The wheat stem rust resistance gene *Sr43* encodes an unusual protein kinase

Guotai Yu
KAUST https://orcid.org/0000-0002-8447-2886
Oadi Matny
University of Minnesota
Spyridon Gourdoupis
KAUST https://orcid.org/0000-0001-9396-4033
Ryan Johnson
University of Minnesota
Fatimah R. Aljedaani
KAUST
Ikram Blilou
KAUST https://orcid.org/0000-0001-8003-3782
Catherine H. Gardener
KAUST
Yajuan Yue
John Innes Centre
Ngoni Kangara
John Innes Centre https://orcid.org/0000-0002-8381-0118
Burkhard Steuernagel
John Innes Centre https://orcid.org/0000-0002-8284-7728
Sadiye Hayta
John Innes Centre https://orcid.org/0000-0003-2920-4720
Mark Smedley
John Innes Centre https://orcid.org/0000-0002-6162-4795
Wendy Harwood
John Innes Centre https://orcid.org/0000-0002-4323-9009
Mehran Patpour
Aarhus University https://orcid.org/0000-0002-4914-2697
Shuangye Wu
Kansas State University
Jesse A. Poland
KAUST
Jonathan Jones
The Sainsbury Laboratory https://orcid.org/0000-0002-4953-261X
Lynne Reuber
2Blades Foundation https://orcid.org/0000-0001-7806-2437

Moshe Ronen
Tel Aviv University
Amir Sharon
Tel Aviv University https://orcid.org/0000-0002-6045-9169
Matthew Rouse
University of Minnesota
Steven Xu
USDA-ARS https://orcid.org/0000-0001-7000-2034
Kateřina Holušová
Institute of Experimental Botany of the Czech Academy of Sciences https://orcid.org/0000-0002-4531-7170
Jan Bartos
Institute of Experimental Botany of the Czech Academy of Sciences https://orcid.org/0000-0002-4154-8895
István Molnár
Institute of Experimental Botany of the Czech Academy of Sciences https://orcid.org/0000-0002-7167-9319
Miroslava Karafiátová
Institute of Experimental Botany of the Czech Academy of Sciences https://orcid.org/0000-0003-1177-6472
Łukasz Jaremko
KAUST https://orcid.org/0000-0001-7684-9359
Jaroslav Doležel
Institute of Experimental Botany of the Czech Academy of Sciences https://orcid.org/0000-0002-6263-0492
Brian Steffenson (**bsteffen@umn.edu**)
University of Minnesota https://orcid.org/0000-0001-7961-5363
Brande B. H. Wulff (**brande.wulff@kaust.edu.sa**)
KAUST https://orcid.org/0000-0003-4044-4346

Research Article

Keywords: Stem rust, kinase, wheat, resistance gene

Posted Date: July 5th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1820134/v1

Abstract

To safeguard bread wheat against pests and diseases, breeders have introduced over 200 resistance genes into its genome, thus nearly doubling the number of designated resistance genes in the wheat gene pool. Isolating these genes facilitates their fast-tracking in breeding programs and incorporation into polygene stacks for more durable resistance. We cloned the stem rust resistance gene *Sr43*, which was crossed into bread wheat from the wild grass *Thinopyrum elongatum*. *Sr43* encodes a protein kinase fused to two domains of unknown function. The gene, which is unique to the Triticeae, appears to have arisen through a gene fusion event 6.7 to 11.6 million years ago. Transgenic expression of *Sr43* in wheat conferred high levels of resistance to a wide range of isolates of the pathogen causing stem rust, highlighting the potential value of *Sr43* in resistance breeding and engineering.

Main

Worldwide, ~ 20% of projected bread wheat (*Triticum aestivum*) production is lost to pests and disease every year[1]. The deployment of genetic variation for disease resistance is the most sustainable and environmentally friendly way to protect wheat crops[2]. For over 100 years, breeders have conducted numerous crosses to enrich the wheat gene pool with novel resistance genes. Notably, more than 200 of the 467 currently designated resistance genes in bread wheat have their origin outside of the bread wheat gene pool[3]. However, the deployment of these interspecific resistance genes is often hampered by linkage drag, that is, the co-introduction of deleterious alleles from linked genes. Moreover, single resistance genes tend to be rapidly overcome by the emergence of resistance-breaking pathogen strains[4]. Cloning individual resistance genes would enable their introduction as genetically modified polygene stacks, which are likely to provide more durable resistance[5].

Most of the ~ 283 plant disease resistance genes cloned to date encode either intracellular receptors of the nucleotide binding and leucine-rich repeat (NLR) class or extracellular membrane-anchored receptor-like proteins (RLPs, called RLKs when they contain an intracellular kinase) (Supplementary Table 1)[3,6]. A new group of resistance genes has recently come to light, whose members encode two protein kinases fused as one protein. These tandem kinase genes include *Rpg1*, *Yr15*, *Sr60*, *Sr62*, *Pm24*, *WTK4*, and *Rwt4* (refs. [7–13]). Other resistance genes offer some variation to this architecture with protein kinases fused to a steroidogenic acute regulatory protein-related lipid transfer domain (*Yr36*)[14], a C2 domain and a multi-transmembrane region (*Pm4*)[15], a major sperm protein (*Snn3*)[16], an NLR (*Tsn1*, *Rpg5*, and *Sm1*)[17,18,19], and a von Willebrand factor type A domain (*Lr9*)[20].

These kinase fusion protein-encoding resistance genes appear to be unique to the Triticeae, the clade of grasses that arose 12 million years ago and encompasses the cereals wheat, rye (*Secale cereale*) and barley (*Hordeum vulgare*)[21]. However, the fusion events that gave rise to these genes, far from being rare and isolated, happened multiple times between different classes of kinases and spawned diverse combinations[8,10]. This genomic innovation resulted in resistance against phylogenetically distinct fungal pathogens spanning the ~ 300-million-year-old ascomycete/basidiomycete divide.

Here, we cloned the stem rust resistance gene *Sr43*, which was transferred from tall wheatgrass (*Thinopyrum elongatum*) into bread wheat 45 years ago[22,23]. The dominant resistance gene *Sr43* was introgressed into chromosome 7D of hexaploid wheat (Fig. 1a,b). We mutagenized grains of the *Sr43* introgression line with ethyl methanesulfonate (EMS) and screened 2,244 surviving M_2 families for susceptibility to *Puccinia graminis* f. sp.

tritici (*Pgt*). We identified 23 families segregating for stem rust susceptibility, of which we confirmed 10 independent mutants by progeny testing (Supplementary Table 2) and genotyping (Supplementary Fig. 1 to 11).

