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

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

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

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

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

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

[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

Introductory paragraph

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 for 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 and unique recognition specificity. We show that *Yr5*, which remains effective to a broad range of *Pst* isolates worldwide, is allelic to *YrSP* and paralogous to *Yr7*, both of which have been overcome by multiple *Pst* isolates. 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 enable 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.

Main

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 blast 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 one of >6,000 tested *Pst* isolates worldwide has been found virulent to *Yr5* (Supplementary Table 1[14,16]), 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 US (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[17] 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[15,18,19]. 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 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 suggests that *Yr5* and *YrSP* are encoded by alleles of the same gene, but are paralogous to *Yr7*. 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). Taken together, the mutant and genetic analyses demonstrate that *Yr5* and *YrSP* are allelic, while *Yr7* is encoded by a 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. *Mla10l*[20], *Sr33*[21], *Pm3*[22]), neither *Yr7* nor *Yr5*/*YrSP* encode Coiled Coil domains at the N-terminus (Supplementary Figure 4). Only 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 to the presence of the premature termination codon in exon 3. YrSP still confers functional resistance to *Pst*, although with a different recognition specificity to Yr5. 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 the *Yr5* and *YrSP* alleles 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 WKRY 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[23].

We examined the allelic variation in *Yr7*, *Yr5*, and *YrSP* across eight sequenced tetraploid and hexaploid wheat genomes (Supplementary Table 5). We identified *Yr7* only in Cadenza and Paragon, which are identical-by-descent in this interval (Supplementary File 3, Supplementary Table 6, 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 6).

For *Yr5/YrSP*, we identified three additional alleles in the sequenced hexaploid wheat cultivars (Figure 2b; Supplementary Table 7). 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 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 allele with a truncation in the LRR region distinct from YrSP, in addition to multiple amino acid substitutions across the C-terminus (Supplementary Table 7). This truncated tetraploid allele is reminiscent of YrSP and is expressed in Kronos (see Methods). However, none of these cultivars (Claire, Robigus,

119 Paragon, Cadenza, Svevo, and Kronos) exhibit a *Yr5*/*YrSP* resistance response, suggesting
120 that these amino acid changes and truncations may alter recognition specificity or protein
121 function.
122
123 We designed diagnostic markers for *Yr7*, *Yr5*, and *YrSP* to facilitate their detection and use in
124 breeding. We confirmed their presence in the donor cultivars Thatcher and Lee (*Yr7*),
125 Spaldings Prolific (*YrSP*), and spelt wheat cv. Album (*Yr5*) (Supplementary Tables 8-9;
126 Supplementary Figures 5-6). We tested *Yr7* and *YrSP* markers in a collection of global
127 landraces[24] and European cultivars[25] released over the past one hundred years. *YrSP* was
128 absent from the tested germplasm, except for AvocetS-*YrSP* (Supplementary Table 9). *Yr7* on
129 the otherhand was more prevalent in the germplasm tested and we could track its presence
130 across pedigrees, including Cadenza derived cultivars (Supplementary Tables 8-9;
131 Supplementary Figure 5). We confirmed *Yr5* in the AvocetS-*Yr5* and Lemhi-*Yr5* lines and it
132 was not detected in the other tested lines, consistent with the fact that *Yr5* has not yet been
133 deployed within European breeding programmes (Supplementary Tables 10 and 17 and
134 Supplementary Figure 6). The *Yr5* diagnostic marker will facilitate its deployment, hopefully
135 within a breeding strategy that ensures its effectiveness long-term[26].
136
137 We defined the *Yr7/Yr5/YrSP* syntenic interval across the wheat genomes and related grass
138 species *Aegilops tauschii* (D genome progenitor), *Hordeum vulgare* (barley), *Brachypodium*
139 *distachyon*, and *Oryza sativa* (rice) (Supplementary files 4 and 5, Supplementary Figure 7).
140 We identified both canonical NLRs, as well as BED-NLRs across all genomes and species,
141 except for barley, which only contained canonical NLRs across the syntenic region. The
142 phylogenetic relationship based on the NB-ARC domain suggests a common evolutionary
143 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 would suggest 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 11; 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[27] 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[28], 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[29]; Supplementary Figure 10, Supplementary Table 12). 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 two paralogous BED-NLRs genes. We report an allelic series for the *Yr5/YrSP* gene with five independent alleles, including three full-length BED-NLRs (including *Yr5*) and two truncated versions (including *YrSP*). This wider allelic series could be of functional significance as previously shown for the *Mla* and *Pm3* loci that confer resistance to *Blumeria graminis*[22,30] in barley and wheat,

194 respectively, and the flax *L* locus conferring resistance to *Melampsora lini*[31]. Overall, our
195 results add strong evidence for the importance of the BED-NLR architecture in plant-
196 pathogen interactions. The paralogous and allelic relationship of these three distinct *Yr* loci
197 will inform future hypothesis-driven engineering of novel recognition specificities.

198 **Methods**

199 **MutRenSeq**

200 *Mutant identification*

201 Supplementary Table 3 summarises plant materials and *Pst* isolates used to identify mutants
202 for each *Yr* gene. We used an EMS-mutagenised population in cultivar Cadenza[32] to identify
203 mutants in *Yr7*; whereas EMS-populations in the corresponding AvocetS-*Yr* near isogenic
204 line (NIL) were used to identify *Yr5* and *YrSP* mutants. For *Yr7*, we inoculated M_3 plants
205 from the Cadenza EMS population with *Pst* isolate 08/21 which is virulent to *Yr1, Yr2, Yr3,*
206 *Yr4, Yr6, Yr9, Yr17, Yr27, Yr32, YrRob,* and *YrSol*[33]. We hypothesised that susceptible
207 mutants would carry mutations in *Yr7*. Plants were grown in 192-well trays in a confined
208 glasshouse with no supplementary lights or heat. Inoculations were performed at the one leaf
209 stage (Zadoks 11) with a talc – urediniospore mixture. Trays were kept in darkness at 10 °C
210 and 100% humidity for 24 hours. Infection types (IT) were recorded 21 days post-inoculation
211 (dpi) following the Grassner and Straib scale[34]. Identified susceptible lines were progeny
212 tested to confirm the reliability of the phenotype and DNA from M_4 plants was used for
213 RenSeq (see section below). Similar methods were used for AvocetS-*Yr7*, AvocetS-*Yr5*, and
214 AvocetS-*YrSP* EMS-mutagenised populations with the following exceptions: *Pst* pathotypes
215 108 E141 A+ (University of Sydney Plant Breeding Institute Culture no. 420), 150 E16 A+
216 (Culture no. 598) and 134 E16 A+ (Culture no. 572) were used to evaluate *Yr7*, *Yr5*, and
217 *YrSP* mutants, respectively. EMS-derived susceptible mutants in Lemhi-*Yr5* were previously
218 identified[35] and DNA from M_5 plants was used for RenSeq.

