BED-domain containing immune receptors confer diverse resistance spectra to yellow rust

Clemence Marchal[1*], Jianping Zhang[2,3,4*], Peng Zhang[2], Paul Fenwick[5], Burkhard Steuernagel[1], Nikolai M. Adamski[1], Lesley Boyd[6], Robert McIntosh[2], Brande B.H. Wulff[1], Simon Berry[5], Evans Lagudah[3], Cristobal Uauy[1 †]

[1] John Innes Centre, Norwich Research Park, Norwich NR4 7UH, United Kingdom

[2] University of Sydney, Plant Breeding Institute, Cobbitty, NSW 2570, Australia

[3] Commonwealth Scientific and Industrial Research Organization (CSIRO) Agriculture & Food, Canberra, ACT 2601, Australia

[4] Henan Tianmin Seed Company Limited, Lankao County, 475300, Henan Province, China

[5] Limagrain UK Ltd, Rothwell, Market Rasen, Lincolnshire, LN7 6DT, United Kingdom

[6] NIAB, Huntingdon Road, Cambridge, CB3 0LE, United Kingdom

*Clemence Marchal and Jianping Zhang contributed equally to this work

[†] Correspondence to cristobal.uauy@jic.ac.uk

Crop diseases reduce wheat yields by ~25% globally and thus pose a major threat to global food security[1]. Genetic resistance can reduce crop losses in the field and can be selected through the use of molecular markers. However, genetic resistance often breaks down following changes in pathogen virulence, as experienced with the wheat yellow (stripe) rust fungus *Puccinia striiformis* f. sp. *tritici* (*Pst*)[2]. This highlights the need to (i) identify genes that alone or in combination provide broad-spectrum resistance, and (ii) increase our understanding of the underlying molecular modes of action. Here we report the isolation and characterisation of three major yellow rust resistance genes (*Yr7*, *Yr5*, and *YrSP*) from hexaploid wheat (*Triticum aestivum*), each having a distinct recognition specificity. We show that *Yr5*, which remains effective to a broad range of *Pst* isolates worldwide, is closely related yet distinct from *Yr7*, whereas *YrSP* is a truncated version of *Yr5* with 99.8% sequence identity. All three *Yr* genes belong to a complex resistance gene cluster on chromosome 2B encoding nucleotide-binding and leucine-rich repeat proteins (NLRs) with a non-canonical N-terminal zinc-finger BED domain[3] that is distinct from those found in non-NLR wheat proteins. We developed diagnostic markers to accelerate haplotype analysis and for marker-assisted selection to expedite the stacking of the non-allelic *Yr* genes. Our results provide evidence that the BED-NLR gene architecture can provide effective field-based resistance to important fungal diseases such as wheat yellow rust.

In plant immunity, NLRs act as intracellular immune receptors that upon pathogen recognition trigger a series of signalling steps that ultimately lead to cell death, thus preventing the spread of infection[4,5]. The NB-ARC domain is the hallmark of NLRs which in most cases include leucine-rich repeats (LRRs) at the C-terminus. Recent *in silico* analyses have identified NLRs with additional 'integrated' domains[6–8], including zinc-finger BED domains (BED-NLRs). The BED domain function within BED-NLRs is unknown, although the BED domain from the non-

NLR DAYSLEEPER protein was shown to bind DNA in *Arabidopsis*[9]. BED-NLRs are widespread across Angiosperm genomes[6–8] and this gene architecture has been shown to confer resistance to bacterial blight in rice (*Xa1*[10,11]).

The genetic relationship between *Yr5* and *Yr7* has been debated for almost 45 years[12,13]. Both genes map to chromosome arm 2BL in hexaploid wheat and were hypothesized to be allelic[14], and closely linked with *YrSP*[15]. Whilst only two of >6,000 tested *Pst* isolates worldwide have been found virulent to *Yr5* (Supplementary Table 1[16,17]), both *Yr7* and *YrSP* have been overcome in the field. For *Yr7*, this is likely due to its wide deployment in cultivars (Supplementary Table 2, Supplementary Figure 1). This highlights the importance of stewardship plans (including diagnostic markers) to deploy *Yr5* in combination with other genes as currently done in the USA (e.g. *Yr5*+*Yr15*; UC Davis breeding programme).

To clone the genes encoding *Yr7*, *Yr5*, and *YrSP*, we identified susceptible ethyl methanesulfonate-derived (EMS) mutants from different genetic backgrounds carrying these genes (Figure 1, Supplementary Tables 3-4). We performed MutRenSeq[18] and isolated a single candidate contig for each of the three genes based on nine, ten, and four independent susceptible mutants, respectively (Figure 1a; Supplementary Figure 2). The three candidate contigs were genetically linked to a common mapping interval, previously identified for the three *Yr* loci[15,19,20]. No recombinant was previously found between *Yr7* and *Yr5* among 143 F_3 progenies[14] and we observed no recombination between *YrSP* and *Yr7* (208 F_3 lines) nor *YrSP* and *Yr5* (256 F_3 lines; Supplementary Table 5). Their closest homologs in the Chinese Spring wheat genome sequence (RefSeq v1.0) all lie within this common genetic interval (Figure 1b; Supplementary Figure 3).

Within each contig we predicted a single open reading frame based on RNA-Seq data. All three predicted *Yr* genes displayed similar exon-intron structures (Figure 1a), although *YrSP* was truncated in exon 3 due to a single base deletion that resulted in a premature termination codon. The 23 mutations identified by MutRenSeq were confirmed by Sanger sequencing and all lead to either an amino acid substitution or a truncation allele (splice junction or termination codon) (Figure 1a; Supplementary Table 4). The DNA sequences of *Yr7* and *Yr5* were 77.9% identical across the complete gene; whereas *YrSP* was a truncated version of *Yr5*, sharing 99.8% identity in the common sequence (Supplementary Files 1 and 2). This high sequence identity between *YrSP* and *Yr5* is on par with that seen for previously characterised allelic series in the wheat *Pm3* (>97% identity)[21] and flax *L* (>90% identity)[22] resistance genes and would suggest that *Yr5* and *YrSP* are allelic. Based on this evidence, we cannot discard the alternative explanations that *Yr5* and *YrSP* are closely linked paralogous genes that arose from a very recent duplication event or that *Yr7* is an allele of *Yr5* that originated from a very diverse haplotype. The absence of recombination between the pairwise populations suggests that *Yr7*, *Yr5*, and *YrSp* are linked in repulsion, but we cannot discriminate between paralogous or allelic relationships. However, the high sequence identity alongside the genetic analyses support the hypothesis that *Yr5* and *YrSP* are derived from a common sequence and most likely constitute alleles, whereas *Yr7* is encoded by a closely related, yet distinct gene.

The Yr7, Yr5, and YrSP proteins contain a zinc-finger BED domain at the N-terminus, followed by the canonical NB-ARC domain. Unlike previously cloned resistance genes in grasses (e.g. *Mla10*[23], *Sr33*[24], *Pm3*[25]), neither *Yr7* nor *Yr5*/*YrSP* encode Coiled Coil domains at the N-terminus (Supplementary Figure 4). Only the Yr7 and Yr5 proteins encode multiple LRR motifs at the C-terminus (Figure 2a; green bars), YrSP having lost most of the LRR region due the premature termination codon in exon 3. YrSP still confers functional resistance to *Pst*,

although with a recognition specificity different from Yr5 (Supplementary Table 1; all isolates virulent to *YrSP* are avirulent to *Yr5*, whereas the two isolates virulent to *Yr5* are avirulent to *YrSP*[16]. Yr7 and Yr5/YrSP are highly conserved in the N-terminus, with a single amino-acid change in the BED domain. This high degree of conservation is eroded downstream of the BED domain (Figure 2a). The BED domain is required for *Yr7*-mediated resistance, as a single amino acid change in mutant line Cad0903 led to a susceptible reaction (Figure 1a). However, recognition specificity is not solely governed by the BED domain, as *Yr5* and *YrSP* have identical BED domain sequences, yet confer resistance to different *Pst* isolates. The highly conserved Yr7 and Yr5/YrSP BED domains could function in a similar way to the integrated WRKY domain in the *Arabidopsis* RRS1-R immune receptor which binds unrelated bacterial effectors yet activates defense response through mechanisms involving other regions of the protein[26].

We examined the variation in *Yr7*, *Yr5*, and *YrSP* across eight sequenced tetraploid and hexaploid wheat genomes (Supplementary Table 6). We identified *Yr7* only in Cadenza and Paragon, which are identical-by-descent in this interval (Supplementary File 3, Supplementary Table 7, and Supplementary Figure 5). Both cultivars are derived from the original source of *Yr7*, tetraploid durum wheat (*T. turgidum* ssp. *durum*) cultivar Iumillo and its hexaploid derivative Thatcher (Supplementary Figure 5). None of the three sequenced tetraploid accessions (Svevo, Kronos, Zavitan) carry *Yr7* (Supplementary Table 7).

