nature biotechnology

LETTERS

https://doi.org/10.1038/s41587-018-0007-9

Resistance gene cloning from a wild crop relative by sequence capture and association genetics

Sanu Arora[1,15], Burkhard Steuernagel[1,15], Kumar Gaurav[1], Sutha Chandramohan[2], Yunming Long[3], Oadi Matny[4], Ryan Johnson[4], Jacob Enk[5], Sambasivam Periyannan[2], Narinder Singh[6], M. Asyraf Md Hatta[1,7], Naveenkumar Athiyannan[2,8], Jitender Cheema[1], Guotai Yu[1], Ngonidzashe Kangara[1], Sreya Ghosh[1], Les J. Szabo[9], Jesse Poland[6], Harbans Bariana[10], Jonathan D. G. Jones[11], Alison R. Bentley[12], Mick Ayliffe[2], Eric Olson[13], Steven S. Xu[14], Brian J. Steffenson[4], Evans Lagudah[2] and Brande B. H. Wulff[1*]

Disease resistance (R) genes from wild relatives could be used to engineer broad-spectrum resistance in domesticated crops. We combined association genetics with R gene enrichment sequencing (AgRenSeq) to exploit pan-genome variation in wild diploid wheat and rapidly clone four stem rust resistance genes. AgRenSeq enables R gene cloning in any crop that has a diverse germplasm panel.

Most resistance (R) genes encode intracellular nucleotide binding/leucine-rich repeat (NLR) immune receptor proteins[1]. Domestication and intensive breeding have reduced R gene diversity in crops, rendering them more vulnerable to disease outbreaks[2]. Introgression breeding of single R genes into crops is laborious and associated with co-integration of deleterious genes[3,4], and pathogens can rapidly evolve to overcome R genes when deployed singly[5]. Multiple cloned R genes could be engineered into crops as a stack to avoid linkage drag and delay emergence of virulent pathogens[6], but existing R gene cloning methods require segregating or mutant progenies[7–12], which are difficult to generate for many wild relatives due to poor agronomic traits.

R genes can be cloned with positional cloning or mutational genomics[7,8,13] (Supplementary Table 1). Both methods require the R gene to exist as a single gene in an otherwise susceptible genetic background to the pathogen of interest, which can take numerous generations to achieve. They also require screening of thousands of recombinant or mutant lines (Supplementary Table 1). So, although wild relatives of crop plants often harbor multiple R genes, they have pre-domestication traits that preclude using existing methods to clone R genes.

Genome-wide association studies (GWAS) enable trait correlation in a genetically diverse population by exploiting pre-existing recombination events accumulated in natural populations. The reliance of GWAS on a reference genome, however, complicates the identification of sequences that have significantly diverged from the reference, such as R genes. This limitation has been overcome by performing trait associations on sub-sequences (*k*-mers) that are different in case and control samples to identify variants associated with human disease[14] or bacterial antibiotic resistance[15,16].

We reasoned that *k*-mer-based association genetics combined with R gene enrichment sequencing (AgRenSeq) would enable the discovery and cloning of R genes from a plant diversity panel (Fig. 1). Here, we apply AgRenSeq to a panel of *Aegilops tauschii* accessions that were phenotyped with races of the wheat stem rust pathogen *Puccinia graminis* f. sp. *tritici* (PGT). *Ae. tauschii* is the wild progenitor species of the *Triticum aestivum* (bread wheat) D genome and a valuable source of stem rust resistance (*Sr*) genes that have been introgressed into bread wheat. To test AgRenSeq, we chose *Ae. tauschii* ssp. *strangulata*, which has numerous diverse accessions[17] and has two cloned *Sr* genes (*Sr33* and *Sr45*) that can serve as positive controls[7,9]. We obtained 174 *Ae. tauschii* ssp. *strangulata* accessions that span its habitat around the Caspian Sea (Supplementary Table 2) and included 21 *Ae. tauschii* ssp. *tauschii* accessions as an outgroup.

We designed and tested a sequence capture bait library[18] optimized for *Ae. tauschii* NLRs plus genomic regions encoding 317 single nucleotide polymorphism (SNP) markers distributed across all seven chromosomes (Supplementary Fig. 1 and Supplementary Tables 3 and 4). We applied RenSeq[19] to the diversity panel with this capture library, generated de novo assemblies from the Illumina short-read sequences, and filtered the contigs using NLR-parser[20] to obtain 1,312 to 2,170 (average 1,437) NLR contigs per accession. Of these, 249 to 336 (average 299) encoded full-length NLRs and 1,024 to 1,921 (average 1,137) encoded partial NLRs (Supplementary Table 5). Captured SNP marker sequences were compared among accessions and a set of 151 genetically distinct *Ae. tauschii* ssp. *strangulata* accessions were selected (Fig. 2a, Supplementary Figs. 2 and 3, and Supplementary Table 6). These accessions were phenotyped using six PGT races with different virulence profiles and geographic origins (Supplementary Table 7 and Supplementary Fig. 4). Eight percent to 38% (13 to 58 lines of 151 phenotyped

[1]John Innes Centre, Norwich Research Park, Norwich, UK. [2]Commonwealth Scientific and Industrial Research Organization Agriculture and Food, Canberra, Australian Capital Territory, Australia. [3]Department of Plant Sciences, North Dakota State University, Fargo, ND, USA. [4]Department of Plant Pathology, University of Minnesota, St. Paul, MN, USA. [5]Arbor Biosciences, Ann Arbor, MI, USA. [6]Wheat Genetics Resource Center, Department of Plant Pathology, Kansas State University, Manhattan, KS, USA. [7]Faculty of Agriculture, Universiti Putra Malaysia, Serdang, Malaysia. [8]Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Brisbane, Queensland, Australia. [9]US Department of Agriculture, Agriculture Research Service, Cereal Disease Laboratory, St. Paul, MN, USA. [10]The University of Sydney Plant Breeding Institute, Cobbitty, New South Wales, Australia. [11]The Sainsbury Laboratory, Norwich Research Park, Norwich, UK. [12]The John Bingham Laboratory, NIAB, Cambridge, UK. [13]Michigan State University, East Lansing, MI, USA. [14]US Department of Agriculture, Agriculture Research Service, Northern Crop Science Laboratory, Cereal Crops Research Unit, Red River Valley Agricultural Research Center, Fargo, ND, USA. [15]These authors contributed equally: Sanu Arora, Burkhard Steuernagel. *e-mail: brande.wulff@jic.ac.uk

