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.

An unusual tandem kinase fusion protein confers leaf rust resistance in wheat

Yajun Wang
KAUST
Michael Abrouk
KAUST
Spyridon Gourdoupis
KAUST
Dal-Hoe Koo
Kansas State University
Miroslava Karafiátová
Institute of Experimental Botany
István Molnár
Institute of Experimental Botany
Jaroslav Doležel
Institute of Experimental Botany
Naveenkumar Athiyannan
KAUST
Emile Cavalet-Giorsa
KAUST
Łukasz Jaremko
KAUST
Jesse Poland
KAUST
Simon G. Krattinger (simon.krattinger@kaust.edu.sa)
KAUST

Research Article

Keywords: wheat, disease resistance, kinase fusion protein, wild wheat relative, leaf rust, Lr9, Lr58

Posted Date: July 5th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1807889/v1

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

An unusual tandem kinase fusion protein confers leaf rust resistance in wheat

Yajun Wang[1,2], Michael Abrouk[1,2], Spyridon Gourdoupis[3], Dal-Hoe Koo[4], Miroslava Karafiátová[5], István Molnár[5,6], Kateřina Holušová[5], Jaroslav Doležel[5], Naveenkumar Athiyannan[1,2], Emile Cavalet-Giorsa[1,2], Łukasz Jaremko[3], Jesse Poland[1,2], Simon G. Krattinger[1,2*]

[1]Plant Science Program, Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia

[2]Center for Desert Agriculture, KAUST, Saudi Arabia

[3]Bioscience Program, Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia

[4]Wheat Genetics Resource Center and Department of Plant Pathology, Kansas State University, Manhattan, KS 66506, USA

[5]Institute of Experimental Botany of the Czech Academy of Sciences, Centre of the Region Haná for Biotechnological and Agricultural Research, Olomouc, Czech Republic

[6]Present address: Agricultural Institute, Centre for Agricultural Research, Hungary

*e-mail: simon.krattinger@kaust.edu.sa

Initial paragraph

The introgression of chromosome segments from wild relatives is an established strategy to enrich crop germplasm with novel disease resistance genes[1]. Here, we use mutagenesis and transcriptome sequencing to clone the leaf rust resistance gene *Lr9,* which was introduced into bread wheat from the wild grass species *Aegilops umbellulata*[2]. We establish that *Lr9* encodes an unusual tandem kinase fusion protein. Long-read sequencing of a wheat *Lr9* introgression line and the putative *Ae. umbellulata Lr9* donor enabled us to assemble the ~28.4-Mb *Lr9* translocation and to identify the translocation breakpoint. We likewise cloned *Lr58*, which was reportedly introgressed from *Ae. triuncialis*[3], but has an identical coding sequence compared to *Lr9*. Cytogenetic and haplotype analyses corroborate that the two genes originate from the same translocation event. Our work sheds light on the emerging role of kinase fusion proteins in wheat disease resistance, expanding the repertoire of disease resistance genes for breeding.

Main Text

Bread wheat (*Triticum aestivum;* $2n = 6x = 42$, AABBDD genome) is one of the most widely cultivated crops worldwide[4]. The fungal pathogen *Puccinia triticina* (*Pt*) causes leaf rust, which is among the most widespread and devastating wheat diseases, resulting in annual production losses of ~25 million tons[5,6]. Of the 69 known leaf rust (*Lr*) resistance genes in wheat, 28 were introgressed from the secondary and tertiary gene pools through interspecific hybridizations[7]. The genus *Aegilops* (annual goat grasses) comprises diploid and polyploid species that serve as reservoirs to increase wheat genetic diversity[8,9]. The two resistance genes *Lr9* and *Lr58* were introgressed into bread wheat from the U-genomes of *Ae. umbellulata* (diploid; $2n = 2x = 14$; UU genome) and *Ae. triuncialis* (tetraploid; $2n = 4x = 28$; $U^tU^tC^tC^t$ genome), respectively[2,3]. The *Lr9* introgression is of historic importance, as it was the first disease resistance gene transferred into bread wheat in the 1950s using irradiation, although the *Ae. umbellulata* donor was not specified[2]. The resulting bread wheat line, 'Transfer', carried an *Ae. umbellulata* translocation at the end of chromosome arm 6BL[2,10]. The *Lr58* introgression reportedly occurred spontaneously through homoeologous recombination while backcrossing *Ae. triuncialis* accession TA10438 to the susceptible

bread wheat line WL711[3]. *Lr58* was mapped to chromosome arm 2BL in the resulting backcross line TA5605[3]. Wheat cultivars carrying *Lr9* were released in the late 1960s. *Lr9* is still effective in many wheat growing areas, although virulent *Pt* isolates emerged a few years after the release of the first *Lr9*-carrying wheat cultivars[11]. *Lr58* has not been widely deployed in wheat breeding, and *Pt* races virulent on *Lr58* have been reported[3].

Both *Lr9* and *Lr58* conferred strong resistance against multiple *Pt* isolates in controlled conditions (Fig. 1a, Extended Data Fig. 1). To clone the two genes, we screened ethyl methanesulfonate (EMS)-induced mutant populations in the genetic backgrounds of Thatcher*Lr9* (Transfer/6*Thatcher; 919 M2 families) and TA5605 (*Lr58* carrier; ~7,400 M2 families) for loss of resistance. We identified and validated 17 and 104 susceptible mutants in these two backgrounds, respectively (Fig. 1b, Extended Data Fig. 2). Because alien introgressions often show suppressed recombination[1], we developed MutIsoSeq to clone *Lr9* and *Lr58*. MutIsoSeq combines isoform sequencing (Iso-seq) of wild-type parental lines and transcriptome deep sequencing (RNA-seq) of EMS mutants to identify altered transcripts for candidate genes and does not require genetic mapping (Fig. 1c). First, we generated Iso-seq reads derived from RNA isolated from *Pt*-inoculated Thatcher*Lr9* and TA5605 leaves. We then mapped RNA-seq reads from 10 Thatcher*Lr9* and 10 TA5605 susceptible mutants against their respective parents and identified EMS-type (G/C to A/T) mutations. For both sets of mutants, we only identified one transcript with EMS-type point mutations in all 20 sequenced mutants (Table 1). Surprisingly, the putative *Lr9* and *Lr58* coding sequences were identical suggesting that 1) two identical genes were introgressed independently from two different U-genome species, or 2) *Lr9* and *Lr58* are actually the same original introgression and named differently from pedigree or mapping errors.

The putative 3,504-bp *Lr9*/*Lr58* coding sequence encoded a 1,167-amino acid protein with an N-terminal tandem kinase domain followed by a von Willebrand factor A (vWA) domain and a Vwaint domain in the C-terminus (Fig. 1d). Wheat tandem kinases (WTKs) have recently emerged as a new prominent player

involved in disease resistance in *Triticeae*, the botanical tribe including wheat and barley[12-18]. The fusion of a WTK to another domain, however, is unique and highly unusual. vWA and Vwaint domains often occur together (subsequently referred to as vWA/Vwaint domains) and are thought to participate in protein-protein interactions in plants[19,20]. The vWA-containing copine proteins in *Arabidopsis* and rice are regulators of disease resistance and the *Arabidopsis* copine protein BON1 might be guarded by the NLR immune receptor SNC1, supporting a potential role of vWA-containing proteins in defense[19,21,22]. To reflect the introduced nomenclature, we refer to the *Lr9/Lr58* candidate as *WTK6-vWA* hereafter. A PCR marker derived from *WTK6-vWA* showed complete linkage with disease resistance in F2 populations derived from crosses between Thatcher*Lr9* and the susceptible wheat cultivar Avocet S and TA5605 with Avocet S (Extended Data Figs. 3, 4).

Several lines of evidence indicate that *Lr9* and *Lr58* are from the same original introgression: 1) Previously published *Lr9* and *Lr58* markers showed complete linkage with *WTK6-vWA* in the two mapping populations (Extended Data Figs. 3, 4)[23,24]. 2) Complementation tests using an F2 population from a Thatcher*Lr9* × TA5605 cross had no susceptible individuals (Extended Data Fig. 5), supporting the linkage of *Lr9* and *Lr58*. 3) Fluorescence *in situ* hybridization using a probe derived from *WTK6-vWA* hybridized to chromosome arm 6BL in both Transfer (*Lr9*) and TA5605 (*Lr58*) (Extended Data Fig. 6). Sequencing of the 3.7-kb probe revealed 100% sequence identity between Thatcher*Lr9* and TA5605 but one polymorphism in the putative *Ae. triuncialis Lr58* donor TA10438. Together, these results indicated that *Lr9* and *Lr58* are identical and that *Ae. triuncialis* accession TA10438 is not the *Lr58* donor. As the *Lr9* translocation was produced 51 years before the *Lr58* introgression, *Lr58* is likely *Lr9*. During the generation of TA5605, leaf rust-resistant plants with normal chromosome numbers were selected, assuming homoeologous recombination between WL711 and *Ae. triuncialis* chromosomes[3]. A rare seed contamination containing *Lr9* may have been selected at this step. Alternatively, materials might have been mislabeled at a later date. The TA5605 and TA10438 accessions used here were the original sources, which were deposited in the Wheat Genetics Resource Center[3]. The reports placing *Lr9* and *Lr58* on different

chromosomes may be explained by the non-homologous *Lr9* translocation following irradiation. *WTK6-vWA* produced the strongest BLAST hits on the long arms of homoeologous group 2 chromosomes in wheat, in agreement with previous reports showing that the *Ae. umbellulata* chromosome 6U segment from which *Lr9* originated is syntenic to wheat group 2 chromosomes[8,25]. Molecular markers derived from the bread wheat group 2 chromosomes might thus map to the *Ae. umbellulata* translocation, while linked markers derived from the bread wheat chromosome that received the translocation would map to chromosome arm 6BL. We failed to identify *WTK6-vWA* in sequences of flow-sorted TA5605 2B chromosomes, providing additional evidence that *Lr58* is not located on chromosome 2B.