For *Yr5/YrSP*, we identified three additional haplotypes in the sequenced hexaploid wheat cultivars (Figure 2b; Supplementary Table 8). Cultivar Claire encodes a complete NLR with six amino-acid changes, including one within the NB-ARC domain, and six polymorphisms in the C-terminus compared to Yr5. Cultivars Robigus, Paragon, and Cadenza also encode a full

length NLR that shares common polymorphisms with Claire, in addition to 19 amino acid substitutions across the BED and NB-ARC domains. The presence of the *Yr5*/*YrSP* haplotype in Cadenza (which also carries *Yr7*) further supports the non-allelic relationship of these genes. The C-terminus polymorphisms between Yr5 and the other cultivars is due to a 774 bp insertion in *Yr5*, close to the 3' end, which carries an alternate termination codon (Supplementary File 2). Tetraploid cultivars Kronos and Svevo encode a fifth Yr5/YrSP haplotype with a truncation in the LRR region distinct from YrSP, in addition to multiple amino acid substitutions across the C-terminus (Supplementary Table 8). This truncated tetraploid haplotype is reminiscent of YrSP and is expressed in Kronos (see Methods). However, none of these cultivars (Claire, Robigus, Paragon, Cadenza, Svevo, and Kronos) exhibit a *Yr5*/*YrSP* resistance response, suggesting that these amino acid changes and truncations may alter recognition specificity or protein function. Additional testing of these haplotypes will provide insight into whether they represent a functional allelic series.

We designed diagnostic markers for *Yr7*, *Yr5*, and *YrSP* to facilitate their detection and use in breeding. We confirmed their presence in the donor cultivars Thatcher and Lee (*Yr7*), Spaldings Prolific (*YrSP*), and spelt wheat cv. Album (*Yr5*) (Supplementary Tables 9-10; Supplementary Figures 5-6). We tested the *Yr7* and *YrSP* markers in a collection of global landraces[27] and European cultivars[28] released over the past century. *YrSP* was absent from the tested germplasm, except for AvocetS-*YrSP* (Supplementary Table 10). *Yr7* on the otherhand was more prevalent in the germplasm tested and we could track its presence across pedigrees, including Cadenza-derived cultivars (Supplementary Tables 9-10; Supplementary Figure 5). We confirmed *Yr5* in the AvocetS-*Yr5* and Lemhi-*Yr5* lines, in addition to wheat cultivars in which *Yr5* has been introduced using gel-based flanking markers (Supplementary Table 11 and

Supplementary Figure 6). The *Yr5* diagnostic marker will facilitate its deployment, hopefully within a breeding strategy that ensures its effectiveness long-term[29].

We defined the *Yr7/Yr5/YrSP* syntenic interval across the wheat genomes and related grass species *Aegilops tauschii* (D genome progenitor), *Hordeum vulgare* (barley), *Brachypodium distachyon*, and *Oryza sativa* (rice) (Supplementary files 4 and 5, Supplementary Figure 7). We identified both canonical NLRs, as well as BED-NLRs across all genomes and species, except for barley, which only contained canonical NLRs across the syntenic region. The phylogenetic relationship based on the NB-ARC domain suggests a common evolutionary origin of these integrated domain NLR proteins before the wheat-rice divergence (~50 Mya) and an expansion in the number of NLRs in the A and B genomes of polyploid wheat species (Figure 3a; Supplementary Figure 8). Within the interval we also identified several genes in the A, B, and D genomes that encode two consecutive in-frame BED domains (named BED-I and BED_II; Figure 3b-c, Supplementary Figure 7) followed by the canonical NLR. The BED domains in these genes were fully encoded within a single exon (exons 2 and 3) and in most cases had a four-exon structure (Figure 3c). This is consistent with the three-exon structure of single BED domain genes, such as *Yr7* and *Yr5/YrSP* (BED-I encoded on exon 2). To our knowledge this is the first report of the double BED domain NLR protein structure. The biological function of this molecular innovation remains to be determined, although our data show that the single BED-I structure can confer *Pst* resistance and is required for *Yr7*-mediated resistance.

Among other mechanisms, integrated domains of NLRs are hypothesised to act as decoys for pathogen effector targets[5]. This suggests that the integrated domain might be sequence-related to the host protein targeted by the effector. To identify these potential effector targets in the

host, we retrieved all BED-domain proteins (108) from the hexaploid wheat genome, including 25 BED-NLRs, and additional BED-NLRs located in the syntenic intervals (Supplementary Table 12; Supplementary file 4). We also retrieved the rice Xa1[10,11] and ZBED proteins, the latter being hypothesized to mediate rice resistance to *Magnaporthe oryzae*[7]. We used the split network method implemented in SplitsTree4[30] to represent the relationships between these BED domains (Figure 3d; Supplementary Figure 9). Overall, BED domains are diverse, although there is evidence of a split between BED domains from BED-NLRs and non-NLR proteins (only 7 of 83 non-NLRs clustered with the BED-NLRs). Given that the base of the split is broad, integrated BED-domains most likely derive from multiple integration events. However, Yr7 and Yr5/YrSP both arose from a common integration event that occurred before the *Brachypodium*-wheat divergence (Supplementary Figure 9, purple). This is consistent with the hypothesis that integrated domains might have evolved to strengthen the interaction with pathogen effectors after integration[31], although we cannot exclude the potential role of the BED domains in signalling at this stage.

Among BED-NLRs, BED-I and BED-II constitute two major clades, consistent with their relatively low amino acid conservation (Figure 3b), that are comprised solely of genes from within the *Yr7/Yr5/YrSP* syntenic region. Seven non-NLR BED domain wheat proteins clustered with BED-NLRs. These are most closely related to the *Brachypodium* and rice BED-NLR proteins and were not expressed in RNA-Seq data from a *Yr5* time-course (re-analysis of published data[32]; Supplementary Figure 10, Supplementary Table 13). Similarly, no BED-containing protein was differentially expressed during this infection time-course, consistent with the prediction that effectors alter their targets' activity at the protein level in the integrated-decoy model[5]. We cannot however disprove that these closely related BED-containing proteins are involved in BED-NLR-mediated resistance.

BED-NLRs are frequent in Triticeae, and occur in other monocot and dicot tribes[6–8]. To date a single BED-NLR gene, *Xa1*, has been shown to confer resistance to plant pathogens[10,11]. In the present study, we show that the distinct *Yr7*, *Yr5*, and *YrSP* resistance specificities belong to a complex NLR cluster on chromosome 2B and are encoded by BED-NLRs genes that are linked in repulsion. We report five haplotypes for *Yr5/YrSP*, including three full-length BED-NLRs (including *Yr5*) and two truncated versions (including *YrSP*). These alternative haplotypes could be of functional significance as previously shown for the *Mla* and *Pm3* loci that confer resistance to *Blumeria graminis*[25,33] in barley and wheat, respectively, and the flax *L* locus conferring resistance to *Melampsora lini*[22]. Overall, our results add strong evidence for the importance of the BED-NLR architecture in plant-pathogen interactions. The relationship of these three distinct *Yr* loci will inform future hypothesis-driven engineering of novel recognition specificities.

Methods

MutRenSeq

Mutant identification

Supplementary Table 3 summarises plant materials and *Pst* isolates used to identify mutants for each *Yr* gene. We used an EMS-mutagenised population in cultivar Cadenza[34] to identify mutants in *Yr7* using a forward genetic screen; whereas EMS-populations in the corresponding AvocetS-*Yr* near isogenic lines (NIL) were used to identify *Yr5* and *YrSP* mutants. For *Yr7*, we inoculated M_3 plants from the Cadenza EMS population with *Pst* isolate 08/21 which is virulent to *Yr1, Yr2, Yr3, Yr4, Yr6, Yr9, Yr17, Yr32, YrRob,* and *YrSol*[35]. We hypothesised that susceptible mutants would carry mutations in *Yr7*. Plants were grown in 192-well trays in a confined glasshouse with no supplementary lights or heat. Inoculations were performed at the one leaf stage (Zadoks 11) with a talc-urediniospore mixture. Trays were kept in darkness at 10 °C and 100% humidity for 24 hours. Infection types (IT) were recorded 21 days post-inoculation (dpi) following the Grassner and Straib scale[36]. Identified susceptible lines were progeny tested (twelve to 16 plants per line) to confirm the reliability of the phenotype. DNA from all seven confirmed M_4 plants was used for RenSeq (see section below). Similar methods were used for AvocetS-*Yr7*, AvocetS-*Yr5*, and AvocetS-*YrSP* EMS-mutagenised populations with the following exceptions: *Pst* pathotypes 108 E141A+ (University of Sydney Plant Breeding Institute Culture no. 420), 150 E16A+ (Culture no. 598) and 134 E16A+ (Culture no. 572) were used to evaluate *Yr7*, *Yr5*, and *YrSP* mutants, respectively. The seven EMS-derived susceptible mutants in Lemhi-*Yr5* were previously identified[37] and progeny tested. DNA from M_5 plants from all seven mutants was used for RenSeq.

DNA preparation, resistance gene enrichment and sequencing (RenSeq)

We extracted total genomic DNA from young leaf tissue using the large-scale DNA extraction protocol from the McCouch Lab (https://ricelab.plbr.cornell.edu/dna_extraction) and a previously described method[38]. We checked DNA quality and quantity on a 0.8% agarose gel and with a NanoDrop spectrophotometer (Thermo Scientific). Arbor Biosciences (Ann Arbor, MI, USA) performed the targeted enrichment of NLRs according to the MYbaits protocol using an improved version of the previously published Triticeae bait library available at github.com/steuernb/MutantHunter. Library construction was performed using the TruSeq RNA protocol v2 (Illumina 15026495). Libraries were pooled with one pool of samples for Cadenza mutants and one pool of eight samples for the Lemhi-*Yr5* parent and Lemhi-*Yr5* mutants. AvocetS-*Yr5* and AvocetS-*YrSP* wild-type, together with their respective mutants, were also processed according to the MYbaits protocol and the same bait library was used. All enriched libraries were sequenced on a HiSeq 2500 (Illumina) in High Output mode using 250 bp paired end reads and SBS chemistry. For the Cadenza wild-type, we generated data on an Illumina MiSeq instrument. In addition to the mutants, we also generated RenSeq data for Kronos and Paragon to assess the presence of *Yr5* in Kronos and *Yr7* in Paragon. Details of all the lines sequenced, alongside NCBI accession numbers, are presented in Supplementary Tables 4 and 14.

