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 mode 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 and tested 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 (LRR) 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 function of the BED domains from 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 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 *Yr5* confers resistance to almost all tested PST isolates worldwide, both *Yr7* and *YrSP* have been overcome in the field, and each gene displays a distinct recognition specificity. Wide deployment of *Yr7* is correlated to the increase in the virulence for *Yr7* among PST isolates in the UK (Supplementary Table 1, Supplementary Figure 1).

To clone the genes encoding *Yr5*, *Yr7*, and *YrSP*, we identified susceptible ethyl methanesulfonate-derived (EMS) mutants from different genetic backgrounds carrying these genes (Figure 1, Supplementary Tables 2-3). We performed MutRenSeq[16] 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[15,17,18]. Their closest homologs in the Chinese Spring wheat genome sequence (RefSeq v1.0, https://wheat-urgi.versailles.inra.fr/Seq-Repository/Assemblies) 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 3). 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. 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.

We examined the allelic variation in *Yr7* and *Yr5/YrSP* across eight sequenced tetraploid and hexaploid wheat genomes (Supplementary Table 4). We identified *Yr7* only in Cadenza and Paragon, which are identical-by-descent in this interval (Supplementary File 3, Supplementary Table 5, and Supplementary Figure 4). 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 4). None of the three sequenced tetraploid accessions (Svevo, Kronos, Zavitan) carry *Yr7* (Supplementary Table 5).

For *Yr5/YrSP*, we identified three additional alleles in the sequenced hexaploid wheat cultivars (Figure 2b; Supplementary Table 6). Claire encodes a complete NLR with six amino-acid changes situated outside the three conserved domains (BED, NB-ARC, and LRRs) and six polymorphisms in the C-terminus compared to Yr5. 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 alleles were due to a 774 bp insertion in *Yr5*, close to the 3' end, which carries an alternate termination codon (Supplementary File 2). Tetraploid accessions 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 6). This truncated tetraploid allele is reminiscent of YrSP and is expressed in Kronos (see Methods). However, none of these varieties (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.

We designed diagnostic markers for *Yr5*, *YrSP*, and *Yr7* 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 7-8; Supplementary Figures 4-5). To further define their specificity, we tested the markers in a collection of global landraces[19] and European varieties[20] released over the past one hundred years. *Yr5* was only present in spelt cv. album, AvocetS-*Yr5*, and Lemhi-*Yr5* and was not detected in any other line (Supplementary Table 9) consistent with the fact that *Yr5* has not yet been deployed within European breeding programmes. *YrSP* was absent from the tested germplasm, except for AvocetS-*YrSP* (Supplementary Table 8). *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 7-8; Supplementary Figure 4).

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 file 4, Supplementary Figure 6). 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 7). 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 6) 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

143 structure of single BED domain genes, such as *Yr7* and *Yr5/YrSP* (BED_I encoded on exon
144 2). To our knowledge this is the first report of the double BED domain NLR protein
145 structure. The biological function of this molecular innovation remains to be determined,
146 although our data show that the single BED_I structure can confer PST resistance and is
147 required for *Yr7*-mediated resistance.
148
149 Among other mechanisms, integrated domains of NLRs are hypothesised to act as decoys for
150 pathogen effector targets[5]. This would suggest that the integrated domain might be sequence-
151 related to the host protein targeted by the effector. To identify these potential effector targets
152 in the host, we retrieved all BED-domain proteins (108) from the hexaploid wheat genome,
153 including 25 BED-NLRs, and additional BED-NLRs located in the syntenic intervals
154 (Supplementary Table 10; Supplementary file 4). We also retrieved the rice Xa1[10,11] and
155 ZBED proteins, the latter being hypothesized to mediate rice resistance to
156 *Magnaporthe oryzae*[7]. We used the split network method implemented in SplitsTree4[21] to
157 represent the relationships between these BED domains (Figure 3d; Supplementary Figure 8).
158 We found a major split in the network with almost all wheat non-NLR BED proteins (76 of
159 83; Figure 3d, black) clustering together, while the BED-NLRs proteins of wheat and other
160 analysed species clustered opposite (Figure 3d). This separation is consistent with the
161 hypothesis that integrated domains might have evolved to strengthen the interaction with the
162 effector after integration[22]. Among BED-NLRs, BED_I and BED_II constitute two major
163 clades, consistent with their relatively low amino acid conservation (Figure 3b), that are
164 comprised solely of genes from within the *Yr7/Yr5/YrSP* syntenic region. Seven non-NLR
165 BED domain wheat proteins clustered with BED-NLRs. These are most closely related to the
166 *Brachypodium* and rice BED-NLR proteins and were not expressed in RNA-Seq data from a
167 *Yr5* time-course (re-analysis of published data[23]; Supplementary Figure 9, Supplementary

168 Table 11). Similarly, no BED-containing protein was differentially expressed during this
169 infection time-course, consistent with the prediction that effectors alter their targets' activity
170 at the protein level in the integrated-decoy model[5]. We cannot however disprove that these
171 closely related BED-containing proteins are involved in BED-NLR-mediated resistance.
172
173 BED-NLRs are frequent in Triticeae, and occur in other monocot[8] and dicot tribes[7,24]. Only a
174 single BED-NLR gene, *Xa1*, was previously shown to confer resistance to plant
175 pathogens[10,11]. In the present study, we show that the distinct *Yr5*, *YrSP*, and *Yr7* resistance
176 specificities belong to a complex NLR cluster on chromosome 2B and are encoded by two
177 paralogous BED-NLRs genes. We report an allelic series for the *Yr5/YrSP* gene with five
178 independent alleles, including three full-length BED-NLRs (including *Yr5*) and two truncated
179 versions (including *YrSP*). This wider allelic series could be of functional significance as
180 previously shown for the *Mla* and *Pm3* loci that confer resistance to *Blumeria graminis*[25,26] in
181 barley and wheat, respectively, and the flax *L* locus conferring resistance to *Melampsora*
182 *lini*[27]. Overall, our results add strong evidence for the importance of the BED-NLR
183 architecture in plant-pathogen interactions. The paralogous and allelic relationship of these
184 three distinct *Yr* loci will inform future hypothesis-driven engineering of novel recognition
185 specificities.

Methods

MutRenSeq

Mutant identification

Supplementary Table 2 summarises plant materials and PST isolates used to identify mutants for each *Yr* gene. We used an EMS-mutagenised population in cultivar Cadenza[28] to identify mutants in *Yr7*; whereas EMS-populations in the corresponding AvocetS-*Yr* near isogenic line (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*, *Yr27*, *Yr32*, *YrRob*, and *YrSol*[29]. 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[30]. Identified susceptible lines were progeny tested to confirm the reliability of the phenotype and DNA from 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 E141 A+ (University of Sydney Plant Breeding Institute Culture no. 420), 150 E16 A+ (Culture no. 598) and 134 E16 A+ (Culture no. 572) were used to evaluate *Yr7*, *Yr5*, and *YrSP* mutants, respectively. EMS-derived susceptible mutants in Lemhi-*Yr5* were previously identified[31] and DNA from M_5 plants 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[32]. 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[33] 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 13). 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 3).

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 4, 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

261 above. This confirmed the candidate contig for *Yr7* as we retrieved the missing 5' region
262 including the BED domain. The improved contig now also contained a mutation in the
263 outstanding mutant line Cad0903 (Supplementary Figure 2). The Triticeae bait library does
264 not include integrated domains in its design so they are prone to be missed, especially when
265 located at the ends of an NLR. Sequencing technology could also have accounted for this:
266 MiSeq was used for Cadenza wild-type whereas HiSeq was chosen for Lemhi-*Yr5* and we
267 recovered the 5' region in the latter, although coverage was lower than for the regions
268 encoding canonical domains. In summary, we sequenced nine, ten, and four mutants for *Yr7*,
269 *Yr5*, and *YrSP*, respectively and identified for each target gene a single contig that accounted
270 for all mutants.
271
272 **Candidate contig confirmation and gene annotation**
273 We sequenced the *Yr5*, *Yr7*, and *YrSP* candidate contigs from the mutant lines (annotated in
274 Supplementary Files 1 and 2) to confirm the EMS-derived mutations using primers
275 documented in Supplementary Table 14. We first PCR-amplified the complete locus from the
276 same DNA preparations as the ones submitted for RenSeq with the Phusion® High-Fidelity
277 DNA Polymerase (New England Biolabs) following the suppliers protocol
278 (https://www.neb.com/protocols/0001/01/01/pcr-protocol-m0530). We then carried out
279 nested PCR on the obtained product to generate overlapping 600-1,000 bp amplicons that
280 were purified using the MiniElute kit (Qiagen). The purified PCR products were sequenced
281 by GATC following the LightRun protocol (https://www.gatc-biotech.com/shop/en/lightrun-
282 tube-barcode.html). Resulting sequences were aligned to the wild-type contig using
283 ClustalOmega (https://www.ebi.ac.uk/Tools/msa/clustalo/). This allowed us to curate the *Yr7*
284 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 HISATt2[34] (v2.1) to map RNA-Seq reads available from Cadenza and AvocetS-*Yr5*[23] 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[35] (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 3). 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).

Genetic linkage

We generated a set of F_2 populations to genetically map the candidate contigs (Supplementary Table 2). 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[36] and KASP assays were carried out as described in[37]. R/qtl package[38] 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[15,17,18]) 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[28] 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 14). 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 14). 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 4 and designed KASP primers targeting polymorphisms that were *Yr7*-specific. Three markers were retained after testing on a selected panel of Cadenza-derivatives and varieties that were positive for *Yr7* markers in the literature, including the *Yr7* reference

cultivar Lee (Supplementary Table 7, 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 varieties, as identified by the three KASP markers, displayed a consistent specificity (Supplementary Table 7). Pathology assays were performed as for the screening of the Cadenza mutant population. We retrieved pedigree information for the analysed varieties from the Genetic Resources Information System for Wheat and Triticale database (GRIS, www.wheatpedigree.net) and used the Helium software[39] (v1.17) to illustrate the breeding history of *Yr7* in the UK (Supplementary Figure 4).

We used the three *Yr7* KASP markers to genotype (i) varieties from the AHDB Wheat Recommended List from 2005-2018 (https://cereals.ahdb.org.uk/varieties/ahdb-recommended-lists.aspx); (ii) the Gediflux collection of European bread wheat varieties released between 1920 and 2010[20] and (iii) the core Watkins collection, which represents a global set of wheat landraces collected in the 1930s[19]. KASP assays were carried out as in[37] and results are reported in Supplementary Table 8.