To clone *Sr43*, we performed chromosome flow sorting and sequenced the wheat-*Th. elongatum* recombinant chromosome 7D in the parental line and eight mutants (Extended Data Fig. 1 and Supplementary Tables 3 and 4). Sequence-assembly of the parental line and mapping of the mutant reads identified a 10 kb-window in a scaffold containing a mutation in all eight mutants. To determine the gene structure of the *Sr43* candidate, we (i) conducted transcriptome deep sequencing (RNA-seq) analysis of *Sr43* seedling leaves and mapped the reads to the *Sr43* genomic scaffold (Extended Data Fig. 2a), and (ii) sequenced *Sr43* clones obtained by PCR from a full-length cDNA library. We detected four different splice variants (Extended Data Fig. 2b and Supplementary Tables 5 and 6). Splice variant 1 contained all eight mutations and consisted of 18 exons with a predicted open reading frame of 2,598 bp (Fig. 1c). The eight mutations were all G/C to A/T transitions typical of EMS mutagenesis and introduced non-synonymous changes (seven mutants) or an early-stop codon in the predicted coding sequence (Fig. 1c,d; Supplementary Tables 7 and 8). The probability that all mutants would have a mutation in the same gene by chance alone out of the 5,822 non-redundant genes of chromosome 7D[24] was $4x10^{-6}$, indicating that the identified gene is a good candidate for *Sr43*.

As all identified EMS mutations affected the predominant full-length *Sr43* transcript, (Fig. 1c), we used its predicted 866–amino acid sequence to search for functional domains and homologs. We determined that Sr43 harbors an N-terminal kinase domain and two domains of unknown function (DUFs) in its C terminus (Fig. 1d). Five of the sequenced mutations resided within the kinase domain, with the remaining three mutations affecting either DUF (Fig. 1d).

The closest homolog of the Sr43 kinase domain was the serine/threonine kinase interleukin-1 receptor associated kinase (STKc IRAK) (Supplementary Fig. 12), which indicates that the Sr43 kinase is conserved between animals and plants. Further homology searches suggested that the kinase domain is intact (Supplementary Fig. 13). Mutant 1013a disrupted one of the conserved glycine residues in the glycine-rich loop, suggesting that kinase activity is required for Sr43 function (Supplementary Table 9). The C terminus of Sr43, containing DUF3475 and DUF668, has a similar domain architecture (44% identity) as the N terminus of PHYTOSULFOKINE SIMULATOR (PSI) proteins from Arabidopsis (*Arabidopsis thaliana*), which are critical for plant growth[26]. Unlike Arabidopsis PSI1, Sr43 lacked a putative nuclear localization signal or a putative myristoylation site. Sr43 had no transmembrane domain, as predicted by InterPro. However, we established that Sr43 likely localizes to the nucleus and plastids, as evidenced by the fluorescence detected from the transient expression of a *Sr43-GFP* (green fluorescent protein) construct in *Nicotiana benthamiana* leaf epidermal cells (Extended data Fig. 3).

The domain structure of Sr43 was thus clearly different from that of proteins encoded by the ~ 283 cloned plant resistance genes, which were largely (78%) extracellular or intracellular immune receptors (Supplementary Table 1). To explore the unusual structure of Sr43 in more detail, we used the AlphaFold artificial intelligence–augmented system to generate a 3D model[27] (Supplementary Data 1). We determined that Sr43 adopts a modular structure, with the kinase and the two DUFs separated by flexible linker loops (Fig. 1e). The kinase domain contained α-helices and anti-parallel β-strands, whereas the DUFs were entirely alpha-helical. We compared the predicted structure of the Sr43 protein to those in the Protein Data Bank[28], which predicted a protein kinase–like structure for DUF668. Therefore, we searched for ATP binding sites using the small molecule docking program HADDOCK[25] and identified one high-confidence ATP-binding site in DUF668 (Fig. 1f, Supplementary Table 10, and Supplementary Data 2).

We cloned a 14-kb genomic *Sr43* fragment including 3.2 kb of upstream and 2.5 kb of downstream regulatory sequence (Fig. 2a) (Supplementary Table 11) and introduced the resulting binary construct into the wheat cultivar Fielder. We obtained one primary (T_0) transgenic plant, which carried three copies of the transgene, based on qPCR of the hygromycin selectable marker. We tested homozygous T_1 and T_2 lines against a geographically and phenotypically diverse panel of 11 *Pgt* isolates from North America, the Middle East, Europe, and Africa. In 10 cases, the *Sr43* transgenic and wild-type introgression lines were resistant, whereas the cultivars Chinese Spring (the introgression parent) and Fielder were susceptible (Fig. 2b, Extended Data Fig. 4a,b, and Supplementary Table 12). By contrast, the *Pgt* isolate 75ND717C was intermediately virulent on the *Sr43* introgression and transgenic lines (Fig. 2b). For *Pgt* isolate 69MN399, we compared the phenotype at 21°C and 26°C and noticed a marked reduction in *Sr43*-mediated resistance at the higher temperature, in line with previous observations[29] (Fig. 2c). Taken together, these results confirm (i) the broad-spectrum efficacy of *Sr43* (ref. [29]), (ii) that a 14-kb *Sr43* genomic fragment is sufficient for function, and (iii) that the transgenic line faithfully recapitulates the race-specific and temperature-sensitive resistance of wild-type *Sr43*.

We searched for Sr43 homologs to investigate its evolutionary origin. We identified proteins harboring either the kinase domain or the two DUFs alone across the Poaceae family spanning 60 million years of evolution (Supplementary Tables 13 and 14; Extended Data Figs. 5 and 6). We detected the Sr43 protein domain arrangement only within the *Thinopyrum*, *Triticum*, *Aegilops* and *Secale* genera of the Triticeae tribe, but not within *Hordeum*, suggesting that *Sr43* likely arose between 6.7 and 11.6 million years ago (Fig. 3 and Supplementary Table 15). In those lineages lacking a clear *Sr43* homolog, we mapped genes encoding the kinase and DUFs present in Sr43 to different chromosomes (e.g., *Sorghum bicolor*, *Zea mays*, *T. urartu* and *Ae. sharonensis*) or on the same chromosome but 6 to 36 Mb apart (*Ae. tauschii* and *Setaria italica*), suggesting that the recruitment of the kinase domain to the DUFs at the *Sr43* locus involved an ectopic recombination event (Supplementary Table 15). In *Thinopyrum elongatum*, the ancestral state and *Sr43* were retained as an intraspecies polymorphism; some species of *Aegilops* and *Triticum* retained the ancestral state (e.g., *Ae. tauschii*), whereas others retained the *Sr43* innovation (e.g., the *T. aestivum* and *T. durum* B genomes) (Fig. 3).