MutantHunter pipeline

We adapted the pipeline from https://github.com/steuernb/MutantHunter/ to identify candidate contigs for the targeted *Yr* genes. First, we trimmed the RenSeq-derived reads with trimmomatic[39] using the following parameters: ILLUMINACLIP:TruSeq2-PE.fa:2:30:10 LEADING:30 TRAILING:30 SLIDINGWINDOW:10:20 MINLEN:50 (v0.33). We made *de novo* assemblies of wild-type plant trimmed reads with the CLC assembly cell and default parameters apart from the word size (-w) parameter that we set to 64 (v5.0,

http://www.clcbio.com/products/clc-assembly-cell/) (Supplementary Table 15). We then followed the MutantHunter pipeline detailed at https://github.com/steuernb/MutantHunter/. For Cadenza mutants, we used the following MutantHunter program parameters to identify candidate contigs: -c 20 -n 6 -z 1000. These options require a minimum coverage of 20x for SNPs to be called; at least six susceptible mutants must have a mutation in the same contig to report it as candidate; small deletions were filtered out by setting the number of coherent positions with zero coverage to call a deletion mutant at 1000. The -n parameter was modified accordingly in subsequent runs with the Lemhi-*Yr5* datasets (-n 6).

To identify *Yr5* and *YrSP* contigs from Avocet mutants, we followed the MutantHunter pipeline with all default parameters, except in the use of CLC Genomics Workbench (v10) for reads QC, trimming, *de novo* assembly of Avocet wild-type and mapping all the reads against *de novo* wild-type assembly. Default MutantHunter parameters were used except that –z was set as 100. The parameter –n was set to 2 in the first run and then to 3 in the second run. Two *Yr5* mutants were most likely sibling lines as they carried identical mutations at the same position (Supplementary Figure 2, Supplementary Table 4).

For *Yr7* we identified a single contig with six mutations, however we did not identify mutations in line Cad0903. Upon examination of the *Yr7* candidate contig we predicted that the 5' region was likely to be missing (Supplementary Figure 2). We thus annotated potential NLRs in the Cadenza genome assembly available from the Earlham Institute (Supplementary Table 6, http://opendata.earlham.ac.uk/Triticum_aestivum/EI/v1.1) with the NLR-Annotator program using default parameters (https://github.com/steuernb/NLR-Annotator). We identified an annotated NLR in the Cadenza genome with 100% sequence identity to the *Yr7* candidate contig, which extended beyond our *de novo* assembled sequence. We therefore replaced the

previous candidate contig with the extended Cadenza sequence (100% sequence identity) and mapped the RenSeq reads from Cadenza wild-type and mutants as described above. This confirmed the candidate contig for *Yr7* as we retrieved the missing 5' region including the BED domain. The improved contig now also contained a mutation in the outstanding mutant line Cad0903 (Supplementary Figure 2). The Triticeae bait library does not include integrated domains in its design so they are prone to be missed, especially when located at the ends of an NLR. Sequencing technology could also have accounted for this: MiSeq was used for Cadenza wild-type whereas HiSeq was chosen for Lemhi-*Yr5* and we recovered the 5' region in the latter, although coverage was lower than for the regions encoding canonical domains. In summary, we sequenced nine, ten, and four mutants for *Yr7*, *Yr5*, and *YrSP*, respectively, and identified for each target gene a single contig that accounted for all progeny tested susceptible mutants.

Candidate contig confirmation and gene annotation

We sequenced the *Yr7, Yr5,* and *YrSP* candidate contigs from the mutant lines (annotated in Supplementary Files 1 and 2) to confirm the EMS-derived mutations using primers documented in Supplementary Table 16. We first PCR-amplified the complete locus from the same DNA preparations as the ones submitted for RenSeq with the Phusion® High-Fidelity DNA Polymerase (New England Biolabs) following the suppliers protocol (https://www.neb.com/protocols/0001/01/01/pcr-protocol-m0530). We then carried out nested PCR on the obtained product to generate overlapping 600-1,000 bp amplicons that were purified using the MiniElute kit (Qiagen). The purified PCR products were sequenced by GATC following the LightRun protocol (https://www.gatc-biotech.com/shop/en/lightrun-tube-barcode.html). Resulting sequences were aligned to the wild-type contig using ClustalOmega (https://www.ebi.ac.uk/Tools/msa/clustalo/). This allowed us to curate the *Yr7*

locus in the Cadenza assembly that contained two sets of unknown ('N') bases in its sequence, corresponding to a 39 bp insertion and a 129 bp deletion (Supplementary File 3), and to confirm the presence of the mutations in each mutant line.

We used HISAT2[40] (v2.1) to map RNA-Seq reads available from Cadenza and AvocetS-*Yr5*[32] to the RenSeq *de novo* assemblies with curated loci to define the structure of the genes. We used the following parameters: --no-mixed --no-discordant to map reads in pairs only. We used the --novel-splicesite-outfile to predict splicing sites that we manually scrutinised with the genome visualisation tool IGV[41] (v2.3.79). Predicted coding sequences (CDS) were translated using the ExPASy online tool (https://web.expasy.org/translate/). This allowed us to predict the effect of the mutations on each candidate transcript (Figure 1a; Supplementary Table 4). The long-range primers for both *Yr7* and *Yr5* loci were then used on the corresponding susceptible Avocet NIL mutants to determine whether the genes were present and carried mutations in that background (Figure 1a; Supplementary Files 1 and 2).

Coiled coil domain prediction

To determine whether *Yr7*, *Yr5*, and *YrSP* encode Coiled Coil (CC) domains we used the NCOILS prediction program[42] (v1.0, https://embnet.vital-it.ch/software/COILS_form.html) with the following parameters: MTIK matrix with applying a 2.5-fold weighting of positions a,d. We compared the profiles to those obtained with already characterised CC-NLR encoding genes *Sr33*, *Mla10*, *Pm3* and *RPS5* (Supplementary Figure 4). We also ran the program on Yr7 and Yr5 protein sequences where the BED domain was manually removed to determine whether its integration could have disrupted an existing CC domain. To further investigate whether *Yr7*, *Yr5*, and *YrSP* encode CC domains we performed a BLASTP analysis[43] with their

N-terminal region, from the methionine to the first amino acid encoding the NB-ARC domain, with or without the BED domain (Supplementary Figure 4).

Genetic linkage

We generated a set of F_2 populations to genetically map the candidate contigs (Supplementary Table 3). For *Yr7* we developed an F_2 population based on a cross between the susceptible mutant line Cad0127 to the Cadenza wild-type (population size 139 individuals). For *Yr5* and *YrSP* we developed F_2 populations between AvocetS and the NILs carrying the corresponding *Yr* gene (94 individuals for *YrSP* and 376 for *Yr5*). We extracted DNA from leaf tissue at the seedling stage (Zadoks 11) following a previously published protocol[44] and Kompetitive Allele Specific PCR (KASP) assays were carried out as described[45]. R/qtl package[46] was used to produce the genetic map based on a general likelihood ratio test and genetic distances were calculated from recombination frequencies (v1.41-6).

We used previously published markers linked to *Yr7*, *Yr5*, and *YrSP* (WMS526, WMS501 and WMC175, WMC332, respectively[15,19,20]) in addition to closely linked markers WMS120, WMS191, and WMC360 (based on the GrainGenes database https://wheat.pw.usda.gov/GG3/) to define the physical region on the Chinese Spring assembly RefSeq v1.0 (https://wheat-urgi.versailles.inra.fr/Seq-Repository/Assemblies). Two different approaches were used for genetic mapping depending on the material. For *Yr7*, we used the public data[34] for Cad0127 (www.wheat-tilling.com) to identify nine mutations located within the *Yr7* physical interval based on BLAST analysis against RefSeq v1.0. We used KASP primers when available and manually designed additional ones including an assay targeting the Cad0127 mutation in the *Yr7* candidate contig (Supplementary Table 16). We genotyped the Cad0127 F_2 populations using these nine KASP assays and confirmed genetic linkage between the Cad0127 *Yr7*

candidate mutation and the nine mutations across the physical interval (Supplementary Figure 3).

For *Yr5* and *YrSP*, we first aligned the candidate contigs to the best BLAST hit in an AvocetS RenSeq *de novo* assembly. We then designed KASP primers targeting polymorphisms between these sequences and used them to genotype the corresponding F_2 population (Supplementary Table 16). For both candidate contigs we confirmed genetic linkage with the previously published genetic intervals for these *Yr* genes (Supplementary Figure 3). Allelism tests between *Yr7*, *Yr5*, and *YrSP* are described in the Supplementary Methods.

_Yr7, Yr5,_ and _YrSP_ gene-specific markers

The development of gene-specific markers are described in the Supplementary Methods.