We amplified and Sanger-sequenced the *WTK6-vWA* genic region from the remaining 101 *Lr9/Lr58* (subsequently referred to as *Lr9*) mutants and detected missense or nonsense mutations in all but one mutant in the Thatcher*Lr9* background, which might harbor second-site mutations or a mutation in the regulatory region. We identified 19 premature stop codon mutations, 67 non-redundant amino acid substitutions, four splice-site mutations, and one start codon mutation (Fig 1d, Extended Data Fig. 7, Supplementary Table 1). We further validated the identity of *Lr9* by virus-induced gene silencing (VIGS). Silencing of *WTK6-vWA* increased leaf rust susceptibility (Fig. 1e). Together, the mutant analysis, genetic mapping, and silencing confirm that *WTK6-vWA* is *Lr9*.

The 67 non-redundant amino acid substitution mutations enabled us to establish a detailed map of critical residues required for WTK6-vWA function. One amino acid substitution mapped to the N-terminal loop, 48 in kinase 1, 13 in kinase 2, and 5 in the C-terminal vWA/Vwaint domains. Protein modeling revealed a marked difference in the effects of individual amino acid substitutions between the two kinase domains (Fig. 2). Thirty-one out of 48 amino acid substitutions in kinase 1 were predicted to be surface-localized, while most (8 out of 13) of the substitutions in kinase 2 affected internal residues (Fig. 2, Supplementary Table 1). Substitutions at surface-localized amino acids are likely to interfere with protein function, while substitutions of internal amino acids might affect protein structure and stability[26,27]. Seventeen of the amino

acid substitutions in kinase 1 affected conserved subdomains and loops, including ATP-binding sites, compared to only three amino acids for the conserved kinase 2 subdomains (Extended Data Fig. 7)[28]. In conserved subdomain III, kinase 2 carried an aspartic acid (D) instead of the highly conserved glutamic acid (E) found in kinase 1 (Extended Data Fig. 7). Substitution of the subdomain III glutamic acid in the *Arabidopsis* EVR kinase abrogated kinase activity[29]. Thus, we conclude that the two kinase domains may play different roles in WTK6-vWA.

Most WTK proteins involved in disease resistance in wheat and barley have been suggested to harbor a kinase and a pseudokinase domain[12-18]. *WTK6-vWA* homologs were very rare in bread wheat, with zero to two functional copies identified in wheat reference assemblies (Supplementary Table 2)[30-35]. All bread wheat *WTK6-vWA* homologs represented *Lr9* orthologs located on chromosome arms 2AL and 2BL. The wheat lines ArinaLrFor and Fielder carried the closest Lr9 orthologs on chromosome 2A, with 90.1% amino acid sequence identity to Lr9. We also identified one to eight genes encoding proteins with a single kinase domain followed by vWA/Vwaint domains in individual assemblies of bread wheat cultivars and various wild relatives (Supplementary Table 2)[30-38]. We hypothesize that kinase-vWA proteins play a role in basal plant immunity and are targeted by *Pt* virulence effectors. A duplication of the kinase domain in WTK6-vWA might have resulted in a new function, whereby kinase 2 and the vWA/Vwaint domains serve as decoy[39], while kinase 1 activity is required to initiate a defense response after effector perception. All five amino acid substitutions in the vWA/Vwaint domains were surfaced-localized, indicating that this domain might be crucial for effector recognition (Fig. 2, Supplementary Table 1).

We used a diagnostic Kompetitive Allele Specific PCR (KASP) marker derived from the polymorphism that distinguished *WTK6-vWA* alleles from Thatcher*Lr9* and *Ae. triuncialis* TA10438 to screen a diversity panel comprising 59 *Ae. umbellulata* and 205 *Ae. triuncialis* accessions (Extended Data Fig. 8, Supplementary Table 3). No *Ae. triuncialis* accession carried *Lr9*, and only one *Ae. umbellulata* accession (TA1851) was positive for *Lr9*. We generated contig-level assemblies of Thatcher*Lr9* and the putative *Ae.*

umbellulata Lr9 donor TA1851 using PacBio circular consensus sequencing (CCS)[40] to determine the genomic sequence of *Lr9,* including surrounding regions (Table 2). We assembled a 5.66-Mb and 14.99-Mb *Lr9*-containing contig in Thatcher*Lr9* and TA1851, respectively (Fig. 3a). The *Lr9* gene sequence spanned 10.63 kb and had 14 exons (Extended Data Fig. 9). The *Lr9* gene sequence was identical between Thatcher*Lr9* and TA1851 and the two contigs showed 99.99% sequence identity over 5.58 Mb, confirming that TA1851 is the *Lr9* donor (Fig. 3a). The few polymorphisms mainly represented short insertion/deletions and are considered likely sequencing errors. The Thatcher*Lr9* contig had two large deletions of 24.8 and 46.0 kb compared to TA1851, most likely caused by the irradiation treatment (Fig. 3a). Mapping of TA5605, Thatcher, and *Ae. triuncialis* TA10438 Illumina reads confirmed that TA5605 carries the same introgression as Thatcher*Lr9*, while the mapping rates for Thatcher and TA10438 reads were low (Fig. 3b), thus further supporting that *Lr9* and *Lr58* come from the same original introgression.

To reconstruct the history of the *Lr9* translocation, we first estimated the size of the *Lr9* translocation to be ~35-50 Mb by genomic *in situ* hybridization (Fig. 3c). We then mapped TA1851 *k*-mers to the Thatcher*Lr9* assembly, which identified 13 Thatcher*Lr9* contigs with high *k*-mer mapping rates (Fig. 3d, Supplementary Table 4). The 13 Thatcher*Lr9* contigs were 0.28-5.66 Mb in size and matched five contigs in TA1851. We ordered and oriented the five TA1851 contigs based on a published *Ae. umbellulata* genetic map[41] (Supplementary Table 5). One Thatcher*Lr9* contig spanned a 585-bp gap between two neighboring TA1851 contigs, which left four gaps of unknown size. Most of the gaps between the Thatcher*Lr9* contigs were small (<4 kb) based on the corresponding TA1851 sequence, suggesting that we covered most of the *Lr9* translocation. Thatcher*Lr9* contig Tc*Lr9*ptg0002621 marked the terminal end of the translocation, with *Lr9* locating 6.05 Mb proximal to the end. We identified the proximal breakpoint of the *Ae. umbellulata* translocation in Thatcher*Lr9* contig Tc*Lr9*ptg0023831 (Fig. 3d). *k*-mer mapping rates and dot plots revealed the exact breakpoint between bread wheat chromosome 6B and the *Ae. umbellulata* translocation (Fig. 3e, Extended Data Fig. 10), which was located in the putative promoter region of a NLR gene in bread wheat, while the breakpoint in *Ae. umbellulata* corresponded to a retrotransposon. These results confirmed that the use of

irradiation led to a fusion of non-homologous chromosome segments. The *Ae. umbellulata* translocation spanned ~28.4 Mb, in agreement with the cytogenetic estimate. Bread wheat line Chinese Spring was used as a recurrent parent to generate Transfer, which allowed us to compare the exact positions of the translocation breakpoints in Thatcher*Lr9* (carrying the *Lr9* translocation from Transfer) and the Chinese Spring reference genome[30,32]. The translocation breakpoint was at position 725.61 Mb (RefSeq v2.1), which was 5,580,341 bp proximal to the end of Chinese Spring chromosome 6B, indicating that this 5.58-Mb segment including 87 high-confidence genes is missing in *Lr9*-containing wheat lines (Fig. 3f). Alien introgressions, including the *Lr9*-containing *Ae. umbellulata* translocation, are often associated with reduced yields[42]. Our results show that such yield decline might be caused both by undesirable alleles on the *Ae. umbellulata* introgression (linkage drag) and by the loss of this 5.58-Mb chromosomal segment.

In summary, we identified a novel class of kinase fusion protein involved in disease resistance. Our results indicate that the duplicated kinase domains have diverse functions and we hypothesize that the second kinase and the vWA/Vwaint domains are integrated decoys. A second example of an 'integrated decoy kinase fusion protein' is the protein encoded by the stem rust resistance gene *Sr43* (ref. 43). Sr43 contains a kinase domain fused to two domains of unknown function that might serve as decoys. Together, these two studies highlight the emerging role of unusual kinase fusion proteins in wheat disease resistance[13-18,44-47], which will form the basis to utilize this protein class in breeding.

Acknowledgements

We are grateful to Lingli Zou for technical assistance, Yong Zhou and Hanin Ibrahim Ahmed for discussions on bioinformatics analyses, and the KAUST Bioscience Core Lab for sequencing support. We thank Zdeňka Dubská, Romana Šperková, and Jitka Weiserová for preparation of chromosome samples for flow cytometry, Mahmoud Said and Petr Cápál for chromosome 2B sorting, and John Raupp from the Wheat Genetics Resource Center for providing germplasm. This publication is based upon work supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research

(OSR) under Award No. OSR-CRG2018-3768. Dal-Hoe Koo was supported by WGRC/IUCRC and NSF (Grant No.1822162). István Molnár was supported from Marie Curie Fellowship grant award 'AEGILWHEAT' (H2020-MSCA-IF-2016-746253) and from the Hungarian National Research, Development and Innovation Office (K135057). Miroslava Karafiátová, Kateřina Holušová and Jaroslav Doležel were supported from the ERDF project "Plants as a tool for sustainable global development" (No. CZ.02.1.01/0.0/0.0/16_019/0000827). Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA LM2018140) provided within the program Projects of Large Research, Development and Innovations Infrastructures.