In summary, we cloned the wheat stem rust resistance gene *Sr43*, which encodes a protein kinase fused to two DUFs. Of the 68 Triticeae resistance genes cloned to date, most encode NLRs (n = 41), followed by protein kinase fusion proteins (n = 15) (Supplementary Table 16). Of the latter, seven are tandem kinases, whereas Sr43, Pm4, Snn3, Sm1, Tsn1, Yr36, Rpg5 and Lr9 are single or tandem protein kinases fused to different domains[14–20] (Extended Data Fig. 7 and Supplementary Table 16). Little is known about the function of kinase fusion proteins, but most confer race-specific resistance that is phenotypically indistinguishable from NLR-mediated resistance. Their encoding genes do not fall into the *Lr34/Lr67* category of adult, broad-spectrum and multi-pathogen resistance[30,31]. To explain the role of these kinases in resistance, we sought clues from NLRs whose *modus operandi* is now well understood. NLRs can act as guards that monitor host components targeted by pathogen effectors[32]. These guards detect the interaction between an effector and its target, leading to a conformational change in the NLR that triggers downstream defense responses. This tripartite interaction creates an evolutionary “tug-of war” that imposes selective pressure (i) on the effector to evade detection by the NLR while maintaining its ability to coerce the pathogenicity target, (ii) on the NLR to recognize new effector variants, and (iii) on the pathogenicity target to avoid being disrupted by the effector while maintaining its cellular function. Duplication of the pathogenicity target can release it from this functional constraint and provide a ‘decoy’ for the effector. This diversification may also result in the decoy behaving genetically as the resistance gene[33]. In about 10% of all NLRs,

the decoy has become integrated into the NLR itself[34]. Such a guardee-decoy fusion ensures that both components are inherited as a single operational unit.

By extrapolation, protein kinase fusion proteins may be pathogenicity targets that are guarded by NLRs. All protein kinase fusion proteins have one apparent functional kinase that is fused to a second, typically non-functional, kinase domain but sometimes to an altogether different domain, as in for example Sr43, Lr9 and Pm4 (Extended Data Fig. 7). Perhaps similarly to those NLRs that carry an integrated decoy, this second domain might be an integrated decoy, while the apparent functional kinase exerts the signaling function. Indeed, plants produce various enzymes, including protein kinases with different integrated domains, to catalyze reactions of various substrates. In the case of protein kinase resistance proteins, the integrated domain would define the specific substrates of pathogen Avr proteins, whereas the kinase would catalyze the phosphorylation of either the Avr protein, the integrated domain, itself, or a third signaling partner to trigger downstream defense, possibly via an NLR guard (Extended Data Fig. 8). EMS mutagenesis of *Yr15*, *Pm24*, *Sr62*, and *Sr43* has shown a preponderance of missense mutations affecting the kinase active site or ATP-binding pocket of the apparent functional kinase domain (Extended Data Fig. 7), supporting the notion that kinase-mediated signaling is required for function.

The transgenic expression of *Sr43* in a different background allowed us to confirm the broad-spectrum efficacy of *Sr43*, highlighting its potential value in resistance breeding. However, it is possible to obtain gain-of-virulence pathogen mutants that have lost *AvrSr43* function under laboratory conditions[35]. Therefore, *Sr43* should be used in combination with other broad-spectrum resistance genes to maximize its longevity in the field.

Materials And Methods

Mutant collection development

We mutagenized 2,700 seeds of the wheat-*Th. elongatum* introgression line RWG34 containing *Sr43* (ref: [29]). Dry seeds were incubated for 16 h with 200 mL of a 0.8% (w/v) EMS solution with constant shaking on a Roller Mixer (Model SRT1, Stuart Scientific) to ensure maximum homogenous exposure of the seeds to EMS. The excess solution was then removed, and the seeds were washed three times with 400 mL tap water. The M_1 seeds were germinated in the greenhouse, and the seeds of M_2 families (single heads) were collected. Eight seeds per M_2 family were phenotyped with the *Pgt* isolate, race TPMKC. The M_3 seeds derived from susceptible M_2 plants were also tested to confirm that the M_2 susceptible plants were true nonsegregating mutants. To rule out seed contamination, 11 mutants were verified using genotyping-by-sequencing (GBS)[36]. GBS data from the background (Chinese Spring), donor species (*Th. elongatum*), accession PI531737 from USDA ARS GRIN), the Chinese Spring-*Th. elongatum Sr43* introgression line (RWG34) and the mutant lines were mapped to the reference genome sequence of Chinese Spring[37] using BWA mem (version 0.7.12) with standard parameters[38]. Mapping results were sorted and converted to mpileup format using SAMtools[39](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 intervals.

Chromosome flow sorting

Suspensions of mitotic metaphase chromosomes were prepared from root tips of the *Sr43* introgression line and eight EMS mutants as described by Vrána et al. (2000) and Kubaláková et al. (2005)[40,41]. Briefly, root tip meristem cells were synchronized using hydroxyurea, accumulated in metaphase using amiprophos-methyl and fixed in 2%

(v/v) formaldehyde at 5°C for 20 min. Intact chromosomes were released by mechanical homogenization of 100 root tips in 600 µL ice-cold LB01 buffer[42]. GAA microsatellites were labeled on isolated chromosomes by fluorescence *in situ* hybridization in suspension (FISHIS) using 5'-FITC-GAA_7-FITC-3' oligonucleotides (Sigma, Saint Louis, USA) according to Giorgi et al. (2013)[43], and chromosomal DNA was stained with 2 µg/mL DAPI (4',6-diamidino 2-phenylindole). Chromosome analysis and sorting were conducted using a FACSAria II SORP flow cytometer and sorter (Becton Dickinson Immunocytometry Systems, San José, USA). Bivariate flow karyotypes plotting FITC and DAPI fluorescence were acquired for each sample, and chromosomes were sorted at rates of 20–40 particles per second. Two batches of 55,000 and 70,000 copies of chromosome 7D with the *Th. elongatum* chromosome segment carrying *Sr43* were sorted from the wild-type *Sr43*-WT, and one batch of 14,000–66,000 copies was sorted from the eight mutants into PCR tubes containing 40 µL sterile deionized water (Supplementary Table 3).

Chromosome content of flow-sorted fractions was estimated by microscopy observations of 1,500–2,000 chromosomes sorted into a 10-µL drop of PRINS buffer containing 2.5% (w/v) sucrose[44] on a microscope slide. Air-dried chromosomes were labeled by FISH with probes for the pSc119.2 repeat, Afa family repeat and 45S rDNA that allowed identification of all wheat and *Th. elongatum* chromosomes according to Molnár et al. (2016)[45]. To determine chromosome contents and purity in the sorted fractions, at least 100 chromosomes in each flow-sorted sample were classified following the karyotype described by Gaál et al. (2018)[46].

Sequencing and assembly of 7D-7E translocation chromosomes

The two separately flow-sorted chromosomal samples of the wild-type genotypes were used for preparation of two sequencing libraries. Chromosomal samples were treated with proteinase K (60 ng/µl), after which DNA was purified without amplification. Chromosomal samples flow-sorted from the mutants were treated similarly, but their DNA contents were amplified to 2.5–12.6 µg by multiple displacement amplification (MDA) (Supplementary Table 3) using an Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare, Chalfont St. Giles, United Kingdom) as described by Šimková et al. (2008)[47] and sequenced by Novogene (Beijing, China). For the *Sr43* wild-type genotypes, 20 ng of non-amplified DNA was fragmented in a 20-µL volume using a Bioruptor Plus (Diagenode) five times for 30 s on the HIGH setting. Libraries for sequencing were prepared from fragmented DNA using an NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® with the following modifications: (i) size selection was directed for larger final library size (~ 1,000 bp), and (ii) PCR enrichment was done with six PCR cycles. Libraries were sequenced on a HiSeq2500 platform using a HiSeq® Rapid SBS Kit v2 as 250-bp paired-end reads. The raw data were trimmed for low-quality bases using Trimmomatic[48] and assembled into scaffolds with Meraculous[49] (v2.0.5) using 111-bp *k*-mers. Scaffolds shorter than 1 kb were eliminated. The assembly contained 168,523 scaffolds with a total assembly length of 1.29 Gb. Among them, 25,581 scaffolds were longer than 13.9 kb with a total length 631.8 Mb.

