

Aegilops sharonensis genome-assisted identification of stem rust resistance gene *Sr62*

Guotai Yu^{1,2,3}, Oadi Matny⁴, Nicolas Champouret^{5,16}, Burkhard Steuernagel¹, Matthew J. Moscou⁵, Inmaculada Hernández-Pinzón⁵, Phon Green⁵, Sadiye Hayta¹, Mark Smedley¹, Wendy Harwood¹, Ngonidzashe Kangara¹, Yajuan Yue¹, Catherine Gardener¹, Mark J. Banfield¹, Pablo D. Olivera⁴, Cole Welchin⁴, Jamie Simmons⁴, Eitan Millet⁶, Anna Minz-Dub⁶, Moshe Ronen⁶, Raz Avni^{6,7,17}, Amir Sharon⁷, Mehran Patpour⁸, Annemarie F. Justesen⁸, Murukarthick Jayakodi⁹, Axel Himmelbach⁹, Nils Stein^{9,10}, Shuangye Wu¹¹, Jesse Poland¹¹, Jennifer Ens¹², Curtis Pozniak¹², Miroslava Karafiátová¹³, István Molnár^{13,18}, Jaroslav Doležel¹³, Eric R. Ward^{5,14,19}, T. Lynne Reuber^{14,20}, Jonathan D. G. Jones⁵, Martin Mascher^{9,15}, Brian J. Steffenson⁴✉ & Brande B. H. Wulff^{1,2,3}✉

The wild relatives and progenitors of wheat have been widely used as sources of disease resistance (*R*) genes. Molecular identification and characterization of these *R* genes facilitates their manipulation and tracking in breeding programmes. Here, we develop a reference-quality genome assembly of the wild diploid wheat relative *Aegilops sharonensis* and use positional mapping, mutagenesis, RNA-Seq and transgenesis to identify the stem rust resistance gene *Sr62*, which has also been transferred to common wheat. This gene encodes a tandem kinase, homologues of which exist across multiple taxa in the plant kingdom. Stable *Sr62* transgenic wheat lines show high levels of resistance against diverse isolates of the stem rust pathogen, highlighting the utility of *Sr62* for deployment as part of a polygenic stack to maximize the durability of stem rust resistance.

¹John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK. ²Plant Science Program, Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia. ³KAUST Center for Desert Agriculture, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia. ⁴Department of Plant Pathology, University of Minnesota, St. Paul, MN 55108, USA. ⁵The Sainsbury Laboratory, University of East Anglia, Norwich NR4 7UJ, UK. ⁶Institute for Cereal Crops Improvement, Tel Aviv University, Tel Aviv, Israel. ⁷School of Plant Sciences and Food Security, Tel Aviv University, Tel Aviv, Israel. ⁸Department of Agroecology, Aarhus University, Forsøgsvej 1, DK-4200 Slagelse, Denmark. ⁹Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Seeland, Germany. ¹⁰Center of integrated Breeding Research (CiBreed), Department of Crop Sciences, Georg-August-University, Von Siebold Str. 8, 37075 Göttingen, Germany. ¹¹Department of Agronomy, Kansas State University, Manhattan, KS 66506, USA. ¹²Crop Development Centre, Department of Plant Sciences, University of Saskatchewan, Saskatoon, SK, Canada. ¹³Institute of Experimental Botany of the Czech Academy of Sciences, Centre of the Region Haná for Biotechnological and Agricultural Research, 779 00 Olomouc, Czech Republic. ¹⁴2Blades Foundation, Evanston, IL, USA. ¹⁵German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany. ¹⁶Present address: Syngenta Flowers, Downers Grove, IL 60515, USA. ¹⁷Present address: Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Seeland, Germany. ¹⁸Present address: Agricultural Institute, Centre for Agricultural Research, ELKH, Martonvásár 2462, Hungary. ¹⁹Present address: AgBiome, Inc., Research Triangle Park, NC 27709, USA. ²⁰Present address: Alliance Management at Enko Chem, 62 Maritime Dr, Mystic, CT 06355, USA. ✉email: bsteffen@umn.edu; brande.wulff@kaust.edu.sa

Stem rust, caused by the fungal pathogen *Puccinia graminis* f. sp. *tritici* (*Pgt*), is one of the most important diseases of wheat worldwide. Stem rust outbreaks were once common, but programs to eradicate the alternate host of this heteroecious fungus, the common barberry (*Berberis vulgaris*), and breeding for disease-resistant wheat cultivars brought the disease under control across most of Europe and North America by the 1950s¹. Epidemics continued to occur in Australia in the 1970s² and South Africa in the 1980s³. Yet, it was the discovery of stem rust in Uganda in the 1998–1999 growing season on wheat lines carrying *Sr31*, a widely deployed and, until then, fully effective stem rust resistance gene, which marked a new era in stem rust epidemics⁴. The causal strain popularly known as Ug99 (race TTKSK on the basis of a widely accepted North American differential set of genotypes) and its derivatives subsequently spread throughout most of East and South Africa and into Yemen and Iran and evolved to overcome additional stem rust resistance genes, including *Sr24* and *Sr36*^{5,6}. In 2012, a severe outbreak of stem rust not related to the Ug99 lineage occurred in Ethiopia in the widely cultivated wheat cultivar Digalu. The “Digalu” *Pgt* strain was subsequently detected across the Middle East⁷ and in some European countries, including Sweden, Denmark, Germany and the UK^{8,9}. In 2016, a stem rust outbreak in southern Italy affected thousands of hectares of both durum and bread wheat¹⁰, while a separate outbreak in Western Siberia caused 30–40% yield losses across 1–2 million hectares¹¹. Stem rust epiphytotics are predicted to become more common as climate change continues, favouring the northward spread of the fungus¹². This outlook highlights the importance of developing new wheat varieties with broad-spectrum, durable resistance to stem rust.

So far, 58 distinct stem rust resistance genes have been designated in wheat¹³. Just over half are from bread wheat (*Triticum aestivum*), while the remainder were introgressed into wheat from wild and domesticated *Triticum* spp. (eight *Sr* genes), *Aegilops* spp. (10 *Sr* genes), rye (*Secale cereale*; four *Sr* genes), wheatgrass (*Thinopyrum* spp.; four *Sr* genes) and the grass *Dasypyrum villosum* (one *Sr* gene)¹³. Most *Sr* genes are pathogen- and race-specific, but *Sr55/Lr67*, *Sr57/Lr34* confer slow-rusting multipathogen resistance^{14,15}. Thirteen of the 58 designated *Sr* genes have been cloned, including *Sr13*¹⁶, *Sr21*¹⁷, *Sr22*¹⁸, *Sr26*¹⁹, *Sr33*²⁰, *Sr35*²¹, *Sr45*¹⁸, *Sr46*²², *Sr50*²³, *Sr55/Lr67*¹⁴, *Sr57/Lr34*¹⁵, *Sr60*²⁴ and *Sr61*¹⁹. In addition, a race-specific stem rust resistance gene with the temporary designation *SrTA1662* was cloned from *Aegilops tauschii*²⁵. Most race-specific *Sr* genes encode nucleotide-binding and leucine-rich repeat (NLR) proteins, except for *Sr60*, which encodes a tandem kinase²⁴. The non-pathogen-specific resistance genes *Sr57/Lr34* and *Sr55/Lr67* encode a putative ATP-binding cassette (ABC) transporter and a hexose transporter, respectively^{14,15}.

Five *R* genes encoding tandem kinases were cloned from various Poaceae species²⁶. The first, *Rpg1* on chromosome arm 1HS in barley, confers stem rust resistance²⁷. *Yr15* on chromosome arm 1BS, originally from emmer wheat (*Triticum turgidum* ssp. *dicoccoides*) and subsequently introgressed into bread wheat, confers broad-spectrum stripe rust resistance²⁸. *Sr60* on chromosome arm 5AS in diploid wheat *Triticum monococcum* confers race-specific stem rust resistance²⁴. *Pm24* on chromosome arm 1DS is a powdery mildew resistance gene from the Chinese wheat landrace Hulutou²⁹, and the most recent addition, *WTK4* on chromosome arm 7DS, confers powdery mildew resistance in *Ae. tauschii* and synthetic hexaploid wheat derivatives²⁵.

Aegilops sharonensis (genome constitution S^hS^h) is a wild diploid relative of wheat in the Sitopsis section found in present day Israel and southern Lebanon³⁰. The species possesses many traits of agricultural importance, including resistance to major diseases of wheat such as the rusts^{31,32}. However, the presence of

gametocidal genes in the genome of *Ae. sharonensis* that restrict interspecies hybridisation have hampered the introgression of its chromatin into wheat^{33–37}. Indeed, of all the 264 designated resistance genes that have been introgressed into wheat, only three are from *Ae. sharonensis*, namely *Lr56*, *Yr38* and *Sr62*^{13,34}. Therefore, the genetic potential of *Ae. sharonensis* remains largely untapped.

Here, we generate a reference-quality genome assembly of *Aegilops sharonensis* and clone the stem rust resistance gene *Sr62*, earlier designated as *Sr1644-1Sh*³⁸. *Sr62* encodes a tandem protein kinase whose individual kinase domain homologues appear across the plant kingdom. We transform *Sr62* into the susceptible wheat cultivar Fielder and confirm the effectiveness of the gene against a range of geographically distinct *Pgt* isolates.

Results

Sequencing and assembly of the *Aegilops sharonensis* genome.

Ae. sharonensis accession AS_1644 was previously used as the resistant parent in creating a recombinant inbred line population to map the stem rust resistance genes *Sr62* (*Sr1644-1Sh*) and *Sr1644-5Sh* and as a donor to introgress these genes into wheat cultivar Zahir^{37,38}. To facilitate the cloning and characterisation of these genes, we generated a genome assembly of AS_1644. A line derived from this accession, which had been advanced through two generations of single-seed descent, contained residual heterogeneity of less than one single-nucleotide polymorphism (SNP) per 10 kb based on analysis of whole-genome shotgun (WGS) reads mapped to 1,440 conserved ‘benchmarking universal single-copy ortholog’ (BUSCO) genes. We utilized multiple technologies to sequence this inbred AS_1644 line (Supplementary Fig. 1) and assembled the genome using the TRITEX pipeline³⁹. In brief, we performed Illumina sequencing-by-synthesis on WGS short-insert, long mate-pair (LMP), 10X linked read and Hi-C chromatin conformation capture libraries, as well as chromosome flow-sorted short-insert libraries (Supplementary Fig. 2; Supplementary Tables 1 and 2). We derived contig assemblies from the WGS short-insert libraries, scaffolded them sequentially with the LMP and 10X data, and connected them into chromosome pseudomolecules using the Hi-C data.

We obtained a scaffold assembly size of 6.7 Gb and a scaffold N50 and N90 of 12.3 and 1.1 Mb, respectively (Table 1). The assembly size of the chromosome pseudomolecules is 6.3 Gb, including 886 Mb of unfilled gaps (Table 1). The chromosome sizes range from 783 Mb (chromosome 1) to 1,022 Mb (chromosome 2). Unanchored scaffolds account for 420 Mb (Table 1). To assess the assembly quality, we performed BUSCO analysis. The assembly contains 96.5% complete BUSCOs and only 1.3% fragmented and 2.2% missing BUSCOs (Supplementary Fig. 3). The structural integrity of the pseudomolecules was supported by inter- and intrachromosomal Hi-C contact matrices (Supplementary Figs. 4, 5), its concordance with sequence data from flow-sorted individual chromosomes (Supplementary Fig. 6), and by collinearity with an *Ae. sharonensis* consensus genetic linkage map comprising 727 sequence markers and spanning 631 cM³⁸ (Supplementary Fig. 7).

Sequence comparison with high-confidence genes from the Chinese Spring A-subgenome⁴⁰ identified 29,849 *Ae. sharonensis* candidate genes (84.5% of the total) at a cut-off of 90% sequence identity (Supplementary Table 3). This increased to 30,260 candidate genes (88.4%) after comparison to the D-subgenome and 30,626 (85.9%) after comparison to the B-subgenome. This analysis suggests that at least 30,626 high-confidence genes are present in our *Ae. sharonensis* genome assembly. We detected strong collinearity in the distal regions between the *Ae. sharonensis* and Chinese Spring subgenomes but disrupted

Table 1 *Aegilops sharonensis* AS_1644 v 1.0 genome assembly statistics.