Author contributions

Y.W. and S.G.K. designed the overall concept of the research. Y.W. performed molecular experiments and rust inoculations. Y.W. and N.A. generated mapping populations. Y.W., M.A. and E.C. performed assemblies and analysed genomic data. D.K. and J.P. supplied germplasm and performed fluorescence *in situ* hybridization and genomic *in situ* hybridization experiments. S.G. and L.J. performed protein modelling. M.K., I.M. and J.D. performed chromosome flow sorting. K.H. sequenced flow sorted chromosome 2B. Y.W. and S.G.K. wrote the initial version of the manuscript. All authors contributed to subsequent versions and have read and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper

Correspondence and requests for materials should be addressed to S.G.K.

References

1. Wulff, B.B. & Moscou, M.J. Strategies for transferring resistance into wheat: from wide crosses to GM cassettes. *Frontiers in Plant Science* **5**, 692 (2014).
2. Sears, E.R. The transfer of leaf-rust resistance from *Aegilops umbellulata* to wheat. *Brookhaven Symposia in Biology* **9**, 1-21 (1956).
3. Kuraparthy, V. *et al.* A cryptic wheat-*Aegilops triuncialis* translocation with leaf rust resistance gene *Lr58*. *Crop Science* **47**, 1995-2003 (2007).
4. Tadesse, W. *et al.* Genetic gains in wheat breeding and its role in feeding the world. *Crop Breeding, Genetics and Genomics* **1**, e190005 (2019).
5. Savary, S. *et al.* The global burden of pathogens and pests on major food crops. *Nature Ecology & Evolution* **3**, 430-439 (2019).
6. Kolmer, J. Leaf rust of wheat: Pathogen biology, variation and host resistance. *Forests* **4**, 70-84 (2013).
7. Hafeez, A.N. *et al.* Creation and judicious application of a wheat resistance gene atlas. *Molecular Plant* **14**, 1053-1070 (2021).
8. Said, M. *et al.* Development of DNA markers from physically mapped loci in *Aegilops comosa* and *Aegilops umbellulata* using single-gene FISH and chromosome sequences. *Frontiers in Plant Science* **12**, 689031 (2021).
9. Molnar-Lang, M., Ceoloni, C. & Dolezel, J. *Alien introgression in wheat*, (Springer, 2015).
10. Schachermayr, G. *et al.* Identification and localization of molecular markers linked to the *Lr9* leaf rust resistance gene of wheat. *Theoretical and Applied Genetics* **88**, 110-115 (1994).
11. Huerta-Espino, J., Singh, R.P. & Reyna-Martinez, J. First detection of virulence to genes *Lr9* and *Lr25* conferring resistance to leaf rust of wheat caused by *Puccinia triticina* in Mexico. *Plant Disease* **92**, 311 (2008).
12. Brueggeman, R. *et al.* The barley stem rust-resistance gene *Rpg1* is a novel disease-resistance gene with homology to receptor kinases. *Proceedings of the National Academy of Sciences* **99**, 9328-9333 (2002).
13. Klymiuk, V. *et al.* Cloning of the wheat *Yr15* resistance gene sheds light on the plant tandem kinase-pseudokinase family. *Nature Communications* **9**, 3735 (2018).
14. Chen, S.S. *et al.* Wheat gene *Sr60* encodes a protein with two putative kinase domains that confers resistance to stem rust. *New Phytologist* **225**, 948-959 (2020).
15. Lu, P. *et al.* A rare gain of function mutation in a wheat tandem kinase confers resistance to powdery mildew. *Nature Communications* **11**, 680 (2020).
16. Gaurav, K. *et al.* Population genomic analysis of *Aegilops tauschii* identifies targets for bread wheat improvement. *Nature Biotechnology* **40**, 422-431 (2022).
17. Yu, G. *et al. Aegilops sharonensis* genome-assisted identification of stem rust resistance gene *Sr62*. *Nature Communications* **13**, 1607 (2022).
18. Arora, S. *et al.* A wheat kinase and immune receptor form the host-specificity barrier against the blast fungus. *bioRxiv*, https://doi.org/10.1101/2022.01.27.477927 (2022).
19. Liu, J.X., Jambunathan, N. & McNellis, T. Transgenic expression of the von Willebrand A domain of the BONZAI1/COPINE1 protein triggers a lesion-mimic phenotype in *Arabidopsis*. *Planta* **221**, 85-94 (2005).
20. Whittaker, C.A. & Hynes, R.O. Distribution and evolution of von Willebrand/integrin A domains: widely dispersed domains with roles in cell adhesion and elsewhere. *Molecular Biology of the Cell* **13**, 3369-3387 (2002).
21. Yin, X. *et al.* Rice copine genes *OsBON1* and *OsBON3* function as suppressors of broad-spectrum disease resistance. *Plant Biotechnology Journal* **16**, 1476-1487 (2018).
22. Yang, S.H. & Hua, J. A haplotype-specific resistance gene regulated by *BONZAI1* mediates temperature-dependent growth control in *Arabidopsis*. *Plant Cell* **16**, 1060-1071 (2004).

23. Gupta, S.K., Charpe, A., Koul, S., Prabhu, K.V. & Haq, Q.M.R. Development and validation of molecular markers linked to an *Aegilops umbellulata*–derived leaf-rust-resistance gene, *Lr9*, for marker-assisted selection in bread wheat. *Genome* **48**, 823-830 (2005).
24. Kuraparthy, V., Sood, S., Guedira, G.-B. & Gill, B.S. Development of a PCR assay and marker-assisted transfer of leaf rust resistance gene *Lr58* into adapted winter wheats. *Euphytica* **180**, 227-234 (2011).
25. Friebe, B., Jiang, J., Tuleen, N. & Gill, B.S. Standard karyotype of *Triticum umbellulatum* and the characterization of derived chromosome addition and translocation lines in common wheat. *Theoretical and Applied Genetics* **90**, 150-156 (1995).
26. Huang, C., Yoshino-Koh, K. & Tesmer, J.J. A surface of the kinase domain critical for the allosteric activation of G protein-coupled receptor kinases. *Journal of Biological Chemistry* **284**, 17206-17215 (2009).
27. Bajaj, K., Chakrabarti, P. & Varadarajan, R. Mutagenesis-based definitions and probes of residue burial in proteins. *Proceedings of the National Academy of Sciences* **102**, 16221-16226 (2005).
28. Hanks, S.K., Quinn, A.M. & Hunter, T. The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science* **241**, 42-52 (1988).
29. Leslie, M.E., Lewis, M.W., Youn, J.-Y., Daniels, M.J. & Liljegren, S.J. The EVERSHED receptor-like kinase modulates floral organ shedding in *Arabidopsis*. *Development* **137**, 467-476 (2010).
30. International Wheat Genome Sequencing Consortium. Shifting the limits in wheat research and breeding through a fully annotated and anchored reference genome sequence. *Science* **361**, eaar7191 (2018).
31. Walkowiak, S. *et al.* Multiple wheat genomes reveal global variation in modern breeding. *Nature* **588**, 277-283 (2020).
32. Zhu, T. *et al.* Optical maps refine the bread wheat *Triticum aestivum* cv. Chinese Spring genome assembly. *Plant Journal* **107**, 303-314 (2021).
33. Sato, K. *et al.* Chromosome-scale genome assembly of the transformation-amenable common wheat cultivar 'Fielder'. *DNA Research* **28**, dsab008 (2021).
34. Athiyannan, N. *et al.* Long-read genome sequencing of bread wheat facilitates disease resistance gene cloning. *Nature Genetics* **54**, 227-231 (2022).
35. Aury, J.-M. *et al.* Long-read and chromosome-scale assembly of the hexaploid wheat genome achieves high resolution for research and breeding. *GigaScience* **11**, giac034 (2022).
36. Maccaferri, M. *et al.* Durum wheat genome highlights past domestication signatures and future improvement targets. *Nature Genetics* **51**, 885-895 (2019).
37. Zhou, Y. *et al.* Introgressing the *Aegilops tauschii* genome into wheat as a basis for cereal improvement. *Nature Plants* **7**, 774-786 (2021).
38. Li, L.F. *et al.* Genome sequences of five Sitopsis species of *Aegilops* and the origin of polyploid wheat B subgenome. *Molecular Plant* **15**, 488-503 (2022).
39. van der Hoorn, R.A. & Kamoun, S. From Guard to Decoy: a new model for perception of plant pathogen effectors. *Plant Cell* **20**, 2009-2017 (2008).
40. Wenger, A.M. *et al.* Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature Biotechnology* **37**, 1155-1162 (2019).
41. Edae, E.A., Olivera, P.D., Jin, Y., Poland, J.A. & Rouse, M.N. Genotype-by-sequencing facilitates genetic mapping of a stem rust resistance locus in *Aegilops umbellulata*, a wild relative of cultivated wheat. *BMC Genomics* **17**, 1039 (2016).
42. Ortelli, S., Winzeler, M., Winzeler, H. & Nösberger, J. Leaf rust resistance gene *Lr9* and winter wheat yield reduction: II. Leaf gas exchange and root activity. *Crop Science* **36**, 1595-1601 (1996).
43. Yu, G. *et al.* The wheat stem rust resistance gene *Sr43* encodes an unusual protein kinase. *submitted* (2022).