_Yr5_ and _YrSP_ gene-specific markers

We identified a 774 bp insertion in the *Yr5* allele 29 bp upstream of the STOP codon with respect to the Cadenza and Claire alleles. Genomic DNA from *YrSP* confirmed that the insertion was specific to *Yr5*. We used this polymorphism to design primers flanking the insertion and tested them on a subset of the collections mentioned above. We added 32 DNA sample from diverse accessions of *Triticum dicoccoides*, the wild progenitor of domesticated

359 wheat (passport data shown in Supplementary Table 16). We included DNA from *Triticum*
360 *aestivum* ssp. *spelta* var. album[31] (*Yr5* donor) and Spaldings Prolific (*YrSP* donor) to assess
361 their amplification profiles. PCR amplification was conducted using a touchdown
362 programme: 10 cycles, -0.5 °C per cycle starting from 67 °C and the remaining 25 cycles at
363 62 °C. This allowed us to increase the specificity of the reaction. We observed three different
364 profiles on the tested varieties; (i) a 1,281 bp amplicon in *Yr5* positive cultivars, (ii) a 507 bp
365 amplicon in the alternate *Yr5* allele carriers, including AvocetS-*YrSP*, Cadenza, and Claire,
366 and (iii) no amplification in other varieties. We sequenced the different amplicons and
367 confirmed the insertion in *Yr5* compared to the alternate alleles (Supplementary File 2). The
368 lack of amplicons in some varieties most likely respresents the absence of the loci in the
369 tested varieties. For *YrSP*, we aligned the *YrSP* and *Yr5* sequences to design KASP primers
370 targeting the G to C SNP between the two alleles (Supplementary File 2, Supplementary
371 Table 15). We tested the marker by genotyping selected varieties as controls and varieties
372 from the AHDB Wheat Recommended List from 2005-2018 (Supplementary Table 8).
373
374 **_In silico_ allele mining for _Yr7_ and _Yr5_**
375 We used the *Yr7* and *Yr5* sequences to retrieve the best BLAST hits in the *T. aestivum* and *T.*
376 *turgdium* wheat genomes listed in Supplementary Table 4. The best *Yr5* hits shared between
377 93.6 and 99.3% sequence identity, which was comparable to what was observed for alleles
378 derived from the wheat *Pm3* (>97% identity)[26] and flax *L* (>90% identity)[27] genes. *Yr7* was
379 identified only in Paragon and Cadenza (Supplementary Table 5; See Supplementary File 3
380 for curation of the Paragon sequence).
381
382 **Analysis of the _Yr7_ and _Yr5_/_YrSP_ cluster on RefSeq v1.0**
383 *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 4 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

We aligned the translated NB-ARC domains from the NLR-Annotator output with MUSCLE[40] using default parameters (v.3.8.31). We verified and manually curated the alignment with Jalview[41] (v2.10.1). We used Gblocks[42] (v0.91b) with the following parameters: Minimum Number Of Sequences For A Conserved Position: 9; Minimum Number Of Sequences For A Flanking Position: 14; Maximum Number Of Contiguous Nonconserved Positions: 8; Minimum Length Of A Block: 10; Allowed Gap Positions: None;

Use Similarity Matrices: Yes; to eliminate poorly aligned positions. This resulted in 36% of the original 156 positions being taken forward for the phylogeny. We built a Maximum Likelihood tree with the RAxML[43] program and the following parameters: raxmlHPC -f a -x 12345 -p 12345 -N 1000 -m PROTCATJTT -s <input_alignment.fasta> (MPI version v8.2.10). The best scoring tree with associated bootstrap values was visualised and mid-rooted with Dendroscope[44] (v3.5.9). There was clear separation between NLRs belonging to the two different clusters but the sub-clades had less support. One explanation would be that conflicting phylogenetic signals that are due to events such as hybridization, horizontal gene transfer, recombination, or gene duplication and loss might have occurred in the region. Split networks allow nodes that do not represent ancestral species and can thus represent such incompatible and ambiguous signals. We therefore used this method in the following part of the analysis to analyse the relationship between the BED domains.

We used the Neighbour-net method[45] implemented in SplitsTree4[21] (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[46] 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*[23]. We used normalised read counts (Transcript Per Million, TPM) from Ramirez-Gonzalez et al. 2018 to produce the heatmap shown in Supplementary Figure 9 with the pheatmap R package[47] (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[48] (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 ~ Variety + Time + Variety:Time to the reduced model ~ Variety + Time to identify genes that were differentially expressed between the two varieties 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 varieties, 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. Steuernagel, B. *et al.* Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture. *Nat. Biotechnol.* **34,** 652–655 (2016).

17. 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).

18. 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).

19. 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).

20. 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).

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

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

23. 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).

24. Germain, H. & Séguin, A. Innate immunity: has poplar made its BED? *New Phytol.* **189,** 678–687 (2011).

25. 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).

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

27. 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).

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

29. Hubbard, A. J., Fanstone, V. & Bayles, R. A. *UKCPVS 2009 Annual report.* https://cereals.ahdb.org.uk/media/1131303/Annual-Report-UKCPVS-2009.pdf

30. Gassner, G. & Straib, W. *Die Bestimmung der biologischen Rassen des*

Weizengelbrostes Puccinia glumarum f.sp. tritici *Schmidt Erikss. u. Henn.* (1932).

31. 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).

32. 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).

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

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

35. 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).

36. 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).

37. 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).

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

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

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

41. 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).

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

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

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

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

46. Pearce, S. *et al.* Regulation of Zn and Fe transporters by the *GPC 1*gene during early wheat monocarpic senescence. *BMC Plant Biol.* **14,** 368 (2014).

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

48. 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).

49. Jupe, F. *et al.* Identification and localisation of the NB-LRR gene family within the

potato genome. *BMC Genomics* **13,** 75 (2012).

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

51. 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 *Yr5*, *Yr7*, 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 12 (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 having providing us with pre-publication access to the RefSeq v1.0 assembly and gene annotation. 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. This research was supported by the NBI Computing Infrastructure for Science (CiS) group in Norwich, UK.

Figure legends

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

a, Left: Wild-type and selected EMS-derived susceptible mutant lines for *Yr7*, *Yr5*, and *YrSP* (Supplementary Table 2 and 3) inoculated with PST isolate 08/21 (*Yr7*), PST 150 E16 A+ (*Yr5*), or PST 134 E16 A+ (*YrSP*). Right: Candidate gene structures, with mutations in red, and their predicted effects on the translated protein. **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 lines underneath the chromosme. *Yr* loci mapping intervals are defined by the red horizontal lines. 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 6), 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 6, 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 8 includes individual labels.

Supplementary Figure 1: Deployment of *Yr7* varieties 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 1 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, colored rectangles depict the motifs identified by NLR-Annotator (each motif is specific to a conserved NLR domain[49]), 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 14). 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: Pedigrees of selected Thatcher-derived varieties and their *Yr7* allelic status.

Pedigree tree of Thatcher-derived varieties where each circle represents a variety 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. Varieties 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 variety is related to a parent that was also positive for *Yr7*. Figure was generated using the Helium software[39] (v1.17).

Supplementary Figure 5: 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 6: Expansion of BED-NLRs in the Triticeae and presence of conserved BED-BED-NLRs across 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 7. Black lines represent phylogenetically related single NLRs located between the two NLR clusters illustrated in Supplementary Figure 8. Details of genes are reported in Supplementary File 4.

Supplementary Figure 7: 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 8: 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 6, 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: BED-NLRs and BED-containing proteins are not differentially expressed in yellow rust-infected susceptible and resistant varieties.

Heatmap representing the normalised read counts (Transcript Per Million, TPM) from the reanalysis of published RNAseq data[23] 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 8 are indicated by black triangles.

Supplementary Table 1: Harvested weight of known *Yr7* varieties from 1990 to 2016 and *virYr7* prevalence among UK PST isolates.

762 Proportion of harvested *Yr7* wheat varieties in the UK from 1990 to 2016. The prevalence of
763 yellow rust isolates virulent to *Yr7* across this time period is shown in the top row. Original
764 data from NIAB-TAG Seedstats journal (NIAB-TAG Network) and the UK Cereal Pathogen
765 Virulence Survey (http://www.niab.com/pages/id/316/UKCPVS).
766
767 **Supplementary Table 2: Plant materials analysed for the present study with the**
768 **different PST isolates used for the pathology assays.**
769
770 **Supplementary Table 3: Plant material submitted for Resistance gene enrichment**
771 **Sequencing (RenSeq).**
772 From left to right: Mutant line identifier, targeted gene, score when infected with PST
773 according to the Grassner and Straib scale, mutation position, coverage of the mutation (at
774 least 99% of the reads supported the mutant base in the mutant reads), predicted effect of the
775 mutation on the protein sequence, comments. Lines with the same mutations are highlighted
776 with matching colours.
777
778 **Supplementary Table 4: Genome assemblies used in the present study.**
779 Summary of the available genome assemblies[50,51] that were used for the *in silico* allele
780 mining and synteny analysis across rice, *Brachypodium*, barley and different Triticeae
781 accessions.
782
783 **Supplementary Table 5: *In silico* allele mining for *Yr7* and *Yr5*/*YrSP* in available**
784 **genome assemblies for wheat.**
785 Table presents the percentage identity (% ID) of the identified alleles and matching colours
786 illustrate identical haplotypes. Investigated genome assemblies are shown in Supplementary
787 Table 4.
788
789 **Supplementary Table 6: Polymorphisms between Yr5 protein and its identified alleles**.
790 Positions of the polymorphic amino acids across the five Yr5/YrSP proteins. Polymorphisms
791 falling into the BED and NB-ARC domains are shown in red and blue, respectively.
792
793 **Supplementary Table 7: Presence/absence of *Yr7* alleles in a selected panel of Cadenza-**
794 **derivatives and associated responses to different PST isolates (avirulent to *Yr7*: PST**
795 **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 varieties might carry other resistance genes that can lead to a 1 – 1 – 1 profile (e.g. Cadenza). Varieties 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 varieties (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 8: 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 varieties 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 varieties (1920-2010)[20], (iv) a core set of the Watkins collection, which represent a set of global bread wheat landraces collected in the 1920-30s[19]. Most of the putative *Yr7* carriers were positive for all the *Yr7* markers apart from Aztec, Chablis and Cranbrook. Chablis was susceptible to the PST isolates that were avirulent to *Yr7* so it probably does not carry the gene. Regarding the 2005-2018-UK_RL results were consistent across already tested varieties: Cadenza, Cordiale, Cubanita, Grafton and Skyfall were already positive in Supplementary Table 7. Energise, Freiston, Gallant, Oakley and Revelation were negative on both panels as well. Results were thus consistent across different sources of DNA. *Yr7*-containing varieties are not prevalent in the 2005-2018 Recommended List set, however, this gene is present in Skyfall, which is currently one of the most harvested varieties in the UK (Supplementary Table 1). 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 1: *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 Tritical - 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 varieties.