Assembly characteristics	Values
Assembly size	6.7 Gb
^a Scaffold N50	12.3 Mb
^a Scaffold N90	1.1 Mb
Pseudomolecule size	6.3 Gb
Unfilled gaps	886 Mb
Chromosome 1S ^{sh}	783 Mb
Chromosome 2S ^{sh}	1022 Mb
Chromosome 3S ^{sh}	972 Mb
Chromosome 4S ^{sh}	827 Mb
Chromosome 5S ^{sh}	868 Mb
Chromosome 6S ^{sh}	807 Mb
Chromosome 7S ^{sh}	1016 Mb
Unassigned to a chromosome	420 Mb
Complete BUSCOs	0.965
Fragmented BUSCOs	0.013
Missed BUSCOs	0.022

^aScaffolds <1 kb were excluded.

collinearity in the centromeric regions (shown for the D-subgenome in Supplementary Fig. 8).

Positional mapping of *Sr62*. Resistance to stem rust was previously transferred into the wheat cultivar Zahir as a compensating Robertsonian translocation between *Ae. sharonensis* chromosome 1S^{sh} and the closely related wheat chromosome 1BL/1DL (1S^{sh}S-1S^{sh}L-1BL/1S^{sh}S-1S^{sh}L-1DL)³⁷ (Fig. 1a). Independently of this, QTL mapping located *Sr62* on the short arm of chromosome 1S^{sh} in an *Ae. sharonensis* F₆ recombinant inbred line population from a cross between accessions 2189 (susceptible) and 1644 (resistant)³⁸. In the same population, a second stem rust resistance locus, *Sr1644-5Sh*, was localized to the long arm of chromosome 5S^{sh}. To genetically separate these two *Sr* loci, we genotyped plants from F_{2:3} families and identified one plant, designated 803, that was heterozygous for markers diagnostic for *Sr62* on chromosome 1 and homozygous for a marker diagnostic of the susceptible haplotype at *Sr1644-5Sh* on chromosome 5 (Supplementary Fig. 9a; Supplementary Table 4). We phenotyped and genotyped 49 F_{3:4} plants derived from plant 803 and observed a 3:1 segregation ratio for resistance and susceptibility and congruency between resistance and PCR molecular markers diagnostic for accession 1644 on chromosome 1 (Supplementary Fig. 9b; Supplementary Table 4). These results indicated that we had genetically isolated *Sr62* and that the gene is dominant in *Ae. sharonensis*. Furthermore, the markers delimited the position of *Sr62* to a 12-cM interval between proximal marker C23635_CAPS and distal marker C25971_CAPS (Supplementary Fig. 9c; Supplementary Data 1 and Supplementary Tables 5–7).

To fine-map *Sr62*, we developed a large population segregating for *Sr62* by selfing F_{4:5} and F_{5:6} plants derived from plant 803 and heterozygous for the *Sr62* interval on chromosome 1. We genotyped 4,638 plants from this population with markers flanking *Sr62*, revealing 12 recombinants between C11308_KASP and C67147_KASP (Fig. 1b; Supplementary Fig. 9d; Supplementary Table 8). By generating more markers in this interval and phenotyping progeny of the 12 recombinants, we mapped *Sr62* to the region between markers S741_KASP-7 and C03246_CAPS (Supplementary Table 8), corresponding to a physical interval of 480 kb on the short arm of *Ae. sharonensis* chromosome 1 (Fig. 1c).

We performed RNA-Seq on AS_1644 and mapped the reads to the genome assembly. This identified seven transcribed genes in

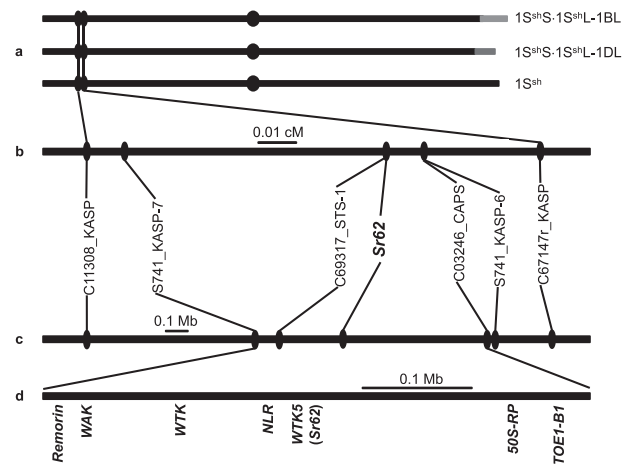


Fig. 1 Positional mapping restricts *Sr62* to a 480 kb interval on chromosome 1S^{sh}. **a** Wheat-*Ae. sharonensis* translocation chromosomes and *Ae. sharonensis* chromosome 1S^{sh}. **b** Genetic map of the region harbouring *Sr62* on the short arm of *Ae. sharonensis* chromosome 1S^{sh}. **c** Physical map of the region around *Sr62*. **d** Genes in the interval genetically delimiting the presence of *Sr62*. WTK is presumably an ortholog of *Pm24*²⁹.

the interval with homology to the genes encoding a remorin family protein, a wall-associated kinase (WAK), two wheat tandem protein kinases (dubbed WTK5 and WTK), an NLR, a 50S ribosomal protein (50S-RP) and a target of Eat1-B1 protein (TOE1-B1) (Fig. 1d).

Identification of an *Sr62* candidate by EMS mutagenesis and RNA-Seq alignment. To identify *Sr62* among the candidate genes in the mapping interval, we performed EMS mutagenesis of Zahir-1644 wheat-*Ae. sharonensis* introgression lines, in which most of wheat chromosomes 1D or 1B were replaced by chromosome 1S^{sh} of *Ae. sharonensis* accession 1644³⁷ (Fig. 1a). We mutagenized 3,025 Zahir-1644 seeds with 0.75% EMS. Eight or more seeds from the surviving 1649 M₂ families were screened for susceptible mutants using the *Pgt* isolate Ug99 (race TTKSK). Thirty families segregating for resistance and susceptibility were identified and tested in the M₃ generation, revealing 14 independent susceptible mutants. Genotyping-by-sequencing⁴¹ of Zahir, the two Zahir-1644 introgression lines and 10 EMS-derived mutants allowed us to rule out cross-contamination from other wheat cultivars for this mutant subset and to determine the translocation type, 1S^{sh}S-1S^{sh}L-1BL or 1S^{sh}S-1S^{sh}L-1DL (Supplementary Figs. 10–20).

We constructed a full-length cDNA library for AS_1644, sequenced this library on the Illumina sequencing platform (generating 98 million 150-bp paired-end reads) and assembled these data to obtain transcripts of the seven genes in the mapping interval. In parallel, we sequenced the leaf transcriptomes of Zahir-1644 and the 14 mutants by generating ≥91 million 150-bp paired-end RNA-Seq reads per sample. We mapped the mutant RNA reads to the transcripts of the seven genes. This procedure, which we termed MutRNA-Seq (Fig. 2), identified eight point-mutations in WTK5 among seven of the 14 mutants. All of the mutations were G/C-to-A/T transition mutations, typical of EMS⁴². We predicted the open reading frame of WTK5 and found that seven of the mutations introduced non-synonymous changes, whereas one introduced an early stop codon (Fig. 3a, b). In contrast, WTK, the NLR, the 50S ribosomal protein and TOE1-B1 had two or no mutations. For the WAK and remorin genes, the expression levels were too low to reliably call mutations (Supplementary Table 9 and Supplementary Data 2).

We developed a formula (function (4), see ‘Methods’ section) to test if *WTK5* was the candidate among the seven genes. We graphically displayed the minimum number of mutants required to successfully identify a candidate gene by mutational genomics as a function of the number of genes investigated (Supplementary Fig. 21). We considered typical scenarios encountered in mutational genomics studies, such as when scrutinizing all genes in a discrete mapping interval (e.g., ten genes); a whole gene family (e.g., all 3,200 NLR loci in hexaploid wheat⁴³); a whole chromosome, as obtained by chromosome flow sorting (i.e.,

~5100 genes); or all ~107,000 genes in the hexaploid wheat genome⁴⁰. This analysis indicated that the minimum number of independent mutants required to identify a candidate gene with a 2000-bp coding DNA sequence (CDS) at $p = 0.01$ is 3, 5, 5, or 6, respectively, for the abovementioned scenarios. This increases to 8, 16, 16, and 20 mutants, respectively, when dealing with two complementation groups. For our *Sr62* mapping interval, which contains seven genes, the probability of obtaining 7 out of 14 mutants with a mutation in the sequence of at least one of the seven genes being investigated based on the *WTK5* CDS (2,223 bp) would be 9.0×10^{-5} . In other words, *WTK5* emerged as the best candidate among the seven genes, prompting further functional analysis.

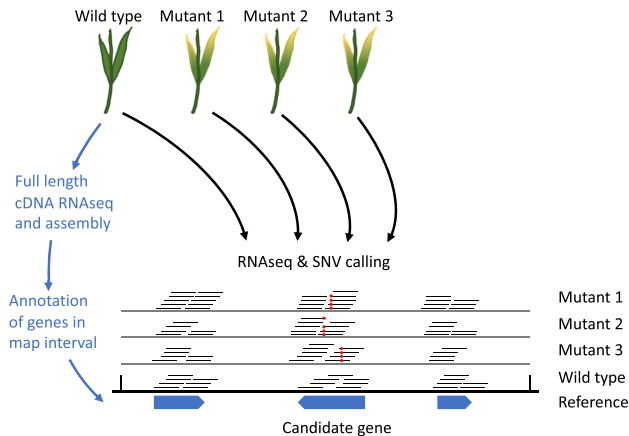


Fig. 2 Candidate gene identification by mutagenesis and transcriptome sequencing (MutRNA-Seq). RNA-Seq reads from the wild-type parent and independently derived EMS mutants are mapped to a reference genome sequence. Annotated genes are inspected (within a mapping interval, if available) for a gene exhibiting a preponderance of single-nucleotide variants (SNVs, red dots) across the mutants.

The *Sr62* candidate confers stem rust resistance in transgenic wheat. By mapping the leaf transcriptome RNA-Seq reads to the genome assembly of *Ae. sharonensis*, we predicted that the *WTK5* transcript spans 18,384 bp and contains 11 introns (Fig. 3a; Supplementary Fig. 22). To engineer a binary construct containing *WTK5*, primer pairs (Supplementary Table 10) were designed to amplify two parts of the genomic DNA sequence that cover most of the native gene, including 2.8 kb of putative promoter sequence 5' of the predicted start codon and 2.0 kb of putative terminator region 3' of the predicted stop codon but excluding 11.4 kb of the middle of the 12.4-kb intron (Fig. 3a). The two PCR products were separately cloned into a binary vector via three-way ligation; the resulting recombinated *WTK5* spans 11.9 kb (Fig. 3c). The construct was verified by Sanger sequencing and transformed into wheat cultivar Fielder. We obtained three independent primary transgenic lines (T_0), which, based on qRT-PCR of the selectable marker, were predicted to contain one copy of the transgene (two lines) or four copies of the transgene (one line). We advanced these hemizygous lines to the next generation to obtain homozygous lines. All three