219

220 *DNA preparation, resistance gene enrichment and sequencing (RenSeq)*

221 We extracted total genomic DNA from young leaf tissue using the large-scale DNA
222 extraction protocol from the McCouch Lab (https://ricelab.plbr.cornell.edu/dna_extraction)

and a previously described method[36]. 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 3 and 12.

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[37] 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 14). 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. The 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 5, 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 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 15. 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

297 sequence, corresponding to a 39 bp insertion and a 129 bp deletion (Supplementary File 3),
298 and to confirm the presence of the mutations in each mutant line.
299 We used HISATt2[38] (v2.1) to map RNA-Seq reads available from Cadenza and AvocetS-
300 *Yr5*[29] to the RenSeq *de novo* assemblies with curated loci to define the structure of the genes.
301 We used the following parameters: --no-mixed --no-discordant to map reads in pairs only.
302 We used the --novel-splicesite-outfile to predict splicing sites that we manually scrutinised
303 with the genome visualisation tool IGV[39] (v2.3.79). Predicted coding sequences (CDS) were
304 translated using the ExPASy online tool (https://web.expasy.org/translate/). This allowed us
305 to predict the effect of the mutations on each candidate transcript (Figure 1a; Supplementary
306 Table 4). The long-range primers for both *Yr7* and *Yr5* loci were then used on the
307 corresponding susceptible Avocet NIL mutants to determine whether the genes were present
308 and carried mutations in that background (Figure 1a; Supplementary Files 1 and 2).
309
310 **Coiled coil domain prediction**
311 To determine whether *Yr7*, *Yr5*, and *YrSP* encode Coiled Coil (CC) domains we used the
312 NCOILS prediction program[40] (v1.0, https://embnet.vital-it.ch/software/COILS_form.html)
313 with the following parameters: MTIK matrix with applying a 2.5-fold weighting of positions
314 a,d. We compared the profiles to those obtained with already characterised CC-NLR
315 encoding genes *Sr33*, *Mla10*, *Pm3* and *RPS5* (Supplementary Figure 4). We also ran the
316 program on Yr7 and Yr5 protein sequences where the BED domain was manually removed to
317 determine whether its integration could have disrupted an existing CC domain. To further
318 investigate whether *Yr7*, *Yr5*, and *YrSP* encode CC domains we performed a BLASTP
319 analysis[41] with their N-terminal region, from the methionine to the first amino acid encoding
320 the NB-ARC domain, with or without the BED domain (Supplementary Figure 4).
321

322 **Genetic linkage**

323 We generated a set of F_2 populations to genetically map the candidate contigs
324 (Supplementary Table 3). For *Yr7* we developed an F_2 population based on a cross between
325 the susceptible mutant line Cad0127 to the Cadenza wild-type (population size 139
326 individuals). For *Yr5* and *YrSP* we developed F_2 populations between AvocetS and the NILs
327 carrying the corresponding *Yr* gene (94 individuals for *YrSP* and 376 for *Yr5*). We extracted
328 DNA from leaf tissue at the seedling stage (Zadoks 11) following a previously published
329 protocol[42] and Kompetitive Allele Specific PCR (KASP) assays were carried out as described
330 in[43]. R/qtl package[44] was used to produce the genetic map based on a general likelihood ratio
331 test and genetic distances were calculated from recombination frequencies (v1.41-6).
332
333 We used previously published markers linked to *Yr7*, *Yr5*, and *YrSP* (WMS526, WMS501
334 and WMC175, WMC332, respectively[15,18,19]) in addition to closely linked markers WMS120,
335 WMS191, and WMC360 (based on the GrainGenes database
336 https://wheat.pw.usda.gov/GG3/) to define the physical region on the Chinese Spring
337 assembly RefSeq v1.0 (https://wheat-urgi.versailles.inra.fr/Seq-Repository/Assemblies). Two
338 different approaches were used for genetic mapping depending on the material. For *Yr7*, we
339 used the public data[32] for Cad0127 (www.wheat-tilling.com) to identify nine mutations
340 located within the *Yr7* physical interval based on BLAST analysis against RefSeq v1.0. We
341 used KASP primers when available and manually designed additional ones including an
342 assay targeting the Cad0127 mutation in the *Yr7* candidate contig (Supplementary Table 15).
343 We genotyped the Cad0127 F_2 populations using these nine KASP assays and confirmed
344 genetic linkage between the Cad0127 *Yr7* candidate mutation and the nine mutations across
345 the physical interval (Supplementary Figure 3).
346

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 15). For both candidate contigs we confirmed genetic linkage with the previously published genetic intervals for these *Yr* genes (Supplementary Figure 3).