Supplementary Table 9: **Presence/absence of *Yr5* alleles in a subset of previously studied collections.**

A subset of the aforementioned collection was investigated for the *Yr5* presence. "Yes" in the *Yr5* column refers to amplification of the 1,281 bp amplicon with the *Yr5*-Insertion primers (Supplementary Figure 5). "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 8. "Yes" in the no amplification column refers to identification of a profile similar to the one found for AvocetS in Supplementary Figure 5.

Supplementary Table 10: 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 7.

Supplementary Table 11: Transcripts per Million-normalised read counts from the re-analysis of published RNA-Seq data[23] and associated differential expression analysis performed with DESeq2.

Supplementary Table 12: Sequencing details of RenSeq data generated in this study.

Supplementary Table 13: *De novo* assemblies generated from the corresponding RenSeq data.

Supplementary Table 14: 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 15: Diagnostic markers for *Yr5*, *Yr7*, and *YrSP*.

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

Supplementary Table 16: 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 14). 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 14). 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.**

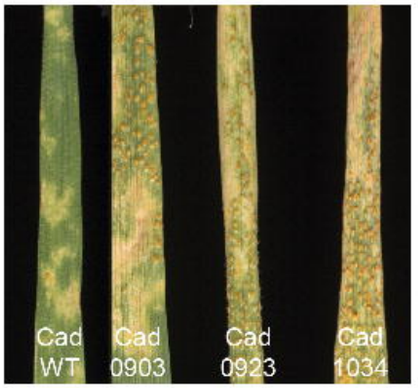
Comments show the position of the unknown bases ("N") in the "Yr7_with_Ns" sequence. Curation based on Sanger sequencing data is shown in bold black in the "curated_Yr7" sequence with the 39 bp insertion and 129 bp deletion. Allele mining for *Yr7* in the Paragon assembly showed that a similar assembly 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 4). 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

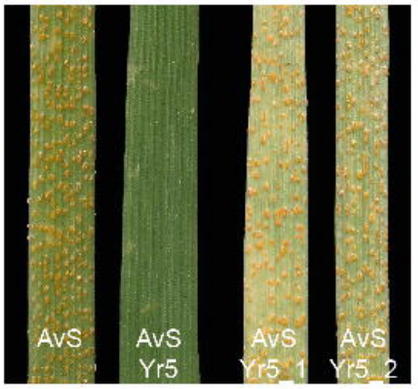
mapped to an additional Cadenza NLR. This provides evidence for the presence of the identical gene in Paragon which is supported by phenotypic data.

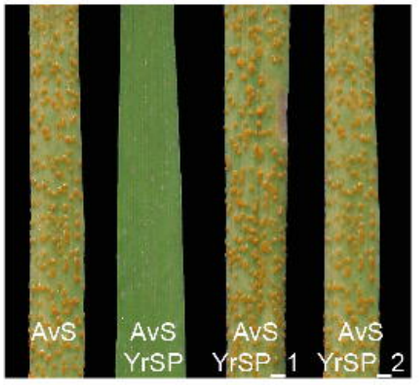
Supplementary File 4: Syntenic region across different grasses (Supplementary Table 4) and the NLR loci identified with NLR-Annotator. See Methods for a detailed explanation of the analysis and Supplementary Figure 6 for an illustration.

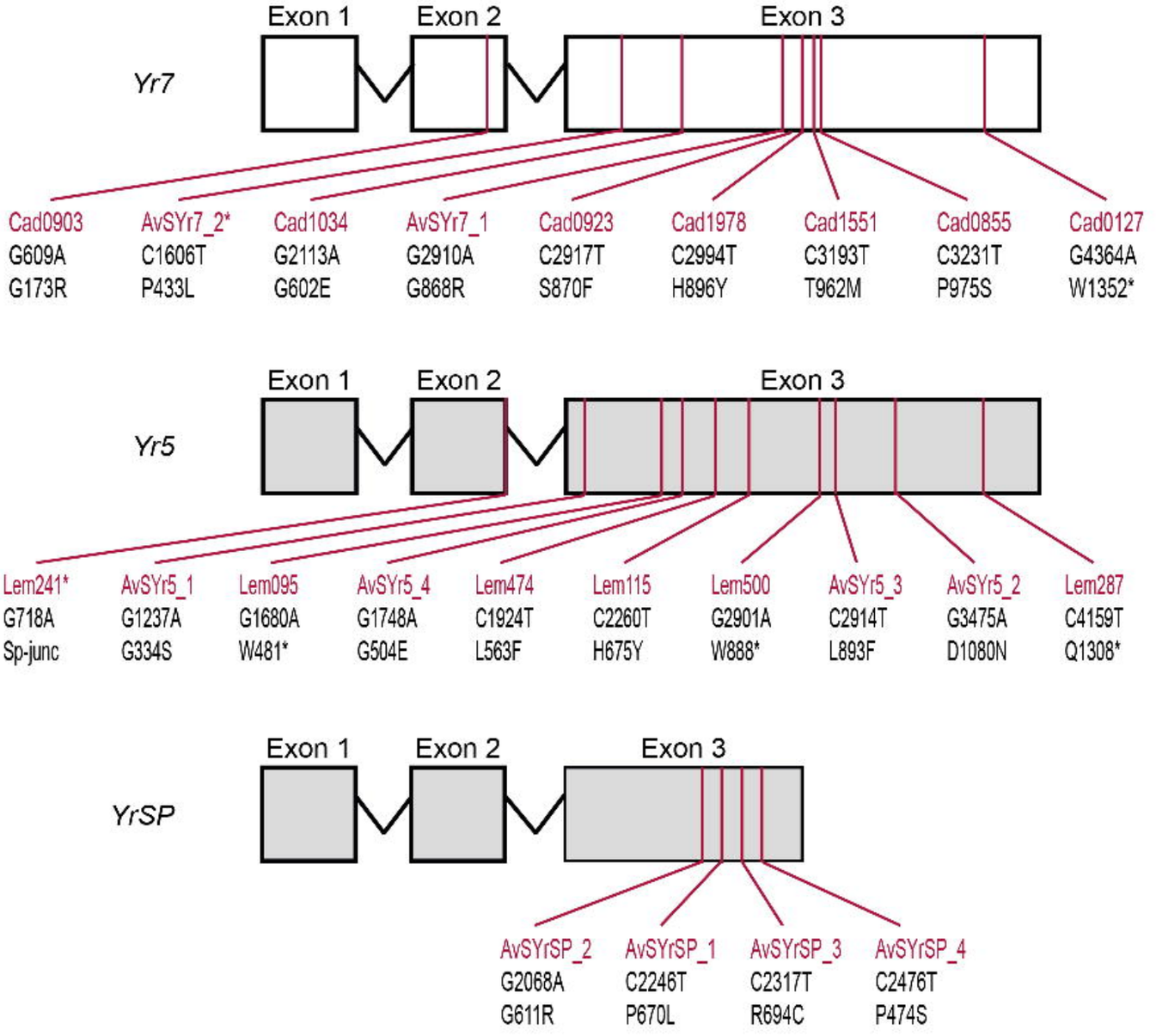
Supplementary File 5: Curated sequences of BED-NLRs from chromosome 2B and Ta_2D7. Exons are highlighted with different colours (yellow, green, blue, pink). Amino acids encoding the BED domain are shown in red and those encoding the NB-ARC domain are in blue. LRR repeats identified with NLR Annotator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