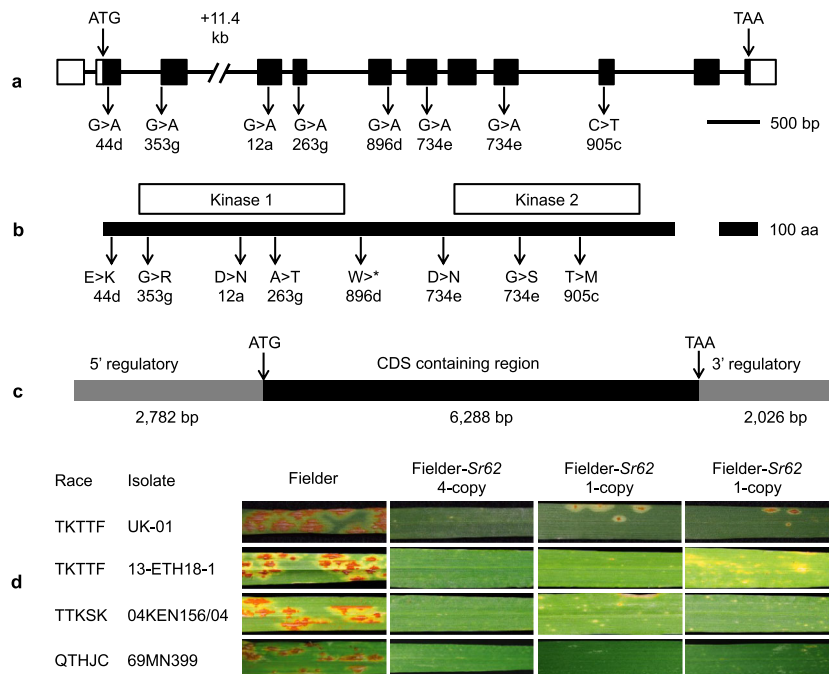


Fig. 3 Functional validation of *Sr62* by EMS mutagenesis and transformation into wheat. **a** Structure of *Sr62*, with predicted nucleotide change caused by EMS-derived loss-of-function mutations. Boxes represent exons and lines represent introns with white boxes representing untranslated regions and black boxes representing the predicted open reading frame. The 11.4-kb portion of the third intron excluded from the binary construct is indicated. **b** Schematic representation of the *Sr62* protein, with the position of the two protein kinase domains and the predicted amino-acid changes caused by the EMS mutations indicated. **c** The *Sr62* sequence used for transformation of wheat cultivar Fielder. CDS, coding DNA sequence. **d** Reactions of three homozygous independent transgenic lines to four *Pgt* isolates. The copy number of the hygromycin selectable marker in T_0 plants is indicated.

lines conferred resistance to *Pgt* stem rust races TTKSK (isolate 04KEN156/04 from Kenya), TKTTF (isolate 13-ETH18-1 from Ethiopia), TKTTF (isolate UK-01 from the UK) and QTHJC (isolate 69MN399 from the US) (Fig. 3d), whereas the null plants were all as susceptible as the parent cv. Fielder (Supplementary Fig. 23). We also tested the line with four copies of the selectable marker against an additional eight *Pgt* isolates/races from Israel (three isolates), Italy (three isolates), Kenya (one isolate) and Ethiopia (one isolate) and found high levels of resistance (Supplementary Table 11; Supplementary Fig. 24).

Both protein kinase domains are required for *Sr62* function.

We mapped the mutations in *Sr62* relative to the two protein kinase domains. The seven amino acid substitutions are spread evenly throughout the predicted amino acid sequence, with three mutations in the Kinase 1 domain and two in Kinase 2 (Fig. 3b). Moreover, the premature stop codon mutant 896d leads to a predicted truncated protein lacking Kinase 2. Based on sequence alignment to previously characterized plant protein kinases, *Sr62* is predicted to encode a protein with two serine/threonine kinase domains (Supplementary Fig. 25). In mutant 353g, a conserved glycine residue at amino acid position 57 was substituted with an arginine proximal to the conserved ATP-binding site of Kinase 1, whereas the aspartic acid-to-asparagine substitution at amino acid position 177 (mutant 12a) is located in the catalytic site of Kinase 1 (Supplementary Data 3 and 4). Moreover, the substitutions at positions 57 and 177 are predicted to be intolerant (Supplementary Table 12). We developed a 3D model of the structure of *Sr62* using the Phyre2 web portal⁴⁴ and CCP4MG⁴⁵ and mapped these two amino acid substitutions onto the model. Interestingly, the Aspartate177Asparagine mutation is in a critical residue for kinase function (Supplementary Fig. 26). In active kinases, this Aspartate is involved in binding the phosphate of

ATP, and acts as the catalytic residue⁴⁶ functioning as a base acceptor for the proton transfer (Supplementary Fig. 26). The PDB coordinates of the comparative model can be found in the source data of Supplementary Fig. 26. We searched *Sr62* for the presence of eight conserved amino acids that are diagnostic for protein kinases⁴⁷. All eight of these amino acids are found in Kinase 1, whereas five out of the eight are present in Kinase 2. Based on this analysis, *Sr62* has a predicted kinase-pseudokinase structure, similar to *Pm24*. However, our mutant analysis suggests that both protein kinase domains are required for *Sr62* function. One of the amino acid substitutions (G539S) in Kinase 2 is predicted to be intolerant (Supplementary Table 12).

Sequence relationship between *Sr62* and other tandem kinases in grasses.

To explore the relationship between *Sr62* and other tandem kinases in major cereal crop species and wild grasses, we performed phylogenetic analysis of tandem kinases in hexaploid wheat (*Triticum aestivum*), durum wheat (*Triticum durum*), barley (*Hordeum vulgare*), rice (*Oryza sativa*), sorghum (*Sorghum bicolor*), maize (*Zea mays*) and the wild wheat relatives *Ae. tauschii* and *Ae. sharonensis* (Fig. 4a). The resistance genes *Sr62*, *Pm24*, *Rpg1*, *WTK4* and *Yr15* belong to the most populous clade, while *Sr60* sits in a small and separate, but closely related, clade (Fig. 4a). The closest neighbour of *Sr62* is *Pm24*, which appears to be a wheat chromosome 1D orthologue of *Ae. sharonensis* *WTK* situated 193 kb away from *Sr62* (Fig. 1d). However, the identity between *Sr62* and *Pm24* is low (<65%) at both the CDS and amino-acid sequence levels (Supplementary Table 13). We next generated a phylogenetic tree with the two protein kinase domains from each tandem kinase separated from each other. In this tree, the two domains of *Rpg1* and *WTK4* are located near each other in Clade 4 (Fig. 4b). By contrast, the two domains of *Sr62* and *Pm24* sit in different clades (Clades 4 and 5), but both *Sr62*-K2 and *Pm24*-K2 sit in Clade 4, and

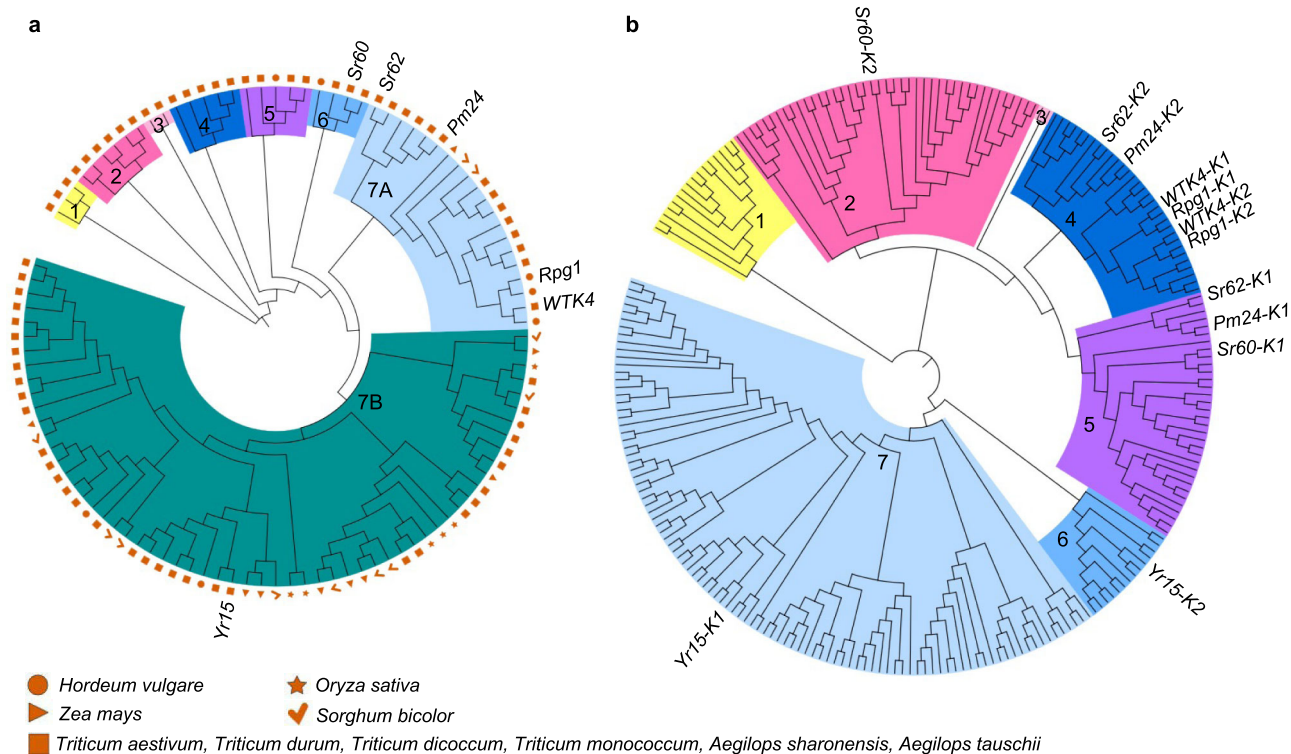


Fig. 4 Phylogenetic relationship between tandem kinases from cereal crop and wild grasses. A total of 99 predicted tandem kinases were retrieved from the genomes of bread wheat, durum wheat, maize, barley, sorghum, rice, *Ae. tauschii* and *Ae. sharonensis*, along with the five cloned tandem kinase disease resistance genes. Phylogenetic clades and subclades are indicated by different colours and labelled with numbers. **a** Phylogeny based on the whole tandem kinase coding sequence. **b** Phylogeny based on the individual protein kinase domain coding sequences.

both *Sr62-K1* and *Pm24-K1* sit in Clade 5. The two domains of *Sr60* and *Yr15* are also in distinct clades (Clades 6 and 7 for *Yr15* and Clades 2 and 5 for *Sr60*). Based on pair-wise alignment using the Needleman–Wunsch method⁴⁸, the two *Sr62* kinase domains share 51% identity. None of the five cloned tandem kinase genes contain kinase domains with more than 80% nucleotide identity.

Using a Hidden Markov Model-based classification approach for protein kinases developed by Lehti-Shiu & Shiu⁴⁹, we investigated the individual protein kinase domains in *Rpg1*, *Pm24*, *Yr15*, *Sr60*, *WTK4* and *Sr62*. *Sr62* has a DLSV-DLSV configuration as do *Rpg1* and *Pm24*, whereas the *Sr60* protein kinase domains belong to the DLSV (domain 1) and CR4L (domain 2) subfamilies, and the *Yr15* domains belong to the WAK (domain 1) and RLCK-VIII (domain 2) subfamilies.

The *Sr62* kinase domains are of an ancient origin. We performed coding sequence homology analysis of *Sr62* using Gramene, a resource for comparative genomics across the plant kingdom^{50,51}. Homologues were detected in five phyla: Tracheophyta, Bryophyta, Marchantiophyta, Chlorophyta and Rhodophyta (Supplementary Data 5). Thus, the homologues of the *Sr62* kinase domains are present in species ranging from unicellular green and red algae to mosses, liverworts and crop species including cereals (wheat, barley, rice, maize and sorghum), brassicas, potato, tobacco and coffee. The number of homologues varies widely, from three in the red alga *Chondrus crispus* to 65 in the wild tobacco *Nicotiana attenuata*. The *Triticum* species *T. aestivum* (hexaploid), *T. durum* (tetraploid), *Triticum*

dicoccum (tetraploid) and *Triticum urartu* (diploid) have 26, 34, 19 and 21 homologues, respectively (Supplementary Data 5). These data suggest that the origin of the *Sr62* kinase domains predates the diversification of plants.

The synteny around *Sr62* is specific to closely related grasses.

We conducted synteny analysis extending to ten genes on either side of *Sr62* in *Ae. sharonensis* and compared this region across nine Poaceae genomes spanning 60 million years of evolution⁵². The synteny block contains an F-box protein, a glutamyl-tRNA reductase, a pentatricopeptide repeat-containing protein, a remorin family protein, a protein kinase, an NLR and a TOE1-B1 (Fig. 5). The synteny is well conserved within the *Triticum* and *Aegilops* species, and to a lesser extent with barley, where it has undergone extensive rearrangements relative to *Triticum* and *Aegilops*. However, the block appears to be absent from *Brachypodium*, rice, sorghum and maize, suggesting that it arose between 11.6 and 35 million years ago⁵² (Fig. 5).