Candidate gene identification

Eight susceptible mutants derived from independent M_2 families were selected for MutChromSeq mapping[15]. The raw reads from the eight mutants were individually mapped to the 10-kb chopped scaffold fragments using BWA[38] (version 0.7.12) and SAMtools[39] (version 1.8). One fragment was identified as having a single nucleotide mutation in all mutants. We calculated the probability of this being the candidate gene using formula number 4 developed by Yu et al. 2022 (ref. [10]), with 2598 bp of *Sr43* coding sequence (CDS), assuming the average gene CDS is 1000 bp in

length, and that chromosome 7D has 5822 genes. All identified mutations were G-to-A or C-to-T transition mutations, which are typical of EMS mutagenesis.

RNA extraction and Sr43 annotation

Total RNA was extracted from the Chinese Spring-*Th. elongatum Sr43* introgression line with an RNeasy Plant Mini Kit (Cat No./ID: 74904, Qiagen, Valencia, CA, USA) following the manufacturer's protocol and digested with Dnase I (Roche). RNA-seq was performed by Novogene. The RNA-seq reads were trimmed with Trimmomatic (http://www.usadellab.org/cms/?page=trimmomatic). Hisat2 (ref. [50]) (version 2.1.0) was used to map the short reads onto the *Sr43* genomic sequence. The SAM output file was converted into a BAM file using SAMtools[39] (version 1.8) (http://www.htslib.org/) and sorted according to their position along the *Sr43* genomic sequence and indexed for visualization by IGV (https://software.broadinstitute.org/software/igv/). To determine the alternative splicing of *Sr43*, we constructed a full-length cDNA library using a SMARTer® PCR cDNA Synthesis kit (Cat. # 634926, Clontech/TaKaRa). Transcripts corresponding to each of the four splice variants were identified by Sanger sequencing of 20 clones obtained from transformation of long-range PCR on the full-length cDNA library with primers specific to the *Sr43* 5' and 3' ends (Supplementary Table 5).

Engineering of the Sr43 binary vector construct

Based on the gene annotation, three overlapping segments of the *Sr43* gene were PCR-amplified (Supplementary Table 11) with high-fidelity Q5 DNA polymerase (NEB, Ipswich, MA, USA) following the manufacturer's instructions. The PCR products were purified with a QIAquick PCR Purification kit (QIAGEN, LLC, Germantown, MD 20874, USA) and A-tailed using *Taq* DNA polymerase before being cloned into the pCR2.1 vector (TOPO PCR Cloning Kits-K202020, Thermo Fisher Scientific). The positive clones were digested with three sets of restriction enzymes, *Not*I, *Not*I-*Pvu*I and *Pvu*I-*Pme*I (NEB, Ipswich, MA, USA), to generate *Sr43* fragment parts 1, 2 and 3, respectively. The digested fragments were gel-purified and then parts 2 and 3 were combined in a three-way ligation reaction with the binary vector pGGG-AH-NotI/PmeI[10] digested with *Not*I and *Pme*I, using T4 DNA ligase (M0202S, NEB, Ipswich, MA, USA). Subsequently, the binary construct was linearized with *Not*I, and part 1 was dropped in. A positive clone with part 1 in the correct orientation, pGGG-*Sr43*, was verified by Sanger sequencing. pGGG-Sr43 is available from Addgene under accession number 186974.

Wheat transformation

The binary construct pGGG-Sr43 was transformed into wheat cv. Fielder using Agrobacterium (*Agrobacterium tumefaciens*)-mediated transformation[51]. The *Sr43* copy number was determined by testing the copy number of the hygromycin selectable marker in T_0 and T_1 plants by iDNA Genetics (Norwich, UK) using qPCR[52].

Stem rust phenotyping

The stem rust tests were carried out in a greenhouse or in growth chambers. The greenhouse/growth chambers were maintained at 21°C with a 14-h-light period 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 sowing, at a rate of approximately 0.12 mg of spores per plant. After a 16-h incubation period in the dark with high humidity, inoculated seedlings were returned to the greenhouse/growth chamber and then scored for reaction to stem rust 12–14 days later. The infection types (IT) were recorded using the Stakman scale[53]. For temperature sensitivity tests, the high temperature was set to 26°C. The *Pgt* races used in this study were TPMKC (isolate 74MN1409) from North

America; QTHJC (isolates 75ND717C and 69MN399 from USA); TKTTF (isolate ET11a-18 from Ethiopia); TTKSK (isolate 04KEN156/04) from Kenya; TKTSC (isolate IS-#2079), TTTTF (isolate IS-#2127) and TTTTC (isolate IS-#2135) from Israel; TKTTF (isolate FR68-20) from France; TTRTF (isolate IT16a-18) from Italy; TKTTF (isolate UK-01) from the UK; and TRTTF (isolate 14GEO189-1) from Georgia (Supplementary Table 12).

Protein homology searches

We used InterPro v. 88.0 to search for protein family domains in Sr43, e.g., a transmembrane domain[54]. To check for the presence of myristoylation sites and nuclear localization signals, we used Myristoylator[55] and NLS mapper[56], respectively (accessed 14 April 2022).

Sr43 protein 3D modeling and ATP-binding site prediction

We used the open source code of AlphaFold v2.0 (ref: [27]) and the supercomputer of King Abdullah University of Science and Technology, Shaheen II (https://www.hpc.kaust.edu.sa/) via the multi-node system Ibex (https://www.hpc.kaust.edu.sa/ibex). We input the amino acid sequence of Sr43, and the output was five unrelaxed, five relaxed and five ranked models in .pdb format. We utilized the ranked_1.pdb model that contains the predictions with the highest confidence with the best Local Distance Difference Test (lDDT) score standing at 70.76. We next input the ranked_1.pdb model obtained from Alphafold and each domain separately into the protein structure comparison server Dali[28].

We used HADDOCK2.4, a web server for small molecule binding site screening[25], to screen the DUF2 domain for potential ATP-binding sites. The input files consisted of the DUF2 domain of Sr43 after removing all loops from the .pdb file and ATP in .pdb format. The settings used were:

Define randomly ambiguous interaction restraints from accessible residues -> ON

Number of structures for rigid body docking -> 10,000

Number of structures for semi-flexible refinement -> 400

Number of structures for the final refinement -> 400

Clustering method (RMSD or Fraction of Common Contacts (FCC)) -> RMSD

RMSD Cutoff for clustering (Recommended: 7.5A for RMSD, 0.60 for FCC) -> 2.0

Evdw 1 -> 1.0

Eelec 3 -> 0.1

Initial temperature for second TAD cooling step with flexible side-chain at the Interface -> 500

Initial temperature for third TAD cooling step with fully flexible interface -> 300

Number of MD steps for rigid body high temperature TAD -> 0

Number of MD steps during first rigid body cooling stage -> 0

The output files were 10 clusters of different predicted ATP-binding sites. The cluster with the best prediction score (Z-Score) was cluster 6.