a

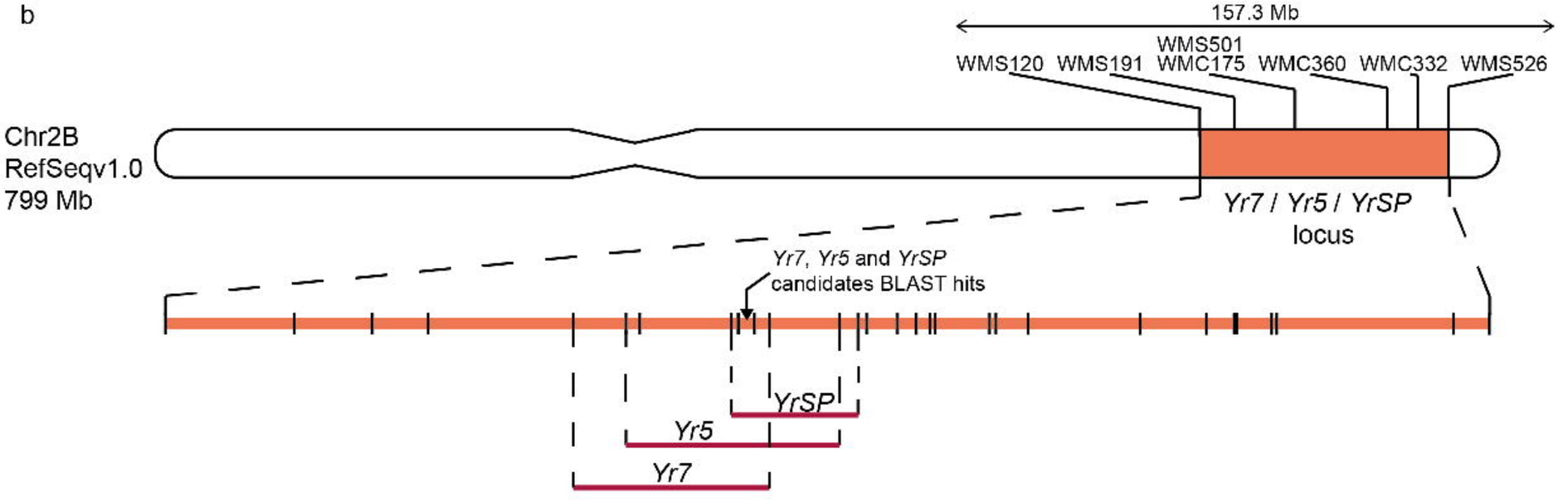

Cad WT
Cad 0903
Cad 0923
Cad 1034

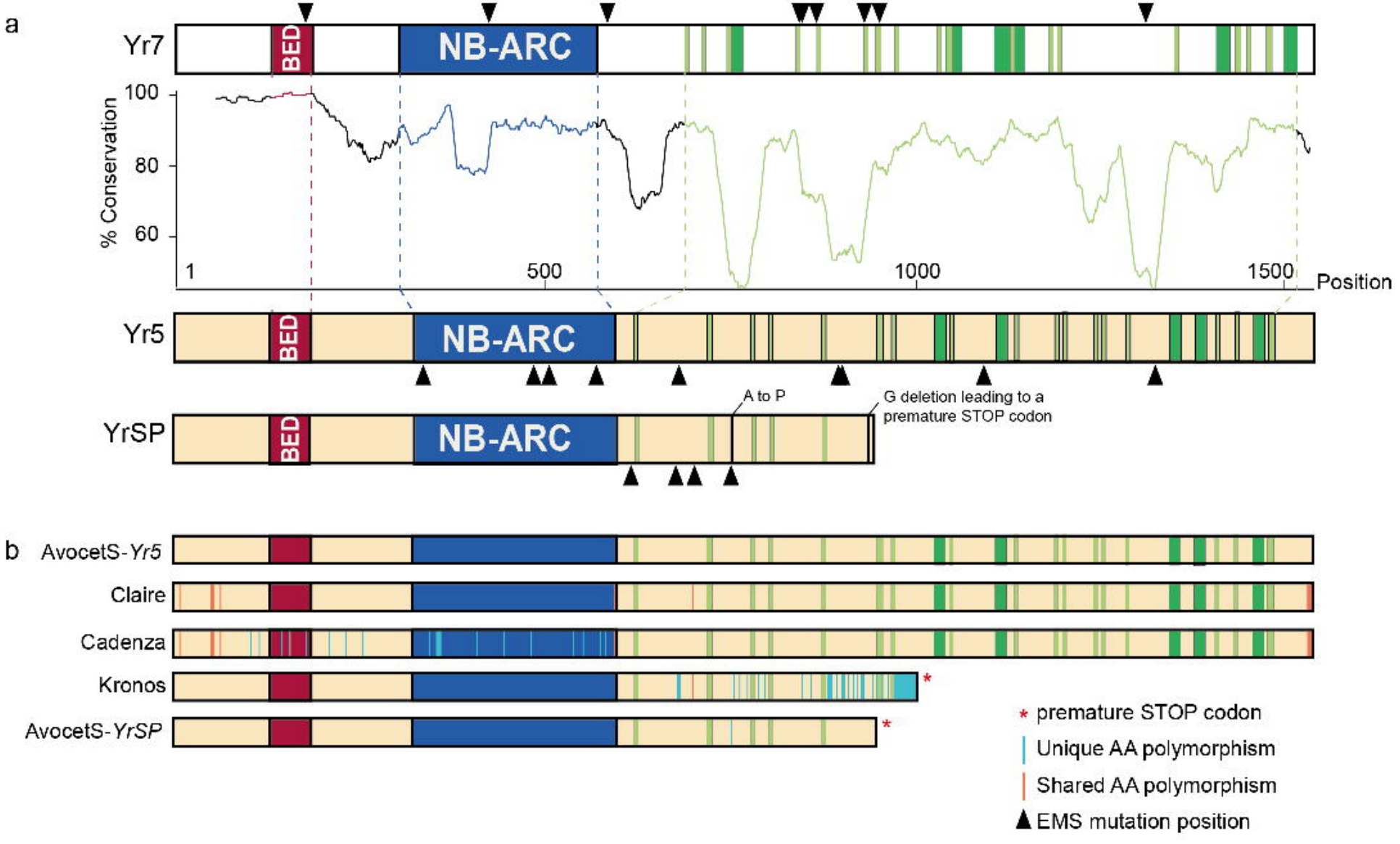

AvS
AvS Yr5
AvS Yr5_1
AvS Yr5_2


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

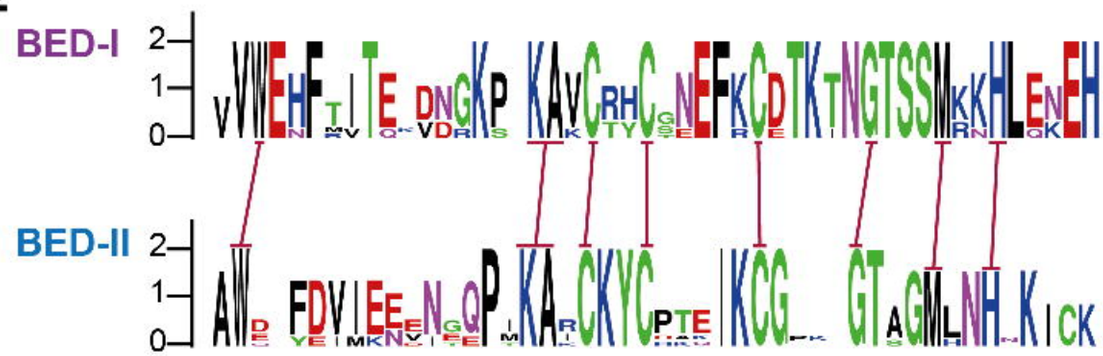

157.3 Mb
WMS501
WMS120
WMS191
WMC175
WMC360
WMC332
WMS526
Chr2B
RefSeqv1.0
799 Mb
Yr7 / Yr5 / YrSP
locus
Yr7, Yr5 and YrSP
candidates BLAST hits
YrSP
Yr5
Yr7


a
Yr7
BED
NB-ARC
100
80
60
% Conservation
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