Discussion

We cloned the *Ae. sharonensis* major-effect, dominant stem rust resistance gene *Sr62* and found that it encodes a tandem kinase. To date, more than 300 disease resistance genes have been cloned in plants, most of which (189 of 310) contain NLRs⁵³. Only a few, from various Poaceae species, encode tandem kinases: *Rpg1*, *Yr15*, *Pm24*, *Sr60* and *WTK4*, all of which were identified in the Triticeae^{24,25,27–29}. These five genes, as well as *Sr62*, contain two protein kinase domains. In *Rpg1* and *WTK4*, the two protein

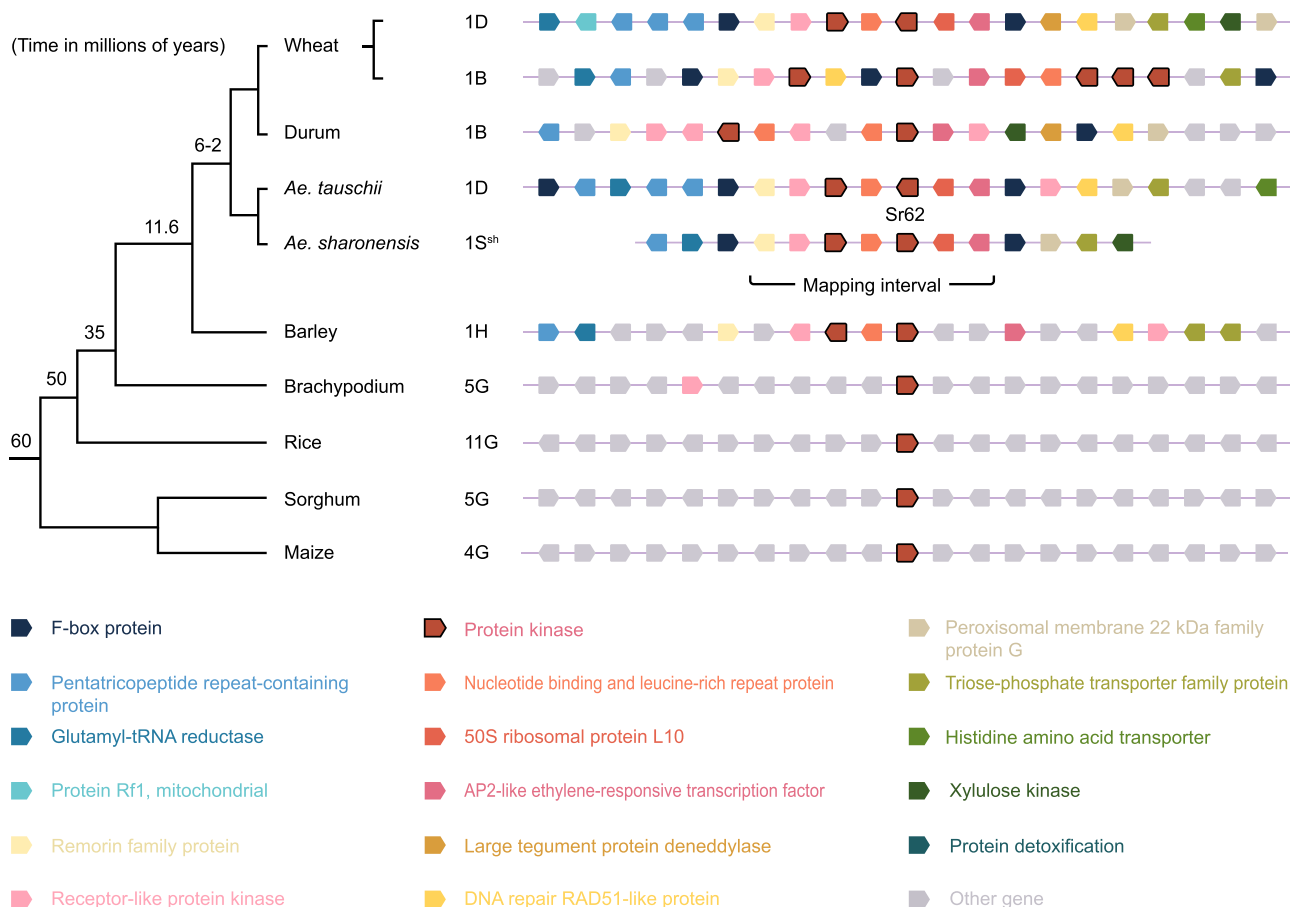


Fig. 5 Synteny around *Sr62*. Genomic regions containing genes orthologous to *Sr62* along with surrounding genes reveal micro-synteny. The syntenic block is well conserved within the *Triticum* spp., *Aegilops* spp., and barley, but appears to be absent from *Brachypodium*, rice, sorghum and maize. The synteny alignment was generated through Gramene, except for *Ae. sharonensis*, which was added manually.

kinase domains are close to each other in the same phylogenetic clade (Fig. 4).

Based on sequence conservation of the key amino acid residues for protein kinase function in the two kinase domains, *Yr15* and *Pm24* have been classified as encoding tandem kinase-pseudokinases, *Sr60* as encoding a tandem kinase-kinase and *Rpg1* encoding a tandem pseudokinase-kinase²⁴. Here, based on amino-acid alignment with plant kinases, we identified two absolutely conserved amino acids (residues 57 and 177) in the binding site and catalytic site of Kinase 1 of *Sr62* (Supplementary Fig. 26). Two EMS-induced susceptible mutants, 353g and 12a, have non-synonymous mutations that alter these two key residues (Supplementary Data 3 and 4), implying that the function of Kinase 1 of *Sr62* is critical for stem rust resistance. Another susceptible mutant (896d) carries an early stop codon resulting in a truncated protein predicted to only contain the Kinase 1 domain, suggesting that an intact Kinase 2 is also required for *Sr62* function.

For more than 20 years, the guard model has provided a useful framework for understanding the molecular mechanism and evolution of plant resistance genes^{54,55}. According to this model, plant resistance proteins guard the pathogenicity targets (guardees) of pathogen effector molecules. The interaction of an effector with the pathogenicity target is detected by the guard, leading to a conformational change that triggers signalling, resulting in downstream defence responses. In NLR proteins, the C-terminal leucine-rich repeats provide the guarding function, while the N-terminal nucleotide-binding and coiled-coil or TIR domains confer the signalling capacity^{56–58}. In the absence of a resistance gene, the interaction between guardee and effector protein promotes pathogen growth, resulting in a susceptible phenotype. All three interactors—guard, guardee and effector—are subject to diversifying selection, but for the guardee, this can be constrained by the requirement to maintain cellular function. Duplication of the guardee can release it from this constraint and provide a ‘decoy’ for the effector⁵⁹. Early experimental support for the guard hypothesis came from the study of *Arabidopsis thaliana* RIN4, which is guarded by the NLRs RPM1 and RPS4 and targeted by the bacterial effectors AvrRpm1 and AvrRpt2^{55,60–62}, and tomato *Pto*, a serine/threonine protein kinase that is guarded by the NLR protein Prf and targeted by the bacterial effector avrPto^{55,63,64}. Somewhat unusually, the *Pto* gene acts genetically as the resistance gene and is part of a complex of six paralogues within a 60-kb region within which *Prf* is embedded⁶⁵. In *A. thaliana*, the PBL2 (kinase)–RKS1 (pseudokinase)–ZAR1 (NLR) complex triggers immunity upon detection of the *Xanthomonas campestris* effector AvrAC⁶⁶.

Perhaps like the *Pto* protein kinase or the PBL2/RKS1 kinase/pseudokinase, *Sr62* is also a pathogenicity target guarded by an NLR. In our EMS mutational genomics experiment targeting *Sr62*, only seven of 14 susceptible mutants carried non-synonymous or missense mutations in the tandem kinase (*Sr62*). This suggests that we obtained second-site mutations in one or multiple genes required for *Sr62* function. Likewise, EMS mutagenesis of *Pm24* (which is syntenic to *WTK* in the *Sr62* haplotype) resulted in 11 mutations in *Pm24* out of 26 susceptible mutants²⁹. This propensity for second-site mutations is unusual; typically, mutagenesis screens targeting major dominant *R* genes yield <20% second-site suppressors¹⁸. Identification of the second-site suppressors of *Pm24* and *Sr62* could provide insight into the mechanism of wheat tandem kinase resistance genes. *Sr62* and *Pm24* both lie adjacent to an NLR gene (Fig. 1). The NLR adjacent to *Pm24* did not confer powdery mildew resistance when transformed into a susceptible wheat cultivar²⁹; however, this does not exclude the possibility that this physically linked NLR is involved in *Pm24* function. We identified two EMS

mutations in the NLR next to *Sr62* (out of the 14 susceptible mutants) of which one was non-synonymous. Further work is required to determine whether the linked NLR functions in concert with the *Sr62* and *Pm24* tandem kinases in a mechanism similar, for example, to the requirement of tomato *Prf* (NLR) for *Pto* function^{63–65}. Alternatively, *Sr62* could directly perceive the presence of its corresponding effector and activate a downstream signal cascade to confer resistance independent of the involvement of any NLR, as has been proposed for other WTK proteins²⁶. In this model, the pseudokinase acts as an effector decoy target that works in concert with the active kinase to initiate defense signalling²⁶.

We found that the protein kinase domains in *Sr62* belong to the Pelle/DLSV subfamily, which is present in Streptophytes, including angiosperms, gymnosperms, bryophytes, and Step-treophyte algae⁶⁷. The DLSV is highly expanded in angiosperms and is the largest protein kinase family and induced in response to biotic stress in *Arabidopsis thaliana*⁶⁷. Our working hypothesis is that diverse plant pathogens have evolved effectors that target the DLSV protein kinase family in order to impair immune signalling. A major question is whether the tandem kinases are simply decoys⁵⁹ or have a specific function in immunity (or another process).

To facilitate the practical exploitation of *Ae. sharonensis*, we developed a high-quality reference genome based on WGS sequencing and chromosome flow sorting. The *Ae. sharonensis* N50 scaffold size of 12.3 Mb compares well with those of other assembled Triticeae genomes, including barley (N50, 1.4 Mb)⁶⁸, *Ae. tauschii* (N50, 11.4 Mb)⁶⁹, durum wheat (N50, 6.0 Mb)⁷⁰ and bread wheat (N50, 22.8 Mb)⁴⁰. The assembly contains 96.5% complete BUSCOs, which is similar to the 95.7% of rice (*Osetiva* v7.0)⁷¹, but higher than the 86.4% of the first version of the barley reference genome (Hvulgare IBSC_PGSR r1)⁷².

Using flow cytometry, Eilam et al. determined the nuclear DNA content (1C value) of *Ae. sharonensis* to be 7.52 pg⁷³, which is equivalent to a genome size of 7.35 Gb. We constructed chromosome pseudomolecules covering 6.3 Gb. Compared to the assembled genome sizes of barley (4.98 Gb)⁷⁴, *Ae. tauschii* (4.0 Gb)⁶⁹ and *T. urartu* (4.79 Gb)⁷⁵, this is the largest diploid Triticeae genome assembled to date. The 2S^{sh} and 7S^{sh} chromosomes are each physically longer than 1 Gb (Supplementary Fig. 5). Interestingly, the short arm of 7S^{sh} (as defined by synteny to other Triticeae) appears to be physically longer than the long arm (Supplementary Fig. 5).

The D-subgenome of hexaploid wheat had a slightly higher percentage of high-confidence genes that aligned with *Ae. sharonensis* (88.4%) than the A- and B-subgenomes (84.5% and 85.9%, respectively; Supplementary Table 3). This supports a closer relationship between *Ae. sharonensis* and the D-subgenome, as previously reported based on gene tree topologies^{76,77}. Further supporting this close evolutionary relationship, extensive haplotype collinearities were observed between *Ae. sharonensis* and wheat D-subgenome chromosomes⁷⁸. Therefore, future *Ae. sharonensis* introgressions into wheat should be directed to the D-subgenome to reduce the likelihood of genetic imbalance. However, gametocidal genes in *Ae. sharonensis* make it difficult to develop introgression lines for every chromosome^{37,79}. Our *Ae. sharonensis* reference genome will support ongoing efforts to clone these gametocidal genes^{35,80,81}, perhaps leading to tools for accelerating introgression of *Ae. sharonensis* chromatin into wheat. However, because of the hybridization barrier imposed by the gametocidal genes themselves, *Ae. sharonensis* remains a largely unexploited source of resistance to major diseases of wheat³⁰. The reference genome presented here will aid in the molecular cloning of such resistance genes, allowing their incorporation into genetically modified (GM) polygene stacks.