Sr43 protein localization

To generate the 35S:*Sr43-GFP* construct, the open reading frame of *Sr43* splice version 1 was synthesized and ligated into pDONR221 (Twist Bioscience, California, US). The entry clone was then introduced into the binary vector pB7FGW2,0 by single Gateway LR reaction (Invitrogen). The 35S:*Sr43-GFP* was transferred into *A. tumefaciens* strain GV3101 and infiltrated into tobacco leaves as described in Aljeedani et al., 2021 (ref: [57]). The GFP signal was excited at 488 nm and detected between 500 and 535 nm. Images were acquired using a Leica SP8 inverse Stellaris FALCON with an HC PL APO 63x1,2 W CORR UVIS CS2 objective.

Phylogenetic analysis

We constructed a phylogenetic tree based on the aligned protein sequences of 100 best hits (ID ≥ 75%) of the kinase domain sequence and Sr43 DUF region sequence against the NCBI protein database. The phylogenetic tree (neighbor-joining method) for whole proteins and domains was computed with Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) and drawn with iTOL (https://itol.embl.de/).

Data availability

The datasets generated during and/or analyzed in the current study are publicly available as follows. The sequence reads were deposited in the European Nucleotide Archive under project numbers PRJEB52878 (GBS data), PRJEB51958 (chromosome flow sorted data), and PRJEB52088 (RNA-seq data). The *Sr43* gene and transcript sequence were deposited in NCBI Genbank under accession number ON237711. The *Sr43* chromosome assembly has been deposited in Zenodo with DOI 10.5281/zenodo.6777941. The following public databases/datasets were used in the study: Chinese Spring reference genome[37], Gramene (http://www.gramene.org/), https://ensembl.gramene.org/Multi/Tools/Blast, https://wheat.pw.usda.gov/GG3/blast, BLAST non-redundant protein sequence (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastx&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome), Taxonomy Browser (https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=1437183), AlphaFold[27] (https://alphafold.ebi.ac.uk), Dali[28] (http://ekhidna2.biocenter.helsinki.fi/dali/), and HADDOCK[25] (https://www.bonvinlab.org/education/HADDOCK-binding-sites/. Source data are provided with this paper.

Declarations

Acknowledgments

We thank Yajun Wang (KAUST, Saudi Arabia) for help with phenotyping; Yue Jin (USDA, Minnesota, USA) for use of *Pgt* isolates 74MN1409 and 69MN399; Michiel van Slageren (Kew, UK) for help with species nomenclature; and Zdeňka Dubská, Romana Šperková, Jitka Weiserová, for preparation of chromosome samples for flow cytometry and Mahmoud Said and Petr Cápál for chromosome sorting (all IEB, Czech Republic). This research was supported by the NBI Research Computing group and the Informatics Platform at the John Innes Centre, UK and financed by funding 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); Marie Curie Fellowship grant award 'AEGILWHEAT' (H2020-MSCA-IF-2016-746253) and the Hungarian National Research, Development and Innovation Office (K135057) to I.M.; ERDF project "Plants as a tool

for sustainable global development" (No. CZ.02.1.01/0.0/0.0/16_019/0000827) to J.B., K.H., M.K. and J.D.; King Abdullah University of Science and Technology to B.B.H.W., Ł.J. and I.B.; the Lieberman-Okinow Endowment at the University of Minnesota to B.J.S.; the Gordon and Betty Moore Foundation through grant GBMF4725 to the 2Blades Foundation and J.D.G.J.; and the Gatsby Charitable Foundation to J.D.G.J.

Author contributions

Generated the mutant population: G.Y.; screened and verified mutants: R.J., B.J.S. and N.K.; genotyped mutants: B.S., S.W., J.P. Performed chromosome flow sorting and amplification: I.M., M.K. and J.D.; chromosome assembly: G.Y., K.H. and J.B. Performed bioinformatics analyses to identify *Sr43*: G.Y. Extracted RNA, determined *Sr43* gene structure and alternative splicing: G.Y. Annotated Sr43: G.Y., S.G. and Ł.J. Performed 3D modeling analyses: S.G. and Ł.J. Developed the binary vector: M.S.; engineered *Sr43* construct: G.Y. Transformed *Sr43* into wheat: S.H. and W.H. Phenotyped the transgenic lines: G.Y., O.M., M.P. and M.N.R. Determined Sr43 localization: F.R.A. and I.B. Performed phylogenetics and synteny: G.Y. Provided germplasm, rust cultures or scientific support and advice: O.M., S.X., M.N.R., A.S., J.D.G.J., C.G., Y.Y. and T.L.R. Conceived study: B.B.H.W., G.Y. and B.J.S. Drafted manuscript: G.Y. and B.B.H.W. All co-authors read and approved the final manuscript. Authors are grouped by institution in the author list, except for the first three and last four authors.

Competing interests

G.Y. and B.B.H.W. are inventors on the US provisional patent application 63/250,413 filed by 2Blades and relating to the use of *Sr43* for stem rust resistance in transgenic wheat. T.L.R. was employed by the 2Blades Foundation, which co-funded the work.

References

1. Savary, S. et al. The global burden of pathogens and pests on major food crops. Nat. Ecol. Evol. 3, 430–439 (2019).
2. van Esse, P. et al. Genetic modification to improve disease resistance in crops. New Phytol. 225, 70–86 (2020).
3. Hafeez, A. N. et al. Creation and judicious application of a wheat resistance gene atlas. Mol. Plant 14, 1053–1070 (2021).
4. McDonald. B. A., & Linde C. Pathogen population genetics, evolutionary potential, and durable resistance. Annu. Rev. Phytopathol. 40, 349–379 (2002).
5. Luo, M. et al. A five-transgene cassette confers broad-spectrum resistance to a fungal rust pathogen in wheat. Nat. Biotechnol. 39, 561–566 (2021).
6. 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).
7. 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).
8. 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).
9. Chen, S. et al. Wheat gene *Sr60* encodes a protein with two putative kinase domains that confers resistance to stem rust. New Phytol. 225, 948–959 (2020).