_Yr7_ gene-specific markers

We aligned the *Yr7* sequence with the best BLAST hits in the genomes listed on Supplementary Table 5 and designed KASP primers targeting polymorphisms that were *Yr7*-specific. Three markers were retained after testing on a selected panel of Cadenza-derivatives and cultivars that were positive for *Yr7* markers in the literature, including the *Yr7* reference cultivar Lee (Supplementary Table 8, 8 and 15). The panel of Cadenza-derivatives was phenotyped with three *Pst* isolates: *Pst* 08/21 (*Yr7*-avirulent), *Pst* 15/151 (*Yr7*-avirulent – virulent to *Yr1, 2, 3, 4, 6, 9, 17, 25, 32, Rendezvous, Sp, Robigus, Solstice*) and *Pst* 14/106 (*Yr7*-virulent, virulent to *Yr1, 2, 3, 4, 6, 7, 9, 17, 25, 32, Sp, Robigus, Solstice ,Warrior, Ambition, Cadenza, KWS Sterling, Apache*) to determine whether *Yr7*-positive cultivars, as identified by the three KASP markers, displayed a consistent specificity (Supplementary Table 8). Pathology assays were performed as for the screening of the Cadenza mutant population. We retrieved pedigree information for the analysed cultivars from the Genetic Resources Information System for Wheat and Triticale database (GRIS, www.wheatpedigree.net) and used the Helium software[45] (v1.17) to illustrate the breeding history of *Yr7* in the UK (Supplementary Figure 5).

We used the three *Yr7* KASP markers to genotype (i) cultivars from the AHDB Wheat Recommended List from 2005-2018 (https://cereals.ahdb.org.uk/varieties/ahdb-

372 recommended-lists.aspx); (ii) the Gediflux collection of European bread wheat cultivars
373 released between 1920 and 2010[25] and (iii) the core Watkins collection, which represents a
374 global set of wheat landraces collected in the 1930s[24]. Results are reported in Supplementary
375 Table 9.
376
377 **_Yr5_ and _YrSP_ gene-specific markers**
378 We identified a 774 bp insertion in the *Yr5* allele 29 bp upstream of the STOP codon with
379 respect to the Cadenza and Claire alleles. Genomic DNA from *YrSP* confirmed that the
380 insertion was specific to *Yr5*. We used this polymorphism to design primers flanking the
381 insertion and tested them on a subset of the collections mentioned above. We added 32 DNA
382 sample from diverse accessions of *Triticum dicoccoides,* the wild progenitor of domesticated
383 wheat (passport data shown in Supplementary Table 17). We included DNA from *Triticum*
384 *aestivum* ssp. *spelta* var. album[35] (*Yr5* donor) and Spaldings Prolific (*YrSP* donor) to assess
385 their amplification profiles. PCR amplification was conducted using a touchdown
386 programme: 10 cycles, -0.5 °C per cycle starting from 67 °C and the remaining 25 cycles at
387 62 °C. This allowed us to increase the specificity of the reaction. We observed three different
388 profiles on the tested varieties; (i) a 1,281 bp amplicon in *Yr5* positive cultivars, (ii) a 507 bp
389 amplicon in the alternate *Yr5* allele carriers, including AvocetS-*YrSP*, Cadenza, and Claire,
390 and (iii) no amplification in other varieties. We sequenced the different amplicons and
391 confirmed the insertion in *Yr5* compared to the alternate alleles (Supplementary File 2). The
392 lack of amplicons in some varieties most likely respresents the absence of the loci in the
393 tested varieties. For *YrSP*, we aligned the *YrSP* and *Yr5* sequences to design KASP primers
394 targeting the G to C SNP between the two alleles (Supplementary File 2, Supplementary
395 Table 16). We tested the marker by genotyping selected cultivars as controls and cultivars
396 from the AHDB Wheat Recommended List from 2005-2018 (Supplementary Table 9).

397

398 **_In silico_ allele mining for _Yr7_ and _Yr5_**

399 We used the *Yr7* and *Yr5* sequences to retrieve the best BLAST hits in the *T. aestivum* and *T.*
400 *turgdium* wheat genomes listed in Supplementary Table 5. The best *Yr5* hits shared between
401 93.6 and 99.3% sequence identity, which was comparable to what was observed for alleles
402 derived from the wheat *Pm3* (>97% identity)[46] and flax *L* (>90% identity)[31] genes. *Yr7* was
403 identified only in Paragon and Cadenza (Supplementary Table 6; See Supplementary File 3
404 for curation of the Paragon sequence).

405

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

407 *Definition of syntenic regions across grass genomes*

408 We used NLR-Annotator to identify putative NLR loci on RefSeq v1.0 chromosome 2B and
409 identified the best BLAST hits to *Yr7* and *Yr5* on RefSeq v1.0. Additional BED-NLRs and
410 canonical NLRs were annotated in close physical proximity to these best BLAST hits.
411 Therefore, to better define the NLR cluster we selected ten non-NLR genes located both
412 distal and proximal to the region, and identified orthologs in barley, *Brachypodium*, and rice
413 in *EnsemblPlants* (https://plants.ensembl.org/). We used different % ID cutoffs for each
414 species (>92% for barley, >84% for *Brachypodium*, and >76% for rice) and determined the
415 syntenic region when at least three consecutive orthologues were found. A similar approach
416 was conducted for *Triticum* ssp and *Ae. tauschii* (Supplementary file 4).

417

418 *Definition of the NLR content of the syntenic region*

419 We extracted the previously defined syntenic region from the grass genomes listed in
420 Supplementary Table 5 and annotated NLR loci with NLR-Annotator. We maintained
421 previously defined gene models where possible, but also defined new gene models that were