Several recently developed technologies facilitate the cloning of disease resistance genes in plants¹³. NLR or WGS sequencing combined with association mapping have been successfully applied for the rapid cloning of *Sr46*, *SrTA1662* and *WTK4* in *Ae. tauschii*^{22,25}. Mutagenesis combined with NLR sequencing (MutRenSeq) or chromosome sequencing (MutChromSeq) allowed the rapid cloning of the wheat resistance genes *Sr22*, *Sr26*, *Sr45*, *Sr61*, *Yr5a*, *Yr5b*, *Yr7* and *Pm2* and the barley gene *Rph1*^{18,19,82,83}. Mutagenesis combined with mapping, chromosome flow sorting and de novo generation of a cultivar-specific reference-quality single-chromosome assembly facilitated the cloning of *Lr22a* and *Pm21*^{84,85}. Here we used a combination of whole-genome de novo assembly, positional mapping and RNA sequencing of multiple EMS-derived mutants to clone *Sr62*. The assembled reference genome of *Ae. sharonensis* facilitated fine mapping without BAC libraries to delimit *Sr62* to a 480-kb interval based on screening 9,276 products of meiosis. We identified seven genes in the interval with homology to a remorin, WAK, WTK, NLR, WTK5, 50 S and TOE1-B1 gene. By applying RNA-Seq to 14 EMS-derived mutants, we identified *WTK5* as the best candidate. The transformation of *Sr62* into cv. Fielder confirmed that *Sr62* is sufficient to confer resistance to stem rust, indicating that MutRNA-Seq is a powerful tool for gene identification.

The effectiveness of mutational genomics approaches, including MutRenSeq^{18,19} and the MutRNA-Seq method developed in this study, relies on multiple factors, including (i) the number of genes controlling a phenotype, (ii) the delimitation of the target gene to a physical map interval, chromosome or gene family and (iii) the number of mutants obtained. To aid in the future experimental design of mutational genomics studies in hexaploid wheat, we calculated the minimum number of mutants required to confidently (at $p=0.01$) identify the correct gene. When considering a discrete map interval, a gene family, a whole chromosome or indeed all the genes in the wheat genome, this ranges from three to six mutants, but it increases to eight to 20 mutants if two complementation groups are revealed by mutagenesis (Supplementary Fig. 21). In practice, most gene cloning studies require a combination of genetic mapping with gene knockout and/or gain-of-function experiments. Extending MutRNA-Seq to other plant species requires sexual reproduction and the ability to obtain mutants. Obtaining mutants is favoured by an agronomy that promotes a large seed set and facile plant husbandry. Polyploidy can also be considered an advantage. This makes it easier to achieve an effective mutagen dose without killing the emerging seedlings or causing sterility. In addition, polyploid plants typically tolerate a 4-fold higher mutation density compared to diploids⁴², which reduces the size of the population that needs screening.

Advances in genomics and bioinformatics, such as those described here, are fuelling an exponential growth in the discovery and cloning of disease resistance genes in wheat and its wild relatives¹³. This is providing exciting opportunities for engineering broad-spectrum and durable disease resistance into wheat. The current major obstacle is no longer technical but imposed by the socio-political stalemate on the acceptance of GM wheat⁸⁶. However, in May 2021 the Argentine biotechnology company Bioceres Crop Solutions and the Latin American food production and high street franchise Havanna announced the marketing of Alfajores biscuits made from a GM wheat containing the *hahb-4* gene from sunflower conferring drought tolerance⁸⁷. In November of the same year, the Brazilian National Biosafety Commission announced the approval of flour made from *hahb-4* wheat⁸⁸. Hopefully this and other efforts will lower the barrier for introducing other GM traits into wheat, such as for disease resistance.

Methods

Phenotyping. The stem rust tests with TTKSK were carried out in the BSL-3 containment facility at the University of Minnesota. The greenhouse was maintained at 19–22 °C with a 14-h photoperiod and approximately 40% relative humidity. Plants were inoculated with *P. graminis* f. sp. *tritici* when the second leaf was fully expanded, 10–12 days after planting, at a rate of ~0.12 mg of spores per plant. The inoculated plants were then placed in mist chambers in the dark overnight at near 100% relative humidity and 22 °C for 16 h. After the 16-h incubation period in the dark, fluorescent lamps were turned on with the misting continuing for an additional 2 h. After that, the misters were turned off and the plants allowed to slowly dry under the lights. Plants were moved back to the greenhouse and then scored for reaction to stem rust 12 days later. The infection types (IT) were recorded using the Stakman scale⁸⁹.

DNA extraction and sequencing. Leaf tissue from a single plant of *Ae. sharonensis* accession 1644 (line number BW_24933) was collected at the 7-leaf stage, and the DNA was extracted using the CTAB method for large quantities of DNA⁹⁰. PCR-free short insert libraries (450 bp and 800 bp) and long mate pair (MP) (3 kb and 6 kb) libraries were generated and sequenced at Novogene, while the 9 kb long MP library was generated and sequenced at the Roy J. Carver Biotechnology Center, University of Illinois.

DNA for Hi-C and 10X was extracted as outlined in Jupe et al.⁹¹ Briefly, nuclei were extracted from up to 1 g of fresh leaf tissue by homogenization in 10 ml of nuclei isolation buffer, filtered through cell strainers and separated from debris using a Percoll layer. The extracted nuclei were embedded in low-melting agarose plugs and exposed to lysis buffer with proteinase K and RNase A. DNA was released by digesting the agarose with Agarase enzyme (New England Biolabs, Ipswich, MA, USA) and analysed by pulsed-field gel electrophoresis.

High-molecular-weight (HMW) genomic DNA (>40 kb) was isolated from the agarose plugs using pulsed-field electrophoresis on a Blue Pippin instrument (Sage Science) following the high-pass protocol with minor modifications. The size and integrity of the recovered HMW DNA was evaluated on a TapeStation 2200 (Agilent) and quantified by fluorometry (Qubit 2.0). One 10X sequencing library was prepared following the Chromium Genome Library protocol v2 (10X Genomics) and sequenced across two lanes of HiSeqX with 150-bp paired-end (PE) reads (Illumina), which produced ~827 million reads (~33× coverage). Long Ranger (10X Genomics) was used to generate FASTQ files for analysis.

Flow cytometric analysis and sorting. Suspensions of mitotic metaphase chromosomes were prepared from root tips of *Ae. sharonensis* accession 1644 as described by Vrána et al.⁹² and Kubaláková et al.⁹³. Briefly, root tip meristem cells were synchronized using hydroxyurea, accumulated in metaphase using amiprophos-methyl and mildly fixed in formaldehyde. Intact chromosomes were released by mechanical homogenization of 100 root tips in 600 μ l ice-cold LB01 buffer. Microsatellites GAA and ACG were labelled on isolated chromosomes by fluorescence in situ hybridization in suspension (FISHIS) using 5'-FITC-GAA7-FITC-3' and 5'-FITC-ACG7-FITC-3' oligonucleotides (Sigma, Saint Louis, USA) according to Giorgi et al.⁹⁴, and chromosomal DNA was stained by DAPI (4',6-diamidino 2-phenylindole) at 2 μ g/ml.

Chromosome analysis and sorting were performed using a FACSria II SORP flow cytometer and sorter (Becton Dickinson Immunocytometry Systems, San Jose, USA). Bivariate flow karyotypes FITC vs. DAPI fluorescence were acquired for each sample, and chromosomes were sorted at a rate of 1500–2000 particles per second. Two batches of 25,000–76,000 copies of each chromosome (chromosomes 1Sh and 6Sh were sorted together) were sorted into PCR tubes containing 40 μ l sterile deionized water.

The chromosome contents of flow-sorted fractions were estimated by microscopic observation of 1500–2000 chromosomes sorted into a 10- μ l drop of PRINS buffer containing 2.5% sucrose⁹⁵ on a microscopic slide. Air-dried chromosomes were labelled by FISH with probes for pSc119.2 repeat, GAA n microsatellite and 45S rDNA according to Molnár et al.⁹⁶. At least 100 chromosomes were classified following the karyotype described by Zhang et al.⁹⁷ and Badaeva et al.⁹⁸ to determine the chromosome content of flow-sorted samples and to assign the populations observed on bivariate flow karyotypes to particular chromosomes.

Flow-sorted chromosome samples were treated with proteinase K, after which their DNA was purified and amplified by multiple displacement amplification (MDA) using an Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare, Chalfont St. Giles, United Kingdom) as described by Šimková et al.⁹⁹.

Genome assembly. Chromosome-scale sequence assembly was performed using the TRITEX assembly pipeline as described by Monat et al.³⁹. Libraries containing a ~450-bp insert size and sequenced with 250-bp paired-end reads (PE450 libraries) were merged with BBMerge¹⁰⁰ and error-corrected with BFC¹⁰¹. Corrected libraries were used for iterative assembly with Minia3 (k -mer sizes: 100, 200, 250, 300, 350, 400, 450)¹⁰². Assembled unitigs were scaffolded with PE800, MP3, MP6 and MP9 data using SOAPdenovo¹⁰³. Internal gaps in scaffolds were closed with GapCloser¹⁰³. Chromosome-scale sequence scaffolds (pseudomolecules) were constructed with R scripts of the TRITEX pipeline using linkage information

afforded by 10X linked-reads, Hi-C data, flow-sorting data and a genetic map³⁸. As the genetic map of *Ae. sharonensis* was less dense than the POPSEQ maps of bread wheat¹⁰⁴ and barley¹⁰⁵, we modified the TRITEX workflow by adopting an iterative approach for pseudomolecule construction. In the first iteration, scaffolds were assigned to chromosomes using the Hi-C map and the genetic map as a guide. Hi-C data were then used to order scaffolds within chromosomes so that approximate chromosomal locations for most scaffolds were known. In the second iteration, this ordering of scaffolds was used to guide the construction of super-scaffolds with 10X linked reads, accepting only scaffold joins supported by both 10X reads and proximity in the Hi-C map. The 10X super-scaffolds were ordered along the chromosomes with Hi-C data. Flow-sorting data were used to corroborate Hi-C-based chromosome assignments and to correct errors. Finally, manual correction of chimeric scaffolds and refinements to the order and orientation were performed by visual inspection of Hi-C contact matrices. The gene content quality control analysis was conducted with BUSCO (v4.06, viridiplantae_orthodb10).

Initial Sr62 mapping with 803 family. We developed several F_{3,4} families from the cross between accession 2189 (susceptible) and accession 1644 (resistant)³⁸ by single-seed descent. One of these families, 803, segregated 3 resistant to 1 susceptible plant. Using this family, we developed a linkage map for Sr62 and using 192 plants, delimited Sr62 to between CAPS markers C24499_CAPS and C23635_CAPS (Supplementary Fig. 9).

Preliminary mapping was performed using genetic markers developed from the *Ae. sharonensis* 1644 genome that was based on 1.15e9 Illumina 100 bp paired-end reads⁷⁶. WGS of *Ae. sharonensis* 2189 (7.6e8 Illumina 100 bp paired-end reads) were aligned to the *Ae. sharonensis* 1644 genome using BWA (BWA version 0.7.17). Single nucleotide variations were identified using BCFtools (version 1.2) and filtered using vcutils.pl. False positive SNPs were minimized based on variations identified with self-alignment of *Ae. sharonensis* 1644 reads. A total of 2,608,758 SNPs were identified based on filtering for positions with unambiguous read support (i.e. homozygous variations). Using the barley consensus genetic map¹⁰⁶, putative orthologs were identified in the *Ae. sharonensis* 1644 genome and SNVs selected for the development of Sequenom assays. Putative single nucleotide polymorphisms between *Ae. sharonensis* accessions 1644 and 2189 were extracted with 80 bp flanking sequence.

These sequences were used as templates for primer design using MassARRAY software v3.1 for the multiplexing of two 28 SNP assays (a total of 56 SNP assays). Sequenom genotyping was carried out at the Iowa State University Genomic Technologies Facility (Ames, IA, USA). All SNPs and WGS contig source information for Sequenom markers are detailed in Supplementary Data 1.

