHQS, FHJL, HCJR, PHRT, and PHTT) were used for evaluation. R, Resistant; S, Susceptible. **b, c, d** GISH images of wheat lines Kern *Lr47*, L2 (BC_2F_2), and L8 (BC_1F_2). The magnified images show the *Ae. speltoides* chromosome segments introgressed into wheat chromosome 7A (*Ae. speltoides* chromatin is painted in green and marked with red arrowheads). **e, f, g** FISH images from wheat lines Kern *Lr47*, L2 (BC_2F_2), and L8 (BC_1F_2). Probes pSc119.2 (green), pTa535 (red), and pTa713 (yellow) were used. Yellow arrows indicate the wheat-*Ae. speltoides* translocated 7A chromosomes, and magnified images of these chromosomes are shown in the insets. Scale bar = 10 μm.

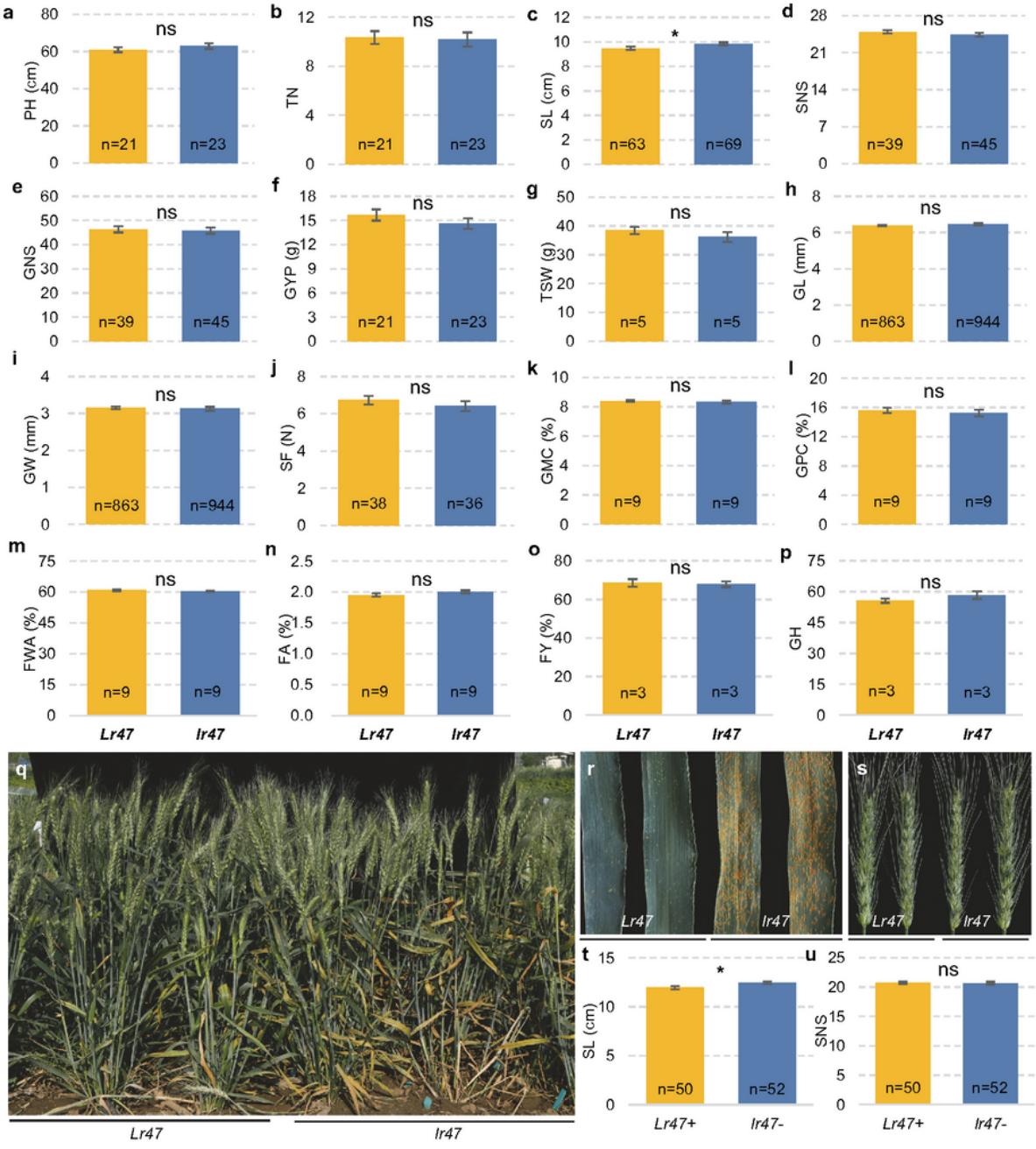


Figure 6

Statistical analysis of agronomic and quality traits in YM21+*Lr47*-1 and its sister control line. a-p Plants grown in a controlled walk-in growth chamber at 24 °C day/22 °C night with a 16 h light/8 h dark photoperiod were phenotyped for the following traits: **a** plant height (PH); **b** tillers number (TN); **c** spike length (SL); **d** spikelet number per spike (SNS); **e** grain number per spike (GNS); **f** grain yield per plant (GYP); **g** thousand-seed weight (TSW); **h** grain length (GL); **i** grain width (GW); **j** shearing force (SF) in

Newtons (N); **k** grain moisture content (GMC); **l** grain protein content (GPC); **m** flour water absorption (FWA); **n** flour ash (FA); **o** flour yield (FY); and **p** grain hardness (GH). **q-u** Plants grown in the field. **q** Visual phenotypes of YM21+*Lr47*-1 and its sister line in the field. **r** Infection types of YM21+*Lr47*-1 and the sister control line. **s** Close-up views of spikes. **t** Spike length (SL). **u**Grain number per spike (GNS). Error bars are standard errors of the means. ns = not significant (P >0.05), * = P < 0.05. *Lr47*, resistant *Lr47* allele present (YM21+*Lr47*-1); *lr47*, sister control line lacking the alien chromosome segment.