44. Fu, D. *et al.* A kinase-START gene confers temperature-dependent resistance to wheat stripe rust. *Science* **323**, 1357-1360 (2009).
45. Faris, J.D. *et al.* A unique wheat disease resistance-like gene governs effector-triggered susceptibility to necrotrophic pathogens. *Proceedings of the National Academy of Sciences* **107**, 13544-13549 (2010).
46. Zhang, Z. *et al.* A protein kinase–major sperm protein gene hijacked by a necrotrophic fungal pathogen triggers disease susceptibility in wheat. *Plant Journal* **106**, 720-732 (2021).
47. Sánchez-Martín, J. *et al.* Wheat *Pm4* resistance to powdery mildew is controlled by alternative splice variants encoding chimeric proteins. *Nature Plants* **7**, 327-341 (2021).
48. Wang, Y. *et al.* Orthologous receptor kinases quantitatively affect the host status of barley to leaf rust fungi. *Nature Plants* **5**, 1129-1135 (2019).
49. Watson, A. *et al.* Speed breeding is a powerful tool to accelerate crop research and breeding. *Nature Plants* **4**, 23-29 (2018).
50. Wicker, T., Matthews, D.E. & Keller, B. TREP: a database for Triticeae repetitive elements. *Trends in Plant Science* **7**, 561-562 (2002).
51. Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One* **11**, e0163962 (2016).
52. Bushnell, B. BBMap: a fast, accurate, splice-aware aligner. (Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States), 2014).
53. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
54. Sánchez-Martín, J. *et al.* Rapid gene isolation in barley and wheat by mutant chromosome sequencing. *Genome Biology* **17**, 221 (2016).
55. Chen, C. *et al.* TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant* **13**, 1194-1202 (2020).
56. Vrána, J. *et al.* Flow sorting of mitotic chromosomes in common wheat (*Triticum aestivum* L.). *Genetics* **156**, 2033-2041 (2000).
57. Doležel, J., Binarová, P. & Lucretti, S. Analysis of nuclear DNA content in plant cells by flow cytometry. *Biologia Plantarum* **31**, 113-120 (1989).
58. Giorgi, D. *et al.* FISHIS: fluorescence in situ hybridization in suspension and chromosome flow sorting made easy. *PLoS One* **8**, e57994 (2013).
59. Kubaláková, M., Macas, J. & Doležel, J. Mapping of repeated DNA sequences in plant chromosomes by PRINS and C-PRINS. *Theoretical and Applied Genetics* **94**, 758-763 (1997).
60. Chapman, J.A. *et al.* Meraculous: *de novo* genome assembly with short paired-end reads. *PLoS One* **6**, e23501 (2011).
61. Green, M.R. & Sambrook, J. Nested polymerase chain reaction (PCR). *Cold Spring Harbor Protocols* **2019**, pdb. prot095182 (2019).
62. Yuan, C. *et al.* A high throughput barley stripe mosaic virus vector for virus induced gene silencing in monocots and dicots. *PLoS One* **6**, e26468 (2011).
63. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
64. Koo, D.-H., Liu, W., Friebe, B. & Gill, B.S. Homoeologous recombination in the presence of *Ph1* gene in wheat. *Chromosoma* **126**, 531-540 (2017).
65. Koo, D.-H. *et al.* Extrachromosomal circular DNA-based amplification and transmission of herbicide resistance in crop weed *Amaranthus palmeri*. *Proceedings of the National Academy of Sciences* **115**, 3332-3337 (2018).
66. Ma, S. *et al.* WheatOmics: A platform combining multiple omics data to accelerate functional genomics studies in wheat. *Molecular Plant* **14**, 1965-1968 (2021).
67. Driguez, P. *et al.* LeafGo: Leaf to Genome, a quick workflow to produce high-quality de novo plant genomes using long-read sequencing technology. *Genome Biology* **22**, 256 (2021).

68. Cheng, H., Concepcion, G.T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods* **18**, 170-175 (2021).
69. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072-1075 (2013).
70. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100 (2018).
71. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
72. Pedersen, B.S. & Quinlan, A.R. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* **34**, 867-868 (2018).
73. Marcais, G. & Kingsford, C. Jellyfish: A fast k-mer counter. *Tutorialis e Manuais* **1**, 1-8 (2012).

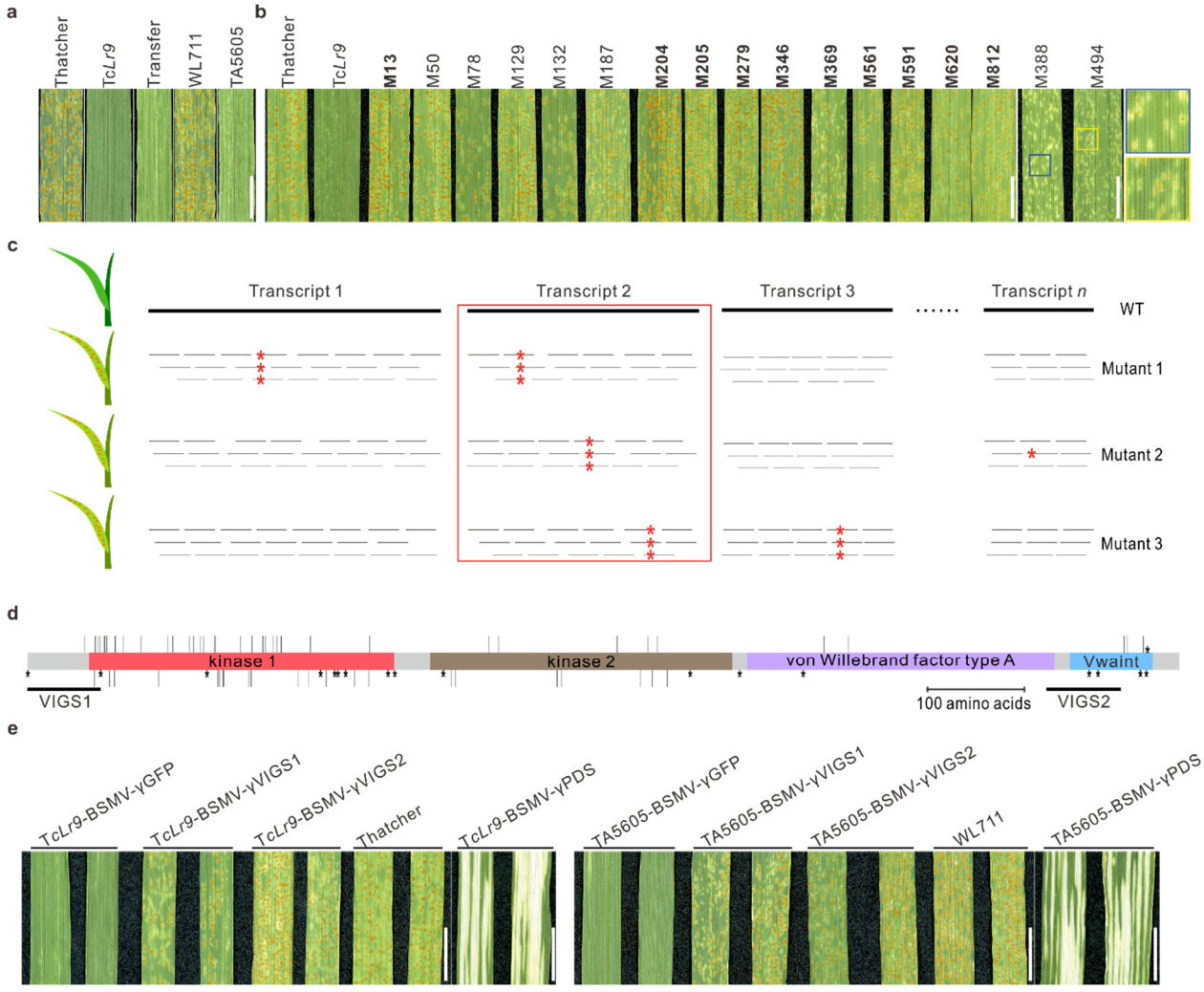


Fig. 1. *Lr9* and *Lr58* encode an unusual tandem kinase fusion protein. a, Phenotypic effect of *Lr9* and *Lr58* against *Puccinia triticina*. Thatcher*Lr9* (Tc*Lr9*) and Transfer represent *Lr9*-carrying bread wheat lines, Thatcher is the susceptible control. TA5605 is a *Lr58*-carring bread wheat backcross line in the genetic backgound of the susceptible parent WL711. Scale bar = 1 cm. **b,** Representative images showing the phenotype of susceptible Thatcher*Lr9* mutants. Mutants in bold were used for MutIsoSeq. Blue and yellow boxes show magnified sections of the partially susceptible mutants M388 and M494, respectively. Scale bar = 1 cm. **c,** Schematic represenation of MutIsoSeq. Isoform sequencing is used to obtain full-length transcript sequences of a wild-type (WT) parent. RNA sequencing reads generated from EMS mutants are mapped against the wild-type full-length transcripts and single nucleotide polymorphisms (SNPs) are called. Transcripts showing an EMS-type point mutation (red asterisks) in all the mutants are considered as candidates (transcript 2 in this example). **d,** The prediced WTK6-vWA protein structure, Salmon red: kinase 1, light brown: kinase 2, purple: vWA domain, light blue: Vwaint domain. Vertical bars represent 67 EMS-induced amino acid substistutions, upper bars = surface-localized residues, lower bars = internal residues, black asterisks = premature stop codon and start codon mutations. VIGS1 and VIGS2 indicate the positions of the two virus-induced gene silencing (VIGS) constructs. **e,** Representative images showing the results of the VIGS experiment using barley stripe mosaic virus (BSMV). BSMV-γVIGS1 and BSMV-γVIGS2 = two silencing constructs targeting *WTK6-vWA,* BSMV-γGFP = control with a GFP silencing construct, BSMV-γPDS = silencing construct targeting the phytoene desaturase genes. Scale bar = 1 cm.