In silico mining for _Yr7_ and _Yr5_

We used the *Yr7* and *Yr5* sequences to retrieve the best BLAST hits in the *T. aestivum* and *T. turgdium* wheat genomes listed in Supplementary Table 6. The best *Yr5* hits shared between 93.6 and 99.3% sequence identity, which was comparable to what was observed for alleles derived from the wheat *Pm3* (>97% identity)[21] and flax *L* (>90% identity)[22] genes. *Yr7* was identified only in Paragon and Cadenza (Supplementary Table 7; See Supplementary File 3 for curation of the Paragon sequence).

Analysis of the _Yr7_ and _Yr5_/_YrSP_ cluster on RefSeq v1.0

Definition of syntenic regions across grass genomes

We used NLR-Annotator to identify putative NLR loci on RefSeq v1.0 chromosome 2B and identified the best BLAST hits to *Yr7* and *Yr5* on RefSeq v1.0. Additional BED-NLRs and

canonical NLRs were annotated in close physical proximity to these best BLAST hits. Therefore, to better define the NLR cluster we selected ten non-NLR genes located both distal and proximal to the region, and identified orthologs in barley, *Brachypodium*, and rice in *EnsemblPlants* (https://plants.ensembl.org/). We used different % ID cutoffs for each species (>92% for barley, >84% for *Brachypodium*, and >76% for rice) and determined the syntenic region when at least three consecutive orthologues were found. A similar approach was conducted for *Triticum* ssp and *Ae. tauschii* (Supplementary file 4).

Definition of the NLR content of the syntenic region

We extracted the previously defined syntenic region from the grass genomes listed in Supplementary Table 6 and annotated NLR loci with NLR-Annotator. We maintained previously defined gene models where possible, but also defined new gene models that were further analysed through a BLASTx analysis to confirm the NLR domains (Supplementary Files 4 and 5). The presence of BED domains in these NLRs was also confirmed by CD-Search (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi).

Phylogenetic and neighbour network analyses

Methods for the phylogenetic analyses are described in the Supplementary Methods.

Transcriptome analysis

Methods for the transcriptomic analyses are described in the Supplementary Methods.

References

1. Oerke, E. C. Crop losses to pests. *J. Agric. Sci.* **144,** 31–43 (2006).
2. Hubbard, A. *et al.* Field pathogenomics reveals the emergence of a diverse wheat yellow rust population. *Genome Biol.* **16,** 23 (2015).
3. Aravind, L. The BED finger, a novel DNA-binding domain in chromatin-boundary-

element-binding proteins and transposases. *Trends Biochem. Sci.* **25,** 421–423 (2000).

4. Jones, J. D. G. & Dangl, J. L. The plant immune system. *Nature* **444,** 323–329 (2006).

5. 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* (2018). doi:10.1105/tpc.17.00579

6. Sarris, P. F., Cevik, V., Dagdas, G., Jones, J. D. G. & Krasileva, K. V. Comparative analysis of plant immune receptor architectures uncovers host proteins likely targeted by pathogens. *BMC Biol.* **14,** 8 (2016).

7. Kroj, T., Chanclud, E., Michel-Romiti, C., Grand, X. & Morel, J.-B. Integration of decoy domains derived from protein targets of pathogen effectors into plant immune receptors is widespread. *New Phytol.* **210,** 618–626 (2016).

8. Bailey, P. C. *et al.* Dominant integration locus drives continuous diversification of plant immune receptors with exogenous domain fusions. *Genome Biol.* **19,** 23 (2018).

9. Bundock, P. & Hooykaas, P. An *Arabidopsis* hAT-like transposase is essential for plant development. *Nature* **436,** 282–284 (2005).

10. Yoshimura, S. *et al.* Expression of *Xa1*, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 1663–1668 (1998).

11. Das, B., Sengupta, S., Prasad, M. & Ghose, T. Genetic diversity of the conserved motifs of six bacterial leaf blight resistance genes in a set of rice landraces. *BMC Genet.* **15,** 82 (2014).

12. Law, C. N. Genetic control of yellow rust resistance in *T. spelta* Album. *Plant Breed. Institute, Cambridge, Annu. Rep.* **1975,** 108–109 (1976).

13. Johnson, R. & Dyck, P. L. Resistance to yellow rust in *Triticum spelta* var. Album and bread wheat cultivars Thatcher and Lee. *Colloq. l'INRA* (1984).

14. Zhang, P., McIntosh, R. A., Hoxha, S. & Dong, C. M. Wheat stripe rust resistance genes *Yr5* and *Yr7* are allelic. *Theor. Appl. Genet.* **120,** 25–29 (2009).

15. Feng, J. Y. *et al.* Molecular mapping of *YrSP* and its relationship with other genes for stripe rust resistance in wheat chromosome 2BL. *Phytopathology* **105,** 1206–1213 (2015).

16. Wellings, C. R. & McIntosh, R. A. *Puccinia striiformis f.sp. tritici* in Australasia: pathogenic changes during the first 10 years. *Plant Pathol.* **39,** 316–325 (1990).

17. Zhan, G. *et al.* Virulence and molecular diversity of the *Puccinia striiformis* f. sp. *tritici* population in Xinjiang in relation to other regions of western China. *Plant Dis.* **100,** 99–107 (2016).

18. Steuernagel, B. *et al.* Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture. *Nat. Biotechnol.* **34,** 652–655 (2016).

19. Sun, Q., Wei, Y., Ni, Z., Xie, C. & Yang, T. Microsatellite marker for yellow rust resistance gene *Yr5* in wheat introgressed from spelt wheat. *Plant Breed.* **121,** 539–541 (2002).

20. Yao, Z. J. *et al.* The molecular tagging of the yellow rust resistance gene *Yr7* in wheat transferred from differential host Lee using microsatellite markers. *Sci. Agric. Sin.* **39,** 1146–1152 (2006).

21. Brunner, S. *et al.* Intragenic allele pyramiding combines different specificities of wheat *Pm3* resistance alleles. *Plant J.* **64,** 433–445 (2010).

22. Ellis, J. G., Lawrence, G. J., Luck, J. E. & Dodds, P. N. Identification of regions in alleles of the flax rust resistance gene *L* that determine differences in gene-for-gene specificity. *Plant Cell* **11,** 495–506 (1999).

23. Bai, S. *et al.* Structure-function analysis of barley NLR immune receptor MLA10 reveals its cell compartment specific activity in cell death and disease resistance. *PLoS Pathog.* **8,** e1002752 (2012).

24. Periyannan, S. *et al.* The gene *Sr33*, an ortholog of barley *Mla* genes, encodes resistance to wheat stem rust race Ug99. *Science (80-.).* **341,** 786–788 (2013).

25. Srichumpa, P., Brunner, S., Keller, B. & Yahiaoui, N. Allelic series of four powdery mildew resistance genes at the *Pm3* locus in hexaploid bread wheat. *Plant Physiol.* **139,** 885–895 (2005).

26. Sarris, P. F. *et al.* A plant immune receptor detects pathogen effectors that target WRKY transcription factors. *Cell* **161,** 1089–1100 (2015).

27. Wingen, L. U. *et al.* Establishing the A. E. Watkins landrace cultivar collection as a resource for systematic gene discovery in bread wheat. *Theor. Appl. Genet.* **127,** 1831–1842 (2014).

28. Reeves, J. C. *et al.* Changes over time in the genetic diversity of four major European crops - a report from the Gediflux Framework 5 project. *Genet. Var. plant breeding. Proc. 17th EUCARPIA Gen. Congr. Tulln, Austria, 8-11 Sept. 2004* 3–7 (2004).

29. Ellis, J. G., Lagudah, E. S., Spielmeyer, W. & Dodds, P. N. The past, present and future of breeding rust resistant wheat. *Front Plant Sci* **5,** 641 (2014).

30. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23,** 254–267 (2006).

31. Ellis, J. G. Integrated decoys and effector traps: how to catch a plant pathogen. *BMC Biol.* **14,** 13 (2016).

32. Dobon, A., Bunting, D. C. E., Cabrera-Quio, L. E., Uauy, C. & Saunders, D. G. O. The host-pathogen interaction between wheat and yellow rust induces temporally coordinated waves of gene expression. *BMC Genomics* **17,** 380 (2016).

33. Seeholzer, S. *et al.* Diversity at the *Mla* powdery mildew resistance locus from cultivated barley reveals sites of positive selection. *Mol. Plant-Microbe Interact.* **23,** 497–509 (2010).

34. Krasileva, K. V *et al.* Uncovering hidden variation in polyploid wheat. *Proc. Natl. Acad. Sci. U. S. A.* **6,** E913–E921 (2017).

35. Hubbard, A. J., Fanstone, V. & Bayles, R. A. *UKCPVS 2009 Annual report*.

36. Gassner, G. & Straib, W. *Die Bestimmung der biologischen Rassen des Weizengelbrostes* Puccinia glumarum f.sp. tritici *Schmidt Erikss. u. Henn*. (1932).

37. McGrann, G. R. D. *et al.* Genomic and genetic analysis of the wheat race-specific yellow rust resistance gene *Yr5*. *J. Plant Sci. Mol. Breed.* **3,** (2014).

38. Lagudah, E. S., Appels, R., Brown, A. H. D. & McNeil, D. The molecular–genetic analysis of *Triticum tauschii*, the D-genome donor to hexaploid wheat. *Genome* **34,** 375–386 (1991).