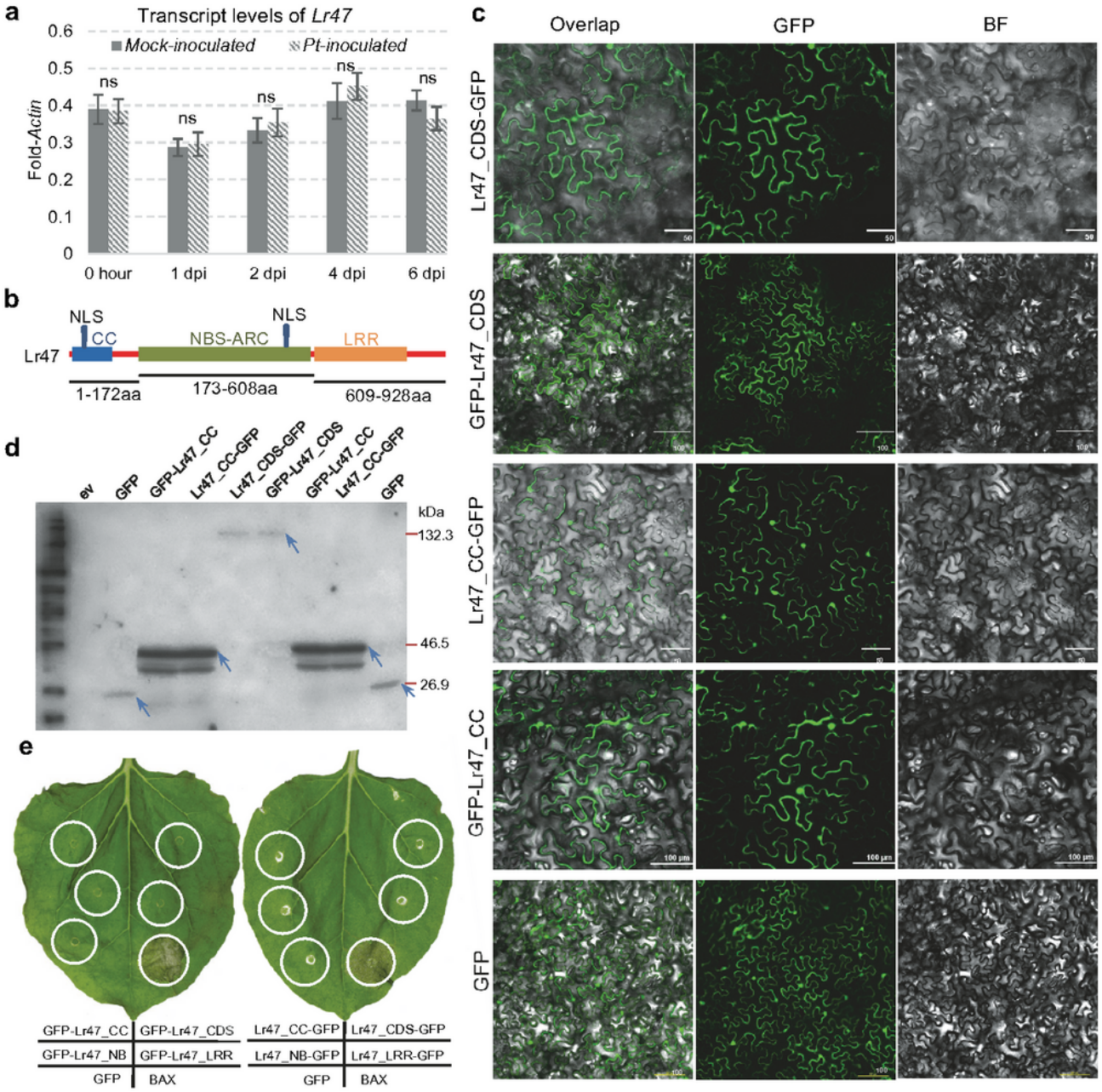


Figure 7

Characterization of Lr47. a Transcript levels of *Lr47* in mock-inoculated and *Pt*-inoculated plants. Leaves were collected at five time points: 0 h, 1 dpi, 2 dpi, 4 dpi, and 6 dpi. Transcript levels were expressed as fold-*Actin* (n = 7). Error bars are standard errors of the mean. ns = not significant. **b** Schematic diagram of conserved domains in the Lr47 protein. **c** Subcellular localization of the GFP- Lr47 fusion protein in tobacco leaves. Scale bars represent 50 or 100 μm. BF = bright field; GFP = green fluorescent protein. **d** GFP-horseradish peroxidase Western blot showing proteins expressed at expected sizes for all constructs. Target bands are highlighted by blue arrowheads. ev = empty vector. **e** Macroscopic cell death in *N. benthamiana* leaves 48 hpi with *Agrobacterium tumefaciens* carrying Lr47 constructs. GFP, negative control; BAX, positive control.

Supplementary Files

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

- rs3.pdf
- Lr47SupplementaryFiguresandTables.pdf
- SupplementaryTable5.xlsx
- SupplementaryTable11.xlsx