422 further analysed through a BLASTx analysis to confirm the NLR domains (Supplementary
423 Files 4 and 5). The presence of BED domains in these NLRs was also confirmed by CD-
424 Search (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi).
425
426 **Phylogenetic and neighbour network analyses**
427 We aligned the translated NB-ARC domains from the NLR-Annotator output with
428 MUSCLE[47] using default parameters (v.3.8.31). We verified and manually curated the
429 alignment with Jalview[48] (v2.10.1). We used Gblocks[49] (v0.91b) with the following
430 parameters: Minimum Number Of Sequences For A Conserved Position: 9; Minimum
431 Number Of Sequences For A Flanking Position: 14; Maximum Number Of Contiguous
432 Nonconserved Positions: 8; Minimum Length Of A Block: 10; Allowed Gap Positions: None;
433 Use Similarity Matrices: Yes; to eliminate poorly aligned positions. This resulted in 36% of
434 the original 156 positions being taken forward for the phylogeny. We built a Maximum
435 Likelihood tree with the RAxML[50] program and the following parameters: raxmlHPC -f a -x
436 12345 -p 12345 -N 1000 -m PROTCATJTT -s <input_alignment.fasta> (MPI version
437 v8.2.10). The best scoring tree with associated bootstrap values was visualised and mid-
438 rooted with Dendroscope[51] (v3.5.9). There was clear separation between NLRs belonging to
439 the two different clusters but the sub-clades had less support. One explanation would be that
440 conflicting phylogenetic signals that are due to events such as hybridization, horizontal gene
441 transfer, recombination, or gene duplication and loss might have occurred in the region. Split
442 networks allow nodes that do not represent ancestral species and can thus represent such
443 incompatible and ambiguous signals. We therefore used this method in the following part of
444 the analysis to analyse the relationship between the BED domains.
445

We used the Neighbour-net method[52] implemented in SplitsTree4[27] (v4.16) to analyse the relationships between BED domains from NLR and non-NLR proteins. First we retrieved all BED-containing proteins from RefSeq v1.0 using the following steps: we used hmmer (v3.1b2, http://hmmer.org/) to identify conserved domains in protein sequences from RefSeq v1.0. We applied a cut-off of 0.01 on i-evalue to filter out any irrelevant identified domains. We separated the set between NLR and non-NLRs based on the presence of the NB-ARC and sequence homology for single BED proteins. BED domains were extracted from the corresponding protein sequences based on the hmmer output and were verified on the CD-search database. Alignments of the BED domains were performed in the same way as for NB-ARC domains and were used to generate a neighbour network in SplitsTree4 based on the uncorrected P distance matrix.

Transcriptome analysis

Kronos analysis

We reanalysed RNA-Seq data from cultivar Kronos[53] to determine whether the Kronos *Yr5* allele was expressed. We followed the same strategy as that described to define the *Yr7* and *Yr5* gene structures (candidate contig confirmation and gene annotation section). We generated a *de novo* assembly of the Kronos NLR repertoire from Kronos RenSeq data and used it as a reference when mapping read data from one replicate of the wild-type Kronos at heading stage. Read depths up to 30x were present for the *Yr5* allele which allowed confirmation of its expression. Likewise, the RNA-Seq reads confirmed the gene structure, which is similar to *YrSP*, and the premature termination codon in Kronos *Yr5*. Whether this allele confers resistance against *Pst* remains to be elucidated.

Re-analysis of RNA-Seq data in Dobon et al., 2016

We used RNA-Seq data previously published by Dobon and colleagues[18]. Briefly, two RNA-Seq time-courses were used based on samples taken from leaves at 0, 1, 2, 3, 5, 7, 9, and 11 dpi for the susceptible cultivar Vuka and 0, 1, 2, 3, and 5 dpi for the resistant AvocetS-*Yr5*[29]. We used normalised read counts (Transcript Per Million, TPM) from Ramirez-Gonzalez et al. 2018 to produce the heatmap shown in Supplementary Figure 10 with the pheatmap R package[54] (v1.0.8). Transcripts were clustered according to their expression profile as defined by a Euclidean distance matrix and hierarchical clustering. Transcripts were considered expressed if their average TPM was ≥0.5 TPM in at least one time point. We used the DESeq2 R package[55] (v1.18.1) to conduct a differential expression analysis. We performed two comparisons: (1) we used a likelihood ratio test to compare the full model ~ Cultivar + Time + Cultivar:Time to the reduced model ~ Cultivar + Time to identify genes that were differentially expressed between the two cultivars at a given time point after 0 dpi (workflow: https://www.bioconductor.org/help/workflows/rnaseqGene/); (2) Investigation of both time courses in Vuka and AvocetS-*Yr5* independently to generate all of the comparisons between 0 dpi and any given time point, following the standard DESeq2 pipeline. Genes were considered as differentially expressed genes if they showed an adjusted p-value < 0.05 and a log2 fold change of 2 or higher. Most BED-containing proteins and BED-NLRs were not expressed in the analysed data. No pattern was observed for those that were expressed: differences were observed between cultivars, but these were independent of the presence of the yellow rust pathogen.

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. 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).

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

18. 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).

19. 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).

20. 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).

21. 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).

22. 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).

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

24. 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).

25. 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).

26. 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).

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

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

29. 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).

30. 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).

31. 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).

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

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

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

35. 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).

36. 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).

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

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

39. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer

(IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14,** 178–192 (2013).

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

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

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

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

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

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

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

47. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).

48. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25,** 1189–1191 (2009).

49. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17,** 540–552 (2000).

50. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22,** 2688–2690 (2006).

51. Huson, D. H. & Scornavacca, C. Dendroscope 3: An interactive yool for rooted phylogenetic trees and networks. *Syst. Biol.* **61,** 1061–1067 (2012).

52. Bryant, D. & Moulton, V. Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. *Mol. Biol. Evol.* **21,** 255–265 (2003).

53. Pearce, S. *et al.* Regulation of Zn and Fe transporters by the GPC1gene during early wheat monocarpic senescence. *BMC Plant Biol.* **14,** 368 (2014).

54. Kolde, R. Pheatmap: pretty heatmaps. *R package version* (2015).

55. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).

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

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

58. Avni, R. *et al*. Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science (80-.).* **357,** 93–97 (2017).

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

Author contributions

CM performed the experiments to clone *Yr7* and *Yr5* and the subsequent analyses of their loci and BED domains, designed the gene-specific markers, analysed the genotype data in the studied panels, and designed and made the figures. JZ performed the experiments to clone *YrSP*, confirm the *Yr7* and *Yr5* genes in AvocetS-*Yr7* and AvocetS-*Yr5* mutants, and identified the full length of *Yr5* and *YrSP* with their respective regulatory elements. CM and 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 13 (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 community feedback on the initial preprint submission which helped improve the manuscript.