10. Yu, G., et al. *Aegilops sharonensis* genome-assisted identification of stem rust resistance gene *Sr62*. Nat. Commun. 13, 1607 (2022).
11. 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).
12. Gaurav, K. et al. Population genomic analysis of *Aegilops tauschiii* identifies targets for bread wheat improvement. Nat. Biotechnol. 40, 422–431 (2022).
13. Arora, S. et al. A wheat kinase and immune receptor form the host-specificity, bioRxiv 2022.01.27.477927.
14. Fu, D. et al. A Kinase-START gene confers temperature-dependent resistance to wheat stripe rust. Science 323, 1357–1360 (2009).
15. Sánchez-Martín, J. et al. Wheat *Pm4* resistance to powdery mildew is controlled by alternative splice variants encoding chimeric proteins. Nat. Plants 7, 327–341 (2021).
16. Zhang, et al. A protein kinase–major sperm protein gene hijacked by a necrotrophic fungal pathogen triggers disease susceptibility in wheat. Plant J. 106, 720–732 (2021).
17. Faris, J. D et al. A unique wheat disease resistance-like gene governs effector-triggered susceptibility to necrotrophic pathogens. Proc. Natl. Acad. Sci. U.S.A. 107, 13544–13549 (2010).
18. Arora, D., et al., Allele characterization of genes required for rpg4-mediated wheat stem rust resistance identifies *Rpg5* as the R gene. Phytopathology. 103, 1153–1161 (2013).
19. Walkowiak, S., et al. Multiple wheat genomes reveal global variation in modern breeding. Nature 588, 277–283 (2020).
20. Wang, Y. et al. An unusual tandem kinase fusion protein confers leaf rust resistance in wheat. (unpublished)
21. Chalupska, D. et al. *Acc* homoeoloci and the evolution of wheat genomes. Proc. Natl. Acad. Sci. U.S.A. 105, 9691–9696 (2008).
22. Knott, D. R. et al. Transfer to wheat and homoeology of an *Agropyron elongatum* chromosome carrying resistance to stem rust. Can. J. Genet. Cytol. 19, 75–79 (1977).
23. Kibiridge-Sebunya, I. & Knott, D. R. Transfer of stem rust resistance to wheat from an *Agropyron* chromosome having a gametocidal effect. Can. J. Genet. Cytol. 25, 215–221 (1983).
24. International Wheat Genome Sequencing Consortium (IWGSC). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science. 345, 6194 (2014).
25. van Zundert, G. C. P. et al, The HADDOCK2.2 Web Server: User-Friendly Integrative Modelling of Biomolecular Complexes, 428, 720–725 (2016).
26. Stührwohldt, N., et al. The PSI family of nuclear proteins is required for growth in Arabidopsis. Plant Mol. Biol. 86, 289–302 (2014).
27. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021).
28. Holm, L. & Rosenström P. Dali server: conservation mapping in 3D. Nucleic Acids Res. 38, W545-W549 (2010).
29. Niu, Z. et al. Development and characterization of wheat lines carrying stem rust resistance gene *Sr43* derived from *Thinopyrum ponticum*. Theor. Appl. Genet. 127, 969–80 (2014).
30. Krattinger, S. G. et al. A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. Science 323, 1360–1363 (2009).
31. Moore, J. W. et al. A recently evolved hexose transporter variant confers resistance to multiple pathogens in wheat. Nat. Genet. 47, 1494–1498 (2015).

32. Jones, J., & Dangl, J. The plant immune system. Nature 444, 323–329 (2006).
33. 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).
34. Cesari, S. et al. A novel conserved mechanism for plant NLR protein pairs: the “integrated decoy” hypothesis. Front. Plant Sci. 5, (2014).
35. Kangara, N. et al. Mutagenesis of *Puccinia graminis* f. sp. *tritici* and selection of gain-of-virulence mutants. Front. Plant Sci. 11, (2020).
36. Poland, J. A. & Rife, T. W. Genotyping-by-sequencing for plant breeding and genetics. Plant Genome 5, 92–102 (2012).
37. International Wheat Genome Sequencing Consortium (IWGSC). Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 361, eaar7191 (2018).
38. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows- Wheeler Transform. Bioinformatics 25, 1754–1760 (2009).
39. Li, H. et al. 1000 Genome Project Data Processing Subgroup, The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25, 2078–2079 (2009).
40. Vrána, J. et al. Flow sorting of mitotic chromosomes in common wheat (*Triticum aestivum* L.). Genetics 156, 2033–2041 (2000).
41. Kubaláková, M. et al. Flow karyotyping and chromosome sorting in bread wheat (*Triticum aestivum* L.). Theor. Appl. Genet. 104, 1362–1372 (2002).
42. Doležel, J. et al., Analysis of nuclear DNA content in plant cells by flow cytometry. Biol. Plant. 31, 113–120 (1989).
43. Giorgi, D. et al. FISHIS: fluorescence *in situ* hybridization in suspension and chromosome flow sorting made easy. PLoS ONE 8, e57994 (2013).
44. Kubaláková, M. et al., Mapping of repeated DNA sequences in plant chromosomes by PRINS and C-PRINS. Theor. Appl. Genet. 94, 758–763 (1997).
45. 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).
46. Gaál, E., et al. Identification of COS markers specific for *Thinopyrum elongatum* chromosomes preliminary revealed high level of macrosyntenic relationship between the wheat and *Th. elongatum* genomes. PLoS ONE 13, 12 (2018).
47. Šimková, H. et al. Coupling amplified DNA from flow-sorted chromosomes to high-density SNP mapping in barley. BMC Genomics 9, 294 (2008).
48. Bolger, A. M. Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30, 2114–2120 (2014).
49. Chapman, J. A. et al. A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. Genome Biol. 16, 26 (2015).
50. Kim, D. et al. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. 37, 907–915 (2019).
51. Hayta, S. et al. An efficient and reproducible *Agrobacterium*-mediated transformation method for hexaploid wheat (*Triticum aestivum* L.). Plant Methods 15, 121 (2019).

52. Bartlett, J. G. et al. High-throughput *Agrobacterium*-mediated barley transformation. Plant Methods 4, 22 (2008).
53. Stakman E. C. et al. Identification of physiologic races of *Puccinia graminis* var. *tritici*. US Department of Agriculture, Agricultural Research Service. pp. Publ. E617. (1962)
54. Blum, M. et al., The InterPro protein families and domains database: 20 years on. Nucleic Acids Research (2020).
55. Bologna G., et al. N-terminal myristoylation predictions by ensembles of neural networks. Proteomics. 4,1626–1632 (2004).
56. Kosugi, S. et al. Systematic identification of yeast cell cycle-dependent nucleocytoplasmic shuttling proteins by prediction of composite motifs. Proc. Natl. Acad. Sci. U.S.A. 106, 10171–10176 (2009).
57. Aljedaani, F., et al., A semi-in vivo transcriptional assay to dissect plant defense regulatory modules. Methods Mol. Biol. 2328, 203–214 (2021).