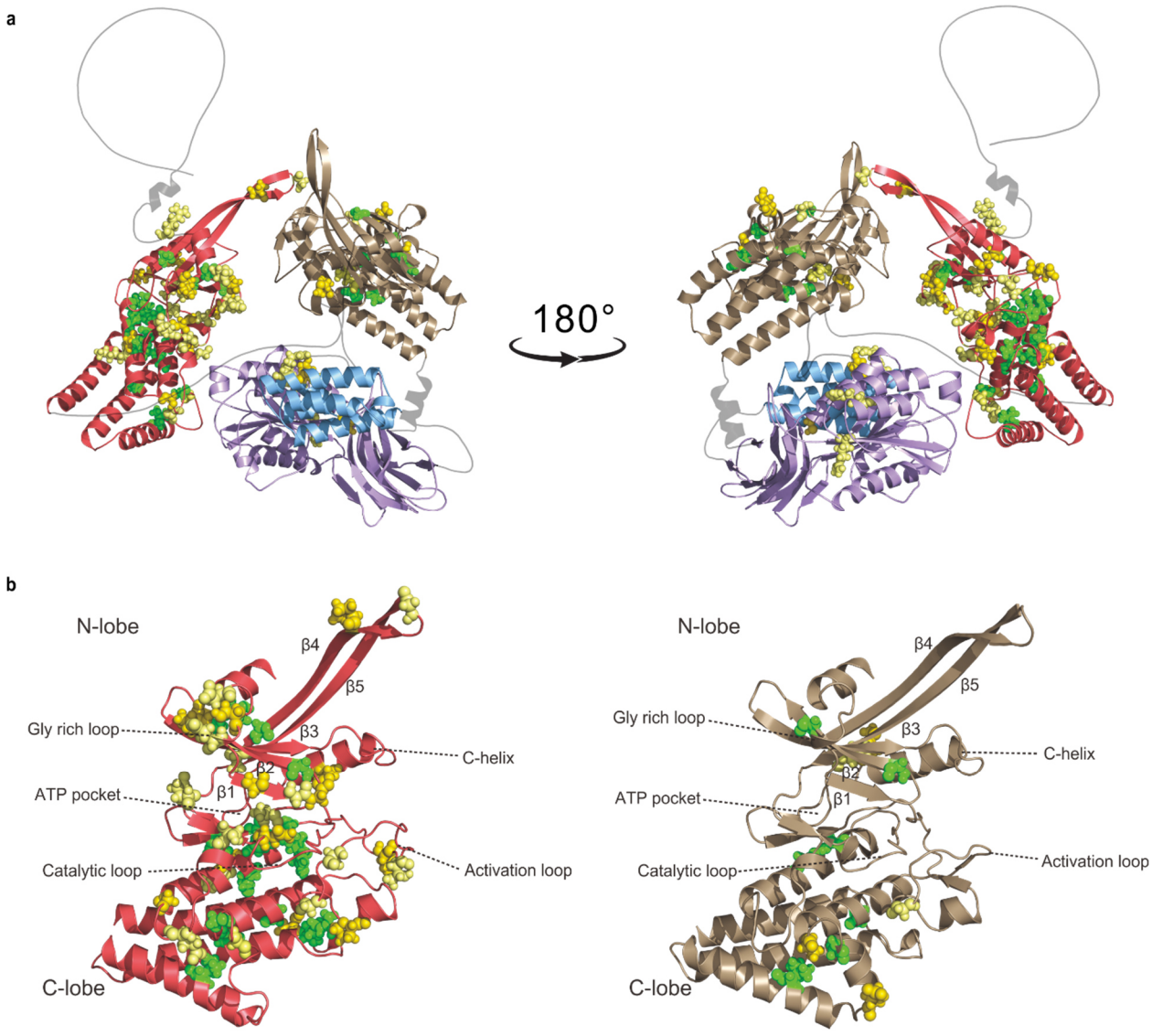


Fig. 2 Protein structure prediction and mutation analysis of WTK6-vWA. a, The protein structure of WTK6-vWA as predicted by AlphaFold v2.0. **b,** Kinase 1 and kinase 2 as predicted by AlphaFold v2.0, indicating the basic secondary functional features. The colors of the domains are identical to Fig. 1d. Salmon red: kinase 1, light brown: kinase 2, purple: vWA domain, light blue: Vwaint domain. Green spheres indicate internal amino acid substitutions resulting from the EMS mutagenesis. Light and dark yellow spheres represent surface-localized amino acid substitutions.

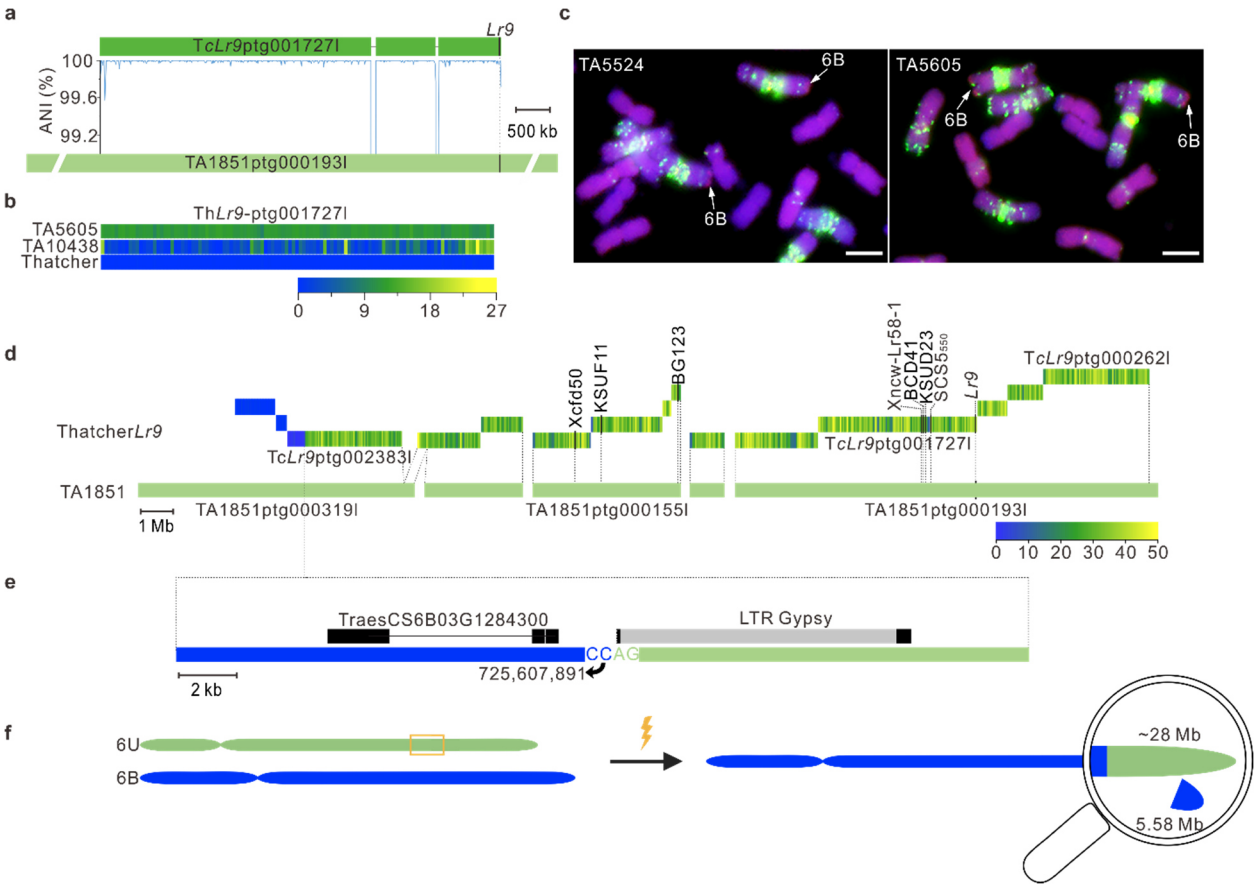


Fig. 3. Assembly of the *Lr9* translocation. a, *Lr9*-carrying contigs identified in the Thatcher*Lr9* (Tc*Lr9*) and *Ae. umbellulata* (TA1851) assemblies. The chart in the middle shows the average nucleotide identity (ANI) calculated in non-overlapping sliding windows of 10 kb. Lines in the Tc*Lr9* contig indicate irradiation-induced deletions. **b,** Heat map showing Illumina read coverage (in non-overlapping sliding windows of 50 kb) from TA5605, Thatcher, and *Ae. triuncialis* accession TA10438 mapped against the Thatcher*Lr9* contig. **c,** Genomic *in situ* hybridization (GISH) patterns of partial mitotic metaphase spreads of TA5524 (Transfer) and TA5605 using *Ae. umbellulata* genomic DNA (visualized in red). The *Ae. umbellulata* translocation on chromosome arm 6BL is indicated by arrows. GAA repeats (green signals) were used to identify the homologous chromosome pairs. Scale bars = 5 µm. **d,** Thirteen Thatcher*Lr9* (Tc*Lr9*) contigs were identified that showed increased coverage of *Ae. umbellulata* derived *k*-mers. The heat map in the Thatcher*Lr9* contigs represents the *k*-mer coverage in sliding windows of 50 kb. The corresponding TA1851 contigs are indicated below the Thatcher*Lr9* contigs (ordered and oriented according to an *Ae. umbellulata* genetic map). The positions of published *Lr9* and *Lr58* genetic markers are indicated. **e,** Schematic representation showing the translocation breakpoint between *Ae. umbellulata* (represented in green) and bread wheat chromsome 6BL (represented in blue). The black bars in the Gypsy retroelement correspond to the long terminal repeats (LTR). The nucleotide position 725,607,891 indicates the corresponding position in the Chinese Spring reference assembly (IWGSC v2.1). **f,** Schematic representation of the non-homologous chromosome translocation between *Ae. umbellulata* chromosome 6U and bread wheat chromosome 6BL. The irriadiation led to chromosome breaks and random chromosome fusions. During the translocation, a 5.58 Mb bread wheat chromosome segment was lost.