We thank the International Wheat Genome Sequencing Consortium for having providing us with 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.

668 **Figure legends**

669 **Figure 1**: ***Yr5* and *YrSP* are allelic, but paralogous to *Yr7*.**

670 **a**, Left: Wild-type and selected EMS-derived susceptible mutant lines for *Yr7*, *Yr5*, and *YrSP*
671 (Supplementary Table 3 and 3) inoculated with *Pst* isolate 08/21 (*Yr7*), *Pst* 150 E16 A+
672 (*Yr5*), or *Pst* 134 E16 A+ (*YrSP*). Right: Candidate gene structures, with mutations in red,
673 and their predicted effects on the translated protein. **b**, Schematic representation of the
674 physical interval of the *Yr* loci. The *Yr7/Yr5/YrSP* locus is shown in orange on chromomsome
675 2B with previously published SSR markers in black. Markers developed in this study to
676 confirm the genetic linkage between the phenotype and the candidate contigs are shown as
677 black lines underneath the chromosme. *Yr* loci mapping intervals are defined by the red
678 horizontal lines. A more detailed genetic map is shown in Supplementary Figure 3.

679

680 **Figure 2: *Yr7* and *Yr5/YrSP* encode integrated BED-domain immune receptor genes.**

681 **a**, Schematic representation of the Yr7, Yr5, and YrSP protein domain organisation. BED
682 domains are highlighted in red, NB-ARC domains are in blue, LRR motifs from NLR-
683 Annotator are in dark green, and manually annotated LRR motifs (xxLxLxx) are in light
684 green. Black triangles represent the EMS-induced mutations within the protein sequence. The
685 plot shows the degree of amino acid conservation (50 amino acid rolling average) between
686 Yr7 and Yr5 proteins, based on the conservation diagram produced by Jalview (2.10.1) from
687 the protein alignment. Regions that correspond to the conserved domains have matching
688 colours. The amino acid changes between Yr5 and YrSP are annotated in black on the YrSP
689 protein. **b**, Five Yr5/YrSP haplotypes were identified in this study. Polymorphisms are
690 highlighted across the protein sequence with orange vertical bars for polymorphisms shared
691 by at least two haplotypes and blue vertical bars for polymorphisms that are unique to the

692 corresponding haplotype. Matching colours across protein structures illustrate 100%
693 sequence conservation.

694 **Figure 3: BED domains from BED-NLRs and non-NLR proteins are distinct.**

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

Percentage of total harvested weight of wheat cultivar carrying *Yr7* (green) and the proportion of *Pst* isolates that are 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[56]), 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* allele 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 15). 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[21], Mla10[20], Pm3[22] and RPS5[57]. 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[20–22]. 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* allelic 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[45] (v1.17).

Supplementary Figure 6: Diagnostic genetic marker for *Yr5*.

The *Yr5*-specific insertion was used to generate a PCR amplification product of 1,281 bp for *Yr5* or a shorter amplicon for the absence of the insertion in *YrSP*, Claire, and Paragon (507 bp). *Yr5* positive lines include the *Yr5* spelt donor and *Yr5* near-isogenic lines AvocetS-*Yr5* and Lemhi-*Yr5*. *YrSP* donor Spaldings Prolific and *YrSP* near-isogenic lines AvocetS-*YrSP* carry the shorter alternate allele, similar to the Claire, Cadenza and Paragon alleles identified in Figure 2. Negative controls include AvocetS and H_2O. Size marker is shown on the left.

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[29] 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 one isolate from Australia, PST 360 E137 A+[14]. 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* cultivarsfrom 1990 to 2016 and *virYr7* prevalence 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 with the different *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, score 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: Genome assemblies used in the present study.

Summary of the available genome assemblies[58,59] that were used for the *in silico* allele mining and synteny analysis across rice, *Brachypodium*, barley and different Triticeae accessions.

Supplementary Table 6: *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 5.

Supplementary Table 7: Polymorphisms between Yr5 protein and its identified alleles.

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 8: 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.

878

879 **Supplementary Table 9: Presence/absence of *Yr7* and *YrSP* in different wheat**
880 **collections.** We used Vuka, AvocetS and Solstice as negative controls for the presence of *Yr7*
881 and *YrSP* and AvocetS-*Yr* near-isogenic lines as controls for the corresponding *Yr* gene. We
882 genotypied different collections: (i) a set of potential *Yr7* carriers based on literature research,
883 (ii) a set of cultivars that belonged to the UK AHDB Recommended List
884 (https://cereals.ahdb.org.uk/varieties/ahdb-recommended-lists.aspx) between 2005 and 2018
885 (labelled 2005-2018-UK_RL), (iii) the Gediflux collection that includes modern European
886 bread wheat cultivars (1920-2010)[25], (iv) a core set of the Watkins collection, which
887 represent a set of global bread wheat landraces collected in the 1920-30s[24]. Most of the
888 putative *Yr7* carriers were positive for all the *Yr7* markers apart from Aztec, Chablis and
889 Cranbrook. Chablis was susceptible to the *Pst* isolates that were avirulent to *Yr7* so it
890 probably does not carry the gene. Regarding the 2005-2018-UK_RL results were consistent
891 across already tested cultivars: Cadenza, Cordiale, Cubanita, Grafton and Skyfall were
892 already positive in Supplementary Table 8. Energise, Freiston, Gallant, Oakley and
893 Revelation were negative on both panels as well. Results were thus consistent across different
894 sources of DNA. *Yr7*-containing cultivars are not prevalent in the 2005-2018 Recommended
895 List set, however, this gene is present in Skyfall, which is currently one of the most harvested
896 cultivars in the UK (Supplementary Table 2). We tested the *YrSP* marker on this set and it
897 was positive only for AvocetS-*YrSP*. The frequency of *Yr7* was relatively low in the Gediflux
898 panel (4%). This is consistent with results in Supplementary Table 2: *Yr7* deployment started
899 in the UK in 1992 with Cadenza and it was rarely used prior to that date. The same was
900 observed in the subset of the Watkins collection (10%) where landraces that were positive for
901 *Yr7* all originated from India and the Mediterranean basin. *Yr7* was introgressed into
902 Thatcher (released in 1936) from Iumillo, which originated from Spain and North-Africa
903 (Genetic Resources Information System for Wheat and Tritical -
904 http://www.wheatpedigree.net/). Iumillo is likely to be pre-1920s and these landraces are all
905 bread wheats so they might have inherited it from another source. However, there is no
906 evidence for *Yr7* coming from another source than Iumillo in the modern bread wheat
907 cultivars.