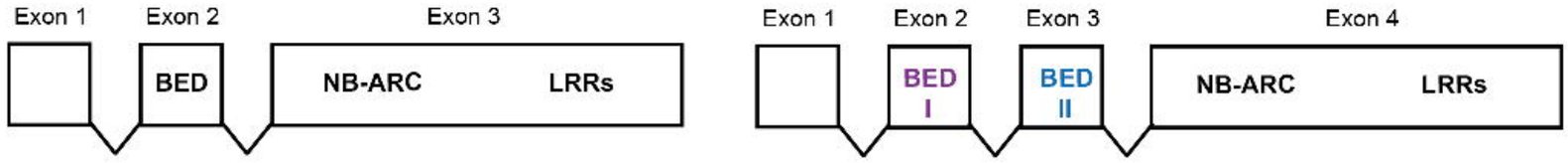
a

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

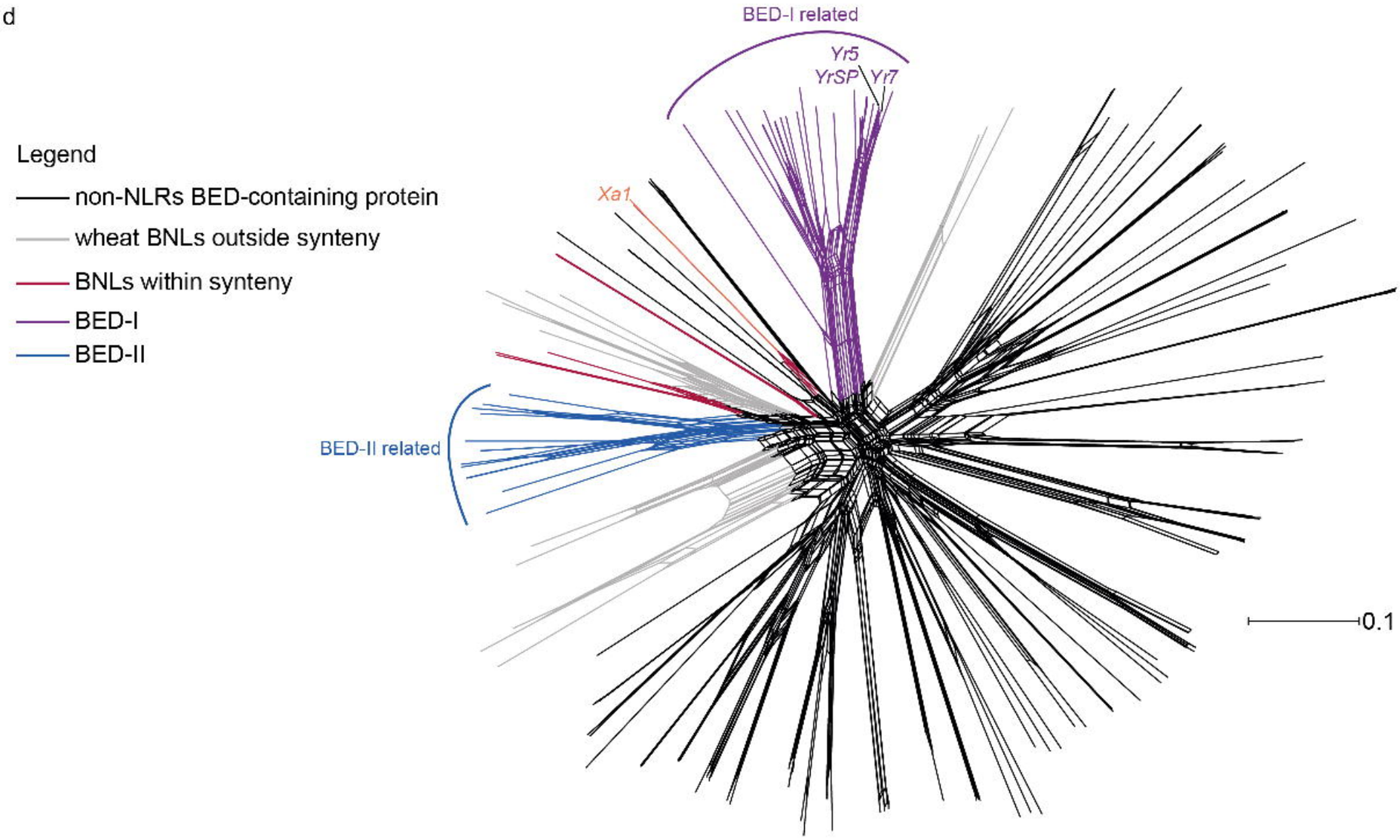
b

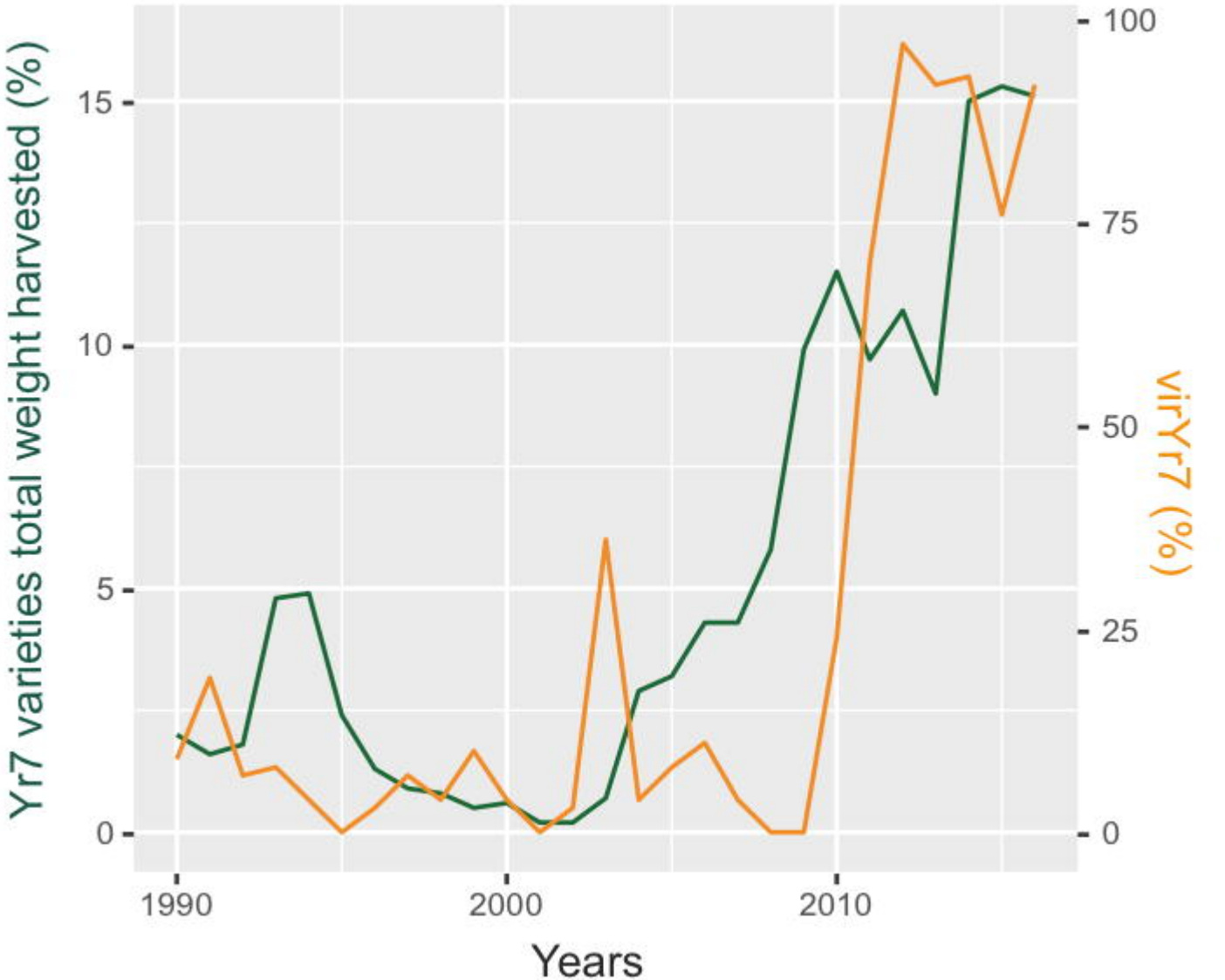


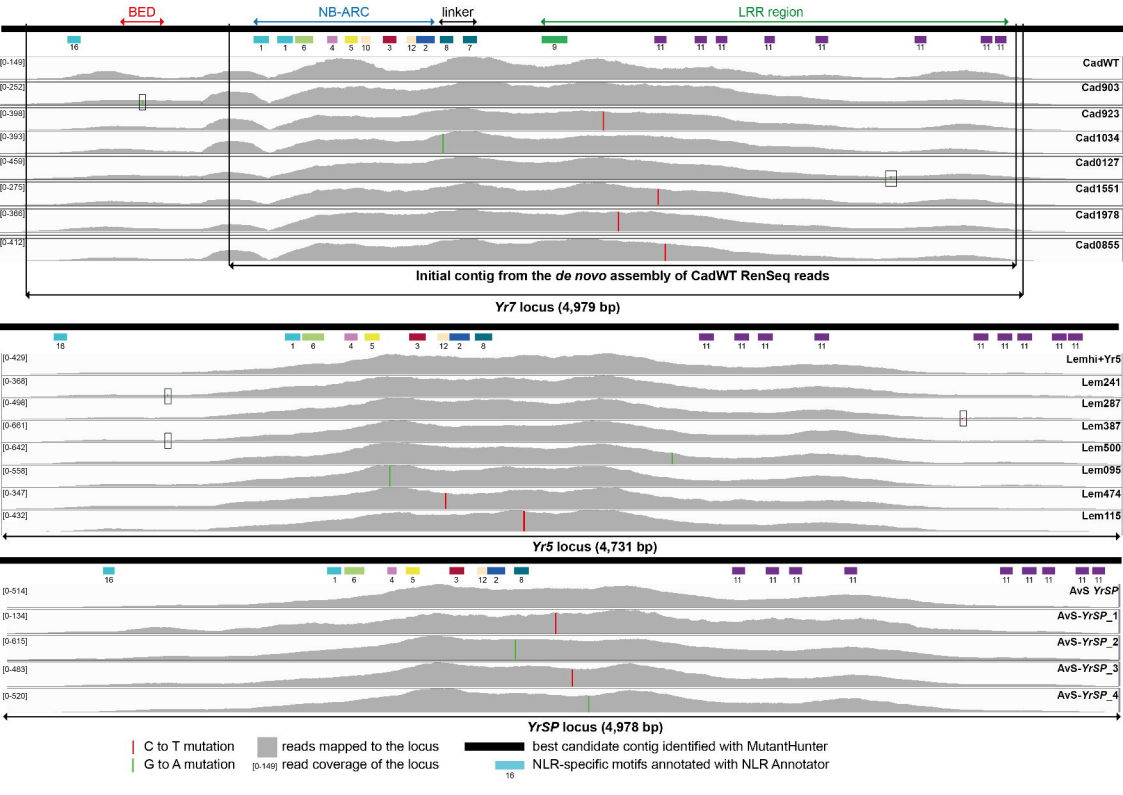
c