High-resolution mapping of Sr62. We first screened 1,304 plants (2608 gametes) from the 803 family and identified 47 recombinants between the two flanking markers C24499_SBE and C23635_CAPS. These were restricted to six key recombinants with the STS marker C11837_CAPS. When the whole-genome shotgun sequencing and assembly became available, KASP (<https://www.biosearchtech.com/Supplementaryport/education/kasp-genotyping-reagents/how-does-kasp-work>) markers were developed by identifying SNPs between the genomic scaffold sequences of 1644 and those of the susceptible parent 2189³⁸. Two of these markers, C122784_KASP and C2468909_KASP, were used to screen the progeny of 3,342 plants (6684 gametes) derived from heterozygous 803 family individuals. All markers (Supplementary Tables 5, 6, 7 and 10) were designed using Primer3 Input (<http://bioinfo.ut.ee/primer3-0.4.0/>).

RNA extraction and Sr62 annotation. Total RNA was extracted from *Ae. sharonensis* accession 1644 with a RNeasy Plant Mini Kit (Cat No./ID: 74904, Qiagen, Valencia, CA, USA) following the manufacturer's protocol and digested with DNase (Roche). RNA-Seq was performed by Novogene. The RNA-Seq reads were trimmed with Trimmomatic (version 0.32, <http://www.usadellab.org/cms/?page=trimmomatic>). Hisat2¹⁰⁷ (version 2.1.0) was used to map the short reads onto the Sr62 reference sequence. The SAM output file was converted into a BAM file using SAMtools¹⁰⁸ (version 1.8) (<http://www.htslib.org/>) and sorted according to their position in the reference and indexed for visualization by IGV (version 2.8.13, <https://software.broadinstitute.org/software/igv/>). The full-length cDNA library was constructed using SMARTer[®] PCR cDNA Synthesis kit (Cat. # 634926, Clontech/TaKaRa) and sequenced on Illumina platform. The sequence reads were assembled using CLC Assembly Cell v 5.0.0 (<https://digitalinsights.qiagen.com/products-overview/discovery-insights-portfolio/analysis-and-visualization/qiagen-clc-assembly-cell/>).

Mutant development. To further determine the candidate gene for Sr62, we mutagenized 3,025 seeds of an introgression line containing Sr62 derived from *Ae. sharonensis* accession 1644 in the hexaploid bread wheat (*T. aestivum*) cultivar Zahir background (Zahir-1644)³⁷. Dry seeds were treated for 16 h with 200 ml of 0.75% EMS solution while being rolled on a Roller Mixer (Model SRT1, Stuart Scientific) to ensure maximum homogenous exposure of the seeds to the EMS. The excess solution was then removed, and the seeds were washed three times with 400 ml tap water. The M₁ seeds were grown in the greenhouse, and the seeds of M₂ families (single heads) were collected. Eight seeds per family were phenotyped with

Pgt isolate 04KEN156/04, race TTKSK. The M₃ seeds derived from susceptible M₂ plants were also tested to confirm that the M₂ susceptible plants were true mutants. To rule out seed contamination a subset of the mutants was verified using genotyping-by-sequencing (GBS)¹⁰⁹.

GBS data from the background (Zahir), donor (*Ae. sharonensis*, accession 1644) and introgression lines were mapped to the reference sequence of Chinese Spring¹¹⁰ using BWA mem (version 0.7.12) with standard parameters¹¹⁰. Mappings were sorted and converted to mpileup format using SAMtools¹⁰⁸ (version 0.1.19). The mpileup files were examined with a custom script to calculate the percentage of SNPs from the donor that were shared with the introgression line per given interval. Several interval lengths were tested; a clear signal was observed for 10 Mb.

Germplasm. Seeds of wheat cultivar Zahir (DPRM0080), the Zahir-1644 introgression lines 1S^{sh}S-1S^{sh}L-1BL (DPRM0081) and 1S^{sh}S-1S^{sh}L-1DL (DPRM0092), six of the EMS-induced mutants (DPRM0082, DPRM0083, DPRM0084, DPRM0085, DPRM0086, DPRM0087), *Ae. sharonensis* accession 1644 (DPRM0088) and the three transgenic lines (DPRM0089, DPRM0090, DPRM0091) are available from the Germplasm Resources Unit, John Innes Centre, Norwich, UK (<https://www.jic.ac.uk/research-impact/germplasm-resource-unit/>).

RNA mapping. Fourteen susceptible mutants derived from independent M₂ families were selected for RNA-seq. Total RNA was extracted from 14-day-old seedlings of the susceptible mutants and the wild-type Zahir-1644 parent line. The raw reads from the 14 mutants were mapped to the CDS of the seven genes from the Sr62 map interval using BWA¹¹⁰ (version 0.7.12) and SAMtools¹⁰⁸ (version 1.8). One CDS in the interval was identified as having a single nucleotide mutation in seven of the 14 mutants (Supplementary Table 9). All identified mutations were G-to-A or C-to-T transition mutations, which are typical of EMS mutagenesis. The coverage of the remaining two genes (remorin and WAK) was too low to call mutations (Supplementary Table 9 and Supplementary Data 2). Reads per kilobases (RPK) and transcript per million (TPM) were calculated by BLASTing FASTA files converted from the assorted BAM files¹⁰⁸ with the CDSs. The amino acid substitution tolerance/intolerance was analysed with THE web-based program SIFT¹¹¹ (version 6.2.1) https://sift.bii.a-star.edu.sg/www/SIFT_seq_submit2.html.

Formula development. We assume that there are n mutants in total and that k mutants have a mutation in the candidate gene CDS. The CDS length (in bp) is L_c and the mutation rate is R , and therefore the probability that the candidate CDS is present in a single mutant is RL_c . Thus, the probability (p_c) for the event, where k out of n mutants have a mutation in the candidate CDS, is a binomial distribution¹¹², expressed as:

$$p_c = \binom{n}{k} (RL_c)^k (1 - RL_c)^{n-k} \quad (1)$$

If there were m genes being investigated (such as those in the genetic mapping interval of positional cloning) and their CDS were the same length as the CDS of the candidate gene, the probability (p) that the event occurs at one or more of the genes being investigated would be:

$$p = 1 - (1 - p_c)^m \quad (2)$$

Because CDS lengths vary among genes, it is necessary to introduce the concept of the effective number (m_e). The effective number of the gene's CDS relative to the candidate gene's CDS is the total CDS length (L_i) of the genes divided by the length (L_c) of the candidate gene CDS.

$$m_e = \frac{L_i}{L_c} \quad (3)$$

Therefore, the probability of obtaining k out of n mutants with a mutation in the sequence of at least one of the genes being investigated is:

$$p = 1 - \left(1 - \binom{n}{k} (RL_c)^k (1 - RL_c)^{n-k}\right)^{\frac{L_i}{L_c}} \quad (4)$$

The CDS of most cloned plant disease resistance genes are ~2000 bp long^{24,25,27–29}, while the average CDS length in wheat is 1,000 bp⁷⁵. The probabilities displayed in Supplementary Fig. 21 were calculated based on the following assumptions: (i) the candidate gene has a 2000-bp CDS, (ii) every NLR has a 2000-bp CDS, (iii) the average CDS length of a gene on one chromosome or throughout the genome is 1000 bp, (iv) all mutants are assumed to have a mutation where the trait is controlled by one gene, and half of the mutants have a mutation where the trait is controlled by two genes. The probability for any candidate gene can be calculated given the specific genome mutation rate, candidate gene CDS length, total number of mutants, number of mutants with mutation in the CDS, and effective number of genes being investigated. In practice, the effective number can be approximated based on total number of genes being investigated by multiplying the average length (L_a) of the gene CDS and then dividing by the candidate-gene CDS length.

$$m_e = \frac{mL_a}{L_c} \quad (5)$$

Engineering Sr62 binary construct. The 5' and 3' halves of Sr62 containing 2782 bp of the putative promoter and 2026 bp of the putative terminator sequence, respectively, were PCR-amplified so as to leave out 11.4 kb of the third intron (Fig. 3a). The amplification was made with the high fidelity Q5 DNA polymerase (NEB, Ipswich, MA, USA) following the manufacturer's instructions. The PCR products were purified with QIAquick PCR Purification kit (QIAGEN, LLC, Germantown, MD 20874, USA). The purified fragments were tailed with nucleotide A using Taq DNA polymerase and cloned into the pCR2.1 vector (TOPO PCR Cloning Kits-K202020, Thermo Fisher Scientific). The positive clones were multiplied and the plasmid DNAs were digested with two pairs of restriction enzymes, NotI + EcoRI (NEB, Ipswich, MA, USA) for the 5' fragment and EcoRI + PmeI (NEB, Ipswich, MA, USA) for the 3' fragment. The two digested fragments were gel purified and then ligated into the binary vector pGGG-AH-NotI/PmeI¹¹³ linearized with NotI and PmeI in a three-way ligation using T4 ligase (M0202S, NEB, Ipswich, MA, USA). A positive clone, pGGG-Sr62, was bulked and verified by Sanger sequencing. pGGG-Sr62 is available from Addgene under the name pGGG-Sr1644-1Sh, accession number 164087.

Wheat transformation. The binary construct pGGG-Sr62 was used in Agrobacterium-mediated transformation of wheat cv. Fielder as described by Hayta et al.¹¹³. The copy number of the hygromycin selectable marker (as a proxy for the Sr62 copy number) in T₀ and T₁ plants was determined by iDNA Genetics (Norwich, UK) using qPCR, as described in Bartlett et al.¹¹⁴.

Evaluation of resistance to stem rust in transgenic wheat seedlings. The primary transgenic plants (T₀) were tested for stem rust response with the UK-01 isolate⁹ and T₁ plants were further tested with multiples isolates/races including TTKSK (isolate Ug99, 04KEN156/04 from Kenya), TKTTF (isolate 13-ETH18-1 from Ethiopia), TKTTF (isolate UK-01 from the United Kingdom), QTHJC (isolate 69MN399 from the USA), TKTSC (isolate #2079 from Israel), TTTTF (isolate #2127 from Israel), TTTTC (isolate #2135 from Israel), TTKTT (isolate KE184a/18, from Kenya), TKTTF (isolate ET11a/18, from Ethiopia), TTKTF (isolate, IT200a/18, from Italy), and TTRTF (isolate IT16a/18, from Italy) (Supplementary Table 11).

Homology searching. BLAST analysis against TGAC CS42 v1 gene models was performed using the Sr62 CDS as a query (<http://eg37-plants.ensembl.org>). The best hit (TRIAE_CS42_1BS_TGACv1_050726_AA0174710.1) was used for homology searching using Gramene, a comparative resource for plants^{50,51} (<https://www.gramene.org/>). The taxonomy for selected species/subspecies was retrieved from Taxonomy Browser (<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?tid=1437183>). The taxonomy was organized by Kingdom, Phylum, class, order and species. Each selected species was blasted with the Sr62 protein sequence or CDS at the NCBI webpage (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

Phylogenetic analysis of plant protein kinase domains. The CDS of the putative protein kinase genes previously extracted from the wheat cv. Chinese Spring reference genome (IWGSC, 2018) for phylogenetic analysis of Yr15²⁸ were used as queries for BLAST analysis against the Chinese Spring, durum, barley, rice, sorghum and maize genomes (<http://plants.ensembl.org/index.html>). For *T. monococcum*, *T. dicoccum*, *Ae. tauschii* and *Ae. sharonensis*, only the cloned gene sequences were added to the phylogenetic trees. The sequence of the Sr62 tandem kinase was determined based on the BLAST results with the CDS (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastx&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome, non-redundant protein sequence-nr). The retrieved CDS were used to perform BLAST analysis against the NCBI protein database (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastx&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome, non-redundant protein sequence-nr) and manually checked for the kinase domain. Genes with two or three complete protein kinase domains were selected for phylogenetic analysis. A total of 105 genes were used for analysis: 51 from Chinese Spring, 15 from durum, seven from barley, six from rice, 11 from sorghum, nine from maize and the six cloned Poaceae genes previously discussed (including Sr62). For kinase domain analysis, all 105 tandem kinase sequences were split into two or three kinase domain sequences and were used for phylogenetic analysis based on domain sequences. A phylogenetic tree (neighbour-joining tree) for whole genes and domains was computed with Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) and drawn with iTOL (<https://itol.embl.de/>). Hidden Markov Model classification was performed using hmmscan (v3.1b2). HMMs were obtained from the Supplementary Data set of Lehti-Shiu and Shiu⁴⁹.