Table 1. Number of transcripts with *n* EMS-induced point mutations in the ten sequenced Thatcher*Lr9* and TA5605 mutants.

	Thatcher*Lr9*	TA5605
Number of transcripts mutated in 0 lines	67,974	124,101
Number of transcripts mutated in 1 line	9,073	18,949
Number of transcripts mutated in 2 lines	765	2,137
Number of transcripts mutated in 3 lines	68	214
Number of transcripts mutated in 4 lines	2	22
Number of transcripts mutated in 5 lines	1	3
Number of transcripts mutated in 6 lines	0	0
Number of transcripts mutated in 7 lines	0	0
Number of transcripts mutated in 8 lines	0	1*
Number of transcripts mutated in 9 lines	0	0
Number of transcripts mutated in 10 lines	1	0

*Two *Lr58* mutants carried a mutation at the same position. These two mutants were missed in the original run because the parameter '-z' (number of mutant lines that are allowed to have a single nucleotide variant in the same position) was set as 1. Point mutations were found in the remaining two mutants in another run with the parameter '-z' set to 2.

444 **Table 2.** Statistics of the Thatcher*Lr9* and TA1851 assemblies.

Genomic feature	Thatcher*Lr9*	TA1851
Length of HiFi assembly	14.49 Gb	4.25 Gb
Number of contigs	7,686	1,379
Length of contig N50	4.52 Mb	11.15 Mb
Length of contig N90	1.2 Mb	3.18 Mb
Largest contig	30.59 Mb	57.09 Mb
GC content	46.13%	47.11%

445

Methods

Plant material

Bread wheat accessions Transfer (TA5524), WL711, TA5605, *Ae. umbellulata* accession TA1851 and *Ae. triuncialis* accession TA10438 were obtained from the Wheat Genetics Resource Center (WGRC). Thatcher*Lr9* (Transfer/6*Thatcher) is a near-isogenic line carrying *Lr9* from Transfer in the genetic background of the susceptible wheat line Thatcher. Thatcher*Lr9* and TA5605 were crossed with leaf rust susceptible wheat cultivar Avocet S to develop segregating F2 populations for *Lr9* (136 F2 plants) and *Lr58* (128 F2 plants), respectively. A F2 populations derived from a cross between Thatcher*Lr9* and TA5605 (133 F2 plants) was also generated.

Leaf rust inoculations

Puccinia triticina isolates B9414, 93012, 94015, 95012, 96007, and 96237 were propagated on seedlings of the susceptible wheat cultivar Thatcher. Freshly harvested urediniospores were used for inoculation experiments. Inoculations were performed using a high-pressure air sprayer or a settling tower. For the spray inoculation, urediniospores were suspended in FC-43 oil (3M Fluorinert FC-43) and sprayed onto plants using a glass sprayer connected to a high-pressure air pipe. For the settling tower inoculation, plants were grown for 12 days. Then, the second leaf was removed and the first leaf was pinned to the soil with the adaxial side facing upwards. A mixture of 10 mg urediniospores and 300 mg lycopodium powder (Sigma-Aldrich, 19108) was blown over the plants in a settling tower[48]. Inoculated plants were placed in an inoculation box equipped with a humidifier overnight and transferred to a walk-in Conviron growth room (16/8 hours day/night, 18/16°C). Symptoms were evaluated 12 days after inoculation. Infected leaves were scanned using an Epson Perfection V600 Photo scanner.

EMS mutagenesis and mutant screening

Grains of wheat lines Thatcher*Lr9* and TA5605 were evenly divided and soaked in two one-liter flasks with 200 ml water at 4°C for 16 h. Grains were washed three times in distilled water. Then, 200 ml of 0.45% and 0.55% EMS solutions (Sigma-Aldrich, M0880) were added and grains were incubated on a shaker at 80 rpm at room temperature for 16 hours. Next, grains were washed three times for 45 minutes each, and then transferred into mesh bags and put under running tap water for 30 minutes. After the washing, grains were put into 55×28 cm black plastic trays with blotting paper at the bottom, which were later covered with plastic lids and kept at 4°C for 36 hours. The treated grains were planted into 18-well trays filled with Stender soil, 12 to 14 grains per well. Plants were grown in a greenhouse under speed-breeding conditions (22/2 hours day/night, 21/18°C)[49]. We performed three EMS experiments for TA5605 (with around 2,000, 4,000, and 5,000 grains per experiment), and one for Thatcher*Lr9* (~2,000 grains). Spikes from the first and second batches of TA5605 M1 plants were bulk harvested. Single spikes were harvested from the third batch of TA5605 M1 plants and from Thacher*Lr9* M1 plants. For the mutant screening, M2 grains from a single spike were planted in one well of 18-well or 24-well trays and grown in a greenhouse (22/2 hours day/night, 21/18°C) or growth room (16/8 hours day/night, 20°C) for three to four weeks. Batches of six trays were spray-inoculated with 30 mg fresh urediniospores of *Pt* isolate B9414 suspended in 20 ml FC-43 oil, and then kept in an inoculation box overnight. Plants were transferred and grown in a walk-in Conviron growth room for 10 to 14 days. Susceptible plants were transferred into single pots. In total, 919 and ~7,400 M2 families were screened for Thatcher*Lr9* and TA5605, respectively.

Iso-seq library preparation and sequencing

Total RNA from infected Thatcher*Lr9* and TA5605 seedling leaves (48, 72, 96, and 120 hours post inoculation with the avirulent *Pt* isolate B9414) was extracted using the Maxwell RSC Plant RNA Kit. To prepare the Iso-seq libraries, 300 ng of total RNA was used as input for complementary DNA (cDNA) synthesis. Each sample was first barcoded (Supplementary Table 6) and then subjected to cDNA amplification using 12 cycles. Purified cDNAs were pooled in equal molarity and then subjected to library

preparation using the SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, 100-938-900) following the Iso-Seq protocol. Libraries were prepared for sequencing by annealing primer v4 with the Sequel II Binding Kit 2.1 and the Internal Control Kit 1.0 (Pacific Biosciences, 101-843-000). Two SMRT Cells 8M (Pacific Biosciences, 101-389-001), one for Thatcher*Lr9* and one for TA5605, were sequenced on the PacBio Sequel II system using the Sequencing Kit 2.0 (Pacific Biosciences, 101-820-200).

RNA-seq of susceptible mutants

Total RNA from the *Pt*-inoculated fourth to sixth leaves of ten susceptible Thacher*Lr9* M2 mutants and ten TA5605 M3 mutants was extracted using the Maxwell RSC Plant RNA Kit. RNA-seq was performed as a service at Novogene. In brief, mRNA samples were enriched using oligo(dT) beads, then used to prepare strand-specific cDNA libraries. Sequencing was done on Illumina Novaseq 6000, which produced an average of 233.2 and 262.7 million paired-end reads (2×150 bp) for Thacher*Lr9* and TA5605 mutants, respectively.

MutIsoSeq

The principle of MutIsoSeq is to use full-length transcripts produced with PacBio Iso-seq as a reference, map RNA-seq reads of mutants to the reference, and identify transcripts that carry multiple independent EMS-type point mutations (Fig. 1c). The workflow of MutIsoSeq is illustrated in Supplementary Figure 1. The raw data from PacBio Sequel II were processed following the Iso-seq pipeline (v.3.; https://github.com/PacificBiosciences/IsoSeq) for the generation of CCS reads and full-length, non-concatemer (FLNC) reads. FLNC reads were clustered and merged into 'clustered.hq.fasta.gz' files containing 300,066 and 283,330 high-quality transcripts for Thatcher*Lr9* and TA5605, respectively. Transposable elements were masked by Repeatmasker (v4.0.7) (http://www.repeatmasker.org) using the TREP database[50] (v.2019) as external library, which resulted in 6.29% and 6.08 % masked sequences,