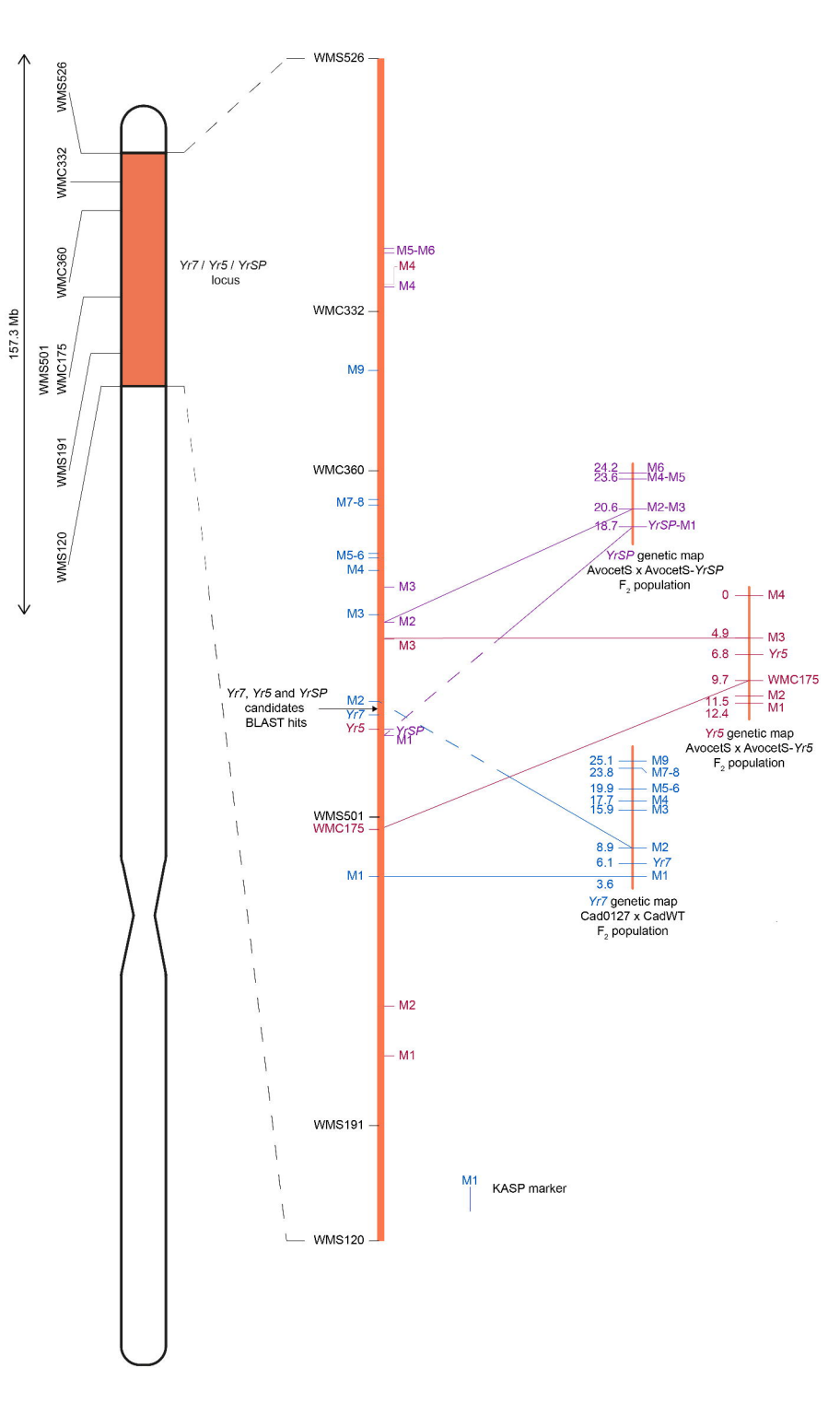


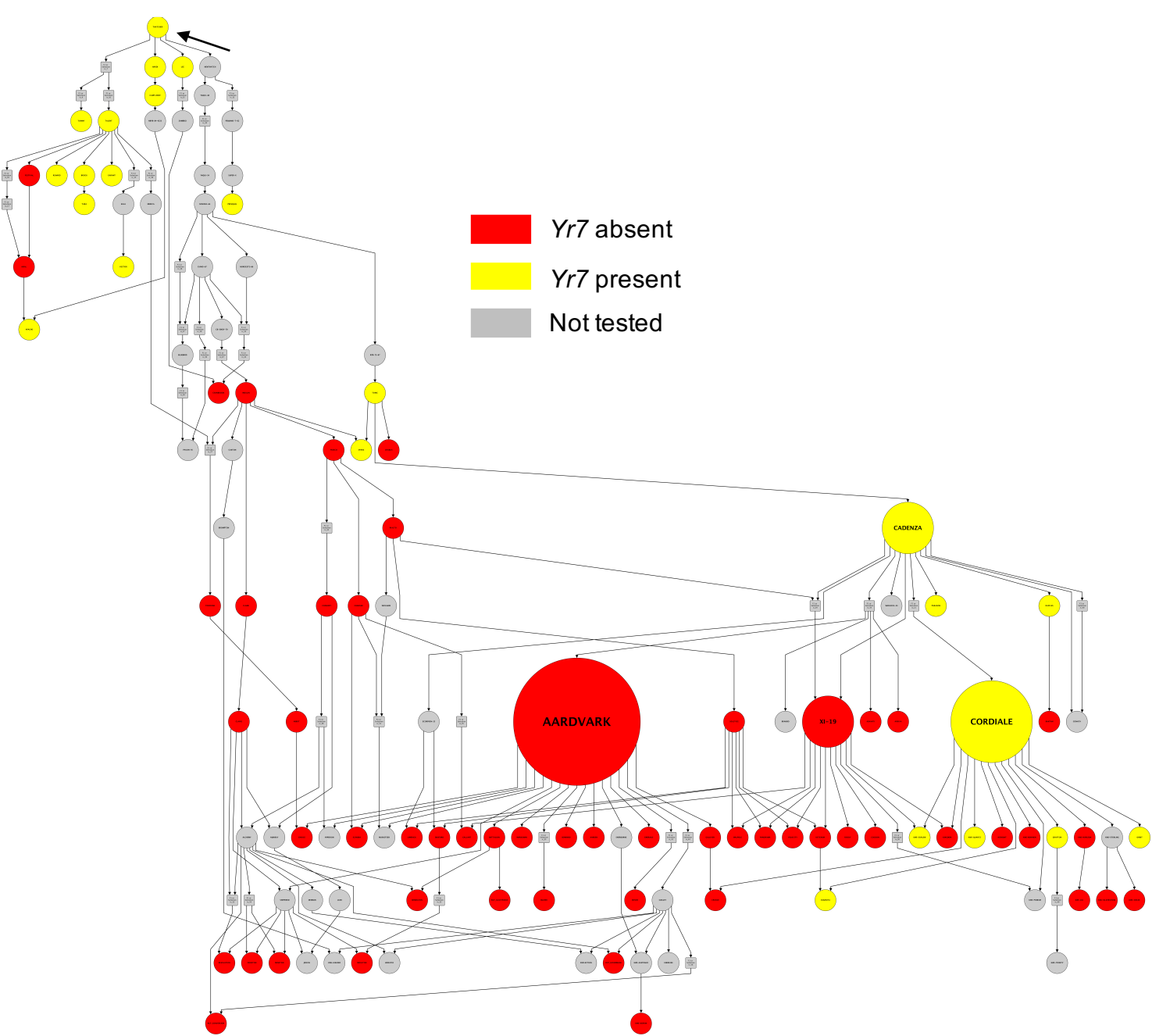
d

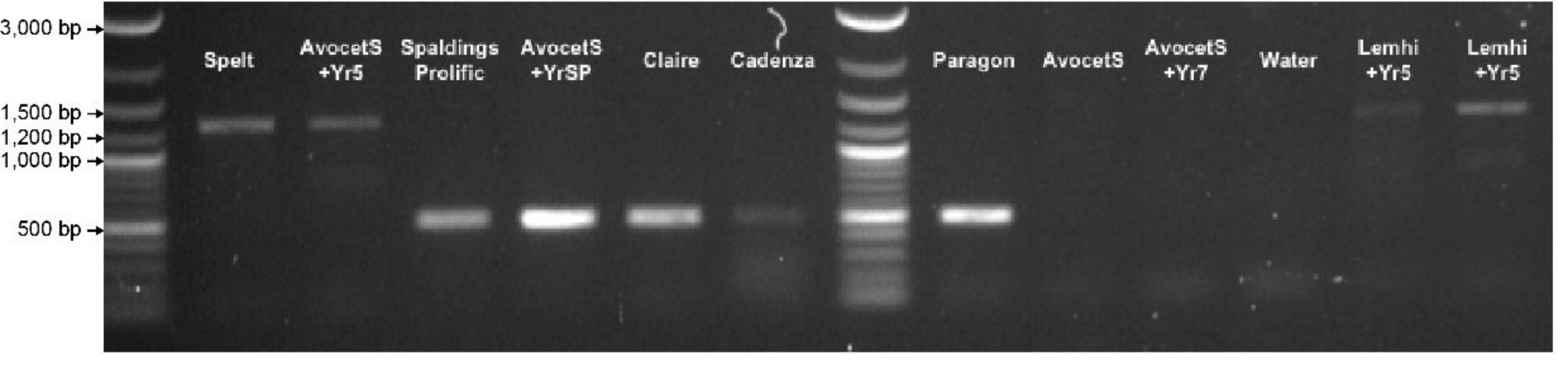


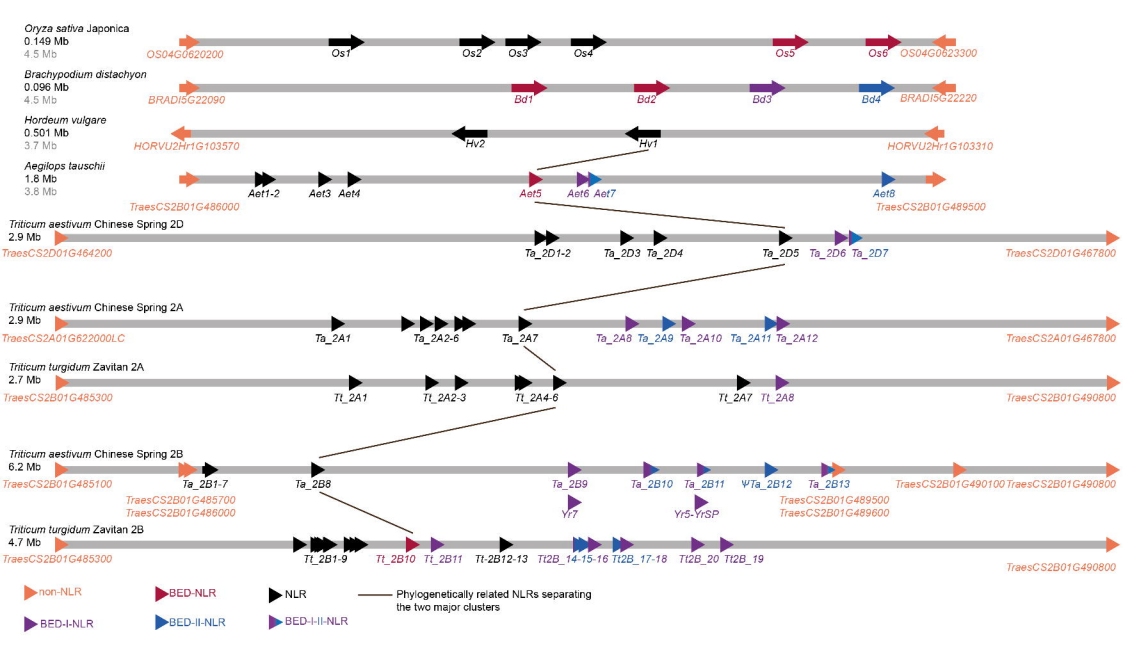

Yr7 varieties total weight harvested (%)
15
10
5
0
100
75
50
25
0
VirYr7 (%)
1990
2000
2010
Years