491 39. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina
492 sequence data. *Bioinformatics* **30,** 2114–2120 (2014).

493 40. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low
494 memory requirements. *Nat. Methods* **12,** 357–360 (2015).

495 41. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer
496 (IGV): high-performance genomics data visualization and exploration. *Brief.*
497 *Bioinform.* **14,** 178–192 (2013).

498 42. Lupas, A., Dyke, M. Van & Stock, J. Predicting coiled coils from protein sequences.
499 *Science (80-.).* **252,** 1162–1164 (1991).

500 43. Altschul, S. F. *et al.* Protein database searches using compositionally adjusted
501 substitution matrices. *FEBS J.* **272,** 5101–9 (2005).

502 44. Pallotta, M. A. *et al.* Marker assisted wheat breeding in the southern region of
503 Australia. in Pogna et al. *Proceedings of 10th International Wheat Genet Symposium*
504 *Instituto Sperimentale per la Cerealcoltura, Rome* 789–791 (2003).

505 45. Ramirez-Gonzalez, R. H. *et al.* RNA-Seq bulked segregant analysis enables the
506 identification of high-resolution genetic markers for breeding in hexaploid wheat.
507 *Plant Biotechnol J* **13,** 613–624 (2015).

508 46. Broman, K. W., Wu, H., Sen, S. & Churchill, G. A. R/qtl: QTL mapping in
509 experimental crosses. *Bioinformatics* **19,** 889–890 (2003).

510 47. Jupe, F. *et al.* Identification and localisation of the NB-LRR gene family within the
511 potato genome. *BMC Genomics* **13,** 75 (2012).

512 48. Warren, R. F., Henk, A., Mowery, P., Holub, E. & Innes, R. W. A mutation within the
513 leucine-rich repeat domain of the arabidopsis disease resistance gene RPS5 partially
514 suppresses multiple bacterial and downy mildew resistance genes. *Plant Cell* **10,**
515 1439–1452 (1998).

516 49. Shaw, P. D., Graham, M., Kennedy, J., Milne, I. & Marshall, D. F. Helium:
517 visualization of large scale plant pedigrees. *BMC Bioinformatics* **15,** 259 (2014).

518 50. Avni, R. *et al.* Wild emmer genome architecture and diversity elucidate wheat
519 evolution and domestication. *Science.* **357,** 93–97 (2017).

520 51. Luo, M.-C. *et al.* Genome sequence of the progenitor of the wheat D genome *Aegilops*
521 *tauschii*. *Nature* **551,** 498 (2017).

522

523 **Author contributions**

524 CM performed the experiments to clone *Yr7* and *Yr5* and the subsequent analyses of their loci
525 and BED domains, designed the gene-specific markers, analysed the genotype data in the
526 studied panels, and designed and made the figures. JZ performed the experiments to
527 clone *YrSP*, confirm the *Yr7* and *Yr5* genes in AvocetS-*Yr7* and AvocetS-*Yr5* mutants, and
528 identified the full length of *Yr5* and *YrSP* with their respective regulatory elements. CM and
529 JZ developed the gene-specific markers. PZ and RM performed the EMS treatment, isolation,

and confirmation of *Yr7*, *Yr5*, and *YrSP* mutants in AvocetS NILs. PF performed the pathology work on the Cadenza *Yr7* mutants and the mapping populations. BS helped with the NLR annotator analysis and provided the bait library for target enrichment and sequencing of NLRs, NMA provided DNA samples for allelic variation studies and LB provided Lemhi-*Yr5* mutants. RM, EL, PZ, BW, SB, and CU conceived, designed, and supervised the research. CM and CU wrote the manuscript. JZ, PZ, RM, BW, NMA, LB and EL provided edits.

Data availability

All sequencing data has been deposited in the NCBI Short Reads Archive under accession numbers listed in Supplementary Table 14 (SRP139043). Cadenza (*Yr7*) and Lemhi (*Yr5*) mutants are available through the JIC Germplasm Resource Unit (www.seedstor.ac.uk).

Competing interests

A patent application based on this work has been filed (United Kingdom Patent Application No. 1805865.1).

Acknowledgements

This work was supported by the UK Biotechnology and Biological Sciences Research Council Designing Future Wheat programme BB/P016855/1 and the Grains Research and Development Corporation, Australia. CM was funded by a PhD studentship from Group Limagrain and JZ is funded by PhD scholarships from the National Science Foundation (NSF) and the Monsanto Beachell-Borlaug International Scholars Programs (MBBISP). We thank the International Wheat Genome Sequencing Consortium for allowing pre-publication access to the RefSeq v1.0 assembly and gene annotation. We thank Jorge Dubcovsky and Xiaoqin Zhang (University of California, Davis) for providing *Yr5* cultivars. We thank the John Innes Centre Horticultural Services and Limagrain Rothwell staff for management of the wheat populations. Also Sebastian Specel (Limagrain; Clermont-Ferrand) and Richard Goram (JIC) for their help in designing and running KASP assays, and Sami Hoxha (The University of Sydney) for technical assistance. This research was supported by the NBI Computing Infrastructure for Science (CiS) group in Norwich, UK.

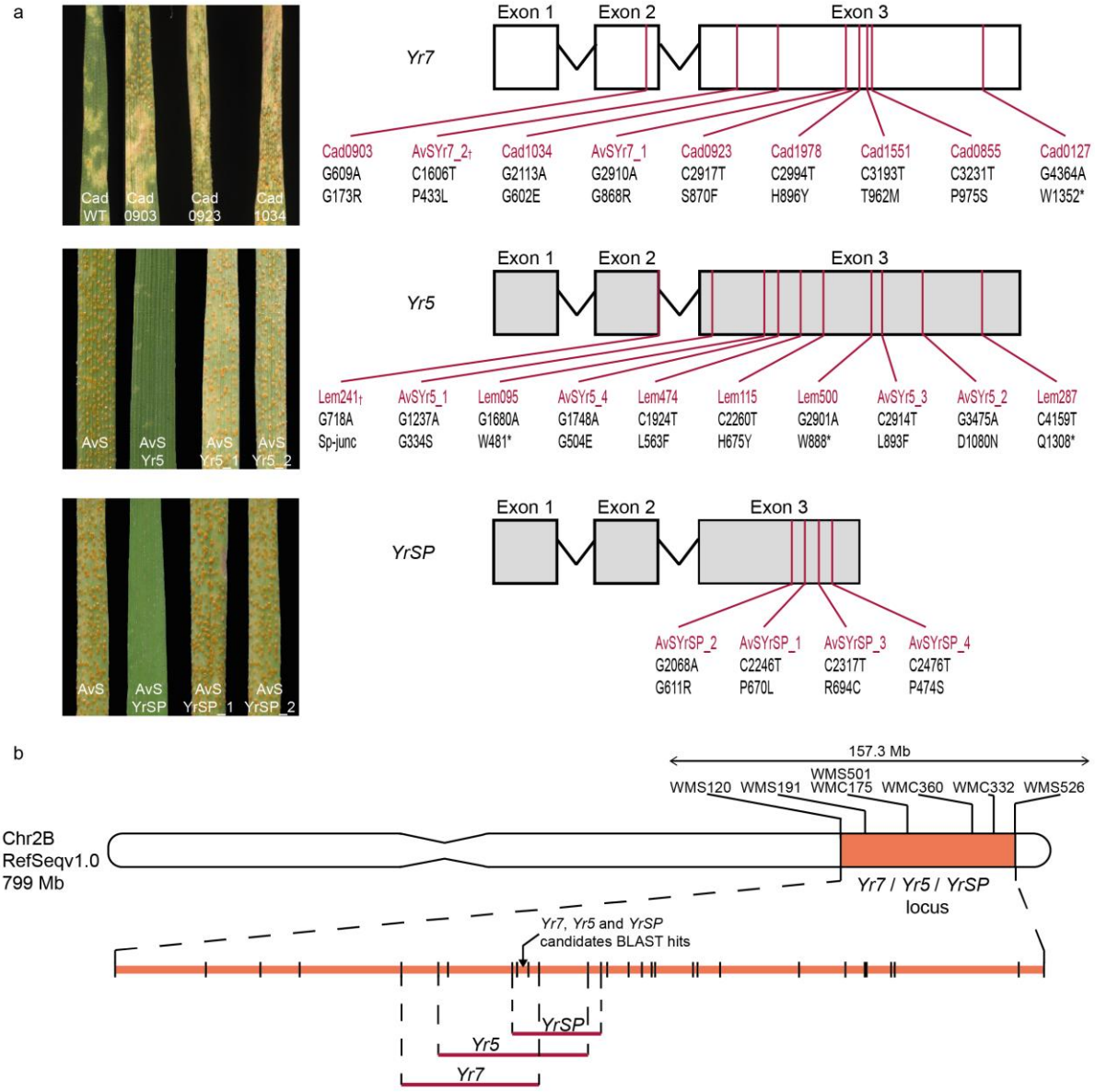

Figure 1: ***Yr5* and *YrSP* are closely related sequences and distinct from *Yr7*.**
a, Left: Wild-type and selected EMS-derived susceptible mutant lines for *Yr7*, *Yr5*, and *YrSP* (Supplementary Tables 3 and 4) inoculated with *Pst* isolate 08/21 (*Yr7*), *Pst* 150 E16A+ (*Yr5*), or *Pst* 134 E16A+ (*YrSP*). Right: Candidate gene structures, with mutations in red, and their predicted effects on the translated protein. Crosses show mutations shared by two independent mutant lines (Supplementary Table 4). **b**, Schematic representation of the physical interval of the *Yr* loci. The *Yr7/Yr5/YrSP* locus is shown in orange on chromomsome 2B with previously published SSR markers in black. Markers developed in this study to confirm the genetic linkage between the phenotype and the candidate contigs are shown as black vertical lines in the expanded 157.3 Mb interval. *Yr* loci mapping intervals are defined by the red horizontal lines below the expanded chromosome. A more detailed genetic map is shown in Supplementary Figure 3.