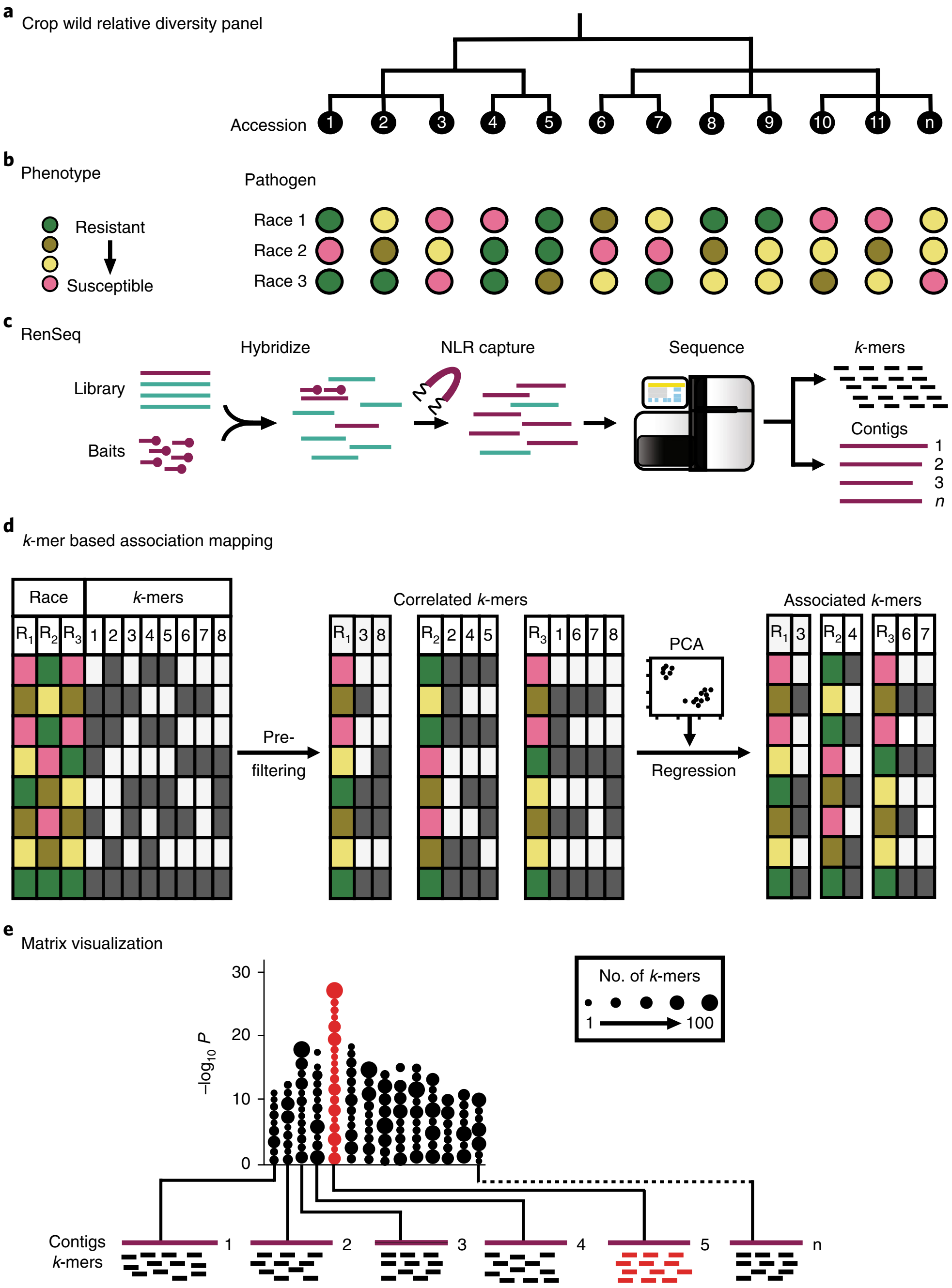

Fig. 1 | Combining association genetics and R gene enrichment sequencing (AgRenSeq) for R gene cloning. a–c, A genetically diverse panel of accessions (**a**) is phenotyped with different pathogen races (**b**), and subjected to RenSeq followed by assembly of the NLR repertoire and extraction of NLR *k*-mers for each accession (**c**). **d**, *k*-mers are pre-filtered based on the correlation of their presence/absence to the level of resistance or susceptibility in the phenotyped panel. Each pre-filtered *k*-mer is given a *P* value based on its ability to predict the phenotype using linear regression, with PCA dimensions as covariates to control for population structure. Phenotypes are color-coded as in **b**, and the presence and absence of *k*-mers is indicated by dark gray and white, respectively. **e**, *k*-mers are then plotted in an association matrix according to their sequence identity to NLRs from a given accession (*x* axis) and the measure of their association with phenotype (*y* axis) (see Fig. 2). A candidate R gene contig is illustrated by a red-dot column.

accessions) showed resistance depending on the PGT race used (Fig. 2b, Supplementary Fig. 5, and Supplementary Table 8).

Next, we carried out a *k*-mer-based AgRenSeq analysis to identify *Sr* genes in the phenotyped panel. *k*-mer sequences were pre-filtered based on the correlation of their presence/absence to the level of resistance or susceptibility observed in the panel. A linear regression model was then fitted to each filtered *k*-mer to predict the phenotype while controlling the population structure using principal component analysis (PCA) of the SNP marker matrix, and the negative log of *P* value obtained was taken as the measure of association for each filtered *k*-mer (Fig. 1d). In previous *k*-mer-based GWAS, associated *k*-mers were used to build a local assembly[14] or reconstruct diverged haplotypes through a direct base-by-base *k*-mer frequency-guided extension process[15]. However, we reasoned that local assembly approaches using only those *k*-mers that are strongly linked to the trait would not generate complete NLR contigs,