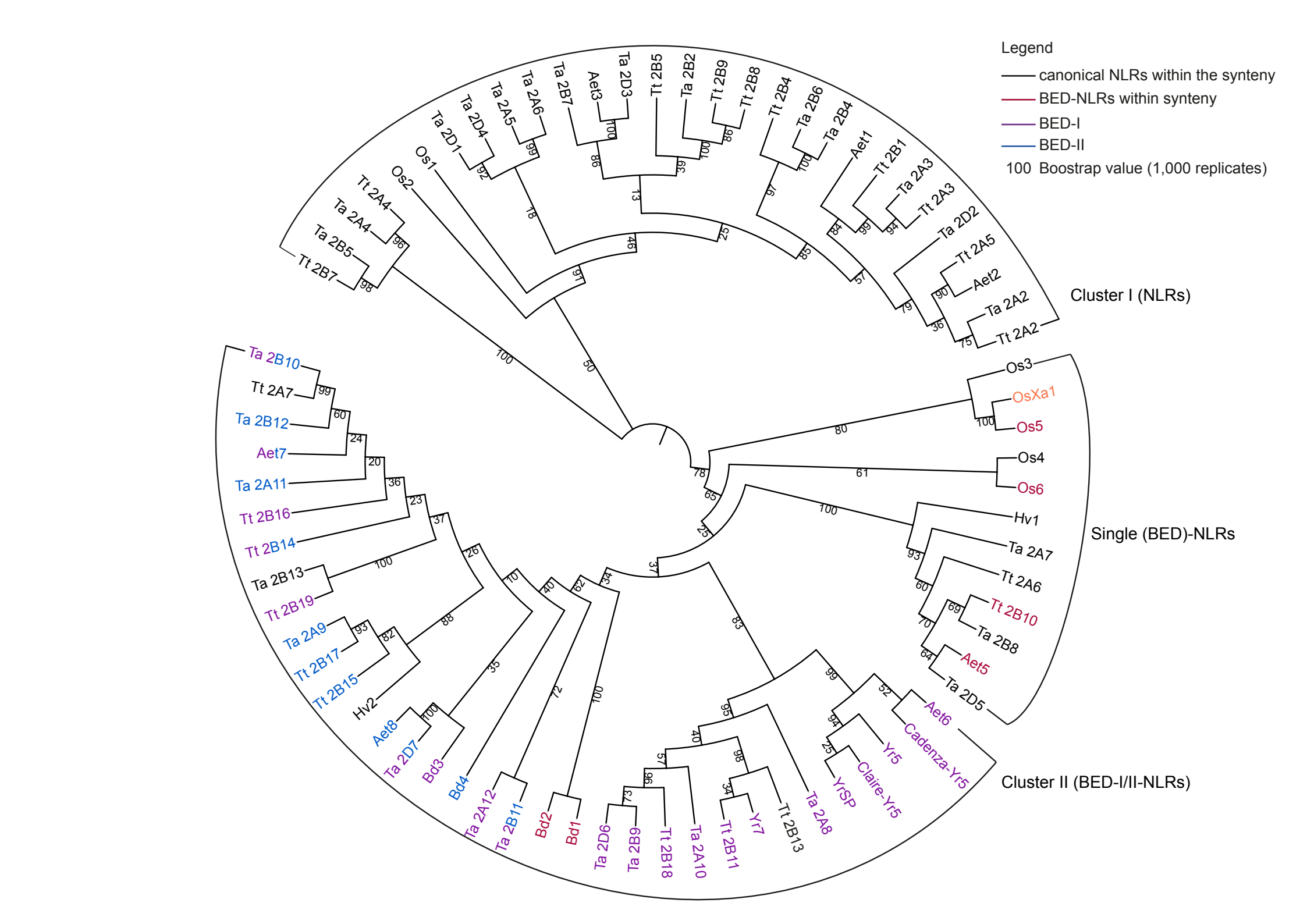

BED
NB-ARC
linker
LRR region
CadWT
Cad903
Cad923
Cad1034
Cad0127
Cad1551
Cad1978
Cad0855
Initial contig from the de novo assembly of CadWT RenSeq reads
Yr7 locus (4,979 bp)
Lemhi+Yr5
Lem241
Lem287
Lem387
Lem500
Lem095
Lem474
Lem115
Yr5 locus (4,731 bp)
AvS YrSP
AvS-YrSP_1
AvS-YrSP_2
AvS-YrSP_3
AvS-YrSP_4
YrSP locus (4,978 bp)
C to T mutation
G to A mutation
reads mapped to the locus
[0-149] read coverage of the locus
best candidate contig identified with MutantHunter
NLR-specific motifs annotated with NLR Annotator

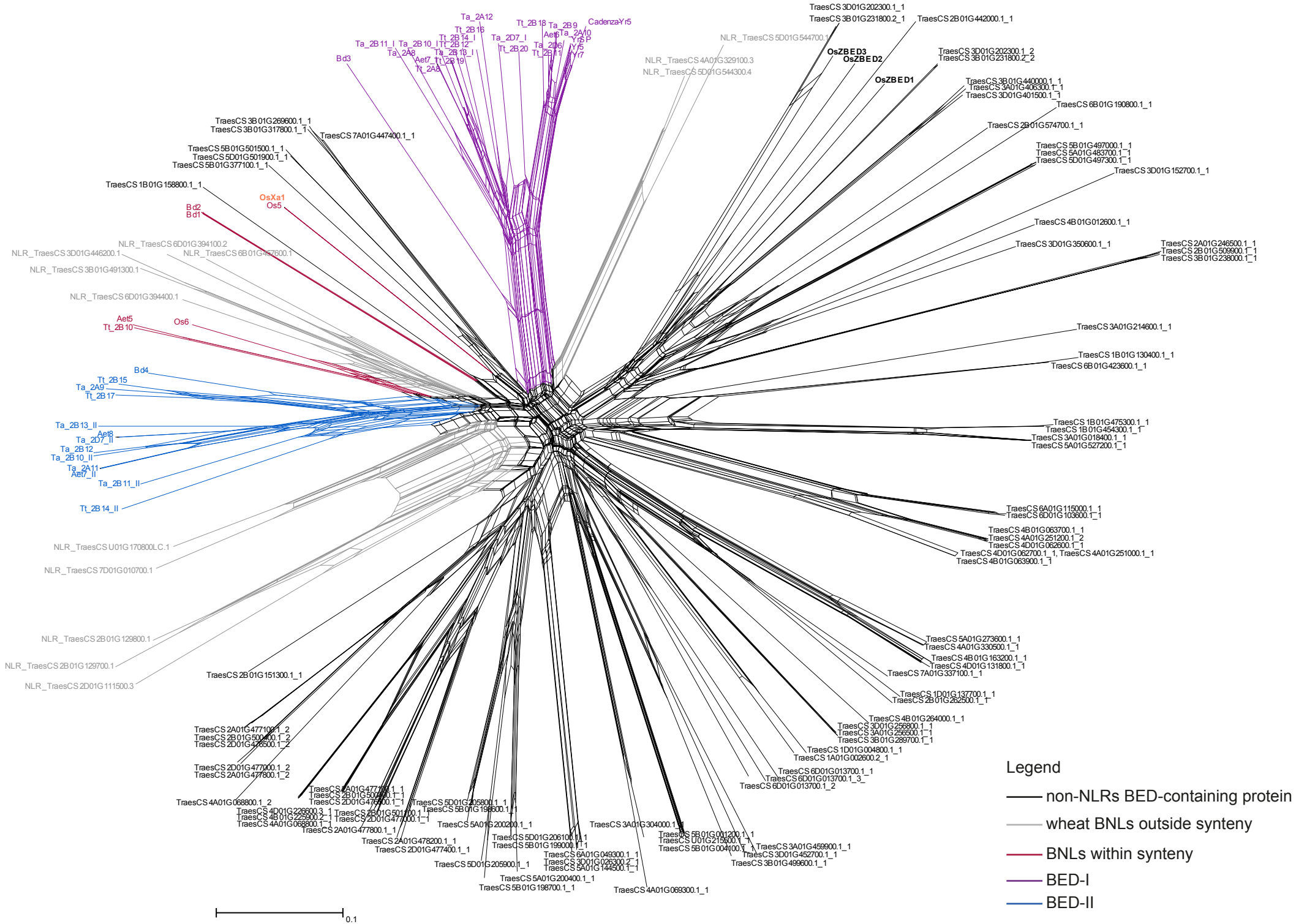

157.3 Mb
WMS526
WMC332
WMC360
WMS501
WMC175
WMS191
WMS120
Yr7 / Yr5 / YrSP
locus
WMS526
M5-M6
M4
M4
WMC332
M9
WMC360
M7-8
M5-6
M4
M3
M3
M2
M3
M2
Yr7
Yr5
YrSP
M1
WMS501
WMC175
M1
M2
M1
WMS191
WMS120
Yr7, Yr5 and YrSP
candidates
BLAST hits
24.2
23.6
M6
M4-M5
20.6
18.7
M2-M3
YrSP-M1
YrSP genetic map
AvocetS x AvocetS-YrSP
F2 population
0
M4
4.9
M3
6.8
Yr5
9.7
WMC175
11.5
M2
12.4
M1
Yr5 genetic map
AvocetS x AvocetS-Yr5
F2 population
25.1
23.8
M9
M7-8
19.9
17.7
15.9
M5-6
M4
M3
8.9
6.1
3.6
M2
Yr7
M1
Yr7 genetic map
Cad0127 x CadWT
F2 population
M1
KASP marker


Yr7 absent
Yr7 present
Not tested
CADENZA
AARDVARK
XI-19
CORDIALE


3,000 bp
1,500 bp
1,200 bp
1,000 bp
500 bp
Spelt
AvocetS +Yr5
Spaldings Prolific
AvocetS +YrSP
Claire
Cadenza
Paragon
AvocetS
AvocetS +Yr7
Water
Lemhi +Yr5
Lemhi +Yr5


Oryza sativa Japonica
0.149 Mb
4.5 Mb
OS04G0620200
Os1
Os2
Os3
Os4
Os5
Os6
OS04G0623300
Brachypodium distachyon
0.096 Mb
4.5 Mb
BRADI5G22090
Bd1
Bd2
Bd3
Bd4
BRADI5G22220
Hordeum vulgare
0.501 Mb
3.7 Mb
HORVU2Hr1G103570
Hv2
Hv1
HORVU2Hr1G103310
Aegilops tauschii
1.8 Mb
3.8 Mb
TraesCS2B01G486000
Aet1-2
Aet3
Aet4
Aet5
Aet6
Aet7
Aet8
TraesCS2B01G489500
Triticum aestivum Chinese Spring 2D
2.9 Mb
TraesCS2D01G464200
Ta_2D1-2
Ta_2D3
Ta_2D4
Ta_2D5
Ta_2D6
Ta_2D7
TraesCS2D01G467800
Triticum aestivum Chinese Spring 2A
2.9 Mb
TraesCS2A01G622000LC
Ta_2A1
Ta_2A2-6
Ta_2A7
Ta_2A8
Ta_2A9
Ta_2A10
Ta_2A11
Ta_2A12
TraesCS2A01G467800
Triticum turgidum Zavitan 2A
2.7 Mb
TraesCS2B01G485300
Tt_2A1
Tt_2A2-3
Tt_2A4-6
Tt_2A7
Tt_2A8
TraesCS2B01G490800
Triticum aestivum Chinese Spring 2B
6.2 Mb
TraesCS2B01G485100
Ta_2B1-7
Ta_2B8
Ta_2B9
Ta_2B10
Ta_2B11
ΨTa_2B12
Ta_2B13
TraesCS2B01G490100
TraesCS2B01G490800
TraesCS2B01G485700
TraesCS2B01G486000
Yr7
Yr5-YrSP
TraesCS2B01G489500
TraesCS2B01G489600
Triticum turgidum Zavitan 2B
4.7 Mb
TraesCS2B01G485300
Tt_2B1-9
Tt_2B10
Tt_2B11
Tt-2B12-13
Tt2B_14-15-16
Tt2B_17-18
Tt2B_20
Tt2B_19
TraesCS2B01G490800
non-NLR
BED-NLR
NLR
Phylogenetically related NLRs separating the two major clusters
BED-I-NLR
BED-II-NLR
BED-I-II-NLR