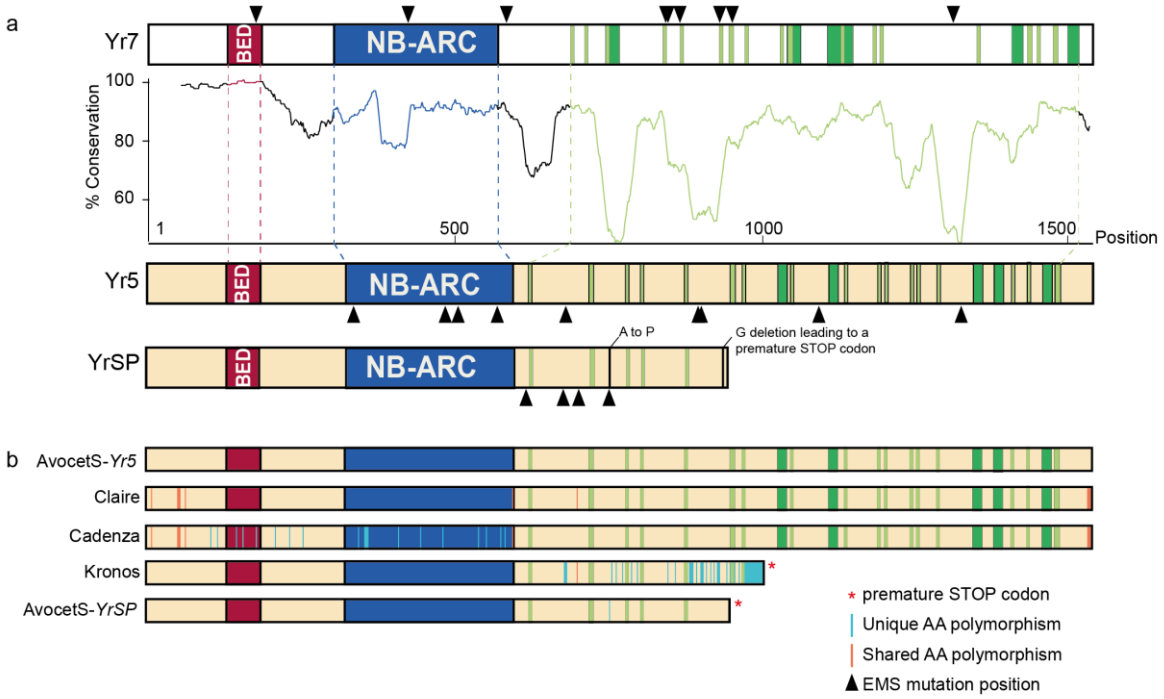

Figure 2: *Yr7* and *Yr5/YrSP* encode integrated BED-domain immune receptor genes. **a**, Schematic representation of the Yr7, Yr5, and YrSP protein domain organisation. BED domains are highlighted in red, NB-ARC domains are in blue, LRR motifs from NLR-Annotator are in dark green, and manually annotated LRR motifs (xxLxLxx) are in light green. Black triangles represent the EMS-induced mutations within the protein sequence. The plot shows the degree of amino acid conservation (50 amino acid rolling average) between Yr7 and Yr5 proteins, based on the conservation diagram produced by Jalview (2.10.1) from the protein alignment. Regions that correspond to the conserved domains have matching colours. The amino acid changes between Yr5 and YrSP are annotated in black on the YrSP protein. **b**, Five Yr5/YrSP haplotypes were identified in this study. Polymorphisms are highlighted across the protein sequence with orange vertical bars for polymorphisms shared by at least two haplotypes and blue vertical bars for polymorphisms that are unique to the corresponding haplotype. Matching colours across protein structures illustrate 100% sequence conservation.

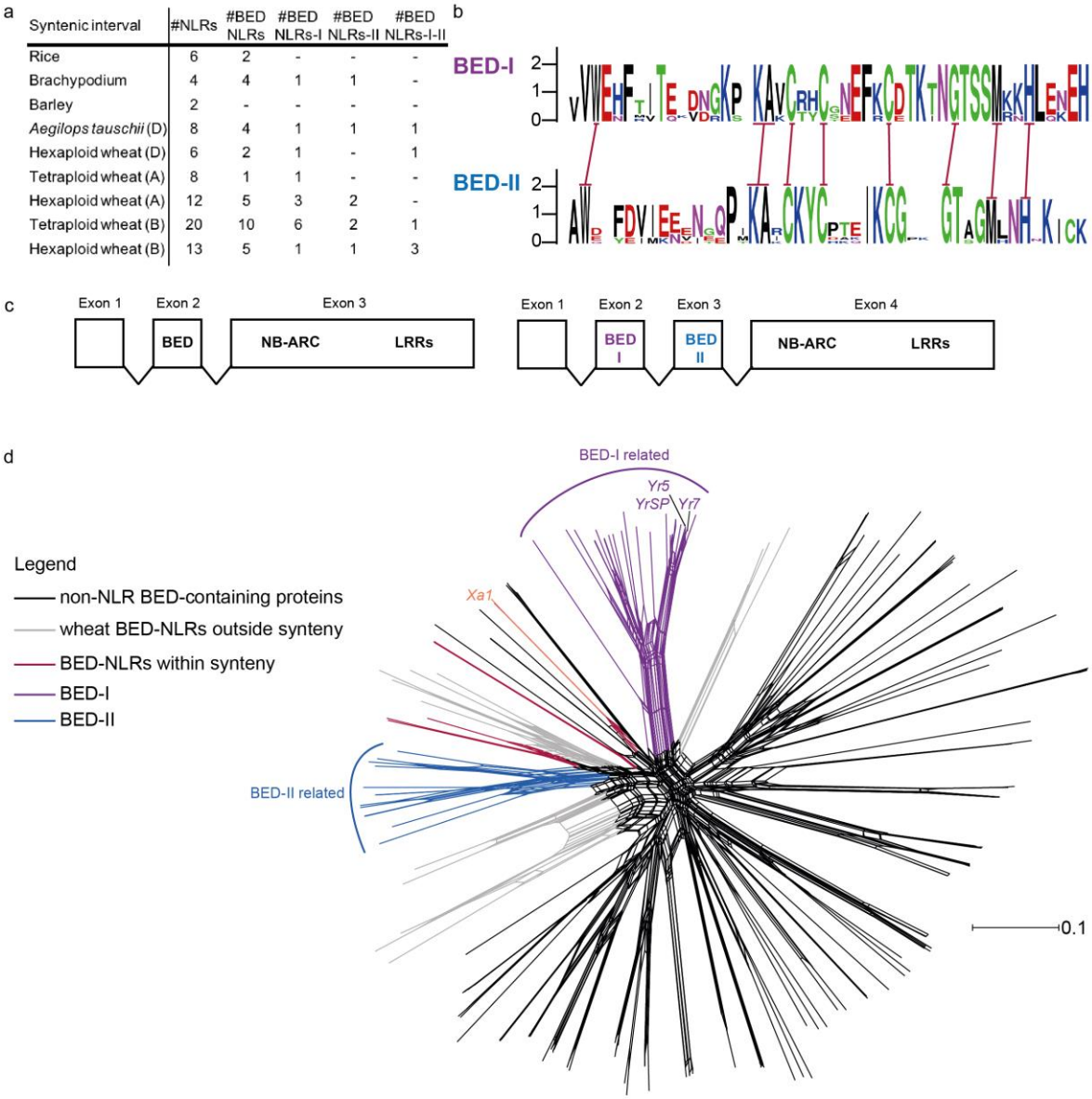

Figure 3: BED domains from BED-NLRs and non-NLR proteins are distinct.
a, Numbers of NLRs in the syntenic regions across grass genomes (see Supplementary Figure 7), including BED-NLRs. **b**, WebLogo (http://weblogo.berkeley.edu/logo.cgi) diagram showing that the BED-I and BED-II domains are distinct, with only the highly conserved residues that define the BED domain (red bars) being conserved between the two types. **c**, Gene structure most commonly observed for BED-NLRs and BED-BED-NLRs within the *Yr7*/*Yr5*/*YrSP* syntenic interval. **d**, Neighbour-net analysis based on uncorrected *P* distances obtained from alignment of 153 BED domains including the 108 BED-containing proteins (including 25 NLRs) from RefSeq v1.0, BED domains from NLRs located in the syntenic region as defined in Supplementary Figure 7, and BED domains from Xa1 and ZBED from rice. BED-I and II clades are highlighted in purple and blue, respectively. BED domains from the syntenic regions not related to either of these types are in red. BED domains derived from non-NLR proteins are in black and BED domains from BED-NLRs outside the syntenic region are in grey. Seven BED domains from non-NLR proteins were close to BED domains from BED-NLRs. Supplementary Figure 9 includes individual labels.