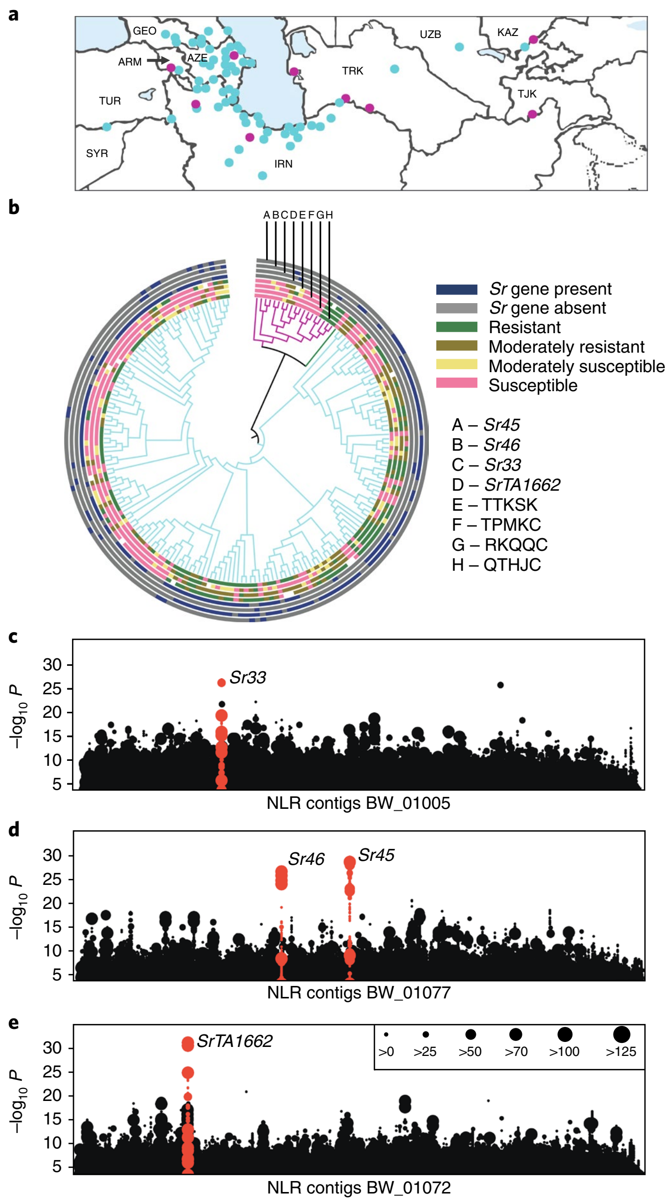

Fig. 2 | Genetic architecture of stem rust resistance in *Ae. tauschii*. **a**, Geographic distribution of *Ae. tauschii* ssp. *strangulata* (cyan) and ssp. *tauschii* (magenta) used in this study. Two accessions from China and two from Pakistan, which fall outside the map, are not shown. ARM, Armenia; AZE, Azerbaijan; GEO, Georgia; IRN, Iran; KAZ, Kazakhstan; SYR, Syria; TJK, Tajikistan; TUR, Turkey; TRK, Turkmenistan; UZB, Uzbekistan. **b**, Phylogenetic tree displaying *Ae. tauschii* ssp. *strangulata* (173 accessions, cyan) and ssp. *tauschii* (19 accessions, magenta) with an intermediate accession (dark green). Stem rust phenotypes and *Sr* genotypes are displayed by concentric circles around the tree. **c**–**e**, Identification of *Sr33*, *Sr45*, *Sr46*, and *SrTA1662* by AgRenSeq using PGT races RKQQC, TTKSK, and QTHJC, respectively. The number of accessions used for each phenotype is provided in Supplementary Table 8. Each dot column on the *x* axis represents an NLR contig from the RenSeq assembly of a single accession (BW_01005, BW_01077 or BW_01072) containing the respective *Sr* gene. Each dot on the *y* axis represents one or more RenSeq *k*-mers associated with resistance across the diversity panel to the respective PGT race. The association score is defined as the negative log of *P* value obtained using likelihood ratio test for nested models. Dot columns corresponding to *Sr* genes are colored red. Dot size is proportional to the number of *k*-mers associated with resistance.

especially across the more conserved NB-ARC (nucleotide-binding adaptor shared by APAF-1 R proteins and CED-4) domain. We therefore projected *k*-mers directly onto NLR assemblies generated from read data of resistant accessions to obtain long contigs including full-length NLRs (Fig. 1e).

The North American PGT race RKQQC is avirulent to *Sr33*, but virulent to most other known *Sr* genes[21]. To identify *Sr33*, we mapped *k*-mers associated with RKQQC resistance onto the RenSeq assembly of a resistant accession. A single discrete association peak was identified in a 5.6 kb RenSeq contig that had 100% identity to *Sr33* (Fig. 2c). The same method was then used to map *k*-mers associated with resistance to all five remaining PGT races in the 151 accessions. Three non-redundant, high-confidence candidate *Sr* genes other than *Sr33* were identified (Fig. 2d,e and Supplementary Table 9). One of these corresponded to the *Sr45* candidate gene previously identified by mutagenesis and RenSeq (MutRenSeq) from a wheat—the *Ae. tauschii* BW_01083 introgression line[7]. When transformed into cv. Fielder, we found that this gene conferred resistance, confirming the previous MutRenSeq identification[7] and these AgRenSeq data (Supplementary Fig. 6 and Supplementary Table 10). Two other candidate *Sr* genes were aligned to the reference genome assembly of wheat cv. Chinese Spring RefSeq1.0 (www.wheatgenome.org). The locations of these candidate *Sr* genes coincided with the positions of two *Sr* genes introgressed into wheat from *Ae. tauschii* ssp. *strangulata* (*Sr46*[22] and *SrTA1662*[23]). The *SrTA1662* candidate gene was mapped in a recombinant inbred line population to a 3.8 cM interval shown to encode *SrTA1662* resistance, thus supporting our AgRenSeq data (Supplementary Fig. 7 and Supplementary Table 11). The *SrTA1662* candidate encodes a coiled-coil NLR protein with 83% amino acid identity to *Sr33*.

We also identified a *Sr46* candidate gene by conventional fine mapping in segregating diploid progenitor and wheat populations coupled with the sequencing of candidate genes in this region in three ethyl methanesulphonate-derived mutants that had lost *Sr46* resistance. In two mutants, the same candidate gene contained non-synonymous substitutions, while the third mutant had a deletion of the chromosomal segment encoding this gene (Supplementary Fig. 8a–f). Comparison of the *Sr46* candidate gene identified by AgRenSeq and map-based/mutagenesis cloning showed that they were 100% identical. This gene, hereafter referred to as *Sr46*, encodes a coiled-coil NLR protein that conferred stem rust resistance when expressed as a transgene in the susceptible wheat cv. Fielder (Supplementary Fig. 8g and Supplementary Table 10). From AgRenSeq analysis and sequence alignments of the four cloned *Sr* genes and the diversity panel, *Sr46* and *SrTA1662* were the most prevalent, being found in 42% of the accessions, while *Sr33* and *Sr45* were found in 5% and 7% of accessions, respectively (Fig. 2b).

To investigate how sample size affects AgRenSeq performance, we progressively reduced the panel size by random subsampling. *SrTA1662* was detected with only 80 diverse accessions, whereas 140 accessions were needed to detect all four *Sr* genes (Supplementary Fig. 9). To assess the general impact of population structure on gene detection by AgRenSeq, we performed an experiment in which the distribution of a gene within the population was manipulated. We subsampled the panel while maintaining the frequency of *Sr33* at 5%; in the first case, we ensured that the *Sr33*-containing accessions came from different clades, while in the second case, all *Sr33*-containing accessions came from the same clade. The former increased the signal-to-noise ratio while the latter suppressed the signal (Supplementary Fig. 10). This experiment indicates that rare genes can be detected provided they are distributed across the breadth of diversity within the panel.

We have shown that AgRenSeq can be applied to rapidly discover and clone functional R genes from a diversity panel. Here, we report cloning of four *Sr* genes that have been introgressed into

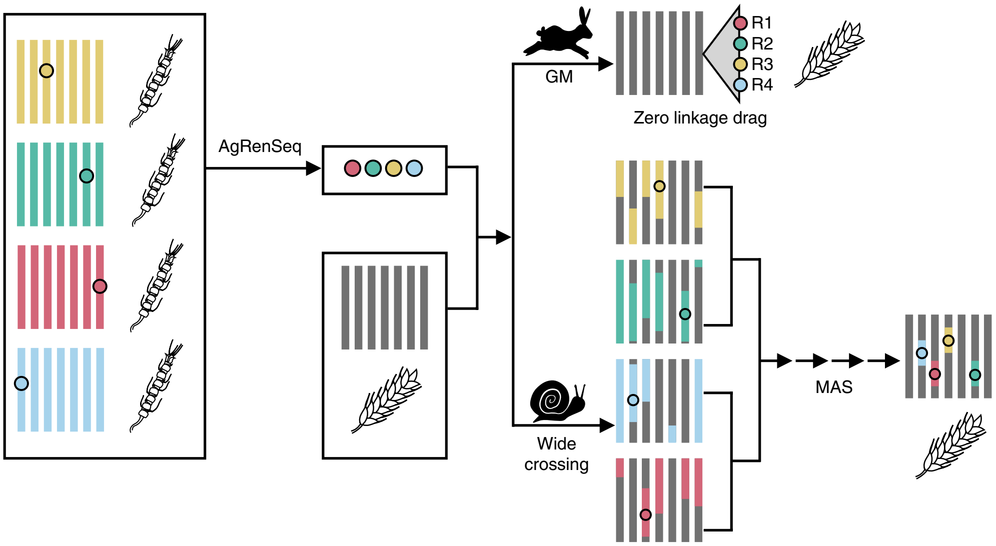

Fig. 3 | AgRenSeq to engineer disease resistance. Discovery and cloning of diverse R genes (colored dots on vertical chromosomes) in a germplasm panel (far left) allows the rapid engineering by transformation of a multi-R gene stack with zero linkage drag (R1 to R4, top right) or facilitates incorporation and stacking of R genes into elite lines and reduction of linkage drag (colored bars around R genes) by multiple backcrossing and marker-assisted selection (MAS; bottom right).

wheat from *Ae. tauschii* ssp. *strangulata* over the past 40 years[7,9,22,23]. *Ae. tauschii* is a rich source of genetic variation for resistance to other diseases and pests of bread wheat besides stem rust, including leaf rust, stripe rust, wheat blast, powdery mildew, Hessian fly, and others (Supplementary Table 12). Our RenSeq-configured diversity panel can be applied as an R gene genotyped resource that can be screened against other pathogens and pests to clone functional R genes.