3D modelling. The 3D protein structure model was constructed with the program Phyre2⁴⁴ (<http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>) and visualized with CCP4MG⁴⁵ (version 2.10.11).

Micro-synteny analysis of Sr62. The Sr62 CDS was used in a BLAST analysis to identify the putative orthologs in *T. aestivum*, *T. durum* (TRITD1Bv1G020740.1),

Ae. tauschii (AET1Gv20143300, and *H. vulgare* (HORVU1Hr1G011730). No clear Sr62 ortholog was detected in *T. aestivum* but a clear ortholog for Pm24 could be identified. The positions of these best hits were used to extract the proximal and distal genes with Gramene^{50,51} (<https://www.gramene.org/>). For rice, *B. distachyon*, maize, and sorghum, the best hit with the putative barley Sr62 ortholog, HORVU1Hr1G011730, was used for the synteny analysis. For *Ae. sharonensis*, the genes around Sr62 were manually annotated based on the RNA-Seq and full-length cDNA data for accession 1644.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The datasets generated during and/or analysed during the current study are publicly available as follows. The sequence reads and the genome assembly were deposited in the European Nucleotide Archive (ENA) under project number PRJEB40322 and PRJEB40049, respectively. The RNA-Seq data for AS_1644, the full-length AS_1644 cDNA library, and Zahir-1644 introgression line wild type and its 14 mutants was deposited at NCBI under project number PRJEB47173 and the transcriptome assembly is available from eDAL (<https://doi.org/10.5447/ipk/2021/21>). The Zahir-1644 and mutant GBS data have been deposited in ENA under project number PRJEB46949. The Sr62 gene and transcript sequence were deposited in NCBI Genbank under accession number MZ826707. The following public databases/datasets were used in the study: Chinese Spring reference genome (IWGSC, 2018), Gramene (<http://www.gramene.org/>), BLAST non-redundant protein sequence (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastx&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome), and Taxonomy Browser (<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=1437183>). Source data are provided with this paper.

Code availability

The scripts used in these analyses have been published in GitHub [https://github.com/steuernb/GBS_introgression_line_analysis] and linked with Zenodo¹¹⁵.

Received: 9 July 2021; Accepted: 24 February 2022;

Published online: 25 March 2022

References

- Peterson, P. D. *Stem Rust of Wheat: From Ancient Enemy to Modern Foe*. Am. Phytopathol. Soc. (APS Press, St. Paul, 2001).
- Le Roux, J. & Rijkenberg, F. H. Occurrence and pathogenicity of *Puccinia graminis* f. sp. *tritici* in South Africa during the period 1981–1985. *Phytophylactica* **19**, 467–472 (1987).
- Addai, D. et al. Potential economic impacts of the wheat stem rust strain Ug99 in Australia, ABARES research report, prepared for the Plant Biosecurity Branch, Department of Agriculture and Water Resources, Canberra (2018).
- Pretorius, Z. A., Singh, R. P., Wagoire, W. W. & Payne, T. S. Detection of virulence to wheat stem rust resistance gene Sr31 in *Puccinia graminis* f. sp. *tritici* in Uganda. *Plant Dis.* **84**, 203 (2000).
- 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).
- 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).
- Olivera, P. et al. Phenotypic and genotypic characterization of race TKTTF of *Puccinia graminis* f. sp. *tritici* that caused a wheat stem rust epidemic in southern Ethiopia in 2013–14. *Phytopathology* **105**, 917–928 (2015).
- Olivera Firpo, P. D. et al. Characterization of *Puccinia graminis* f. sp. *tritici* isolates derived from an unusual wheat stem rust outbreak in Germany in 2013. *Plant Pathol.* **66**, 1258–1266 (2017).
- Lewis, C. M. et al. Potential for re-emergence of wheat stem rust in the United Kingdom. *Commun. Biol.* **1**, 13 (2018).
- Bhattacharya, S. Deadly new wheat disease threatens Europe's crops. *Nature* **542**, 145–146 (2017).
- Shamanin, V. et al. Genetic diversity of spring wheat from Kazakhstan and Russia for resistance to stem rust Ug99. *Euphytica* **212**, 287–296 (2016).
- Prank, M., Kenaley, S. C., Bergstrom, G. C., Acevedo, M. & Mahowald, N. M. Climate change impacts the spread potential of wheat stem rust, a significant crop disease. *Environ. Res. Lett.* **14**, 124053 (2019).
- Hafeez, A. N. et al. Creation and judicious application of a wheat resistance gene atlas. *Mol. Plant* **14**, 1053–1070 (2021).
- Moore, J. W. et al. A recently evolved hexose transporter variant confers resistance to multiple pathogens in wheat. *Nat. Genet.* **47**, 1494–1498 (2015).
- Krattinger, S. G. et al. A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. *Science* **323**, 1360–1363 (2009).

16. 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–E9492 (2017).
17. 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).
18. Steuernagel, B. et al. Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture. *Nat. Biotechnol.* **34**, 652–655 (2016).
19. Zhang, J. et al. A recombinant *Sr26* and *Sr61* disease resistance gene stack in wheat encodes unrelated NLR genes. *Nat. Commun.* **12**, 3378 (2021).
20. Periyannan, S. et al. The gene *Sr33*, an ortholog of barley *Mla* genes, encodes resistance to wheat stem rust race Ug99. *Science* **341**, 786–789 (2013).
21. Saintenac, C. et al. Identification of wheat gene *Sr35* that confers resistance to Ug99 stem rust race group. *Science* **341**, 783–786 (2013).
22. Arora, S. et al. Resistance gene cloning from a wild crop relative by sequence capture and association genetics. *Nat. Biotechnol.* **37**, 139–143 (2019).
23. Mago, R. et al. The wheat *Sr50* gene reveals rich diversity at a cereal disease resistance locus. *Nat. Plants* **1**, 15186 (2015).
24. Chen, S. et al. Wheat gene *Sr60* encodes a protein with two putative kinase domains that confers resistance to stem rust. *N. Phytol.* **225**, 948–959 (2020).
25. Gaurav, K. et al. Population genomic analysis of *Aegilops tauschii* identifies targets for bread wheat improvement. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-021-01058-4> (2021).
26. Klymiuk V., Coaker G., Fahima T., Pozniak C. Tandem protein kinases emerge as new regulators of plant immunity. *Mol. Plant Microbe. Interact.* <https://doi.org/10.1094/MPMI-03-21-0073-CR> (2021).
27. Brueggeman, R. et al. The barley stem rust-resistance gene *Rpg1* is a novel disease-resistance gene with homology to receptor kinases. *Proc. Natl Acad. Sci. USA* **99**, 9328–9333 (2002).
28. Klymiuk, V. et al. Cloning of the wheat *Yr15* resistance gene sheds light on the plant tandem kinase-pseudokinase family. *Nat. Commun.* **9**, 3735 (2018).
29. Lu, P. et al. A rare gain of function mutation in a wheat tandem kinase confers resistance to powdery mildew. *Nat. Commun.* **11**, 680 (2020).
30. Olivera, P. D. & Steffenson, B. J. *Aegilops sharonensis*: Origin, genetics, diversity, and potential for wheat improvement. *Botany* **87**, 740–756 (2009).
31. Olivera, P. D., Anikster, Y., Kolmer, J. A. & Steffenson, B. J. Resistance of Sharon goatgrass (*Aegilops sharonensis*) to fungal diseases of wheat. *Plant Dis.* **91**, 942–950 (2007).
32. Scott, J. C., Manisterski, J., Sela, H., Ben-Yehuda, P. & Steffenson, B. J. Resistance of *Aegilops* species from Israel to widely virulent African and Israeli races of the wheat stem rust pathogen. *Plant Dis.* **98**, 1309–1320 (2014).
33. Tsujimoto, H. & Tsunewaki, K. Gametocidal genes in wheat and its relatives. I. Genetic analyses in common wheat of a gametocidal gene derived from *Aegilops speltoides*. *Can. J. Genet. Cytol.* **26**, 78–84 (1984).
34. Marais, G. F., McCallum, B. & Marais, A. S. Leaf Rust and Stripe Rust Resistance Genes Derived from *Aegilops Sharonensis*. *Euphytica* **149**, 373–380 (2006).
35. Knight, E. et al. Mapping the 'breaker' element of the gametocidal locus proximal to a block of sub-telomeric heterochromatin on the long arm of chromosome 4S^{sh} of *Aegilops sharonensis*. *Theor. Appl. Genet.* **128**, 1049–1059 (2015).
36. Millet, E. et al. Introgression of leaf rust and stripe rust resistance from Sharon goatgrass (*Aegilops sharonensis* Eig) into bread wheat (*Triticum aestivum* L.). *Genome* **57**, 309–316 (2014).
37. Millet, E. et al. Genome targeted introgression of resistance to African stem rust from *Aegilops sharonensis* into bread wheat. *Plant Genome* **10**, 1–11 (2017).
38. Yu, G. et al. Discovery and characterization of two new stem rust resistance genes in *Aegilops sharonensis*. *Theor. Appl. Genet.* **130**, 1207–1222 (2017).
39. Monat, C. et al. TRITEX: chromosome-scale sequence assembly of Triticeae genomes with open-source tools. *Genome Biol.* **20**, 284 (2019).
40. International Wheat Genome Sequencing Consortium (IWGSC). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **361**, eaar7191 (2018).
41. Elshire, R. J. et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **6**, e19379 (2011).
42. Uauy, C., Wulff, B. H. & Dubcovsky, J. Combining traditional mutagenesis with wew high-throughput sequencing and genome editing to reveal hidden variation in polyploid wheat. *Annu Rev. Genet.* **51**, 435–454 (2017).
43. Steuernagel, B. et al. The NLR-annotator tool enables annotation of the intracellular immune receptor repertoire. *Plant Physiol.* **183**, 468–482 (2020).
44. Kelley, L. et al. The Phyre2 web portal for protein modelling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
45. McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. M. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Cryst.* **D67**, 386–394 (2011).
46. Holliday, G. L., Mitchell, J. B. & Thornton, J. M. Understanding the functional roles of amino acid residues in enzyme catalysis. *J. Mol. Biol.* **390**, 560–577 (2009).
47. 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).
48. Needleman, S. B. & Wunsch, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**, 443–453 (1970).
49. Lehti-Shiu, M. D. & Shiu, S.-H. Diversity, classification and function of the plant protein kinase superfamily. *Philos. Trans. R. Soc. B* **367**, 2619–2639 (2012).
50. Tello-Ruiz, M. K. et al. Gramene 2021: Harnessing the power of comparative genomics and pathways for plant research. *Nucleic Acids Res* **49**, D1452–D1463 (2021).
51. Howe, K. L. et al. Ensembl Genomes 2020—enabling non-vertebrate genomic research. *Nucleic Acids Res.* **48**, D689–D695 (2020).
52. Chalupska, D. et al. Acc homoeoloci and the evolution of wheat genomes. *Proc. Natl Acad. Sci. USA* **105**, 9691–9696 (2008).
53. Kourelis, J. & van der Hoorn, R. A. L. Defended to the Nines: 25 Years of resistance gene cloning identifies nine mechanisms for R protein function. *Plant Cell* **30**, 285–299 (2018).
54. Van der Biezen, E. A. & Jones, J. D. G. Plant disease-resistance proteins and the gene-for-gene concept. *Trends Plant Sci.* **23**, 454–456 (1998).
55. Jones, J. D. G. & Dangl, J. L. The plant immune system. *Nature* **444**, 323–329 (2006).
56. DeYoung, B. J. & Innes, R. W. Plant NBS-LRR proteins in pathogen sensing and host defense. *Nat. Immunol.* **7**, 1243–1249 (2006).
57. Cesari, S. Multiple strategies for pathogen perception by plant immune receptors. *N. Phytol.* **219**, 17–24 (2018).
58. van Wersich, S., Tian, L., Hoy, R. & Li, X. Plant NLRs: the whistleblowers of plant immunity. *Plant Commun.* **1**, 100016 (2020).
59. 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).
60. Mackey, D., Holt, B. F. 3rd, Wiig, A. & Dangl, J. L. RIN4 interacts with *Pseudomonas syringae* type III effector molecules and is required for RPM1-mediated resistance in Arabidopsis. *Cell* **108**, 743–754 (2002).
61. Kim, M. G. et al. Two *Pseudomonas syringae* type III effectors inhibit RIN4-regulated basal defense in Arabidopsis. *Cell* **121**, 749–759 (2005).
62. Hofius, D. et al. Autophagic components contribute to hypersensitive cell death in Arabidopsis. *Cell* **137**, 773–783 (2009).
63. Wu, A. J., Andriotis, V. M., Durrant, M. C. & Rathjen, J. P. A patch of surface-exposed residues mediates negative regulation of immune signalling by tomato Pto kinase. *Plant Cell* **16**, 2809–2821 (2004).
64. Mucyn, T. S. et al. The tomato NBARC-LRR protein Prf interacts with Pto kinase in vivo to regulate specific plant immunity. *Plant Cell* **18**, 2792–2806 (2006).
65. Martin, G. B. et al. Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science* **262**, 1432–1435 (1993).
66. Wang, G. et al. The decoy substrate of a pathogen effector and a pseudokinase specify pathogen-induced modified-self recognition and immunity in plants. *Cell Host Microbe* **18**, 285–295 (2015).
67. Lehti-Shiu, M. D., Zou, C., Hanada, K. & Shiu, S. H. Evolutionary history and stress regulation of plant receptor-like kinase/Pelle genes. *Plant Phys.* **150**, 12–26 (2009).
68. Mascher, M. et al. A chromosome conformation capture ordered sequence of the barley genome. *Nature* **544**, 427–433 (2017).
69. Luo, M. C. et al. Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. *Nature* **551**, 498–502 (2017).
70. Maccacferri, M. et al. Durum wheat genome highlights past domestication signatures and future improvement targets. *Nat. Genet.* **51**, 885–895 (2019).
71. Ouyang, S. et al. The TIGR rice genome annotation resource: improvements and new features. *Nucleic Acids Res.* **35** (Database issue), D883–D887 (2007).
72. Beier, S. et al. Construction of a map-based reference genome sequence for barley, *Hordeum vulgare* L. *Sci. Data* **4**, 170044 (2017).
73. Eilam, T. et al. Genome size and genome evolution in diploid Triticeae species. *Genome* **50**, 1029–1037 (2007).
74. International Barley Genome Sequencing Consortium (IBGSC). A physical, genetic and functional sequence assembly of the barley genome. *Nature* **491**, 711–716 (2012).
75. Ling, H. Q. et al. Genome sequence of the progenitor of wheat A subgenome *Triticum urartu*. *Nature* **557**, 424–428 (2018).
76. Marcussen, T. et al. Ancient hybridizations among the ancestral genomes of bread wheat. *Science* **345**, 1250092 (2014).
77. 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).