Legend
canonical NLRs within the synteny
BED-NLRs within synteny
BED-I
BED-II
100 Boostrap value (1,000 replicates)
Cluster I (NLRs)
Single (BED)-NLRs
Cluster II (BED-I/II-NLRs)
Ta 2B10
Tt 2A7
Ta 2B12
Aet7
Ta 2A11
Tt 2B16
Tt 2B14
Ta 2B13
Tt 2B19
Ta 2A9
Tt 2B17
Tt 2B15
Hv2
Aet8
Ta 2D7
Bd3
Bd4
Ta 2A12
Ta 2B11
Bd2
Bd1
Ta 2D6
Ta 2B9
Tt 2B18
Ta 2A10
Tt 2B11
Yr7
Tt 2B13
Ta 2A8
YrSP
Claire-Yr5
Yr5
Cadenza-Yr5
Aet6
Ta 2D5
Aet5
Ta 2B8
Tt 2B10
Tt 2A6
Ta 2A7
Hv1
Os6
Os4
Os5
OsXa1
Os3
Tt 2B7
Ta 2B5
Ta 2A4
Tt 2A4
Os2
Os1
Ta 2D1
Ta 2D4
Ta 2A5
Ta 2A6
Ta 2B7
Aet3
Ta 2D3
Tt 2B5
Ta 2B2
Tt 2B9
Tt 2B8
Tt 2B4
Ta 2B6
Ta 2B4
Aet1
Tt 2B1
Ta 2A3
Tt 2A3
Ta 2D2
Tt 2A5
Aet2
Ta 2A2
Tt 2A2


Ta_2A12
Tt_2B18
Ta_2B9
CadenzaYr5
Tt_2B16
Aet6
Ta_2A10
Ta_2B11_I
Ta_2B10_I
Tt_2B12
Tt_2B14_I
Ta_2D7_I
Yr5P
Ta_2A8
Tt_2B18_I
Ta_2D6
Yr6
Aet7
Tt_2B19
Tt_2B20
Tt_2B11
Yr7
Tt_2A5
Bd3
NLR_TraesCS5D01G544700.1
NLR_TraesCS4A01G329100.3
NLR_TraesCS5D01G544300.4
TraesCS3D01G202300.1_1
TraesCS3B01G231800.2_1
TraesCS2B01G442000.1_1
OsZBED3
OsZBED2
OsZBED1
TraesCS3D01G202300.1_2
TraesCS3B01G231800.2_2
TraesCS3B01G440000.1_1
TraesCS3A01G406300.1_1
TraesCS3D01G401500.1_1
TraesCS6B01G190800.1_1
TraesCS3B01G269600.1_1
TraesCS3B01G317800.1_1
TraesCS7A01G447400.1_1
TraesCS2B01G574700.1_1
TraesCS5B01G501500.1_1
TraesCS5D01G501900.1_1
TraesCS5B01G377100.1_1
TraesCS5B01G497000.1_1
TraesCS5A01G483700.1_1
TraesCS5D01G497300.1_1
TraesCS3D01G152700.1_1
TraesCS1B01G158800.1_1
OsXa1
Os5
Bd2
Bd1
TraesCS4B01G012600.1_1
NLR_TraesCS6D01G394100.2
NLR_TraesCS3D01G446200.1
NLR_TraesCS6B01G457600.1
TraesCS3D01G350600.1_1
TraesCS2A01G246500.1_1
TraesCS2B01G509900.1_1
TraesCS3B01G238000.1_1
NLR_TraesCS3B01G491300.1
NLR_TraesCS6D01G394400.1
Aet5
Tt_2B10
Os6
TraesCS3A01G214600.1_1
TraesCS1B01G130400.1_1
TraesCS6B01G423600.1_1
Bd4
Tt_2B15
Ta_2A9
Tt_2B17
TraesCS1B01G475300.1_1
TraesCS1B01G454300.1_1
TraesCS3A01G018400.1_1
TraesCS5A01G527200.1_1
Ta_2B13_II
Aet8
Ta_2D7_II
Ta_2B12
Ta_2B10_II
Ta_2A11
Aet7_II
Ta_2B11_II
Tt_2B14_II
TraesCS6A01G115000.1_1
TraesCS6D01G103600.1_1
TraesCS4B01G063700.1_1
TraesCS4A01G251200.1_2
TraesCS4D01G062600.1_1
TraesCS4D01G062700.1_1, TraesCS4A01G251000.1_1
TraesCS4B01G063900.1_1
NLR_TraesCSU01G170800LC.1
NLR_TraesCS7D01G010700.1
TraesCS5A01G273600.1_1
TraesCS4A01G330500.1_1
TraesCS4B01G163200.1_1
TraesCS4D01G131800.1_1
NLR_TraesCS2B01G129800.1
NLR_TraesCS2B01G129700.1
NLR_TraesCS2D01G111500.3
TraesCS7A01G337100.1_1
TraesCS2B01G151300.1_1
TraesCS1D01G137700.1_1
TraesCS2B01G262500.1_1
TraesCS4B01G264000.1_1
TraesCS3D01G256800.1_1
TraesCS3A01G256500.1_1
TraesCS3B01G289700.1_1
TraesCS2A01G477100.1_2
TraesCS2B01G500400.1_2
TraesCS2D01G476500.1_2
TraesCS1D01G004800.1_1
TraesCS1A01G002600.2_1
TraesCS2D01G477000.1_2
TraesCS2A01G477800.1_2
TraesCS6D01G013700.1_1
TraesCS6D01G013700.1_3
TraesCS6D01G013700.1_2
TraesCS2A01G477200.1_1
TraesCS2B01G500800.1_1
TraesCS2D01G476500.1_1
TraesCS4A01G068800.1_2
TraesCS5D01G205800.1_1
TraesCS5B01G199600.1_1
TraesCS4D01G226600.3_1
TraesCS4B01G225900.2_1
TraesCS4A01G068800.1_1
TraesCS2B01G501100.1_1
TraesCS2D01G477000.1_1
TraesCS5A01G200200.1_1
TraesCS3A01G304000.1_1
TraesCS2A01G477800.1_1
TraesCS2A01G478200.1_1
TraesCS2D01G477400.1_1
TraesCS5D01G206100.1_1
TraesCS5B01G199000.1_1
TraesCS5B01G001200.1_1
TraesCSU01G215600.1_1
TraesCS5B01G004100.1_1
TraesCS3A01G459900.1_1
TraesCS3D01G452700.1_1
TraesCS3B01G499600.1_1
TraesCS6A01G049300.1_1
TraesCS3D01G026300.2_1
TraesCS5A01G144500.1_1
TraesCS5D01G205900.1_1
TraesCS5A01G200400.1_1
TraesCS5B01G198700.1_1
TraesCS4A01G069300.1_1
0.1
Legend
non-NLRs BED-containing protein
wheat BNLs outside synteny
BNLs within synteny
BED-I
BED-II

Time
Variety
Scale
20
15
10
5
0
Time
0dpi
1dpi
2dpi
3dpi
5dpi
7dpi
9dpi
11dpi
Variety
AvocetS+Yr5
Vuka

TraesCS4B01G163200
TraesCS4B01G063900
TraesCS4D01G062700
TraesCS2A01G246500
TraesCS6B01G190800
TraesCS3B01G231800
TraesCS4A01G251000
Ta_2A10
TraesCS3D01G152700
TraesCS4B01G225900
TraesCS4D01G226600
TraesCS5D01G497300
TraesCS4A01G330500
TraesCS5D01G544300
Ta_2A3
TraesCS2A01G477100
TraesCS2B01G500400
TraesCS4B01G264000
TraesCS1D01G137700
TraesCS3D01G256800
TraesCS5A01G483700
Ta_2B8
TraesCS3A01G018400
TraesCS4D01G131800
TraesCS2B01G501100
Ta_2A12*
Ta_2A11*
Ta_2A7
Ta_2B10
TraesCS3A01G459900
TraesCS3B01G499600
TraesCS1B01G130400
Ta_2A12*
TraesCS5D01G501900
TraesCS5B01G501500
TraesCS6B01G423600
TraesCS5A01G527200
TraesCS2B01G442000
TraesCS2B01G487700
Ta_2A12*
TraesCS6A01G049300
Ta_2A6
TraesCSU01G215500
TraesCS5B01G004100
TraesCS7A01G337100
TraesCS2A01G478200
TraesCS3D01G452700
TraesCS5A01G273600
TraesCS1D01G004800
TraesCS3B01G269600
TraesCS1B01G158800
TraesCS5A01G200400
Ta_2B13
TraesCS7D01G010700
TraesCS5D01G206100
TraesCS5D01G205900
TraesCS5D01G205800
TraesCS5B01G198600
TraesCS5A01G200200
TraesCS5A01G144500
TraesCS4B01G012600
TraesCS4A01G069300
TraesCS3D01G401500
TraesCS3D01G350600
TraesCS3B01G440000
TraesCS3B01G317800
TraesCS3A01G406300
TraesCS3A01G304000
TraesCS3A01G214600
TraesCS2D01G477400
TraesCS2B01G574700
TraesCS2B01G509900
TraesCS2B01G262500
Ta_2A11*
Ta_2A11*
TraesCS5B01G199000
TraesCS1B01G454300
TraesCS1B01G475300
TraesCS5B01G198700
TraesCS5B01G377100
TraesCSU01G170800LC
Ta_2A5
TraesCS3D01G026300
TraesCS6D01G013700
Ta_2B7
TraesCS5B01G001200
TraesCS6D01G394400
TraesCS6D01G394100
TraesCS2B01G129800
TraesCS7A01G447400
Ta_2B11*
Ta_2D5
Ta_2B6
Ta_2D2
TraesCS3B01G238000
Ta_2B12
Ta_2A2
Ta_2A8
TraesCS2D01G477000
TraesCS6D01G103600
Ta_2B4
Ta_2D4
Ta_2D7
Ta_2D3
TraesCS6B01G457600
TraesCS5D01G544700
TraesCS2B01G151300
Ta_2B2
TraesCS6A01G115000
TraesCS3A01G256500
TraesCS3B01G289700
TraesCS2D01G111500
TraesCS2D01G476500
Ta_2D6
TraesCS3D01G202300
TraesCS3B01G491300
TraesCS3D01G446200
TraesCS4A01G068800
TraesCS5B01G497000
TraesCS2B01G129700
Ta_2B9
Ta_2A1
Ta_2B11*
TraesCS1A01G002600
TraesCS4A01G329100
TraesCS4B01G063700
TraesCS4D01G062600
Ta_2B1
Ta_2D1

Vuka_t0
Vuka_1dpi
Vuka_2dpi
Vuka_3dpi
Vuka_5dpi
Vuka_7dpi
Vuka_9dpi
Vuka_11dpi
AvocetS+Yr5_t0
AvocetS+Yr5_1dpi
AvocetS+Yr5_2dpi
AvocetS+Yr5_3dpi
AvocetS+Yr5_5dpi

Legend
non-NLR BED containing proteins
wheat BED-NLRs outside synteny
BED-I
BED-II
non-BED NLRs within synteny
non-NLR BED containig proteins clustering with BED-NLRs on Supplementary Figure 8