Unlike other association studies, AgRenSeq is reference-genome-independent and directly identifies the NLR that confers resistance rather than identifying a genomic region encoding multiple paralogs. AgRenSeq can exploit pan-genome sequence variation in diverse germplasm to isolate uncharacterized R genes without crossing or mutagenesis and can be applied to clone R genes from wild species as long as sufficient biological material can be obtained for sequencing and phenotyping. One limitation of AgRenSeq is the bias introduced by NLR capture, which precludes the identification of atypical R genes.

Germplasm collections are available for soybean, pea, cotton, maize, potato, wheat, barley, rice, banana, and cocoa crops and their wild relatives (Supplementary Table 13). The ability to rapidly clone agriculturally valuable R genes by AgRenSeq further increases the value of these germplasm collections. Multiple cloned R genes could be used for engineering resistance (Fig. 3) or as precise molecular markers for use in breeding programs. Due to the functional epistasis of many R genes, molecular markers enable the incorporation of multiple genes into the same elite line (Fig. 3), which will likely improve the longevity of R genes[5]. Finally, the use of precise markers will reduce linkage drag by enabling selection for the R gene of interest while selecting against the donor background.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41587-018-0007-9.

Received: 29 January 2018; Accepted: 12 December 2018;
Published online: 4 February 2019

References

1. Jones, J. D. & Dangl, J. L. The plant immune system. *Nature* **444**, 323–329 (2006).
2. Dangl, J. L. & Jones, J. D. Plant pathogens and integrated defence responses to infection. *Nature* **411**, 826–833 (2001).
3. Knott, D. R. The genetic nature of mutations of a gene for yellow pigment linked to *Lr19* in 'Agatha' wheat. *Can. J. Genet. Cytol.* **26**, 392–393 (1984).
4. Niu, Z. et al. Development and characterization of wheat lines carrying stem rust resistance gene *Sr43* derived from *Thinopyrum ponticum*. *Theor. Appl. Genet.* **127**, 969–980 (2014).
5. McDonald, B. A. & Linde, C. Pathogen population genetics, evolutionary potential, and durable resistance. *Annu. Rev. Phytopathol.* **40**, 349–379 (2002).
6. Dangl, J. L., Horvath, D. M. & Staskawicz, B. J. Pivoting the plant immune system from dissection to deployment. *Science* **341**, 746–751 (2013).
7. Steuernagel, B. et al. Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture. *Nat. Biotechnol.* **34**, 652–655 (2016).
8. Sánchez-Martín, J. et al. Rapid gene isolation in barley and wheat by mutant chromosome sequencing. *Genome Biol.* **17**, 221 (2016).
9. Periyannan, S. et al. The gene *Sr33*, an ortholog of barley *Mla* genes, encodes resistance to wheat stem rust race Ug99. *Science* **341**, 786–788 (2013).
10. Saintenac, C. et al. Identification of wheat gene *Sr35* that confers resistance to Ug99 stem rust race group. *Science* **341**, 783–786 (2013).
11. Witek, K. et al. Accelerated cloning of a potato late blight-resistance gene using RenSeq and SMRT sequencing. *Nat. Biotechnol.* **34**, 656–660 (2016).
12. Zhao, B. et al. A maize resistance gene functions against bacterial streak disease in rice. *Proc. Natl Acad. Sci. USA* **102**, 15383–15388 (2005).
13. 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).
14. Rahman, A., Hallgrímsdóttir, I., Eisen, M. & Pachter, L. Association mapping from sequencing reads using *k*-mers. *eLife* **7**, e32920 (2018).
15. Audano, P. A., Ravishankar, S. & Vannberg, F. O. Mapping-free variant calling using haplotype reconstruction from k-mer frequencies. *Bioinformatics* **34**, 1659–1665 (2018).
16. Lees, J. A. et al. Sequence element enrichment analysis to determine the genetic basis of bacterial phenotypes. *Nat. Commun.* **7**, 12797 (2016).
17. Jones, H. et al. Strategy for exploiting exotic germplasm using genetic, morphological, and environmental diversity: the *Aegilops tauschii* Coss. example. *Theor. Appl. Genet.* **126**, 1793–1808 (2013).
18. Steuernagel, B., Witek, K., Jones, J. D. G. & Wulff, B. B. H. MutRenSeq: a method for rapid cloning of plant disease resistance genes. *Methods Mol. Biol.* **1659**, 215–229 (2017).
19. Jupe, F. et al. Resistance gene enrichment sequencing (RenSeq) enables reannotation of the *NB-LRR* gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant J.* **76**, 530–544 (2013).

20. Steuernagel, B., Jupe, F., Witek, K., Jones, J. D. & Wulff, B. B. NLR-parser: rapid annotation of plant NLR complements. *Bioinformatics* **31**, 1665–1667 (2015).
21. Rouse, M. N., Olson, E. L., Gill, B. S., Pumphrey, M. O. & Jin, Y. Stem rust resistance in *Aegilops tauschii* germplasm. *Crop Sci.* **51**, 2074–2078 (2011).
22. Yu, G. et al. Identification and mapping of *Sr46* from *Aegilops tauschii* accession CIae 25 conferring resistance to race TTKSK (Ug99) of wheat stem rust pathogen. *Theor. Appl. Genet.* **128**, 431–443 (2015).
23. Olson, E. L. et al. Simultaneous transfer, introgression, and genomic localization of genes for resistance to stem rust race TTKSK (Ug99) from *Aegilops tauschii* to wheat. *Theor. Appl. Genet.* **126**, 1179–1188 (2013).

Acknowledgements

We are grateful to the germplasm banks at Kansas State University, International Center for Agricultural Research in the Dry Areas, USDA-ARS National Small Grains Collection, and the N. I. Vavilov Research Institute of Plant Industry, and to our colleagues J. Raupp, J. Dvorak, and C. Hiebert for providing seed of *Ae. tauschii*. We thank our colleagues Y. Yue and JIC Horticultural Services for plant husbandry, J. Brown for helpful discussions, M. Ambrose and A. Meldrum for help with MTAs, K. Witek for technical assistance with RenSeq, M. Rouse and Y. Jin for PGT isolates, P. Solomon for help with bait source sequences, the International Wheat Sequencing Consortium for pre-publication access to REFSEQ v1.0, and B. McIntosh for critical reading of the manuscript. This research was supported by the NBI Computing Infrastructure for Science (CiS) group and financed by the Two Blades Foundation, USA to B.J.S. and B.B.H.W.; the Biotechnology and Biological Sciences Research Council (BBSRC) to B.B.H.W.; the BBSRC Designing Future Wheat Cross-Institute Strategic Programme to A.R.B. and B.B.H.W.; the Norwich Research Park Translational Fund to B.B.H.W.; the Lieberman-Okinow Endowment at the University of Minnesota to B.J.S.; the Grains Research and Development Corporation, Australia to E.L. and H.B.; the Gordon and Betty Moore Foundation to J.D.G.J.; and in-kind support by Arbor Biosciences.