519 respectively. Then, SeqKit v2.2.0[51] was used to sort all the transcripts from longest to shortest and the fasta
520 files generated during this step were used as references. We used BBMap (v.38.96)[52] (parameter:
521 ambiguous=best subfilter=1 trimreaddescriptions=t nodisk) to map the Illumina reads from the susceptible
522 mutants to the references. To exclude non-specific alignments as much as possible, we used the parameter
523 'subfilter=1', which removes alignments with more than one substitution. Iso-seq captures different
524 isoforms originating from the same gene. If a gene had multiple isoforms, Illumina reads would be randomly
525 assigned to different isoforms by default, which would reduce the average coverage per isoform. We thus
526 used the parameter 'ambiguous=best', which assigns reads to the first best site. Because transcripts were
527 ordered by size, this step resulted in the mapping of most reads to the longest isoform of a gene. The
528 resulting sam files were sorted and indexed by samtools (v1.6)[53]. We followed the MutChromSeq pipeline
529 for the remaining steps (https://github.com/steuernb/MutChromSeq)[54]. Briefly, sorted sam files were
530 converted to the mpileup format using samtools (v1.6)[53]. Individual pileup files were subsequently
531 converted to xml format using Pileup2XML.jar (parameter, -a 0.01 -c 5). Reports were given by
532 MutChromSeq.jar using the parameters -n 10 -c 10 -a 0.01 -z 2 (Lr9-M13 and TA5605-Spt1 were used with
533 the parameter "-w" because we did not use the wild-type Thatcher*Lr9* and TA5605 RNA-seq data)
534 (Supplementary Fig. 2). For both *Lr9* and *Lr58*, only one candidate transcript was identified that had EMS-
535 type G/C-to-A/T transitions in all 20 sequenced mutants. We performed a BLAST search with the identified
536 candidate transcript against the high-quality transcripts using TBtools (v1.09873)[55], which identified 72 and
537 61 transcripts for *Lr9* and *Lr58,* respectively (Supplementary Fig. 3a, b). Only transcripts that encoded a
538 full-length protein spanning all the detected mutants were retained, resulting in 16 transcripts each for *Lr9*
539 and *Lr58*. Twelve and 11 of the retained *Lr9* and *Lr58* transcripts encoded the same 1,167 amino acid
540 protein, respectively, and they mainly differed in the length of the 5' and 3' untranslated regions
541 (Supplementary Fig. 3c). The transcripts encoding this protein version accounted for the majority of the
542 total transcripts (68.15% and 73.26% coverage, which were calculated using the coverage number of
543 transcripts, representing the number of molecules sequenced in the PacBio SMRT cells) (Supplementary
544 Fig. 3a, b). We also found four *Lr9* transcripts (representing 2.54% coverage) and four *Lr58* transcripts

(2.18% coverage) that encoded full-length proteins lacking 12 to 24 amino acids at the N-terminus, and one *Lr58* transcript (0.41% coverage) encoding a full-length protein with 28 extra amino acids. The transcripts encoding the 1,167 amino acid protein were considered as the main protein version because of the high transcript abundance.

Flow cytometric chromosome analysis, sorting and sequencing of chromosome 2B from TA5605

Suspensions of mitotic metaphase chromosomes were prepared from root tips of bread wheat line TA5605 according to Vrána et al[56]. The procedure involved cell cycle synchronization in root tip meristems of hydroponically grown young seedlings and accumulation of dividing cells at metaphase using hydroxyurea and amiprophos-methyl, respectively. The roots were then fixed mildly by formaldehyde and intact chromosomes were released into 600 µl ice-cold LB01 buffer[57] by mechanical homogenization of 100 root tips. GAA repeats in chromosomes were labelled by fluorescence *in situ* hybridization in suspension (FISHIS) using 5'-FITC-GAA7-FITC-3' oligonucleotide probes (Sigma, Saint Louis, USA)[58] and chromosomal DNA was stained by DAPI (4′,6-diamidino 2-phenylindole) at 2 µg/ml. Chromosome analysis and sorting was done using a FACSAria II SORP flow cytometer and sorter (Becton Dickinson Immunocytometry Systems, San José, USA). Dot plots FITC vs. DAPI fluorescence (bivariate flow karyotypes) were acquired and chromosome 2B was sorted at rates of 20 to 40 particles per second. Around 42,000 copies of chromosome 2B were sorted into a PCR tube containing 40 µl sterile deionized water. In order to estimate chromosome composition of flow-sorted fractions 1,500 to 2,000 chromosomes were sorted into 10 µl drops of PRINS buffer containing 2.5% sucrose[59] onto a microscopic slide. Air-dried chromosomes were labelled by FISH using the probes for the pSc119.2 repeat, Afa family repeat and 45S rDNA. Chromosome identity and contamination by other chromosomes were determined by microscopic observation of at least 100 chromosomes following the karyotype[59]. Chromosome 2B was identified according to weak telomeric and strong interstitial pSc119.2 signals observed on the short and long arms, respectively (Supplementary Fig. 4). The purity in the sorted chromosome 2B was 94.7%. Chromosome

fractions were treated with proteinase K and DNA purified. Twenty nanograms (ng) of DNA was fragmented in 20 µl solution using Bioruptor Plus (Diagenode, Denville, USA) five times for 30 s at HIGH setting. Sequencing libraries were prepared using the NEBNext Ultra™ II DNA Library Prep Kit. DNA libraries were sequenced on Illumina NovaSeq6000, which produced about 82 Gb 2×150 bp paired-end reads for TA5605 that were *de novo* assembled using Meraculous (v2.2.6)[60] using default parameters, resulting in a total assembly size of ~727 Mb (101,896 contigs with N50 of 11.6 kb).

Sanger sequencing of EMS mutants

We performed a BLAST search with the *WTK6-vWA* CDS against the genome sequence of bread wheat cultivars[31,33] and found full-length matches on chromosome arms 2AL and 2BL of wheat cultivars ArinaLrFor and Fielder. The corresponding genomic sequences were downloaded and aligned to *WTK6-vWA* full-length transcript/46588 of Thatcher*Lr9* using Geneious Prime (Version 2020.2.4). Primers were designed from the UTRs and the CDS of *WTK6-vWA* and intron sequence of ArinaLrFor and Fielder (Supplementary Table 7). PCR amplifications were done in 20 µl reactions containing 1× GoTaq Green Master Mix (Promega, M7122), ~100 ng genomic DNA, and 100 nM primers. Primer pairs K1-F/K1-R, K2-F2/K2-R1, and vWA-F2/vWA-R1 were used to amplify 1,949-bp, 2,313-bp, and 1,050-bp fragments from the DNA samples of mutants. Amplicons were Sanger-sequenced using primers K1-F/K1-MR/K1-R, K2-F2/K2-MF1/K2-MR/K2-R1, and vWA-F2/vWA-R1, respectively. For the mutants from the last batch of mutant screening (TA5605-Spt124 to Spt186), primer pair K1-F/K1-R was used to amplify the 1,949-bp fragment, and nested PCR[61] primer pairs K2-F1/vWA-R1 (for first round PCR) and K2-F2/vWA-R2 (for second round PCR) were used to amplify a second 3,740-bp fragment from the mutants, which were subsequently Sanger-sequenced using primers K1-exon1-R/K1-MF1/K1-MF2/K1R and K2-F2/K2-F3/K2-F4/K2-R1/vWA-F1/vWA-R2. Sanger sequencing files were aligned to the *WTK6-vWA* CDS and genomic sequence to find polymorphisms using Geneious Prime (Version 2020.2.4). Primer sequences are provided in Supplementary Table 7.

595

596 Virus-induced gene silencing

597 To develop specific virus-induced gene silencing (VIGS) probes, we first performed a BLAST search with
598 the *WTK6-vWA* CDS against various wheat genomes[31]. *WTK6-vWA* sequence stretches with little homology
599 in the wheat assemblies were selected as targets (Fig. 1d). Primers for the two targets and the phytoene
600 desaturase gene (PDS)[62] were used to amplify a 223-bp (VIGS1), 226-bp (VIGS2), and 400-bp (PDS)
601 fragment from the TA5605 cDNA, respectively (Supplementary Table 7). Fragments were cloned into the
602 BSMV-γ (BSMV, barley stripe mosaic virus) vector[62], resulting in constructs BSMV-γVIGS1, BMSV-
603 γVIGS2 and BMSV-γPDS. These constructs together with BSMV-γ (empty vector), BSMV-γGFP[34],
604 BSMV-α and BSMV-β were transformed into *Agrobacterium tumefaciens* strain GV3101. Agrobacteria
605 were cultured overnight at 28°C in lysogeny broth with appropriate antibiotics. Cells were collected by
606 centrifugation, then re-suspended and adjusted to OD_{600} = 0.7. Equal volumes of BSMV-α and BSMV-β
607 were mixed with BSMV-γVIGS1, BSMV-γVIGS2, BSMV-γGFP, BSMV-γ, and BSMV-γPDS,
608 respectively. After 3 h incubation at 28°C, cultures were infiltrated into *Nicotiana benthamiana* leaves.
609 Infiltrated leaves were harvested 5 days after infiltration and homogenized with PBS buffer (Gibco,
610 10010023) containing 1% celite. The buffer containing viral particles was rub inoculated on seedlings of
611 Thatcher*Lr9* and TA5605 at two-leaf stage. After two weeks, when viral symptom were clearly visible,
612 plants were spray-inoculated with *P. triticina* isolate B9414 as described above. Disease symptoms were
613 evaluated and recorded 12 days after inoculation.

614

615 Protein structure prediction

616 To predict the 3D structure of WTK6-vWA, we utilized the open source code of AlphaFold v2.0[63]. The
617 input was the amino-acid sequence of WTK6-vWA. The output consisted of five PDB format text files
618 containing the predicted structure exactly as predicted by the model ("unrelaxed models"), which were later

refined by an Amber relaxation procedure ("relaxed models") to provide the relaxed predicted structure reordering by model confidence based on the Local Distance Difference test (lDDT) score ("ranked models"). We used one of the ranked models with a reported lDDT score of 0.7507 of very good accuracy (lDDT >0.6 being considered as reasonable models and lDDT >0.8 as great models, accuracy-wise). The positions of 67 non-redundant amino acid substitutions were visualized as spheres using PyMOL (v.2.3.0). Substituted residues that did not exhibit any visual surface on the 3D structure model were characterized as internal residues.