78. 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.* <https://doi.org/10.1111/tbj.15664> (2022).
79. Tsujimoto, H. Gametocidal genes in wheat and its relatives. IV. Functional relationships between six gametocidal genes. *Genome* **38**, 283–289 (1995).
80. Friebe, B. et al. Characterization of a knock-out mutation at the *Gc2* locus in wheat. *Chromosoma* **111**, 509–517 (2003).
81. Grewal, S. et al. Comparative mapping and targeted-capture sequencing of the gametocidal loci in *Aegilops sharonensis*. *Plant Genome* **10**, 2 (2017).
82. Sánchez-Martín, J. et al. Rapid gene isolation in barley and wheat by mutant chromosome sequencing. *Genome Biol.* **17**, 221 (2016).
83. Dracatos, P. M. et al. The coiled-coil NLR *Rph1*, confers leaf rust resistance in barley cultivar Sudan. *Plant Physiol.* **179**, 1362–1372 (2019).
84. 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).
85. Xing, L. et al. *Pm21* from *Haynaldia villosa* encodes a CC-NBS-LRR protein conferring powdery mildew resistance in wheat. *Mol. Plant* **4**, 874–878 (2018).
86. Wulff, B. B. H. & Dhugga, K. S. Wheat—the cereal abandoned by GM. *Science* **361**, 451–452 (2018).
87. Argentina first to market with drought-resistant GM wheat. *Nat. Biotechnol.* **39**, 652 (2021).
88. Crop Solutions Announces Regulatory Approval of Drought Tolerant HB4® Wheat by Brazil's CTNBio, <https://www.businesswire.com/news/home/20211111006003/en/Bioceres-Crop-Solutions-Announces-Regulatory-Approval-of-Drought-Tolerant-HB4%C2%AE-Wheat-by-Brazil%E2%80%99s-CTNBio> (2021).
89. Stakman, E. C. Barberry eradication prevents black rust in Western Europe. *U. S. Dep. Agriculture, Dep. Circular* **269**, 1–15 (1923).
90. Yu, G., Hatta, A., Periyannan, S., Lagudah, E. & Wulff, B. B. H. Isolation of wheat genomic DNA for gene mapping and cloning. *Methods Mol. Biol.* **1659**, 207–213 (2017).
91. Jupe, F. et al. The complex architecture and epigenomic impact of plant T-DNA insertions. *PLoS ONE* <https://doi.org/10.1371/journal.pgen.1007819> (2019).
92. Vrána, J. et al. Flow sorting of mitotic chromosomes in common wheat (*Triticum aestivum* L.). *Genetics* **156**, 2033–2041 (2000).
93. Kubaláková, M. et al. Chromosome sorting in tetraploid wheat and its potential for genome analysis. *Genetics* **170**, 823–829 (2005).
94. Giorgi, D. et al. FISHIS: fluorescence in situ hybridization in suspension and chromosome flow sorting made easy. *PLoS ONE* **8**, e57994 (2013).
95. Kubaláková, M., Macas, J. & Doležel, J. Mapping of repeated DNA sequences in plant chromosomes by PRINS and C-PRINS. *Theor. Appl. Genet.* **94**, 758–763 (1997).
96. Molnár, I. et al. Dissecting the U, M, S and C genomes of wild relatives of bread wheat (*Aegilops* spp.) into chromosomes and exploring their synteny with wheat. *Plant J.* **88**, 452–467 (2016).
97. Zhang, Y., Zhu, M.-L. & Dai, S.-L. Analysis of karyotype diversity of 40 Chinese chrysanthemum cultivars. *J. Systematics. Evolution* **51**, 335–352 (2013).
98. Badaeva, E. D., Friebe, B. & Gill, B. S. Genome differentiation in *Aegilops*. 1. Distribution of highly repetitive DNA sequences on chromosomes of diploid species. *Genome* **39**, 293–306 (1996).
99. Šimková, H. et al. Coupling amplified DNA from flow-sorted chromosomes to high-density SNP mapping in barley. *BMC Genomics* **9**, 294 (2008).
100. Bushnell, B., Rood, J. & Singer, E. BBMerge—accurate paired shotgun read merging via overlap. *PLoS ONE* **12**, e0185056 (2017).
101. Li, H. BFC: correcting Illumina sequencing errors. *Bioinformatics* **31**, 2885–2887 (2015).
102. Chikhi, R., Limasset, A. & Medvedev, P. Compacting de Bruijn graphs from sequencing data quickly and in low memory. *Bioinformatics* **32**, i201–i208 (2016).
103. Luo, R. et al. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *Gigascience* **1**, 18 (2012).
104. Chapman, J. A. et al. A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. *Genome Biol.* **16**, 26 (2015).
105. Mascher, M. et al. Barley whole exome capture: a tool for genomic research in the genus *Hordeum* and beyond. *Plant J.* **76**, 494–505 (2013).
106. Muñoz-Amatriáin, M. et al. An improved consensus linkage map of barley based on flow-sorted chromosomes and single nucleotide polymorphism markers. *Plant Genome* **4**, 238–249 (2011).
107. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
108. Li, H. et al. 1000 Genome Project Data Processing Subgroup. The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
109. Poland, J. A. & Rife, T. W. Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome* **5**, 92–102 (2012).
110. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* **25**, 1754–1760 (2009).
111. Sim, N. L. et al. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* **40**, W452–W457 (2012).
112. Wypij, D. In *Wiley StatsRef: Statistics Reference Online* (eds N. Balakrishnan et al.). <https://doi.org/10.1002/9781118445112.stat04852> (2014).
113. Hayta, S. et al. An efficient and reproducible *Agrobacterium*-mediated transformation method for hexaploid wheat (*Triticum aestivum* L.). *Plant Methods* **15**, 121 (2019).
114. Bartlett, J. G. et al. High-throughput *Agrobacterium*-mediated barley transformation. *Plant Methods* **4**, 22 (2008).
115. Yu, G. et al. Reference genome-assisted identification of stem rust resistance gene *Sr62* encoding a tandem kinase. Zenodo, <https://zenodo.org/record/394326594> (2022).

Acknowledgements

We are grateful to the Harold and Adele Lieberman Germplasm Bank (Tel Aviv University), Hazera Seeds Ltd and Limagrain for making germplasm available. We thank Ryan Johnson for phenotyping some of the *Sr62* introgression line mutants, Yue Jin for supplying several of the stem rust cultures, JIC Horticultural Services for plant husbandry, Matthew Heaton for assistance with figure design, Manuela Knauff and Ines Walde for technical assistance on Hi-C library preparation and sequencing, Anne Fiebig for sequence data submission, Jan Vrána, Zdeňka Dubská and Romana Šperková for assistance with chromosome sorting and DNA amplification, Bob McIntosh for review of the draft manuscript, and Willem Boshoff for helpful discussion. This research was supported by the NBI Computing Infrastructure for Science (CiS) group and financed by grants from the 2Blades Foundation, USA, to B.J.S. and B.B.H.W.; the Biotechnology and Biological Sciences Research Council (BBSRC) Designing Future Wheat Cross-Institute Strategic Programme to B.B.H.W. (BBS/E/J/000PR9780); the Lieberman-Okinow Endowment at the University of Minnesota to B.J.S.; Human Frontier Science Program long-term fellowship (LT000218/2011-L) to M.J.M.; the Gordon and Betty Moore Foundation through grant GBMF4725 to the 2Blades Foundation; and the Gatsby Charitable Foundation to J.D.G.J.

Author contributions

Extracted DNA for whole-genome shotgun sequencing and long mate-pair libraries: G.Y. Performed 10X Chromium sequencing: A.M.D., J.E. and C.P.; performed Hi-C sequencing: A.H. and N.S. Performed chromosome flow sorting and amplification: M.K., I.M. and J.D. Assembled the *Ae. sharonensis* genome: M.M.; conducted BUSCO analysis: M.J. Rough-mapped *Sr62*: G.Y., N.C., M.J.M., I.H.P., J.S. and P.D.O.; fine-mapped *Sr62*: G.Y. and O.M. Generated Zahir-1644: E.M. Generated the mutant population: G.Y.; screened for mutants: C.W. and B.J.S.; genotyped mutants: J.P., B.S. and S.W.; developed the formula and calculated the minimum number of mutants required for gene identification: G.Y. Extracted RNA for AS_1644 annotation and MutRNA-Seq: G.Y. Performed bioinformatics analyses to identify *Sr62* and determined its gene structure: G.Y. Engineered the binary construct: G.Y.; developed the binary vector: M.S.; transformed it into wheat: S.H. and W.H.; phenotyped the transgenic lines: G.Y., O.M., N.K., M.R., B.J.S., M.P. and A.J. Performed phylogenetics, synteny and 3D modelling analyses: G.Y. and M.B. Provided scientific support: C.G., Y.Y., R.A. and P.G. Conceived study: B.W., B.J.S., J.J., A.S., E.W. and T.L.R.; drafted manuscript: G.Y. and B.W. All co-authors read and approved the final manuscript. Authors are grouped by institution in the author list, except for the first four and last four authors.

Competing interests

G.Y. and B.B.H.W. are inventors on a US provisional patent application 63/250,413 filed by 2Blades and relating to the use of *Sr62* for stem rust resistance in transgenic wheat. T.L.R. and E.R.W. were employed by the 2Blades Foundation, and E.R.W. continues to serve on the 2Blades board. Both were involved in the conceptualization and design of the research presented, which was cofunded by the 2Blades Foundation. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-29132-8>.

Correspondence and requests for materials should be addressed to Brian J. Steffenson or Brande B. H. Wulff.

Peer review information *Nature Communications* thanks Bruno Contreras-Moreira, Tzion Fahima, Frank You and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022