Author contributions

S.A., A.R.B., J.P., N.S., S.G., E.L., and B.B.H.W. configured the diversity panel. S.A. and J.C. extracted DNA and performed phylogenetic analysis. B.S., S.A., and J.E. designed and tested bait library. O.M., R.J., S.A., N.K., L.J.S., and B.J.S. phenotyped the diversity panel. J.E. prepared enriched libraries. S.A., G.Y., B.S., and B.B.H.W. performed AgRenSeq pilot studies. B.S. designed, implemented, and visualized the AgRenSeq data matrix, while K.G. and S.A. implemented the regression model and sample size power study. S.C., Y.L., S.P., N.A., H.B., M.A., S.S.X., and E.L. map-base cloned *Sr46*. S.P., M.A.M.H., and M.A. made *Sr45* transgenics. E.O. mapped *SrTA1662*. B.B.H.W. conceived the study and drafted the manuscript with S.A., B.S., K.G., J.D.G.J., A.R.B., M.A., E.O., S.S.X., B.J.S., and E.L. All authors read and approved the final manuscript.

Competing interests

Patent applications filed based on this work include PCT/US2019/013430 by B.B.H.W., B.S. and S.A. and PCT/GB2019/050077 by B.B.H.W., B.S., S.A. and K.G.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41587-018-0007-9.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to B.B.H.W.

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

Methods

***Ae. tauschii* diversity panel and DNA extraction.** A set of 195 *Ae. tauschii* accessions were assembled from the Wheat Genetics Resource Center (Kansas State University, USA), The Leibniz Institute of Plant Genetics and Crop Plant Research (Gatersleben, Germany), National Small Grains Collection (Beltsville, Maryland, USA), International Center for Agricultural Research in the Dry Areas (Aleppo, Syria), The Commonwealth Scientific and Industrial Research Organisation (Canberra, Australia), The N. I. Vavilov Institute of Plant Industry (St. Petersburg, Russia), University of California-Davis (Davis, California, USA), and Agriculture and Agri-Food Canada (Winnipeg, Canada). The majority of these accessions were originally collected from diverse collection sites around the Caspian Sea (Fig. 2a and Supplementary Table 2). For each accession, we used seeds obtained by self-pollination of a single plant in a glasshouse at the John Innes Centre in 2016. High-molecular-weight genomic DNA was extracted from leaf tissue using a modified CTAB method[24]. DNA quantification was performed using a NanoDrop spectrophotometer (Thermo Scientific) and the Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies).

Design of *Ae. tauschii* bait library. *NLRs.* We used several available sequence resources from *Ae. tauschii* and related genomes to generate a set of baits with high potential to capture the NLR repertoire across the *Ae. tauschii* species complex. Transcriptome raw data from publicly available datasets from *Ae. tauschii* (ERR420230, ERR420231), *Triticum urartu* (PRJNA191053)[25], *T. turgidum* (PRJNA191054)[25], *Ae. sharonensis* (PRJEB5340)[24], and hexaploid wheat (ERX1053979[7], PRJEB23474) were assembled using Trinity[4] with default parameters. In addition, we added NLR gene models from *T. aestivum*[26,27], *Hordeum vulgare*[28], and *Brachypodium distachyon*[29]. All resources were screened for NLR association using NLR-Parser[20].

KASP markers. A total of 2,110 SNP markers were used in the design of the bait library. These markers were composed of 331 KASP markers selected from an existing screen of six *Ae. tauschii* accessions (Ent-230, Ent-088, Ent-323, Ent-336, Ent-392, Ent-414) available from www.cerealsdb.net[30]; 682 KASP markers designed using the probe sequences from the 820K Axiom array (Affymetrix)[31] based on polymorphism among 14 *Ae. tauschii* accessions (data available from www.cerealsdb.net and as Supplementary Table 2 in the study by Winfield et al.[31]); 363 KASP markers selected from screening of the 6 *Ae. tauschii* accessions (as above) on the 90K iSelect array[32]; 662 probes from the 90K D-genome markers were shortlisted based on D-genome map position and conversion into KASP markers on the NIAB MAGIC population[33] (map available from www.niab.com/magic); 38 KASP markers from a previous characterization of *Ae. tauschii* diversity[17]; and 34 KASP markers used for the development of standardized marker-assisted selection assays (http://www.cerealsdb.uk.net/cerealgenomics/CerealsDBNEW/download_mas.php?URL=/cerealgenomics/CerealsDBNEW/FORM_SNPs_vs_Brachy.php). Only 800 marker sequences could be aligned to the AL8/78 reference sequence (National Center for Biotechnology Information GenBank assembly accession: GCA_000347335.1). The best hit for these markers was elongated to 240 bp. The International Union of Pure and Applied Chemistry base for the SNP was replaced by the allele found in AL8/78.

Pathogenesis-related genes. A set of 23 Triticeae pathogenesis-related (PR) genes were made available by Peter Solomon, Australian National University, Canberra, Australia, and included in the design (although not a subject of the present study).

The resulting resource of NLR genes, *Ae. tauschii* SNP markers, and PR genes were masked for repeats using RepeatMasker (http://repeatmasker.org) and the Triticeae Repeat Database[34]. A set of 58,943 baits were generated from the unmasked sequences. Bait sequences are available at https://github.com/steuernb/AgRenSeq. The program for generating baits is available at https://github.com/steuernb/BaitLibraryBuilder.

Library preparation and sequencing. Illumina libraries with an average insert size of 700 bp were enriched in groups of four by Arbor Biosciences, Michigan, USA, as previously described[18], and sequenced on an Illumina HiSeq with 250 bp PE reads at Novogene, China to generate an average of 1.67 Gb per accession (Supplementary Table 5).

Testing bait library. Bait sequences were aligned to AL8/78 genome sequence using BLASTn[35]. Local alignments were filtered for matches with at least length of 100 and 80% identity. Regions from the genome sequence with matches to at least two baits were extracted including 1,000 bp flanking sequence. Reads from RenSeq of AL8/78 accession were aligned to extracted regions using BWA[36]. Mapping was converted to mpileup format using SAMtools[37], and the percentage of bases in those regions that were covered by RenSeq reads were extracted from the mpileup file using a custom script. Of a total of 1,950,922 positions in regions, 1,952,834 (99.9%) were covered.

NLR assembly parameters. The primary sequence data (Supplementary Table 5) were trimmed using Trimmomatic v0.2[38] and de novo assembled with the CLC Assembly Cell (http://www.clcbio.com/products/clc-assembly-cell/) using word size (-w = 64) with standard parameters. The output of the CLC Assembler was a FASTA file containing all the contig sequences.

Construction of phylogenetic tree and selection of non-redundant accessions. The KASP marker sequences were extracted from the RenSeq assemblies of each accession using BLASTn[36] (Supplementary Table 3), and the sequences were multiple-aligned to score the SNP variation across the accessions. Marker sequences with minor allele frequency less than 5% in the panel and missing data greater than 60% were excluded from the analysis. The resulting genotyping matrix with KASP markers sequences (Supplementary Table 4) was used to build a UPGMA (unweighted pair group method with arithmetic mean) tree with 100 bootstraps using the Bio.Phylo module from the Biopython (http://biopython.org) package[39]. Further, a Python script was used to generate an iTOL (https://itol.embl.de/) compatible tree for rendering and annotation. The Python scripts used for generating the tree are available at https://github.com/arorasanu/KASPTree. In addition, an identity-by-state matrix was computed for all possible pairs of accessions to group those accessions with ≥99% genetic identity. The set of 151 non-redundant accessions (Supplementary Table 6) was selected for association mapping based on genetic identity (99%), variation in phenotype data of accessions within identity group, significance of any accession reported in the literature, and the seed availability in our stock.