Figures

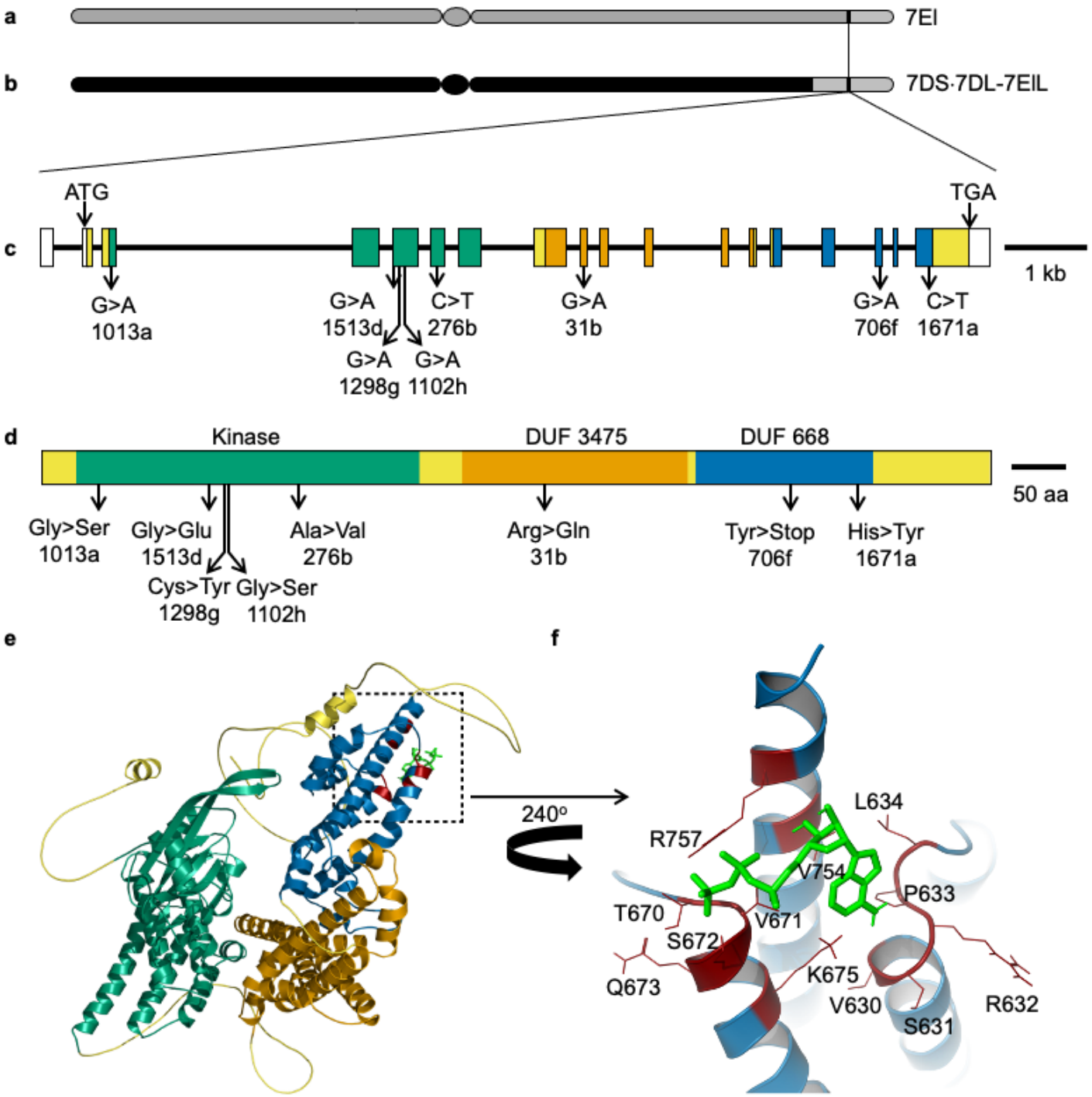

Figure 1

***Sr43* encodes a protein kinase fused to two domains of unknown function.**

a, *Thinopyrum elongatum* chromosome 7. **b**, Schematic diagram of the wheat-*Thinopyrum* translocation chromosome. **c**, Identified EMS mutations along the predicted *Sr43* gene model. **d**, Schematic diagram of Sr43, showing predicted domains and amino acid changes induced by the EMS mutations. **e**, Three-dimensional model of Sr43, as predicted by AlphaFold. Green, kinase; orange and blue, DUF domains; yellow, flexible linkers. **f**, Structural detail (dashed box in panel e) of a high-confidence ATP-binding site (red residues) in DUF668 bound to an ATP molecule, as determined by the small molecule docking program HADDOCK[25]. ATP is depicted as a stick structure (light green) connected to DUF668 residues by hydrogen bonds (red lines).