Supplementary Figure 1: Deployment of *Yr7* cultivars in the field is correlated with an increase in the prevalence of *Pst* isolates virulent to *Yr7* in the UK.

Percentage of total harvested weight of wheat cultivars carrying *Yr7* (green) and the proportion of *Pst* isolates virulent to Yr7 (orange) from 1990 to 2016 in the United Kingdom. See Supplementary Table 2 for a summary of the data.

Supplementary Figure 2: Identification of candidate contigs for the *Yr* loci using MutRenSeq.

View of RenSeq reads from the wild-type and EMS-derived mutants mapped to the best candidate contigs identified with MutantHunter for the three genes targeted in this study. From top to bottom: vertical black lines represent the *Yr* loci, coloured rectangles depict the motifs identified by NLR-Annotator (each motif is specific to a conserved NLR domain[47]), while read coverage (grey histograms) is indicated on the left, e.g. [0 - 149], and the line from which the reads are derived on the right, e.g. CadWT for Cadenza wild-type. Vertical bars represent the position of the SNPs identified between the reads and reference assembly – red shows C to T transitions and green G to A transitions. Black boxes highlight SNP for which the coverage was relatively low, but still higher than the 20x detection threshold. The top view shows the *Yr7* sequence annotated from the Cadenza genome assembly before manual curation (Supplementary File 3). Vertical black lines illustrate the assembled candidate contigs and the one that was formerly *de novo* assembled from Cadenza RenSeq data, lacking the 5' region containing the BED domain and thus the Cad903 mutation. The middle view illustrates the *Yr5* locus annotated from the Lemhi-*Yr5 de novo* assembly. The results are similar to those described above for *Yr7*. The full locus was *de novo* assembled. The bottom view illustrates the *YrSP* locus annotated from the AvocetS-YrSP *de novo* assembly with the four identified susceptible mutants all carrying a mutation in the candidate contig. The full locus was *de novo* assembled.

Supplementary Figure 3: Candidate contigs identified by MutRenSeq are genetically linked to the *Yr* loci mapping interval.

Schematic representation of chromosome 2B from Chinese Spring (RefSeq v1.0) with the positions of published markers linked to the *Yr* loci and surrounding closely linked markers that were used to define their physical position (orange rectangle). The chromosome is depicted as a close-up of the physical locus indicating the positions of KASP markers that were used for genetic mapping (horizontal bars, Supplementary Table 16). Blue colour refers to *Yr7*, red to

Yr5, and purple to *YrSP*. The black arrow points to the NLR cluster containing the best BLAST hits for *Yr7* and *Yr5*/*YrSP* on RefSeq v1.0. Coloured lines link the physical map to the corresponding genetic map for each targeted gene (see Methods). Genetic distances are expressed in centiMorgans (cM).

Supplementary Figure 4: **Yr7, Yr5 and YrSP proteins do not encode for a Coiled-Coil domain in the N-terminus.**

Graphical outputs from the COILS prediction programm in three sliding windows (14, 21, and 28 amino acid, shown in green, blue, and red, respectively) for Yr5 and Yr7 with or without the BED domain (left) and characterised canonical NLRs: Sr33[24], Mla10[23], Pm3[25] and RPS5[48]. The X axis shows the amino acid positions and the Y axis the probability of a coiled coil domain formation. There was no difference in the prediction between the two Yr proteins with or without their BED domain. The 14 amino acid sliding window is the least accurate according to the user manual, consistent with the additional peaks observed in Sr33, Mla10 and Pm3 that were not annotated as CC domains in the corresponding publications[23–25]. Thus, the peak at position 1,200 in Yr5 is unlikely to represent a CC domain. We performed a BLASTP search with the N-terminal region of the Yr5 and Yr7 proteins (from Met to the first amino-acid encoding the NB-ARC) with or without the BED domain and the best hits were proteins predicted to encode BED-NLRs from *Aegilops tauschii*, *Triticum uratu* and *Oryza sativa* (data not shown). Based on the COILS prediction and the BLAST search, we concluded that Yr7 and Yr5/YrSP do not encode CC domains.

Supplementary Figure 5: Pedigrees of selected Thatcher-derived cultivars and their *Yr7* status.

Pedigree tree of Thatcher-derived cultivars where each circle represents a cultivar and the size of the circle is proportional to its prevalence in the tree. Colours illustrate the genotype with red showing the absence of *Yr7* and yellow its presence. Cultivars in grey were not tested or are intermediate crosses. *Yr7* originated from *Triticum durum* cv. Iumillo and was introgressed into hexaploid wheat through Thatcher (indicated by arrow). Each *Yr7* positive cultivar is related to a parent that was also positive for *Yr7*. Figure was generated using the Helium software[49] (v1.17).

Supplementary Figure 6: Illustration of *Yr5*, *YrSP*, and *Yr7* KASP assays.

Graphical output from KlusterCaller from the *Yr5*, *YrSP*, and *Yr7* KASP assays. Each circle represents a sample listed in the corresponding Supplementary Table (7 (*Yr7*), 8 (*YrSP*) and 9 (*Yr5*)). Red and blue colours show the signal for the VIC and FAM tails, respectively, with the corresponding primer sequence (without the tail) below. Pink shows DNA that did not amplify for the *Yr5* marker and both DNA that did not amplify and water controls for *YrSP* and *Yr7* markers. Black shows water control for the *Yr5* KASP assay. Controls cultivars are shown in the matching colour with the amplified signal.

Supplementary Figure 7: Expansion of BED-NLRs in the Triticeae and presence of conserved BED-BED-NLRs aross the syntenic region.

Schematic representation of the physical loci containing *Yr7* and *Yr5*/*YrSP* homologs on RefSeq v1.0 and its syntenic regions. The syntenic region is flanked by conserved non-NLR genes (orange arrows). Black arrows represent canonical NLRs and purple/blue/red arrows represent different types of BED-NLRs based on their BED domain and their relationship identified in Figure 3 and Supplementary Figure 8. Black lines represent phylogenetically related single NLRs located between the two NLR clusters illustrated in Supplementary Figure 9. Details of genes are reported in Supplementary File 4.

Supplementary Figure 8: The *Yr* loci are phylogenetically related to nearby NLRs on RefSeq v1.0 and their orthologs.

Phylogenetic tree based on translated NB-ARC domains from NLR-Annotator. Node labels represent bootstrap values for 1,000 replicates. The tree was rooted at mid-point and visualized with Dendroscope v3.5.9. The colour pattern matches that of Figure 3 to highlight BED-NLRs with different BED domains.

Supplementary Figure 9: Neighbour-net analysis network as shown in Figure 3 with identifiers.

Neighbour-net analysis based on uncorrected *P* distances obtained from alignment of 153 BED domains including the 108 BED-containing proteins (including 25 NLRs) from RefSeq v1.0, BED domains from NLRs located in the syntenic region as defined in Supplementary Figure 7, and BED domains from Xa1 and ZBED from rice. BED-I and II clades are highlighted in purple and blue, respectively. BED domains from the syntenic regions not related to either of these types are in red. BED domains derived from non-NLR proteins are in black and BED

domains from BED-NLRs outside the syntenic region are in grey. Seven BED domains from non-NLR proteins were close to BED domains from BED-NLRs.

Supplementary Figure 10: BED-NLRs and BED-containing proteins are not differentially expressed in yellow rust-infected susceptible and resistant cultivars.

Heatmap representing the normalised read counts (Transcript Per Million, TPM) from the reanalysis of published RNAseq data[32] for all the BED-containing proteins, BED-NLRs and canonical NLRs located in the syntenic region annotated on RefSeq v1.0. Lack of expression is shown in white and expression levels increase from blue to red. Asterisks show cases where several gene models were overlapping with NLR loci identified with NLR Annotator. The colour pattern matches that of Figure 3 to highlight BED-NLRs with different BED domains. Orange labels show the expression of the canonical NLRs located within the syntenic interval. The seven non-NLR BED genes whose BED domain clustered with the ones from BED-NLR proteins in Figure 3 and Supplementary Figure 9 are indicated by black triangles.

Supplementary Table 1: Summary of Pst isolates tested on Yr5 differential lines from 2004 to 2017 in different regions.

Overall, >6,000 isolates from 44 countries displaying >200 different pathotypes were tested on Yr5 materials and no virulence was recorded apart from two isolates from Australia, PST 360 E137 A-/+[16]. Data were obtained from public databases and reports on yellow rust surveillance, whose references are recorded. It is important to note that we report here the number of identified pathotypes for a given region and database. Similar pathotypes could thus have been counted twice if identified in different regions.

Supplementary Table 2: Harvested weight of known *Yr7* cultivars from 1990 to 2016 and prevalence of *Yr7* virulence among UK *Pst* isolates.

Proportion of harvested *Yr7* wheat cultivars in the UK from 1990 to 2016. The prevalence of yellow rust isolates virulent to *Yr7* across this time period is shown in the top row. Original data from NIAB-TAG Seedstats journal (NIAB-TAG Network) and the UK Cereal Pathogen Virulence Survey (http://www.niab.com/pages/id/316/UKCPVS).

Supplementary Table 3: Plant materials analysed for the present study and *Pst* isolates used for the pathology assays.

Supplementary Table 4: Plant material submitted for Resistance Gene Enrichment Sequencing (RenSeq).