***Ae. tauschii* stem rust phenotyping for AgRenSeq.** Two-week-old seedlings were inoculated with PGT spores and infection types (IT) were scored 2 weeks later as previously described[24] based on the size (relative length and width of uredinia, if present) and type (presence of chlorosis and/or necrosis) of uredinia present according to the wheat stem rust scoring system of Stakman et al.[40]. Symbols + and − denote more or less sporulation, respectively, of classically described uredinia. Plants scored as IT = 1 would have slightly larger pustules than plants scored as IT = 1−. Similarly, plants scored as IT = 2+ would have slightly larger pustules than plants scored as IT = 2. These minor differences are captured in Supplementary Fig. 4 and Supplementary Table 8.

***k*-mer presence/absence matrix.** *k*-mers (*k* = 51) were counted in trimmed raw data per accession using Jellyfish[41]. *k*-mers with a count of less than 10 in an accession were discarded immediately. *k*-mer counts from all accessions were integrated to create a presence/absence matrix with one row per *k*-mer and one column per accession. The entries were reduced to 1 (presence) and 0 (absence). Subsequently, the matrix was cleaned for *k*-mers that occur in less than five accessions or occur in all but four or fewer accessions. Programs to process the data were implemented in Java and are published at https://github.com/steuernb/AgRenSeq.

Phenotype scale conversion and filtering of *k*-mers based on NLR annotation. Phenotype scores were converted from Stakman infection types[40] to numeric values between −2 and 2 associating (0, ;, 1−, 1, 1+, 2−, 2, 2+, 3−, 3, 3+, 4) with (2, 1.67, 1.33, 1, 0.67, 0.33, 0, −0.33, −0.67, −1, −1.33, −2). Values from replicates were combined into a mean. For a resistant accession, that is, with the expectation of carrying a R gene, contigs associated with NLR genes were annotated using NLR-Parser v2 (https://github.com/steuernb/MutantHunter)[20]. Only *k*-mers occurring in the NLR-associated contigs were considered for association analysis.

This part of the pipeline was implemented in Java (available at https://github.com/steuernb/AgRenSeq). The pipeline was then translated into Python and coupled to the association mapping steps described below (available at https://github.com/kgaurav1208/AgRenSeq_GLM).

Correlation pre-filtering. For each of the NLR-associated *k*-mers identified in the previous step, if also present in the pre-calculated presence/absence matrix, correlation between the vector of that *k*-mer's presence/absence and the vector of the phenotype scores was calculated according to the following equation:

$$\mathrm{Cor} = \sum_{i=1}^{N} \frac{(K_i - \mu_K) \times (\mathrm{pheno}_i - \mu_{\mathrm{pheno}})}{\sigma_K \times \sigma_{\mathrm{pheno}}}$$

where N represents the number of accessions, K_i represents the presence/absence of the *k*-mer in the *i*th accession, pheno_i represents the phenotype score of the *i*th accession, μ_K and σ_K represent the mean and the standard deviation of the *k*-mer's presence/absence vector, respectively, and μ_{pheno} and σ_{pheno} represent the mean and the standard deviation of the vector of phenotype scores, respectively.

Only those *k*-mers for which the correlation obtained was positive were retained as it makes biological sense and reduces the computational resources required for the next step. The correlation threshold can be set even higher (for example, 0.2) to further reduce the computing time by filtering out the *k*-mers unlikely to be significantly associated with the phenotype.

Linear regression model accounting for population structure. For each filtered *k*-mer, a linear regression model[16] was fitted to predict the phenotype of an accession based on whether it contains the *k*-mer, while using a number of significant PCA dimensions obtained from the SNP marker matrix as covariates to control for the population structure. The regression model for a particular *k*-mer can be described by:

$$\text{pheno} \sim \alpha K + \beta_1 P_1 + \beta_2 P_2 + \ldots + \beta_n P_n$$

where pheno represents the vector of the phenotype scores, *K* represents the presence/absence vector of the *k*-mer, and $P_1, P_2, \ldots, P_n$ represent the *n* most significant PCA dimensions. Fitting the above regression model means finding the coefficients $\alpha, \beta_1, \beta_2, \ldots, \beta_n$ such that the Euclidean distance between the vectors on the left- and right-hand sides of the above expression is minimized. A likelihood ratio test for nested models[42] was then used to obtain a *P* value for each *k*-mer.

The exact number of significant PCA dimensions was chosen heuristically. We observed that ten dimensions over-corrected for population structure in the case of certain rare variants (*Sr33*), which are also not well-distributed across the population structure, while one dimension did not correct sufficiently for population structure in the case of other phenotypes. In the context of this study, three dimensions were found to represent a good trade-off.

Power and sample size. For each phenotype and a sample size in the range of 70–150, random subsampling of accessions was done 50 times, and the above pipeline was run for each subsample[43]. The power was defined as the proportion of runs where the verified true positives (*SrTA1662* for QTHJC, *Sr45* and *Sr46* for TTKSK, and *Sr33* for RKQQC) were detected as the topmost candidate(s) for the corresponding phenotype. The minimum sample size at which the power dips below 80% was considered as the detection threshold.

Generating AgRenSeq plots. AgRenSeq plots were generated using R. In these plots, each integer on the *x* axis corresponds to one contig. Dots on plot are according to the *P* value of *k*-mers within each contig. Dot size is plotted according to number of *k*-mers with the specific *P* value. An example of an R script for plotting is published on https://github.com/steuernb/AgRenSeq.

Phylogenetic tree graphics. The phylogenetic tree with PGT phenotypes and *Sr* gene genotypes displayed in concentric circles (Fig. 2b) was constructed with iTOL v3[44].

Positional cloning of Sr46. Mapping families were developed at both the diploid (*Ae. tauschii*) and hexaploid (*T. aestivum*) levels. An F_3 family from AUS18913 (diploid donor of *Sr46*) × CPI110856 (stem rust susceptible diploid) was phenotyped with the Australian PGT race 34-1,2,3,4,5,6,7 (culture accession 103) at the Plant Breeding Institute, University of Sydney, Australia. At the hexaploid level, selections of resistant lines (R9.3, R11.4), derived from a cross between the synthetic hexaploid Langdon/AUS18913[45] and cv. Meering (stem rust susceptible), which carried only *Sr46*, were backcrossed to Meering or crossed directly with cv. Westonia. The simple sequence repeat markers *gwm1099* and *barc297*, located on the distal end of wheat chromosome 2DS, which flank the *Sr46* locus, were used to select 30 recombinant lines out of 960 F_2 seed from R9.3 × Meering and R11.4 × Westonia (Supplementary Fig. 8a). These recombinants were also phenotyped with PGT accession 103. We used a marker, *psr649*, closely linked (0.1 cM) to *Sr46* from the diploid family to identify a BAC contig, ctg5195, from the *Ae. tauschii* genome Assembly 1.1 (http://aegilops.wheat.ucdavis.edu/ATGSP/). Within this contig, we identified sequences with (1) a truncated NLR that carried only the nucleotide-binding domain and (2) an inosine-5′-monophosphate dehydrogenase (which was previously annotated as 'putative brown planthopper-induced resistance protein 1' and which possesses a CBS domain, a small domain originally identified in cystathionine beta-synthase) that co-segregated with *Sr46* (Supplementary Fig. 8b). A comparative genomics approach was adopted whereby BLAST searches identified an orthologous inosine-5′-monophosphate dehydrogenase gene in *Brachypodium* chromosome 5 (Bradi_5g01180) with an adjacent NB-ARC domain LRR (Bradi_5g01167) and a protein annotated with a LIM domain (Bradi_5g01160) (Supplementary Fig. 8c). Sequence similarity searches with the Bradi_5g01167 NLR against the wheat chromosome 2DS survey sequence identified a small family of related NLRs on three separate contigs, one of which contained adjacent sequences with a LIM domain (IWGSC 2DS survey sequences) (Supplementary Fig. 8d). Amplified fragments generated from consensus regions of these wheat NLRs were used to screen an AUS18913 BAC library[46]. A specific amplicon, primer S46ConsFA (5′-GCTGAGATTGTGCTTCTTCTAG-3′) + primer S46PREVR (5′-CATTGTACGCTCTGTCCAATA-3′), derived from re-sequencing AUS18913 with the consensus NLR co-segregated with *Sr46* and the corresponding BAC clone (2 A/B_60P19) was sequenced (Supplementary Fig. 8e). A single full-length gene with NB-ARC and LRR domains was identified in BAC2A/B_60P19 and further investigated as a candidate gene for *Sr46* (Supplementary Fig. 8f). Ethyl methane sulfonate mutagenesis was performed on line R9.3 according to procedures previously described[47]. Three M_2 mutants were identified (M1BS, M2S, M3B) and progeny tested at M_3 and M_4 generations. Mutant M1BS carried a large deletion of chromosome 2DS, whereas point mutations occurred in M2S and M3B (Supplementary Fig. 8f).