908

909 **Supplementary Table 10**: **Presence/absence of *Yr5* alleles in selected cultivars.**
910 A subset of the aforementioned collection was investigated for the *Yr5* presence. “Yes” in the
911 *Yr5* column refers to amplification of the 1,281 bp amplicon with the *Yr5*-Insertion primers

(Supplementary Figure 6). "Yes" in the *Yr5* alternate alleles column refers to the amplification of the 507 bp amplicon that was identified for AvocetS-*YrSP*, Claire, Cadenza and Paragon in Supplementary Figure 6. "Yes" in the no amplification column refers to identification of a profile similar to the one found for AvocetS in Supplementary Figure 6.

Supplementary Table 11: 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 12: Transcripts per Million-normalised read counts from the re-analysis of published RNA-Seq data[29] and associated differential expression analysis performed with DESeq2.

Supplementary Table 13: Sequencing details of RenSeq data generated in this study.

Supplementary Table 14: *De novo* assemblies generated from the corresponding RenSeq data.

Supplementary Table 15: 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 16: 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 Table 17: Passport data of tested *T. dicoccoides* accessions

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

15). 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 15)**. 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.

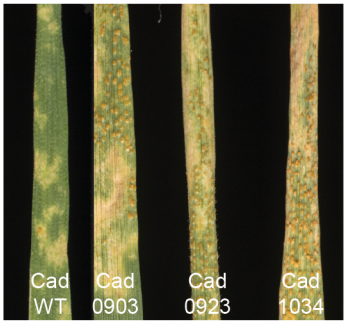
Supplementary File 3: **Curation of the *Yr7* locus in the Cadenza genome assembly based on Sanger sequencing results.**

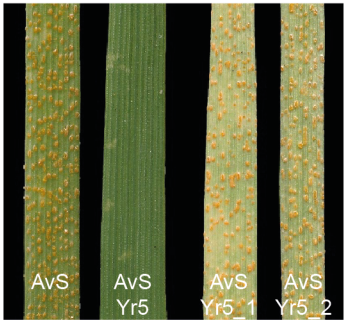
Comments show the position of the unknown bases ("N") in the "Yr7_with_Ns" sequence. Curation based on Sanger sequencing data is shown in bold black in the "curated_Yr7" sequence with the 39 bp insertion and 129 bp deletion. Allele mining for *Yr7* in the Paragon assembly showed that a similar assembly 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 5). 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.

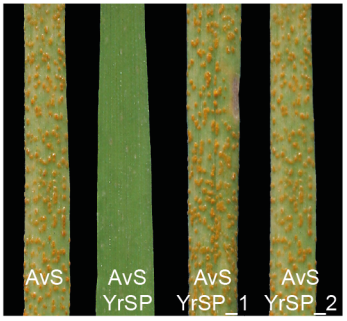
Supplementary File 4: **Syntenic region across different grasses (Supplementary Table 5) and the NLR loci identified with NLR-Annotator**. See Methods for a detailed explanation of the analysis and Supplementary Figure 7 for an illustration.

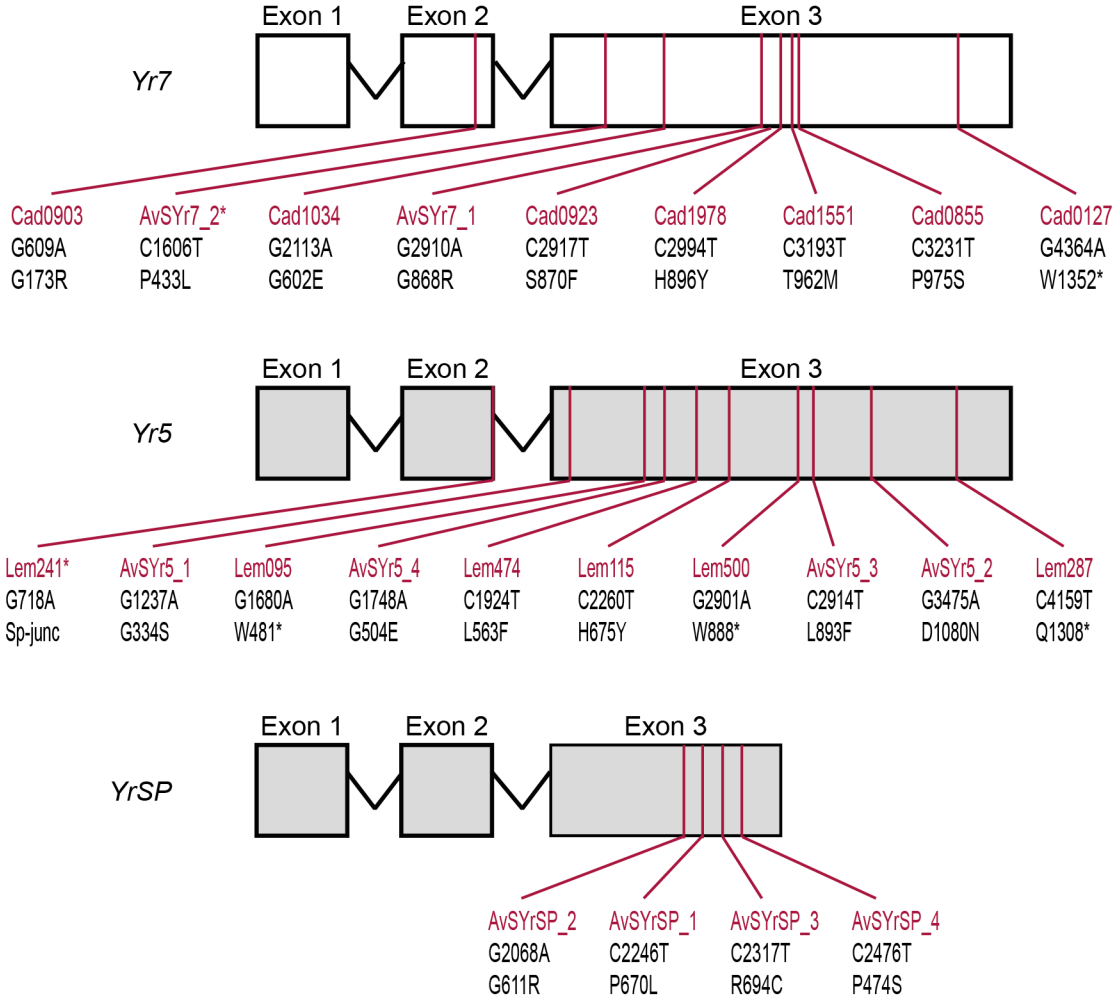
980 **Supplementary File 5**: **Curated sequences of BED-NLRs from chromosome 2B and**
981 **Ta_2D7**. Exons are highlighted with different colours (yellow, green, blue, pink). Amino
982 acids encoding the BED domain are shown in red and those encoding the NB-ARC domain
983 are in blue. LRR repeats identified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

a

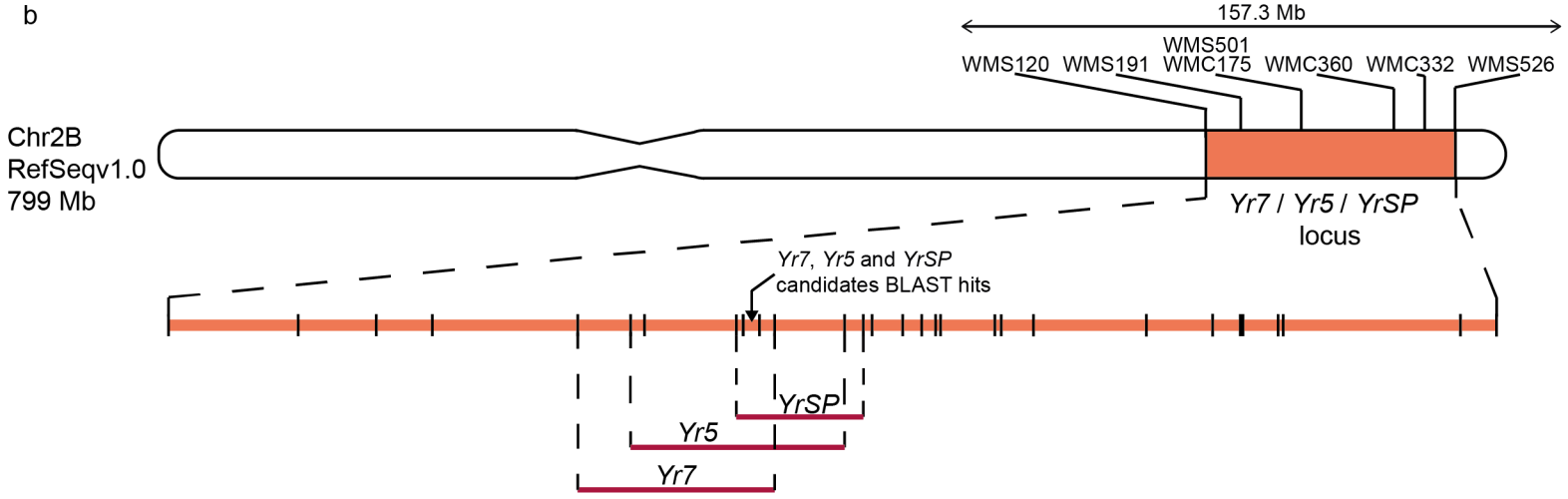

Cad WT
Cad 0903
Cad 0923
Cad 1034

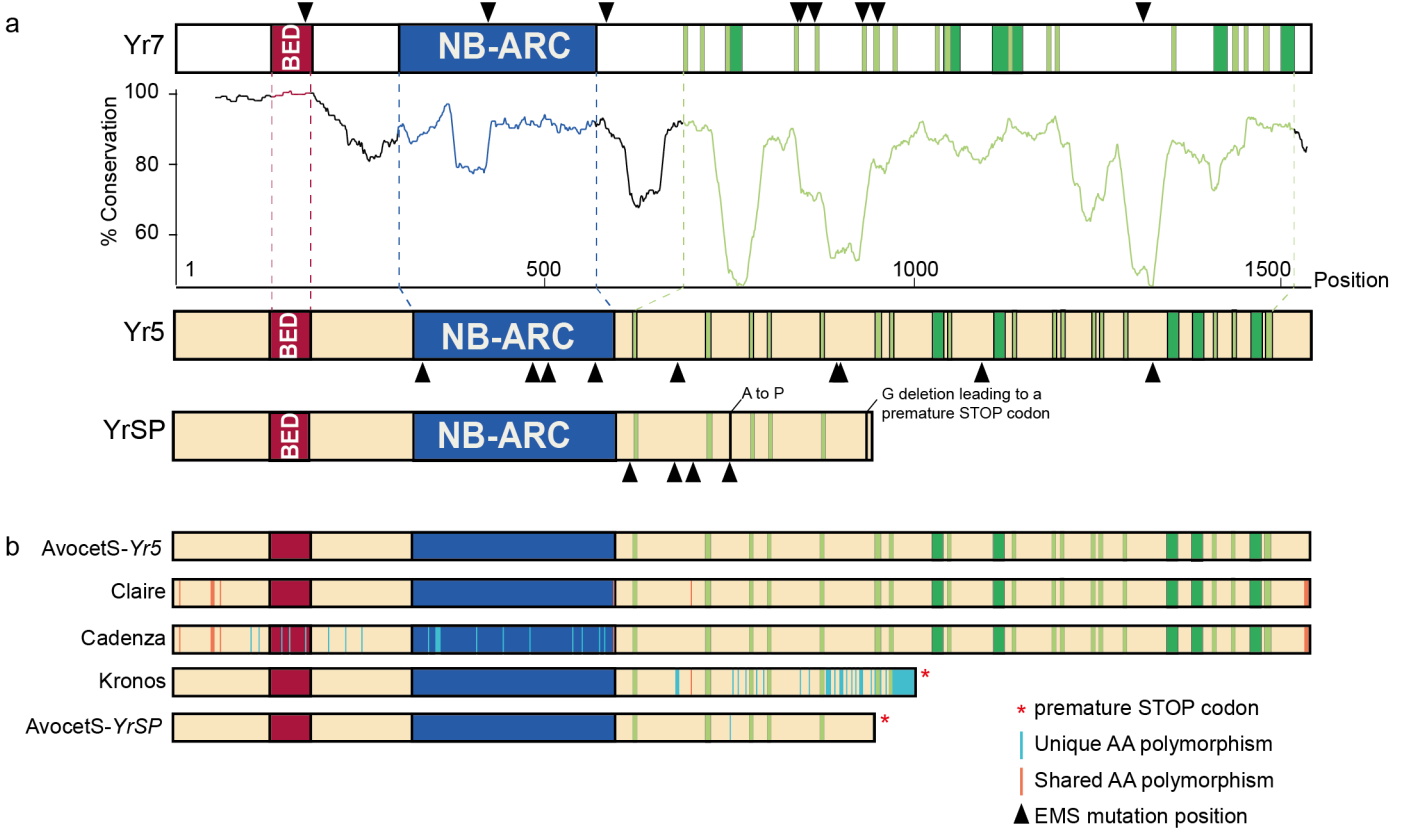

AvS
AvS Yr5
AvS Yr5_1
AvS Yr5_2


AvS
AvS YrSP
AvS YrSP_1
AvS YrSP_2


Exon 1
Exon 2
Exon 3
Yr7
Cad0903 G609A G173R
AvSYr7_2* C1606T P433L
Cad1034 G2113A G602E
AvSYr7_1 G2910A G868R
Cad0923 C2917T S870F
Cad1978 C2994T H896Y
Cad1551 C3193T T962M
Cad0855 C3231T P975S
Cad0127 G4364A W1352*
Yr5
Lem241* G718A Sp-junc
AvSYr5_1 G1237A G334S
Lem095 G1680A W481*
AvSYr5_4 G1748A G504E
Lem474 C1924T L563F
Lem115 C2260T H675Y
Lem500 G2901A W888*
AvSYr5_3 C2914T L893F
AvSYr5_2 G3475A D1080N
Lem287 C4159T Q1308*
YrSP
AvSYrSP_2 G2068A G611R
AvSYrSP_1 C2246T P670L
AvSYrSP_3 C2317T R694C
AvSYrSP_4 C2476T P474S

b

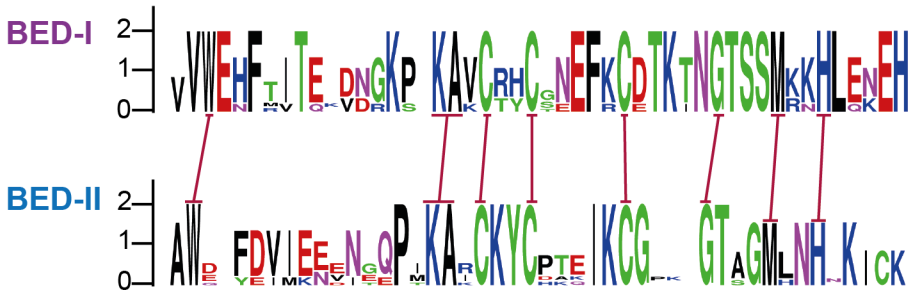

157.3 Mb
WMS501
WMS120
WMS191
WMC175
WMC360
WMC332
WMS526
Chr2B
RefSeqv1.0
799 Mb
Yr7 / Yr5 / YrSP
locus
Yr7, Yr5 and YrSP
candidates BLAST hits
YrSP
Yr5
Yr7


a
Yr7
BED
NB-ARC
% Conservation
100
80
60
1
500
1000
1500
Position
Yr5
BED
NB-ARC
YrSP
BED
NB-ARC
A to P
G deletion leading to a premature STOP codon
b
AvocetS-Yr5
Claire
Cadenza
Kronos
AvocetS-YrSP
* premature STOP codon
Unique AA polymorphism
Shared AA polymorphism
EMS mutation position

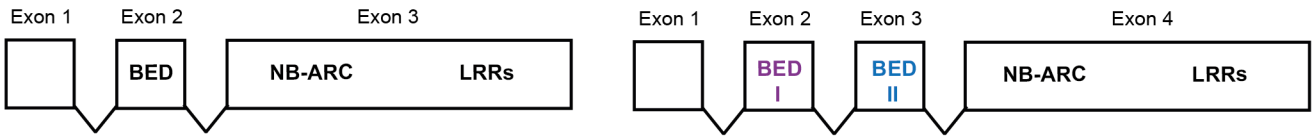
a

Syntenic interval	#NLRs	#BED NLRs	#BED NLRs-I	#BED NLRs-II	#BED NLRs-I-II
Rice	6	2	-	-	-
Brachypodium	4	4	1	1	-
Barley	2	-	-	-	-
Aegilops tauschii (D)	8	4	1	1	1
Hexaploid wheat (D)	6	2	1	-	1
Tetraploid wheat (A)	8	1	1	-	-
Hexaploid wheat (A)	12	5	3	2	-
Tetraploid wheat (B)	20	10	6	2	1
Hexaploid wheat (B)	13	5	1	1	3

b



c



d