Cytogenetic analysis

The procedures for mitotic chromosomes preparation, FISH and GISH were adapted from Koo et al[64,65]. A 3,740-bp fragment was amplified from Thacher*Lr9*, TA5605 and TA10438 using nested PCR[61] primer pairs K2-F1/vWA-R1 (for first round PCR) and K2-F2/vWA-R2 (for second round PCR) and Phusion High-Fidelity DNA Polymerase (New England Biolabs, M0530), which was subsequently Sanger-sequenced using primers K2-F2/K2-MR/K2-MF2/K2-R2/K2R1/LastIntron-F/vWA-F3/vWA-R2. The 3,740-bp fragment from Thacher*Lr9* was cloned into pJET1.2/blunt (Thermo Fisher, K1231), and used in FISH as a probe. Probe DNA was labeled with either digoxigenin-11-dUTP or biotin-16-dUTP following the manufacturer's instructions (Roche, CA). Unlabeled total genomic wheat DNA was used as blocker in GISH experiments. After post-hybridization washes, the probe hybridization signals were detected with Alexafluor 488 streptavidin (Invitrogen, MA) for biotin-labeled probes, and rhodamine-conjugated anti-digoxigenin (Roche) for dig-labeled probes. Images were captured using a Zeiss Axioplan 2 microscope (Carl Zeiss Microscopy LLC, Thornwood, NY) mounted with a cooled CCD camera CoolSNAP HQ2 (Photometrics, Tucson, AZ) and operated with AxioVision 4.8 software. The final contrast of the images were processed using Adobe Photoshop CS5. Primer sequences are provided in Supplementary Table 7.

643 Identification of *WTK6-vWA* homologs

644 *WTK6-vWA* homologous genes were retrieved by BLAST search against whole-genome assemblies of 12
645 bread wheat cultivars, durum wheat, wild emmer wheat, and six *Aegilops* species at WheatOmics 1.0
646 (http://202.194.139.32/blast/blast.html)[31-38,66]. Corresponding sequences were downloaded. Non-annotated
647 sequences were manually annotated by MUSCLE alignment in Geneious Prime (Version 2020.2.4).
648 Translated protein sequences were used as input in the 'NCBI Conserved Domain Search'
649 (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) for protein domain annotation.

650

651 KASP assay

652 A 5 µl reaction containing 2.5 µl of KASP Master Mix (Low ROX KBS-1016-016), 0.07 µl of assay mix
653 (KASP-Lr9-F/KASP-TA10438-F/KASP-R; Supplementary Table 7) and 2.5 µl (50 ng) of DNA was used
654 for the *Lr9* KASP assay. PCR cycling was performed using an ABI QuantStudio 6 Flex Real-Time PCR
655 machine as follows: pre-read at 30°C for 60 s; hold stage at 94°C for 15 min, followed by ten touchdown
656 cycles (94°C for 20 s; touchdown at 61°C, decreasing by 0.6°C per cycle for 60 s), followed by 29 additional
657 cycles (94°C for 20 s; 55°C for 60 s). Plates were read at 30°C for endpoint fluorescent measurement.

658

659 HiFi library preparation, sequencing and assembly of Thatcher*Lr9* and TA1851

660 High molecular weight (HMW) DNA was extracted from young seedlings of Thatcher*Lr9* and TA1851
661 using a modified Qiagen Genomic DNA extraction protocol
662 (https://dx.doi.org/10.17504/protocols.io.bafmibk6)[67]. DNA was sheared to the appropriate size range (15–
663 20 kb) using Megaruptor 3 (Diagenode, Denville, USA) for the construction of PacBio HiFi sequencing
664 libraries. Library preparation was done with the Express Template prep kit 2.0 (100-938-900 + Enzyme
665 Clean up 2.0 (101-932-600)), and size selected with a PippinHT System (Sage Science HTP0001). Final

SMRTbell QC were assessed with Qubit dsDNA High Sensitivity (ThermoFisher Scientific Q33230) and FEMTO Pulse (Agilent Technologies, Inc. P-0003-0817). Sequencing was performed on PacBio Sequel II/IIe systems in CCS mode. The 169.0 Gb of Thatcher*Lr9* PacBio HiFi reads and 130.9 Gb of TA1851 PacBio HiFi reads were assembled using hifiasm (v.0.16.1)[68] with default parameters. The genome assemblies were evaluated using QUAST 5.0.2[69] and BBMap (v.38.96)[52].

Re-sequencing of TA5605, Thatcher and TA10438

Genomic DNA was extracted from leaf samples of bread wheat accessions TA5605, Thatcher and *Ae. triuncialis* accession TA10438 using a CTAB DNA extraction method. Library preparation and sequencing were done as a service by Novogene. In brief, sequencing libraries were generated using the NEBNext® Ultra II DNA Library Prep Kit and sequenced on an Illumina NovaSeq 6000 system. Illumina sequencing resulted in about 564.3 Gb, 542.5 Gb and 561.8 Gb of 2×150 bp paired-end reads for TA5605, Thatcher and TA10438, respectively, corresponding to about 36-fold, 35-fold, and 53-fold genome coverage for TA5605, Thatcher and TA10438, respectively.

Assembly of the *Lr9* translocation

To calculate the average nucleotide identity between contigs Tc*Lr9*ptg001727l (Tatcher*Lr9* contig carrying *Lr9*) and TA1851ptg000193l (TA1851 contig carrying *Lr9*), we aligned them using Minimap2 (v.2.21) (parameter: '-ax asm5')[70]. InDels and single nucleotide polymorphisms were manually evaluated in non-overlapping sliding windows of 10 kb along the two contig. The assemblies of Thatcher*Lr9* and TA1851 were merged and used as a reference for the Illumina read mapping of TA5605, Thatcher, and TA10438. Illumina reads of TA5605, Thatcher, and TA10438 were aligned against the combined PacBio assembly using BBMap (v.38.96) (parameter: ambiguous=random trimreaddescriptions=t pairedonly=t perfectmode=t nodisk). From the output sam file, the read coverage for the contig Tc*Lr9*ptg001727l has

been calculated in non-overlapping sliding windows of 50 kb using bedtools (v2.30.0)[71] and mosdepth (v. 0.3.3)[72]. The heatmap representing the read coverage was generated with TBtools (v1.09873)[55]. We used a *k*-mer-based approach to reconstruct the *Lr9* translocation. Briefly, we counted the canonical 51-mers from the 130.9 Gb TA1851 raw HiFi reads using Jellyfish (v. 2.2.10)[73], then filtered the 51-mers with at least four occurrences. The dump file containing the 51-mers was converted into fasta format using AWK, which generated 2,726,203,170 *k*-mers. Then, the 51-mers were mapped against the Thatcher*Lr9* assembly using BWA mem (v.0.7.17) allowing only perfect matches (parameter: -k 51 -T 51), generating a bam file of *k*-mer mapping, which was imported to Geneious Prime (Version 2022.1.1) and sorted by mean coverage from highest to lowest (Supplementary Table 4). Thirteen contigs with average mapping coverage above 20 were considered as originating from *Ae. umbellulata* (Supplementary Table 4). We performed a BLAST search using non-repetitive sequences of 13 Thatcher*Lr9* contigs against the TA1851 assembly and identified five corresponding contigs in the *Ae. umbellulata* genome. To validate, order and orient the five selected TA1851 contigs, we anchored 3,009 markers (SNP tags) of the *Ae. umbellulata* linkage map[41] to the TA1851 assembly using BBMap (v.38.96) (parameter: ambiguous=random trimreaddescriptions=t subfilter=1 nodisk)[52] (Fig. 3d, Supplementary Table 5). *k*-mer mapping revealed an introgression breakpoint in contig Tc*Lr9*ptg002383l. To further investigate the breakpoint, we extracted 10 kb sequences upstream and downstream of the breakpoint from TA1851ptg000319l, Tc*Lr9*ptg002383l, and the long arm of CS chromosome 6B. Dot plot graphs were generated for 20 kb sequences from TA1851 and CS against Thatcher*Lr9* using Geneious Prime (Version 2020.2.4), which revealed the same breakpoint as with *k*-mer mapping (Extended Data Fig. 10). We amplified the PCR markers $SCS5_{550}$[23] and Xncw-Lr58-1[24] from Thatcher*Lr9* and TA5605 and sequenced the PCR products using Sanger sequencing. Both markers showed the same sequence from Thatcher*Lr9* and TA5605, and they were anchored to contigs Tc*Lr9*ptg001727l and TA1851ptg000193l, respectively. The *Lr58* linked markers BCD410, KSUD23, KUSF11, BG123 and Xcfd50[3] were also anchored to the Thatcher*Lr9* contigs using published probe or SSR sequences (Fig. 3d).

Data availability

The genome assemblies of Thatcher*Lr9* and *Aegilops umbellulata* accession TA1851, the assembly of sorted TA5605 chromosome 2B, the clustered Iso-seq transcripts of Thatcher*Lr9* and TA5605, and the CDS and genomic sequence of *Lr9* are available on the DRYAD database under https://datadryad.org/stash/share/df5K6GT-Gvtn1nDZ3rA-btyTmzp6aD9U_wj8lNxu5sA.

Supplementary Files

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

- ExtendedDataFigures.pdf
- SupplementaryFigures.pdf
- SupplementaryTables.xlsx