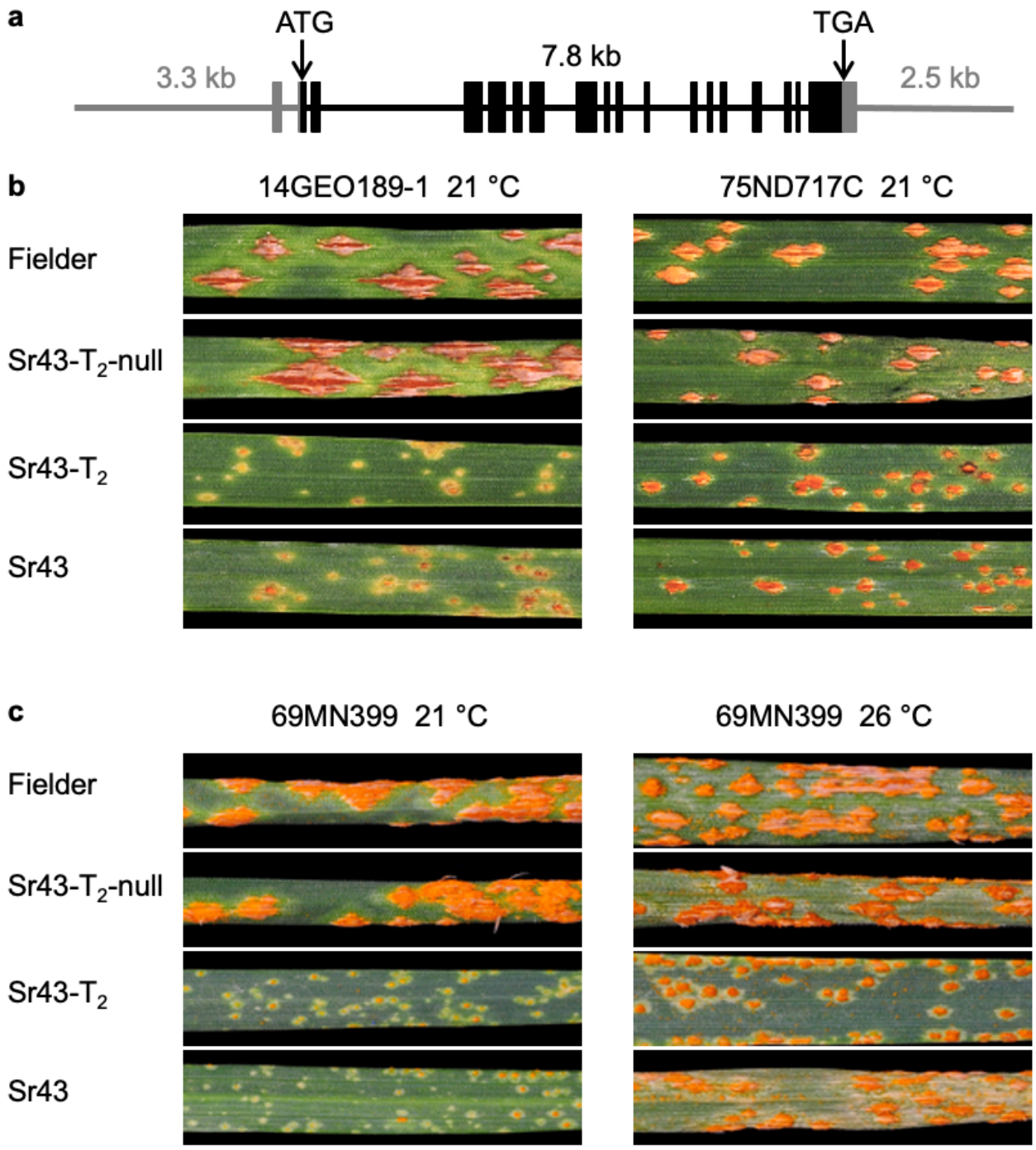

Figure 2

Confirmation of *Sr43* function, race specificity, and temperature sensitivity. a, Schematic diagram of the *Sr43* genomic fragment used for transformation into wheat cv. Fielder. **b**, Representative leaves from seedlings of *Sr43* wild-type and transgenic lines alongside non-transgenic wild-type Fielder and null controls inoculated with *Puccinia graminis* f. sp. *tritici* isolates 14GEO189-1 (avirulent on *Sr43*) and 75ND717C (intermediately virulent on *Sr43*). **C**. Effect of temperature on *Sr43*-mediated resistance to *Pgt* isolate 69MN399.

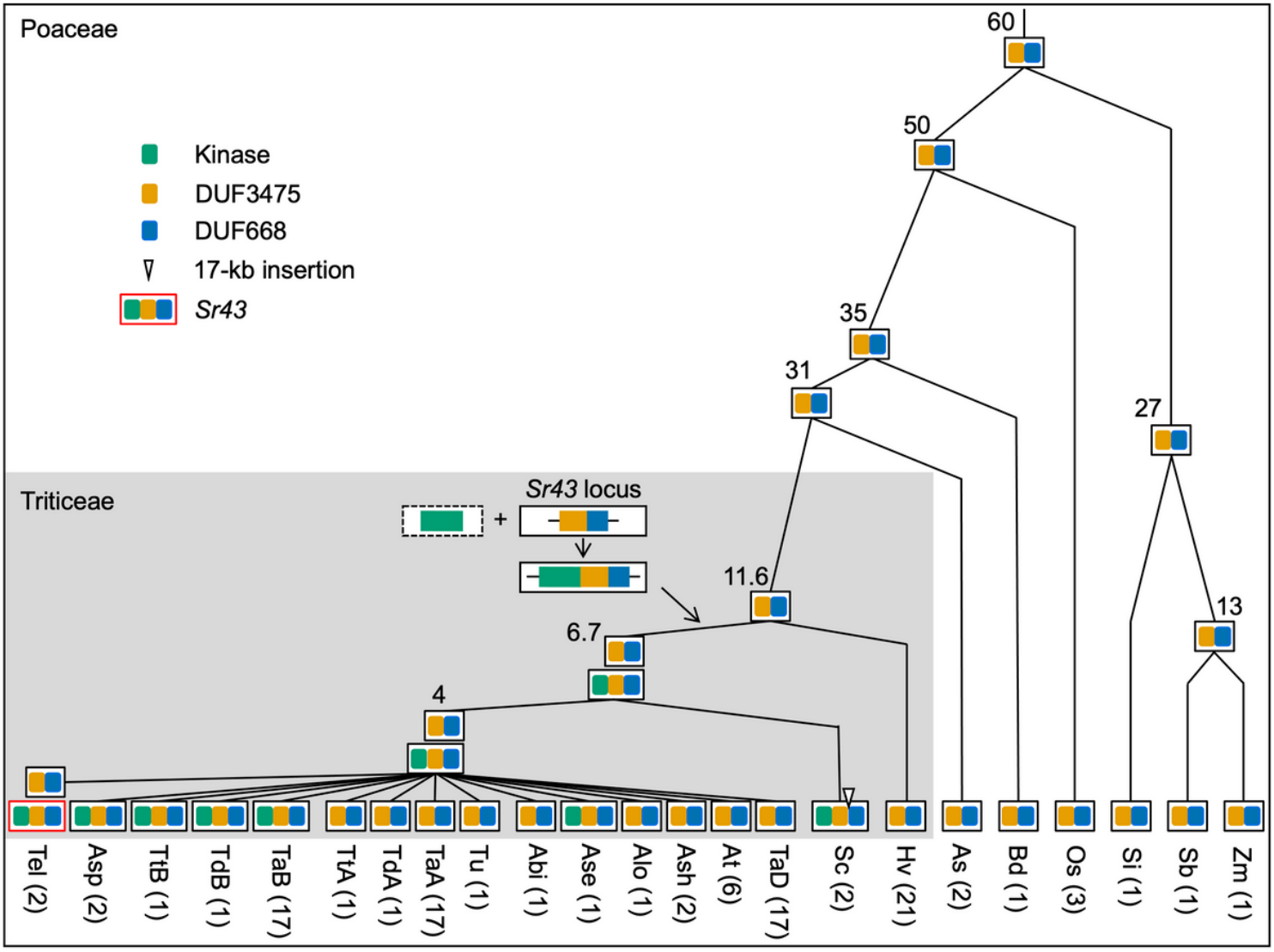

Figure 3

Phylogenetic tree of the Poaceae showing the origin and distribution of *Sr43* orthologs. The age of the last common ancestor in millions of years is indicated at each node based on Chalupska et al.[21]. The Triticeae clade is highlighted in gray. Species are indicated at the bottom and abbreviated as follows: Tel, *Thinopyrum elongatum*; Asp, *Ae. Speltoides*; TtB, *Triticum turgidum* spp. *Durum* B genome; TdB, *T. dicoccoides* B genome; TaB, *T. aestivum* B genome; TtA, *T. turgidum* ssp. Durum A genome; TdA, *T. dicoccoides* A genome; TaA, *T. aestivum* A genome; Tu, *T. urartu*; Abi, *Aegilops bicornis*; Ase, *Ae. searsii*; Alo, *Ae. longissima*; Ash, *Aegilops sharonensis*; At, *Ae. tauschii*; TaD, *T. aestivum* D genome; Sc, *Secale cereale*; Hv, *Hordeum vulgare*; As, *Avena sativa*; Bd, *Brachypodium distachyon*; Os, *Oryza sativa*; Si, *Setaria italica*; Sb, *Sorghum bicolor*; Zm, *Zea mays*. The number of genomes analyzed for each species is indicated in parentheses.

Supplementary Files

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

- Sr43SupplementaryinformationFigs.docx
- Sr43SupplementaryinformationTables.xlsx

- SupplementaryData1Sr43WT.pdb
- SupplementaryData2Sr43WTATP.pdb
- Sr43ExtendedFigs.docx