From left to right: Mutant line identifier, targeted gene, infection type when infected with *Pst* according to the Grassner and Straib scale, mutation position, coverage of the mutation (at least 99% of the reads supported the mutant base in the mutant reads), predicted effect of the mutation on the protein sequence, comments. Lines with the same mutations are highlighted with matching colours.

Supplementary Table 5: Allelism tests in AvocetS-*YrSP* x AvocetS-*Yr5* and AvocetS-*YrSP* x AvocetS-*Yr7* F_3 populations.

For each cross, the same highly resistant plants identified with one *Pst* isolate were highly susceptible for the alternative *Pst* isolate.

Supplementary Table 6: Genome assemblies used in the present study.

Summary of the available genome assemblies[50,51] that were used for the *in silico* allele mining and synteny analysis across rice, *Brachypodium*, barley and different Triticeae accessions.

Supplementary Table 7: *In silico* allele mining for *Yr7* and *Yr5*/*YrSP* in available genome assemblies for wheat.

Table presents the percentage identity (% ID) of the identified alleles and matching colours illustrate identical haplotypes. Investigated genome assemblies are shown in Supplementary Table 6.

Supplementary Table 8: Polymorphisms among Yr5 proteins.

Positions of the polymorphic amino acids across the five Yr5/YrSP proteins. Polymorphisms falling into the BED and NB-ARC domains are shown in red and blue, respectively.

Supplementary Table 9: Presence/absence of *Yr7* alleles in a selected panel of Cadenza-derivatives and associated responses to different *Pst* isolates (avirulent to *Yr7*: *Pst* 15/151 and 08/21; virulent to *Yr7*: 14/106).

Infection types were grouped into two categories: 1 for resistant and 2 for susceptible. We used Vuka as a positive control for inoculation and absence of *Yr7*. The typical response of a *Yr7* carrier would thus be 1 – 1 – 2, although some cultivars might carry other resistance genes that can lead to a 1 – 1 – 1 profile (e.g. Cadenza). Cultivars that were positive for *Yr7* had either

one or the other profile so none of them was susceptible to a *Pst* isolate that is avirulent to *Yr7*. Few cultivars (e.g Bennington, KWS-Kerrin, Brando) were susceptible to one of the two isolates avirulent to *Yr7* in addition to their susceptibility to the *Yr7*-virulent isolate. However, none of them carried the *Yr7* allele.

Supplementary Table 10: Presence/absence of *Yr7* and *YrSP* in different wheat collections. We used Vuka, AvocetS and Solstice as negative controls for the presence of *Yr7* and *YrSP* and AvocetS-*Yr* near-isogenic lines as controls for the corresponding *Yr* gene. We genotypied different collections: (i) a set of potential *Yr7* carriers based on literature research, (ii) a set of cultivars that belonged to the UK AHDB Recommended List (https://cereals.ahdb.org.uk/varieties/ahdb-recommended-lists.aspx) between 2005 and 2018 (labelled 2005-2018-UK_RL), (iii) the Gediflux collection that includes modern European bread wheat cultivars (1920-2010)[28], and (iv) a core set of the Watkins collection, which represents a set of global bread wheat landraces collected in the 1920-30s[27]. Most of the putative *Yr7* carriers, apart from Aztec, Chablis and Cranbrook, were positive for all the *Yr7* markers. Chablis was susceptible to the *Pst* isolates that were avirulent to *Yr7* so it probably does not carry the gene. Given the 2005-2018-UK_RL results were consistent across already tested cultivars: Cadenza, Cordiale, Cubanita, Grafton and Skyfall were already positive in Supplementary Table 9. Energise, Freiston, Gallant, Oakley and Revelation were negative on both panels. Results were thus consistent across different sources of DNA. *Yr7*-containing cultivars are not prevalent in the 2005-2018 Recommended List set, however, this gene is present in Skyfall, which is currently one of the most widely harvested cultivars in the UK (Supplementary Table 2). We tested the *YrSP* marker on this set and it was positive only for AvocetS-*YrSP*. The frequency of *Yr7* was relatively low in the Gediflux panel (4%). This is consistent with results in Supplementary Table 2: *Yr7* deployment started in the UK in 1992 with Cadenza and it was rarely used prior to that date. The same was observed in the subset of the Watkins collection (10%) where landraces that were positive for *Yr7* all originated from India and the Mediterranean basin. *Yr7* was introgressed into Thatcher (released in 1936) from Iumillo, which originated from Spain and North-Africa (Genetic Resources Information System for Wheat and Triticale - http://www.wheatpedigree.net/). Iumillo is likely to be pre-1920s and these landraces are all bread wheats so they might have inherited it from another source. However, there is no evidence for *Yr7* coming from another source than Iumillo in the modern bread wheat cultivars.

Supplementary Table 11: **Presence/absence of *Yr5* alleles in selected cultivars.**

We tested the KASP marker on the *Yr5* and *YrSP* donors spelt cultivar Album and Spaldings Prolific, respectively. We further tested the marker on *Yr5*-introgressed lines in AvocetS and Lemhi backgrounds and cultivars from the University of California, Davis breeding program (Yecora Rojo 515, Redwin 515, UC 1745 515, and Summit 515). We included bread wheat cultivars Claire, Cadenza, and Paragon in which we identified alternate alleles for *Yr5* (Figure 2). We used Iumillo, *Yr7* donor, Marquillo (Marquis x Iumillo), Lemhi, and AvocetS-*Yr7* as negative controls.

Supplementary Table 12: Identified BED-containing proteins in RefSeq v1.0 based on a hmmer scan analysis (see Methods).

Several features are added: number of identified BED domains and the presence of other conserved domains present, the best BLAST hit from the non-redundant database of NCBI with its description and score, and whether the BED domain was related to BED domains from NLR proteins based on the neighbour network shown in Supplementary Figure 8.

Supplementary Table 13: Transcripts per Million-normalised read counts from the re-analysis of published RNA-Seq data[32] **and associated differential expression analysis performed with DESeq2.**

Supplementary Table 14: Sequencing details of RenSeq data generated in this study.

Supplementary Table 15: *De novo* assemblies generated from the corresponding RenSeq data.

Supplementary Table 16: Primers designed to map and clone *Yr7, Yr5,* and *YrSP*.

Note that KASP assays require the addition of the corresponding 5'-tails for the two KASP primers

Supplementary Table 17: Diagnostic markers for *Yr7, Yr5,* and *YrSP*.

Note that KASP assays require the addition of the corresponding 5' -tails for the two KASP primers.

Supplementary File 1: Annotation of the *Yr7* locus in Cadenza with exon/intron structure, positions of mutations and the position of primers for long-range PCR and nested PCRs that were carried out prior to Sanger sequencing (Supplementary Table 16). The file also includes the derived CDS and protein sequences with annotated conserved domains. Amino acids encoding the BED domain are shown in red and those encoding the NB-ARC domain are in blue. LRR repeats identified with NLR Annotator are highlighted in dark green and manually annotated LRR motifs xxLxLxx are underlined and in bold black.

Supplementary File 2: **Annotation of the *Yr5*/*YrSP* locus in Lemhi-*Yr5* and AvocetS-*YrSP*, respectively, with exon/intron structure, the position of mutations and the position of primers for long-range PCR and nested PCRs that were carried out prior to Sanger sequencing (Supplementary Table 16)**. The derived CDS and protein sequences with annotated conserved domains are also shown. Amino acids encoding the BED domain are shown in red and those encoding the NB-ARC domain are in blue. LRR repeats identified with NLR Annotator are highlighted in dark green and manually annotated LRR motifs xxLxLxx are underlined and in bold black. Design of the *Yr5* PCR marker is shown at the end of the file with the insertion that is specific to *Yr5* when compared to *YrSP* and Claire.

Supplementary File 3: **Curation of the *Yr7* locus in the Cadenza genome assembly based on Sanger sequencing results.**

Comments show the position of the unknown bases ("N") in the "Yr7_with_Ns" sequence. Curation based on Sanger sequencing data is shown in bold black in the "curated_Yr7" sequence with the 39 bp insertion and 129 bp deletion. Allele mining for *Yr7* in the Paragon assembly showed that a similar assembly issue might have occurred for this cultivar (same annotation in the "Yr7_Paragon_with_Ns" sequence). This is consistent with the fact that both assemblies were produced with the same pipeline (Supplementary Table 6). We used RenSeq data available for Paragon and performed an alignment as described for the MutRenSeq pipeline against Cadenza NLRs with the curated *Yr7* loci included. A screen capture of the mapping is shown. Only one SNP was identified (75% Cadenza, 25% Paragon). Across the six reads supporting the alternate base, four displayed several SNPs and mapped to an additional Cadenza NLR. This provides evidence for the presence of the identical gene in Paragon which is supported by phenotypic data.

Supplementary File 4: **Syntenic region across different grasses (Supplementary Table 6) and the NLR loci identified with NLR-Annotator**. See Methods for a detailed explanation of the analysis and Supplementary Figure 7 for an illustration.

Supplementary File 5: **Curated sequences of BED-NLRs from chromosome 2B and Ta_2D7**. Exons are highlighted with different colours (yellow, green, blue, pink). Amino acids encoding the BED domain are shown in red and those encoding the NB-ARC domain are in blue. LRR repeats identified with NLR Annotator are highlighted in dark green and manually annotated LRR motifs xxLxLxx are underlined and in bold black.