Validation of *Sr46* and *Sr45* candidate genes by transformation. Validation of the candidate *Sr46* gene, S46NLR- BAC2A/B_60P19, in wheat transformation (cv. Fielder) was conducted according to previous procedures[9] using a gene construct with the genomic clone which included 2 kb of the native promoter region and 1.8 kb of downstream sequences (PC92; Supplementary Table 10). Transgenic plants and sib lines were tested with the Australian PGT race 98–1,2,3,5,6 (virulent on Fielder) on T_0, T_1, and T_2 progenies (Supplementary Fig. 8g).

An NLR gene was previously identified as the candidate gene for *Sr45* based on map position and screening of six independent loss of function mutants derived from ethyl methane sulfonate treatment[3]. However, no transgenic wheats with the candidate gene sequence were generated and tested for rust response to confirm its *Sr45*-mediated stem rust resistance function. A transformation construct containing a native *Sr45* expression cassette (PC110; Supplementary Table 10) was assembled by amplification of a 6,481 bp fragment of genomic DNA including the *Sr45* coding sequence, 885 bp of 5′ regulatory sequence, and 1,508 bp of 3′ regulatory sequence with primers S45F1 (5′-AGTACTGTAATAATTGATTCCGTCG-3′) and S45R5 (5′-GAAATTCCTGCTGCATTGC-3′). The amplified fragment was cloned into the *Not*I site of the binary vector pVecBARII, a derivative of pWBvec8 in which the selectable marker cassette conferring resistance to hygromycin was replaced with sequences that confer resistance to bialaphos[48] to generate construct PC110. In addition, a transformation construct containing an *Sr45* expression cassette was constructed using the Golden Gate Modular Cloning Toolbox for Plants[49]. The *Sr33* promoter and *Sr33* terminator[9] were used to regulate expression of the *Sr45* coding sequence. Synthetic DNA fragments of these three sequences were obtained from a commercial provider each flanked by a pair of divergent *Bsa*I recognition sites for Golden Gate Cloning and four base pair standard fusion sites for TypeIIS assembly defined in the Plant Common Genetic Syntax[50]. These parts were assembled into the pICH47732 level one acceptor[51], and then the gene cassette was cloned into the *Not*I site of the pVec8 binary vector to produce PC147 (Supplementary Table 10).

The two *Sr45* constructs, with native and non-native regulatory elements, were transformed into the rust-susceptible wheat cultivar Fielder. A total of 12 primary transgenic plants were recovered containing construct PC110 and inoculated with PGT race 98-1,2,3,5,6. Seven lines (PC110-1, -2, -4, -5, -7, -10, -12) had an infection type ;1−, four lines (PC110-3, -6, -9, -11) had an infection type 1 while the transgenic line PC110-8 showed a susceptible infection type 3+ (similar to non-transgenic Fielder) as it carried a truncated *Sr45* gene sequence (Supplementary Fig. 6a). In addition, a total of 10 primary transgenic plants carrying PC147 were also tested with PGT race 98-1,2,3,5,6. Nine out of the 10 tested lines showed the typical *Sr45* resistance with an infection type of ;1 (Supplementary Fig. 6b).

Mapping of the *SrTA1662* candidate NLR gene. *SrTA1662* primer design. Sequence generated from 12-plex genotyping-by-sequencing using *Pst*I and *Msp*I enzymes according to Poland et al.[52] were available for the *Ae. tauschii* accession TA1662 and the hexaploid wheat genotype KS05HW14. A custom Python script was used to parse KS05HW14 and TA1662 sequences by barcode and trim barcode sequences. A BLAST database was created from the KS05HW14 and TA1662 sequences. The *SrTA1662* sequence was queried onto each database. Three genotyping-by-sequencing tags from KS05HW14 and one tag from TA1662 with *Pst*I cut sites were aligned to the *SrTA1662* sequence using MUSCLE (https://www.ebi.ac.uk/Tools/msa/muscle/) and targeted for primer design.

PCR conditions. The forward and reverse primer pairs, R1.F3 5′-AGTAAACTCCCGGCTGAGATA-3′ and R1.R3 5′-GCCGAGACTGAACTCAACTC-3′, respectively, amplify a 750 bp fragment in homozygous and heterozygous genotypes carrying *SrTA1662* (*SrTA1662/SrTA1662*; *SrTA1662/srTA1662*), whereas no fragment is amplified in individuals lacking a *SrTA1662* allele (*srTA1662/srTA1662*) (Supplementary Fig. 7a). Reaction conditions for amplification of the *SrTA1662* fragment were as follows: 13.6 µl ddH$_2$0, 2.4 µl 10X reaction buffer, 1.9 µl 2.5 mM dNTPs, 2.0 µl 10 pM of forward primer, 2.0 µl 10 pM of reverse primer, 0.1 µl (0.5 U) *Taq* polymerase, and 2 µl of template DNA (30 ng µl^{-1}). Cycling conditions include an initial denaturation of 95 °C (5 min) followed by 35 cycles of 95 °C (45 s), 63 °C (45 s), and 72 °C (90 s), and a final extension at 72 °C (10 min).

Mapping of the SrTA1662 fragment. The R1.F3 + R1.R3 PCR fragment was mapped to a 3.8 cM interval in 83 individuals from the U6714 BC$_2$F$_1$ mapping population[23] using six SSR markers and one expressed sequence tag–derived marker (Supplementary Fig. 7b and Supplementary Table 11). Marker order and distances were determined using R/qtl[53]. The 750 bp R1.F3 + R1.R3 PCR fragment co-segregated with resistance to PGT race QTHJC conferred by *SrTA1662* on 1DS (that is, zero recombinants were obtained among 83 BC$_2$F$_1$ plants).

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

Code availability

The programs, scripts, and bait library sequences used in this analysis are on Github (https://github.com/steuernb/AgRenSeq, https://github.com/kgaurav1208/AgRenSeq_GLM and https://github.com/arorasanu/KASPTree).

Data availability

Sequence reads were deposited in the European Nucleotide Archive (ENA) under project number PRJEB23912. The *Sr46* and *SrTA1662* loci were deposited at the National Center for Biotechnology Information under accession numbers MG851023 and MG763911. *Ae. tauschii* accessions with the GRU accession numbers in Supplementary Table 2 are available from the Germplasm Resources Unit, John Innes Centre, Norwich, UK (https://www.jic.ac.uk/germplasm/).

References

24. Yu, G. et al. Discovery and characterization of two new stem rust resistance genes in *Aegilops sharonensis*. *Theor. Appl. Genet.* **130**, 1207–1222 (2017).
25. Krasileva, K. V. et al. Separating homeologs by phasing in the tetraploid wheat transcriptome. *Genome Biol.* **14**, R66 (2013).
26. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
27. Choulet, F. et al. Structural and functional partitioning of bread wheat chromosome 3B. *Science* **345**, 1249721 (2014).
28. The International Barley Genome Sequencing Consortium. A physical, genetic and functional sequence assembly of the barley genome. *Nature* **491**, 711–716 (2012).
29. International Brachypodium Initiative Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* **463**, 763–768 (2010).
30. Wilkinson, P. A. et al. CerealsDB 2.0: an integrated resource for plant breeders and scientists. *BMC Bioinformatics* **13**, 219 (2012).
31. Winfield, M. O. et al. High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant. Biotechnol. J.* **14**, 1195–1206 (2016).
32. Wang, S. et al. Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. *Plant. Biotechnol. J.* **12**, 787–796 (2014).
33. Gardner, K. A., Wittern, L. M. & Mackay, I. J. A highly recombined, high-density, eight-founder wheat MAGIC map reveals extensive segregation distortion and genomic locations of introgression segments. *Plant. Biotechnol. J.* **14**, 1406–1417 (2016).
34. Wicker, T. M. D. E. & Keller, B. TREP: a database for Triticeae repetitive elements. *Trends Plant. Sci.* **7**, 561–562 (2002).
35. Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* **7**, 203–214 (2000).
36. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
37. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
38. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
39. Talevich, E., Invergo, B. M., Cock, P. J. & Chapman, B. A. Bio.Phylo: a unified toolkit for processing, analyzing and visualizing phylogenetic trees in Biopython. *BMC Bioinformatics* **13**, 209 (2012).
40. Stakman, E. C., Stewart, D. M. & Loegering, W. Q. *Identification of Physiological Races of* Puccinia graminis *var.* tritici E617 (United States Department of Agriculture, Agricultural Research Service, 1962).
41. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics* **27**, 764–770 (2011).
42. Wilks, S. S. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Stat.* **9**, 60–62 (1938).
43. Brynildsrud, O., Bohlin, J., Scheffer, L. & Eldholm, V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol.* **17**, 238 (2016).
44. Letunic, I. & Bork, P. Interactive tree of life (iTOL)v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245 (2016).
45. Lagudah, E., MacRitchie, F. & Halloran, G. M. Influence of high-molecular-weight glutenin subunits from *Triticum tauschii* on flour quality of synthetic hexaploid wheat. *J. Cereal Sci.* **5**, 129–138 (1987).
46. Moullet, O., Zhang, H. B. & Lagudah, E. S. Construction and characterisation of a large DNA insert library from the D genome of wheat. *Theor. Appl. Genet.* **99**, 305–313 (1999).
47. Mago, R. et al. Generation of loss-of-function mutants for wheat rust disease resistance gene cloning. *Methods Mol. Biol.* **1659**, 199–205 (2017).
48. Wang, M., Li, Z., Matthews, P. R., Upadhyaya, N. M. & Waterhouse, P. M. Improved vectors for *Agrobacterium tumefaciens*-mediated transformation of monocot plants. *Acta Hortic.* **461**, 401–408 (1998).
49. Engler, C. et al. A golden gate modular cloning toolbox for plants. *ACS Synth. Biol.* **3**, 839–843 (2014).
50. Patron, N. J. et al. Standards for plant synthetic biology: a common syntax for exchange of DNA parts. *New Phytol.* **208**, 13–19 (2015).
51. Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A modular cloning system for standardized assembly of multigene constructs. *PLoS ONE* **6**, e16765 (2011).
52. Poland, J. A., Brown, P. J., Sorrells, M. E. & Jannink, J. L. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* **7**, e32253 (2012).
53. Broman, K. W., Wu, H., Sen, S. & Churchill, G. A. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889–890 (2003).

nature research

Corresponding author(s): Brande Wulff, NBT-TR44213B

Last updated by author(s): Nov 20, 2018

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
□	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
□	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
□	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
□	☒	A description of all covariates tested
□	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
□	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	□	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	□	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	□	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	□	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

No software was used during data collection.

Data analysis

Bespoke code for the study is available from Github (links provided in the manuscript).
The softwares used in the study during data analysis are:
- Java Runtime Environments 1.6 or higher
- CLC assembly cell (https://www.qiagenbioinformatics.com/products/clc-assembly-cell/)
- NLR-Parser (github.org/MutantHunter)
- Jellyfish-2.1.4
- RStudio version 1.1.41
- Python programming (https://www.python.org) language version 2.7.13 from Anaconda (https://anaconda.org) built (64-bit) with Biopython (http://biopython.org/)) library v1.68
- BLAST+ command-line tools for alignment version 2.2.28
- iTOL (https://itol.embl.de/)
- Python 3.5.3 for AgRenSeq
- trimmomatic-0.33.jar

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequence reads were deposited in the European Nucleotide Archive (ENA) under project number PRJEB23912. The Sr46 and SrTA1662 loci were deposited at NCBI under accession numbers MG851023 and MG763911. Ae. tauschii accessions with the GRU accession numbers in Supplementary Table 2 are available from the Germplasm Resources Unit, John Innes Centre, Norwich, UK (https://www.jic.ac.uk/germplasm/). The programs, scripts and bait library sequences used in this analysis are on Github (https://github.com/steuernb/AgRenSeq, github/kgaurav1208/AgRenSeq_GLM and https://github.com/arorasanu/KASPTree).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample calculation was performed. We obtained as many Ae. tauschii ssp. strangulata accessions as we could get hold of.
Data exclusions	We excluded genetically redundant accessions - a common practice in GWAS. The data exclusions criteria was not pre-established. We have used the genetic identity of 99% along with other criteria which has been provided in the updated online methods.
Replication	The rust phenotypes were generated in replicates (Please refer to Supplementary table 8)
Randomization	No deliberate randomization was imposed on the phenotyping procedure. The phenotyping was done in a controlled environment where all of the exact same methods were used and scored by two experienced researchers who are co-authors on the manuscript. A single randomization order was used in the phenotyping test. When experiments are run with two or more replicates simultaneously, additional randomization orders of the accessions are generated by Excel. The rust phenotypes are very clear and consistent (see Table 8) such that we seldom have to worry about multiple replicates completed at the same time. Any accession that gave even a slightly equivocal reaction was phenotyped again in a separate experiment. Phenotyping experiments conducted separately over time provide a more robust validation than those with multiple replicates conducted at one time. Moreover, seed of these wild species is limited; thus, we must use caution as to how many replicates for phenotyping can be practically managed. While we agree that multi-replicate phenotyping experiments conducted over time and with different randomizations would be ideal, the fact that we identified very clear peaks associated with functional resistance genes (two known and two new genes) validates our methods and conclusions.
Blinding	The person doing the phenotyping did not have access to the genotype data. So in retrospect, the data collection was blinded.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging