

# *A framework for gene mapping in wheat demonstrated using the Yr7 yellow rust resistance gene*

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Gardiner, L.-J., Bansept-Basler, P., El-Soda, M., Hall, A. and O'Sullivan, D. M. ORCID: <https://orcid.org/0000-0003-4889-056X> (2020) A framework for gene mapping in wheat demonstrated using the Yr7 yellow rust resistance gene. PLoS ONE, 15 (4). e0231157. ISSN 1932-6203 doi: 10.1371/journal.pone.0231157 Available at <https://centaur.reading.ac.uk/88552/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1371/journal.pone.0231157>

Publisher: Public Library of Science

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

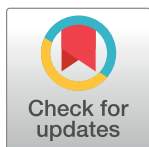
RESEARCH ARTICLE

# A framework for gene mapping in wheat demonstrated using the *Yr7* yellow rust resistance gene

Laura-Jayne Gardiner<sup>1,2</sup>, Pauline Bansept-Basler<sup>3</sup>, Mohamed El-Soda<sup>4</sup>, Anthony Hall<sup>2,5</sup>, Donal M. O'Sullivan<sup>6\*</sup>

**1** IBM Research, Warrington, England, United Kingdom, **2** Earlham Institute, Norwich, England, United Kingdom, **3** Syngenta, Ferme de Moyencourt, Orgerus, France, **4** Department of Genetics, Faculty of Agriculture, Cairo University, Giza, Egypt, **5** School of Biological Sciences, University of East Anglia, Norwich, England, United Kingdom, **6** School of Agriculture, Policy and Development, University of Reading, Reading, England, United Kingdom

\* [d.m.osullivan@reading.ac.uk](mailto:d.m.osullivan@reading.ac.uk)



## OPEN ACCESS

**Citation:** Gardiner L-J, Bansept-Basler P, El-Soda M, Hall A, O'Sullivan DM (2020) A framework for gene mapping in wheat demonstrated using the *Yr7* yellow rust resistance gene. PLoS ONE 15(4): e0231157. <https://doi.org/10.1371/journal.pone.0231157>

**Editor:** Dragan Perovic, Julius Kuhn-Institut, GERMANY

**Received:** October 22, 2019

**Accepted:** March 17, 2020

**Published:** April 15, 2020

**Copyright:** © 2020 Gardiner et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All Supporting Information data is contained within the Supporting Information files. Sequence data underlying mapping-by-sequencing is now available from the European Nucleotide Archive repository (at <https://www.ebi.ac.uk/ena>), study PRJEB36010. Phenotype data has been made available via Supplementary [S3 Table](#) (Avalon x Cadenza DH yellow rust scores following seedling infection with isolate Pst08/21, the “Solstice” race, avirulent on *Yr7*) & [S4 Table](#) (YR-GWAS Panel yellow rust scores following seedling infection with the same

## Abstract

We used three approaches to map the yellow rust resistance gene *Yr7* and identify associated SNPs in wheat. First, we used a traditional QTL mapping approach using a double haploid (DH) population and mapped *Yr7* to a low-recombination region of chromosome 2B. To fine map the QTL, we then used an association mapping panel. Both populations were SNP array genotyped allowing alignment of QTL and genome-wide association scans based on common segregating SNPs. Analysis of the association panel spanning the QTL interval, narrowed the interval down to a single haplotype block. Finally, we used mapping-by-sequencing of resistant and susceptible DH bulks to identify a candidate gene in the interval showing high homology to a previously suggested *Yr7* candidate and to populate the *Yr7* interval with a higher density of polymorphisms. We highlight the power of combining mapping-by-sequencing, delivering a complete list of gene-based segregating polymorphisms in the interval with the high recombination, low LD precision of the association mapping panel. Our mapping-by-sequencing methodology is applicable to any trait and our results validate the approach in wheat, where with a near complete reference genome sequence, we are able to define a small interval containing the causative gene.

## Introduction

Yellow stripe rust (YR) caused by *Puccinia striiformis* is one of the most damaging diseases of wheat with 88% of the world's wheat production now susceptible to infection [1, 2, 3]. Therefore, identifying new YR resistance genes and their associated molecular makers is of great interest in breeding programs using marker assisted selection [4, 5]. Two major YR resistance genes, *Yr5* and *Yr7*, were identified in wheat and mapped to the long arm of chromosome 2B [6, 7]. Intercrossing *Yr5* and *Yr7* near-isogenic lines (NILs) in the genetic background of Avocet S (AVS) revealed distinct segregation patterns for each gene, indicating probable allelism. This hypothesis was confirmed by a recent study which showed that *Yr5* and *Yr7* are distinct

isolate Pst08/21). AxC genotype data is available from CerealsDB.net ([https://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/genotyping\\_data/AvalonxCadenzaDH.csv](https://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/genotyping_data/AvalonxCadenzaDH.csv)) GWAS panel genotype data can be downloaded from niab.com ([https://www.niab.com/sites/default/files/imce\\_uploads/Research/NIAB\\_90K\\_Wagtail\\_genotype\\_data.xlsx](https://www.niab.com/sites/default/files/imce_uploads/Research/NIAB_90K_Wagtail_genotype_data.xlsx)) Pedigree relationships were interrogated using the UK winter wheat pedigree resource <https://www.niab.com/research/research-projects/uk-wheat-varieties-pedigree>.

**Funding:** This work was funded by BBSRC grant BB/J0072607/1 (DOS), a PhD bursary to PBB from the John Oldacre Foundation ERA-CAPS grant BB/N005104/1, BB/N005155/1 (LJG, AH) and BBSRC Designing Future Wheat BB/P016855/1 (AH). The funders provided support in the form of salaries for authors LJG and PB, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The current affiliation of LJG with IBM and PBB with Bayer does not alter our adherence to PLOS ONE policies on sharing data and materials.

paralogues arranged in a complex tandem cluster [6]. Both are a valuable source of resistance, Yr5 being effective against the pathogen in most wheat growing regions of the world [8] and Yr7 being effective under various environmental conditions including high temperature [9].

Genetically dissecting complex responses such as stripe rust resistance is regularly achieved by the analysis of trait-marker associations via QTL and/or genome wide association (GWA) mapping [10,11]. While traditional QTL mapping using bi-parental populations is more powerful, GWA mapping can potentially offer higher mapping resolution, as it benefits from historic recombination events and linkage disequilibrium (LD) [12]. However, GWA mapping can suffer from false positives, leading to false associations, and false negatives, missing true association. Therefore, combining QTL with GWA mapping is increasingly attracting researchers to benefit from both approaches [13]. Recently, next generation sequencing (NGS) technologies, e.g. mapping-by-sequencing (MbS), has evolved as a new approach to explore trait-marker association with greater resolution via developing well-distributed genic and non-genic SNPs [4, 14]. MbS can be combined with bulked segregant analysis to speed the identification of candidate genes [15]. Typically QTL, GWA mapping or MbS define a genetic interval that can contain multiple candidate genes dependent on the resolution of the approach, these candidates can then be refined using homology if candidate genes are known in other species or related traits (demonstrated here) or via knockout using methods such as EMS or gene editing.

Here, we take advantage of using both bi-parental and GWA mapping populations, phenotyped for severity of infection following inoculation with a pure Pst isolate avirulent on Yr7. Both populations were genotyped with the same high-density iSelect SNP array to show the relative power and precision of these population types in locating Yr7. We go on to show that using mapping-by-sequencing of resistant and susceptible bulks from the bi-parental population we can identify an overlapping interval to that defined by QTL analysis and GWA. In addition, we can populate the Yr7 interval with a much higher density of polymorphisms including a series of SNPs from an NBS-LRR gene which would make a logical positional and biological candidate. Performing mapping-by-sequencing of resistant and susceptible bulks using the recent wheat reference genome sequence (RefSeqv1) we show the power of an ordered reference genome. With no prior knowledge of the gene of interest here, we demonstrate that we can refine a genetic interval of 60Mbp containing 589 genes, only 0.55% of the high confidence genes in the genome. Furthermore, with limited information regarding gene function, e.g. in this case our candidate is likely to be an NBS-LRR, we were able to define a candidate gene list of only 10 genes. Our methodology is therefore broadly applicable for mapping genes associated with a wide variety of traits. Finally, given that wheat is hexaploid, it was notable that our method is perfectly capable of distinguishing between the homoeologs, identifying a clear peak on chromosome 2B that is absent from 2A and 2D.

## Materials and methods

### Plant materials

**Avalon x Cadenza doubled-haploid population.** The population of doubled haploid (DH) individuals, derived from F1 progeny of a cross between cultivars 'Avalon' and 'Cadenza', was developed by Clare Ellerbrook, Liz Sayers and the late Tony Worland (John Innes Centre), as part of a Defra funded project led by ADAS. The parents were originally chosen (to contrast for canopy architecture traits) by Steve Parker (CSL), Tony Worland and Darren Lovell (Rothamsted Research). BS-coded SNP genotypes were imported from supplementary data of [16].

**YR-GWAS panel.** The main criterion for the membership of the YR-GWAS panel was the existence of at least one year of adult plant resistance data as part of evaluations carried out

under the auspices of the National and Recommended List trials and UK Cereal Pathogen Virulence Survey. Three hundred and ten winter wheat varieties released between 1990 and 2009 were selected to comprise the core wheat YR-GWAS panel. An additional 9 older winter varieties were included in the YR-GWAS panel based on their importance in the pedigrees of elite wheat germplasm. The full wheat YR-GWAS panel is therefore composed of 319 wheat varieties, mainly elite UK winter wheat varieties, but other European countries are also represented.

**Rust phenotyping and the seedling tests.** *Puccinia striiformis* (Pst) isolate 08/21 (also known as '08/21 Solstice') virulent on Avalon and avirulent on Cadenza was obtained from Amelia Hubbard at the UK Cereal Pathogen Virulence Survey at NIAB. This isolate was used to screen both the AxC DH population and the YR-GWAS panel. The full virulence profile of Pst '08/21 Solstice' is as reported in S4 Table of [17]. Seeds were sown in 96 cell trays, organized in complete randomized block design. Once sown, the trays were placed in a disease-free glasshouse or a growth room prior to inoculation. When the seedlings were between GS11 and GS12, the average height of the seedlings was recorded and plots with average seedling size less than 5 cm, limited seedling number with less than 3, or discoloration were discarded. Following the measurements, the trays were well watered and placed in individual polythene bags. A 1:19 spore: talc mixture was prepared and then 3g of the mixture per tray, equivalent to 0.15g spores and 2.85g of talc, was distributed in an individual glass jar. Using air-blown spore inoculators, each tray was inoculated individually with the contents of a jar. Bags were then sealed and placed in an incubator at 8°C, 24 to 48 hours, in the dark to keep a high humidity level and provide favorable conditions for spore germination. After the incubation, the seedling trays were removed from their bag and placed in a growth room. Two experiments were carried out using the YR-GWAS panel with 2 replications each and one experiment was done with 3 replicates for the AxC DH population. Symptom assessment for the three experiments was carried out at 17 days post inoculation (dpi). The variety Victor was placed in each tray as a susceptible control and to help choosing the most suitable scoring date. The infection type was assessed on the first leaf of the seedling twice independently at 17 dpi for consistency. Disease assessments followed the 0 to 9 infection type (IF) scoring system described by [18]. Lines with an average IF score  $< 4$  were classified resistant, with IF score 4–6 were interpreted as intermediate response, and with IF  $> 6$  were considered susceptible; however for DNA pool construction, more stringent cutoffs of  $< = 2$  for the resistant pool and  $> = 7$  were chosen for the susceptible pool (S1 Table).

**90k iSelect SNP genotyping.** The wheat 90k Illumina iSelect SNP array [19] was used to genotype both populations. In case of the AxC DH population, the iSelect mapping data was derived from [19], whereas genotyping of the YR-GWAS panel revealed 26,015 SNPs.

**QTL mapping using DH population.** Only SNPs on linkage group 2B, i.e. 353 SNPs, which had been scored and were polymorphic in both the AxC and YR panel were used for QTL mapping of seedling test using QTL library in GenStat 17<sup>th</sup> edition (VSN International, Hemel Hempstead, UK). Default settings of single trait linkage analysis, i.e. 10 cM step size, minimum cofactor proximity of 50 cM, 30 cM minimum separation of selected QTL, and a threshold calculated using Li and Ji option [20] with  $\alpha = 0.05$ , corresponding to  $-\log_{10}(P)$  values of 3.9, were used. The QTL was mapped based on interval mapping followed by two rounds of composite interval mapping using cofactors. Finally, to estimate the allelic effect and the explained phenotypic variance, a final QTL model was run.

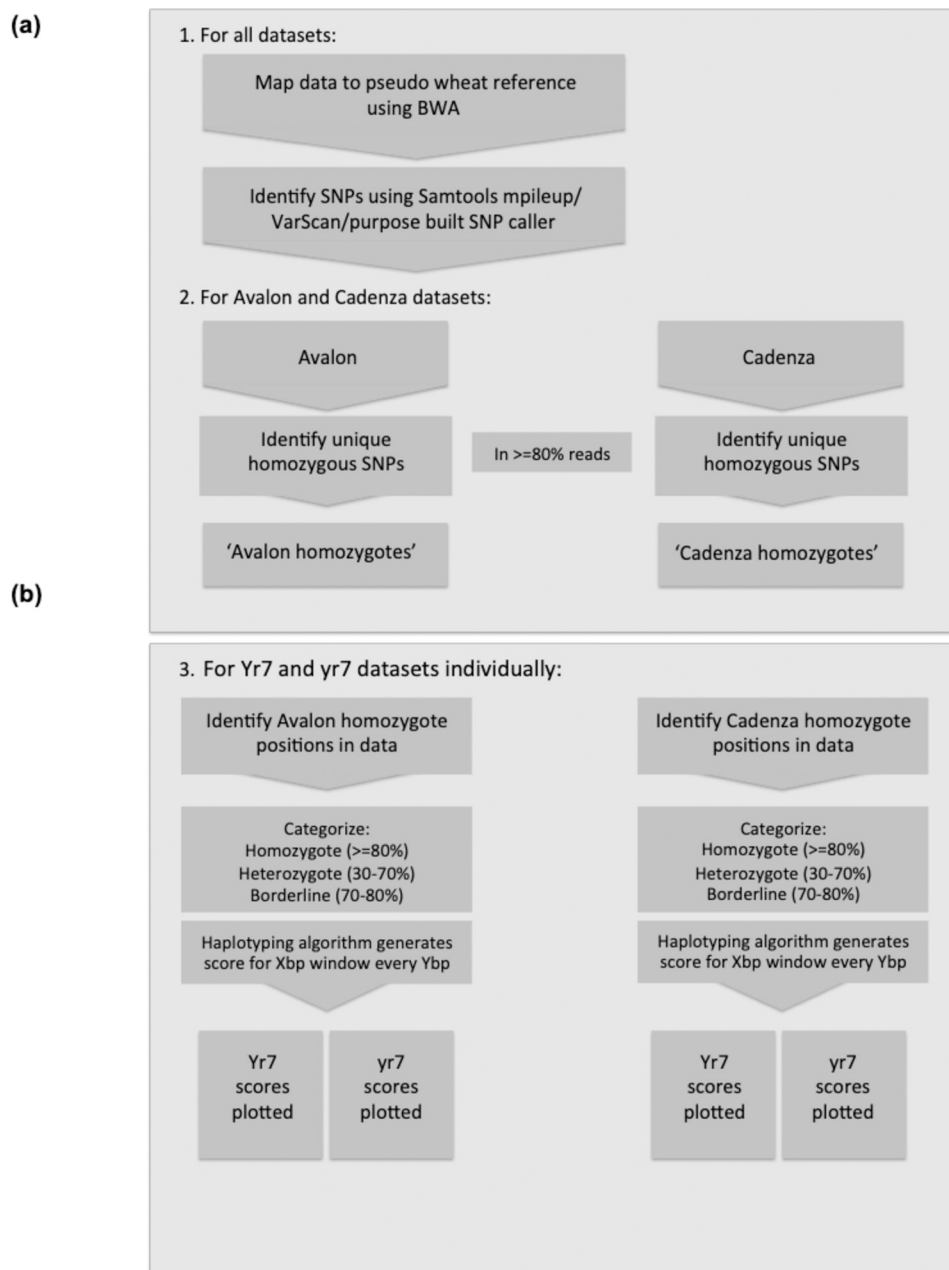
**Association mapping using YR-GWAS panel.** Kinship matrix and principal components were calculated using Genomic Association and Prediction Integrated Tool (GAPIT) statistical package in R software [21, 22]. Association mapping was performed using the GAPIT package in R [22] using compressed mixed linear model (CMLM) approach that account for population parameters [21].

**Gene enrichment combined with mapping-by-syteny.** Resistant and susceptible bulks were created by pooling equal amounts of DNA of 51 individuals with an infection score of  $\geq 7$  (referred to as the *yr7* bulk) and 54 individuals scoring  $\leq 2$  (referred to as the *Yr7* bulk). At cost of a potential small loss in genetic resolution, lines with intermediate scores (2.1–6.9) were omitted from the bulks to ensure that phenotyping errors did not dilute the allelic discrimination potential of the bulks.

We used the NimbleGen SeqCap EZ in solution custom capture probe set (~110 Mb) to enrich genomic DNA from the two bulk segregated pools along with the purebred Avalon and Cadenza parental lines. The design and targets of the capture probe set that is used here is as previously reported [23]. The final design covered the majority of the wheat genes with each probe intended to enrich all 3 homoeologous gene copies in hexaploid wheat. The samples were paired end sequenced using Illumina HiSeq technology as follows; Genomic DNA was purified using Agencourt AMPure XP beads (Beckman Coulter). Samples were quantified using a Qubit double-stranded DNA Broad Range Assay Kit and Qubit fluorometer (Life Technologies). 2.6 µg of genomic DNA, in a total volume of 130 µl, was sheared for 3×60s using a Covaris S2 focused-ultrasonicator (duty cycle 10%, intensity 5, 200 cycles per burst using frequency sweeping). The size distribution of the fragmented DNA was assessed on a Bioanalyser High Sensitivity DNA chip (Agilent). 50 µl (~1 µg) of sheared DNA was used as input for library preparation. End-repair, 3'-adenylation, and adapter ligation were performed according to the Illumina TruSeq DNA Sample Preparation Guide (Revision B, April 2012) without in-line control DNA and without size-selection. Amplification of adapter-ligated DNA (to generate pre-capture libraries), hybridisation to custom wheat NimbleGen sequence capture probes, and washing, recovery and amplification of captured DNA were all carried out according to the NimbleGen Illumina Optimised Plant Sequence Capture User's Guide (version 2, March 2012), with the exception that purification steps were carried out using Agencourt AMPure XP beads instead of spin columns. Final libraries were quantified by Qubit double-stranded DNA High Sensitivity Assay and the size distribution ascertained on a Bioanalyser High Sensitivity DNA chip. The 4 libraries were then pooled in equimolar amounts based on the aforementioned Qubit and Bioanalyser data. Sequencing was carried out on two lanes of an Illumina HiSeq 2000, using version 3 chemistry, generating 2 x 100bp paired end reads.

The pipeline and algorithm for the processing of this sequencing data is summarized below and in Fig 1; it is adapted from a previously described analysis [4, 23]. The reference genome that was implemented to map the sequence data was the Chinese Spring hexaploid wheat RefSeqv1 genome assembly [24]. These chromosomal pseudomolecules represent the A, B and D diploid sub-genomes of hexaploid wheat individually resulting in an effectively diploid reference genome of 21 chromosomes [24, 25]. The sequence data was mapped to the pseudomolecules using the mapping software BWA [26]. A minimum mapping quality score of 10 was implemented and any non-uniquely mapping reads, duplicate reads or unmapped sequencing reads were removed from the analysis.

SNP calling was carried out on the mapped datasets using Samtools mpileup [27] and then VarScan variant detection [28]. SNPs were scored at a minimum quality of 15 and a minimum depth of 10X in Avalon and Cadenza and 20X in the bulk segregant datasets due to their relative average depths of coverage (see results). A maximum depth of 200 was applied and a minimum alternate allele frequency of 10% was used for SNP calls (Table 1). Parental homoeologous homozygous SNPs were called if the SNP allele was observed in 80% or more of the sequencing reads with the reference allele seen in the remainder of the reads. Homozygous SNPs were designated as Avalon- or Cadenza-specific alleles if the homozygous SNP allele was not observed in other parent at that position resulting in 70,895 Avalon and 73,224 Cadenza specific homozygous alleles.



**Fig 1. Processing four sets of enriched sequencing data to identify a mapping interval containing the gene controlling the phenotype of interest.** (a) Standard mapping and SNP calling pipeline to construct “reference genomes” (b) Pipeline implementing an algorithm to score regions of interest by prioritizing long homozygous parental haplotypes for the bulk segregant samples to identify the interval of interest. Shown here using Xbp windows at Ybp intervals (user defined) and standard homozygote, heterozygote and borderline SNP definitions for a diploid—values can all be adjusted throughout the analysis as necessary.

<https://doi.org/10.1371/journal.pone.0231157.g001>

A mapping-by sequencing mutant identification pipeline and homozygous haplotyping algorithm that we presented in a previous study [4, 23], were applied here for the analysis of Yr7 and yr7 bulks (Fig 1B). Of the Avalon specific homozygous alleles, in regions with a minimum of 20X coverage, 53,658 were seen in the Yr7 bulk and 60,008 in the yr7 bulk at the same position, regardless of homozygous or heterozygous status. Similarly of the Cadenza specific



**Table 1. Summary of mapping statistics across the pseudo-chromosome reference.** Mapping depth of coverage of the pseudo-chromosome reference and SNP numbers that were identified for the purebred parental lines Avalon and Cadenza plus the bulks.

Sample	Base pairs mapped (> = 1X)	Base pairs mapped (> = 5X)	Average depth of coverage	Total number of sequencing reads	Number of sequencing reads mapped (post filtering)	% of capture design space mapped	Average depth of coverage (capture design space)	Number of SNP calls (min 10x)
Avalon	3,954,915,256	309,339,249	2.82	152,260,292	116,881,076	93.9	43.2	693,462
Cadenza	3,595,684,321	343,126,954	3.25	155,802,050	122,450,267	94.3	49.5	788,678
Bulk Yr7	5,808,329,705	469,381,798	3.47	274,027,688	210,593,909	96.1	73.1	1,493,665
Bulk yr7	6,763,399,040	523,357,788	3.62	341,681,490	256,156,931	96.5	87.5	1,636,640

<https://doi.org/10.1371/journal.pone.0231157.t001>

homozygous alleles, 51,125 were found in the bulk segregant dataset Yr7 and 54,946 in yr7. These allele positions were then categorized as homozygous, heterozygous or borderline according to the following thresholds; homozygote allele in a minimum of 80% of sequencing reads; heterozygote in 30–70% and borderline in 70–80% (shown in Fig 1B). The scoring algorithm was then implemented to calculate a homozygote score per Xbp window along each pseudo-chromosome at Xbp intervals for these alleles that were found in the bulks and were specific to either the Avalon or Cadenza parents. Window sizes and interval lengths are user defined and adjusted as necessary to suit the dataset and produce the cleanest final plot.

## Results and discussion

### Yr7 maps to a low-recombination region near the centromere of chromosome 2B

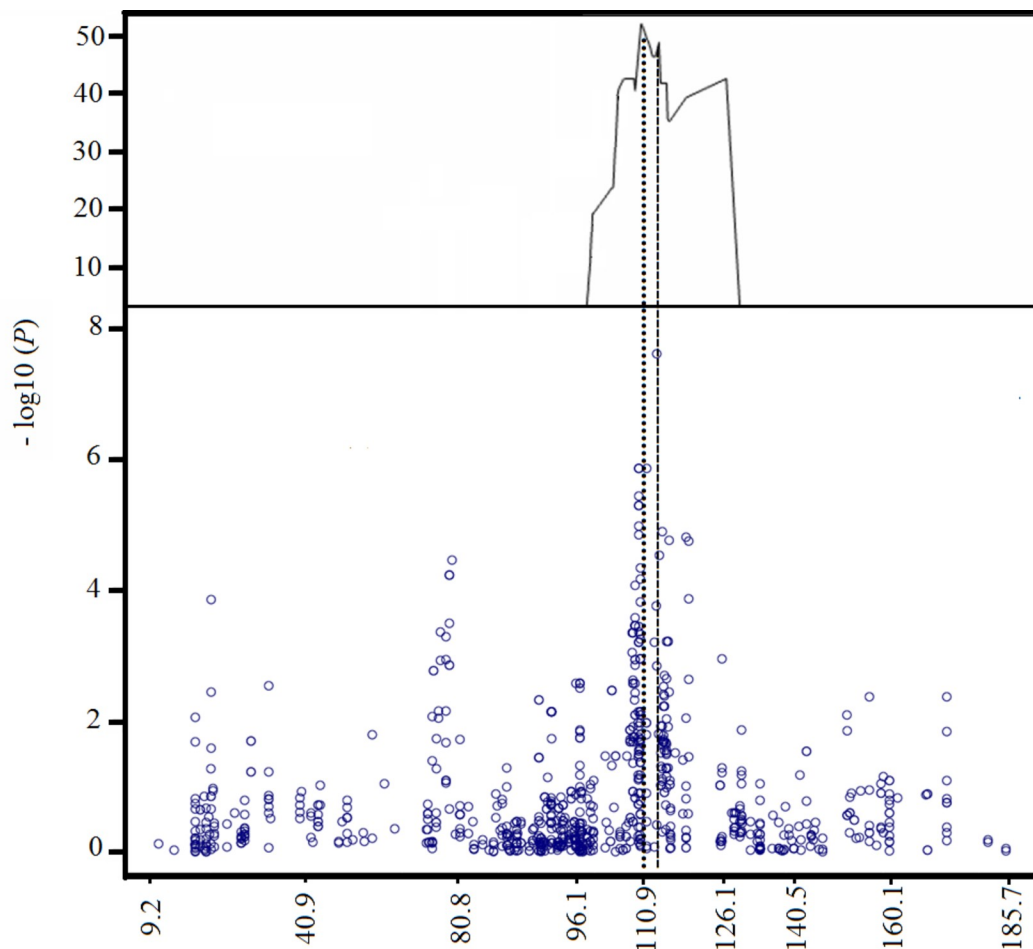
Screening of the AxC DH population with Pst isolate 08/21 in a seedling test revealed a bimodal distribution of infection scores, following an approximately 1:1 segregation ratio—the expected inheritance pattern for a single, major gene (S2 Table).

A QTL scan for loci conferring resistance to Pst 08/21 showed a single major QTL ( $-\log_{10}(P) = 52.6$ ) that explains 70.6% of the variance with a peak at position 110.9 cM on chromosome 2B. Fig 2 shows the CIM QTL trace for chromosome 2B, where it can be noted that the confidence interval spans 29.16 cM. The already large confidence interval, though it spans just 10% of the genetic length of chromosome 2B, contains ~50% of all 2B AxC SNPs and thus is likely to span an enormous physical interval characterized by low recombination rates, which poses a major challenge for the fine mapping of Yr7.

### Genome-wide association mapping reveals historic recombination in the Yr7 region

In order to gain information on how widespread Yr7 is in UK winter wheat germplasm, to see the expression of the gene in adult plants grown in the field and to attempt fine-mapping of the Yr7 gene by another means, the YR Panel of mainly UK winter wheat varieties, which includes both ‘Avalon’ and ‘Cadenza’ as members, was inoculated both in seedling and in adult plant field tests with Pst 08/21. The single significant hit obtained when iSelect SNPs scored in the YR Panel were scanned for association with resistance to Pst 08/21 was located near the centromere on chromosome 2B. The SNP showing the highest probability of association (wsnp\_Ex\_c10071\_16554911) in the GWAS scan is located at 112.9 cM, close to the peak of the QTL interval shown in Fig 2. Closer examination of SNP haplotypes formed by a subset of 46 SNPs informative in both the YR Panel and the AxC population, and located between 107.9 and 119.1cM (see Fig 3) shows that although the yr7 susceptible haplotype represented by ‘Avalon’ (green shading where ‘Avalon’ allele is present) predominates (251/329) over the

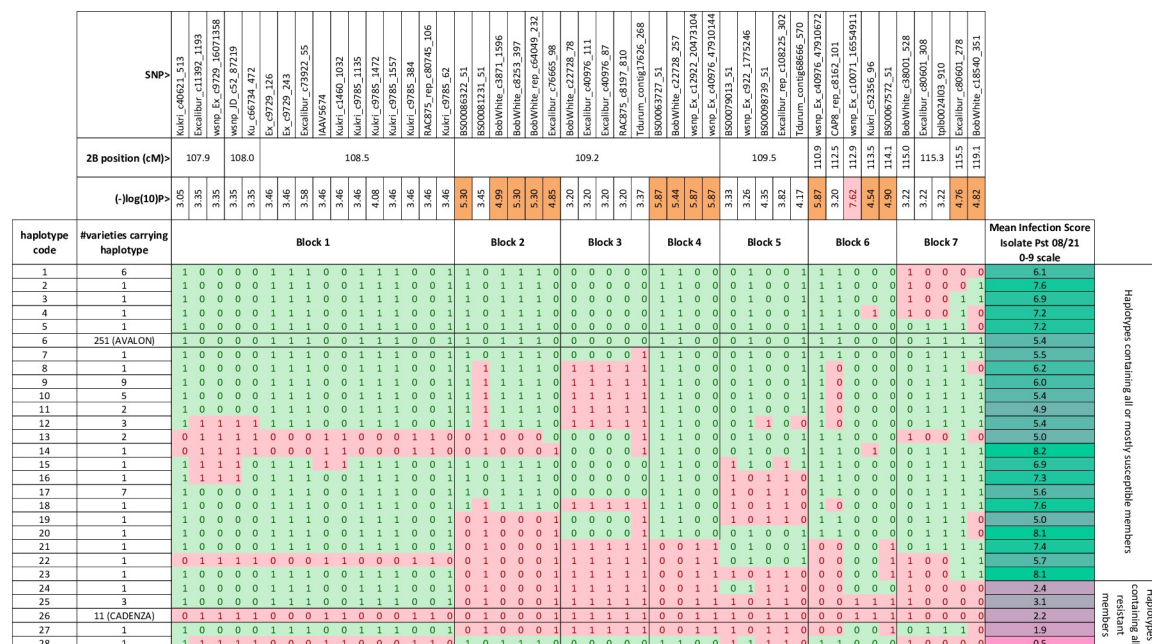




**Fig 2. Genetic mapping of the Yr7 locus.** The uppermost plot shows the CIM QTL scan of chromosome 2B for resistance to yellow rust isolate Pst 08/21 in the AxC DH population. The lower plot shows the GWAS Manhattan plot of chromosome 2B (showing only 353 markers in common with AxC), using AxC map order and distances (cM) to scale the X-axis.

<https://doi.org/10.1371/journal.pone.0231157.g002>

resistant ‘Cadenza’ (red shaded) haplotype, several minor haplotypes, some represented by single cultivars, show evidence of recombination between susceptible-like and resistant-like haplotypes. Vertical lines delineating seven recombination ‘blocks’ have been inserted to highlight where multiple distinct and independent haplotypes show blocks of SNPs, presumably in a state of identity-by-descent, have undergone historical recombination. This historic recombination is useful as it allows us both to partly resolve map order, as in the sub-division of SNPs co-segregating in AxC at position 109.24cM into Blocks 2–4, and to tentatively rule out all but Block 6 as potentially carrying the causal gene, since each of the other Blocks have both fully ‘Cadenza’ and fully ‘Avalon’ partial haplotypes in association with susceptibility to Pst 08/21. The notion that graphical genotypes as displayed in Fig 3 have arisen through identity-by-descent in an ancestral recombining population is supported by pedigree analysis which explains 11 out of 17 suspected occurrences of Yr7 in terms of vertical transmission of haplotypes 25 and 26 through known pedigrees. Haplotype 25 is possessed by varieties **Brock**, Tara and Vault belonging to the same lineage and Haplotype 26 is shared by three separate lineages (1: **Tonic**, Cordiale, Spark, Cadenza; 2: **Ekla**, Vector; 3: **Thatcher**, Tommy). In each case, the founding Yr7 donor in the lineage is highlighted in bold.



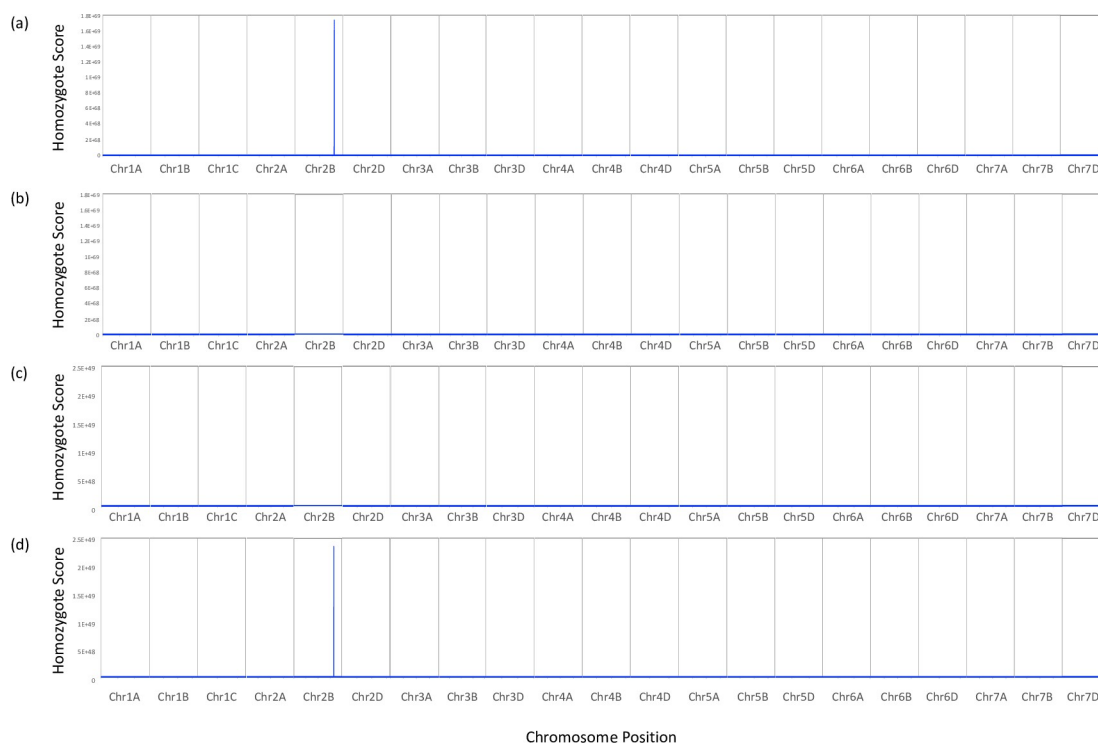
**Fig 3. Haplotype structure of the Yr7 interval in the YR panel.** The 28 distinct haplotypes formed by the 47 SNPs that span the Yr7 confidence interval are displayed in graphical genotype format. For each SNP, the SNP identifier, chromosomal position and P-value for the GWA scan with the Pst 08/21 (Avr-Yr7) isolate are shown.  $-\log_{10}(P)$  values above 4.5 are highlighted in orange and the most significant value, occurring at 113.8cM in Block 6, is highlighted in pink. The commonest susceptible haplotype (Haplotype 6 which includes 'Avalon' and 250 other varieties) is shaded in green and the most common resistant haplotype (Haplotype 26, which includes 'Cadenza' and 10 other varieties) is shaded in pink and all other haplotypes are coloured according to which alleles are shared with these reference haplotypes. Mean IF score from inoculation with Pst 08/21 is shown on the right highlighted according to a colour scale that goes from 0 –pink to 9 –green.

<https://doi.org/10.1371/journal.pone.0231157.g003>

## Homozygous haplotyping allows the application of mapping-by-sequencing to hexaploid wheat

After enrichment, the two bulk segregant pools (Yr7 and yr7) were sequenced and mapped along with the purebred Avalon and Cadenza parental lines; over 3.6Gbp of the RefSeqv1 reference was mapped in each of the four datasets with an average depth of coverage of approximately 3.3X. To focus on the targeted regions for enrichment only, reads were aligned to the sequences that were used for the capture probe set design, known as the capture probe design-space. It was observed that ~95% of this reference was mapped across the four datasets with an average depth of coverage of approximately 45X in Avalon and Cadenza and 80X in the pools Yr7 and yr7 (Table 1). Since the capture probe design-space contains one representative copy of each set of three homoeologous wheat genes this translates to ~15X coverage per wheat sub-genome in Avalon and Cadenza and ~25X per sub-genome in the pools Yr7 and yr7 for the targeted regions.

A mapping-by-sequencing mutant identification pipeline and algorithm that we presented in a previous study [4, 23], was here, applied successfully to a new target (Fig 1). Homozygote scores were calculated for a range of windows 100,000–10,000,000bp along each chromosomal pseudomolecule at 10,000bp intervals for the homozygous Avalon and Cadenza specific alleles that could be found in the bulks. The highest signal:noise ratio was obtained when homozygote scores were calculated per 10,000,000bp window at 10,000bp intervals and these scores were plotted for each bulk in relation to both parents in Fig 4.

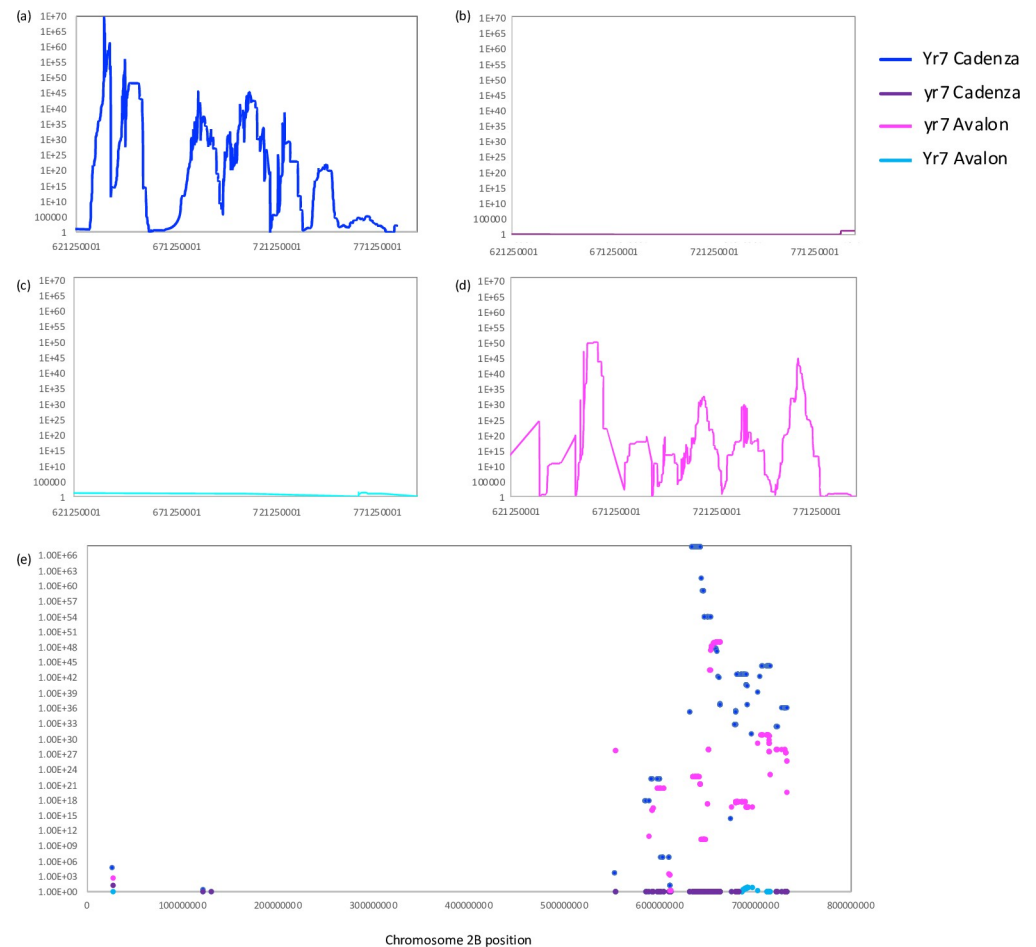


**Fig 4. Homozygosity scores calculated for the Yr7 and yr7 bulked datasets along each chromosomal pseudomolecule.** Scores calculated per 10,000,000bp window along each chromosome at 10,000bp intervals. (a) Scores plotted for 'Cadenza unique homozygote alleles' found in the Yr7 bulk segregated dataset. (b) Scores plotted for 'Avalon unique homozygote alleles' found in the Yr7 bulk segregated dataset. (c) Scores plotted for 'Cadenza unique homozygote alleles' found in the yr7 bulk segregated dataset. (d) Scores plotted for 'Avalon unique homozygote alleles' found in the yr7 bulk segregated dataset.

<https://doi.org/10.1371/journal.pone.0231157.g004>

Fig 4 shows a clear peak interval in chromosome 2B that was seen both in the resistant Yr7 pool highlighting conserved homozygosity for the Cadenza parent and in the susceptible yr7 pool highlighting conserved homozygosity for the Avalon parent. The peak interval is shown in greater detail in Fig 5. This is likely to represent our Yr7 resistance interval of interest as the signal for the yr7 pool in Cadenza is low, as is the signal for Yr7 in Avalon (Fig 3B and 3C). The homozygosity scores successfully highlight mirrored 'Cadenza'-specific and 'Avalon'-specific peaks in the respective Yr7 and yr7 bulks on chromosome 2B with low signal across the rest of the genome. The peak interval encompasses the region from approximately 633,430,001–734,910,001bp on chromosome 2B (homozygosity scores  $> 1e35$ ) which for Yr7 contains 1,262 'Cadenza'-specific SNPs and for yr7 823 'Avalon'-specific SNPs. Fig 5 highlights multiple individual peak regions that make up the apparently single observed peak in Fig 4; the fragmentation of the signal across disjointed nearby segments is likely a product of the exome capture strategy that was used for sequencing with the multiple peaks representing uneven sequence coverage focused primarily on genic sequence. The highest scoring individual peak regions for Yr7 Cadenza represent the intervals 633,430,001–663,920,001bp, 681,790,001–693,030,001bp and 702,660,001–721,190,001bp (total of 60,260,000bp with homozygosity scores  $> 1e40$ ). There are 589 high confidence genes observed in these regions.

BLAST analysis was used to infer common features with the sequence (SNP)-based AxC genetic map and the Chinese Spring based RefSeqv1. 296 iSelect SNP sequences from the Yr7 interval were used in a BLAST search against RefSeqv1 to find their relative positions (E-value  $\geq 1e-5$ , minimum length 50bp, minimum sequence identity 90%). 277 iSelect SNPs



**Fig 5. Homozygosity scores calculated for the Yr7 and yr7 bulked datasets along chromosome 2B.** Scores calculated per 10,000,000bp window along chromosome 2B at 10,000bp intervals. (a) Scores plotted for 'Cadenza unique homozygote alleles' found in the Yr7 bulk segregated dataset. (b) Scores plotted for 'Cadenza unique homozygote alleles' found in the yr7 bulk segregated dataset. (c) Scores plotted for 'Avalon unique homozygote alleles' found in the Yr7 bulk segregated dataset. (d) Scores plotted for 'Avalon unique homozygote alleles' found in the yr7 bulk segregated dataset. (e) Scores reported here only for windows on chromosome 2B containing the iSelect SNPs that could be mapped to pseudo-chromosome 2B. If multiple windows hit the iSelect SNP, the average score across the windows was calculated. Scores plotted for all datasets (a-d) i.e. Yr7/yr7 Avalon and Cadenza specific homozygote scores and datapoints coloured according to the legend.

<https://doi.org/10.1371/journal.pone.0231157.g005>

had hits (93.6%) and of these SNPs; 246 (88.8%) were anchored to chromosomal pseudomolecule 2B. The relative homozygosity scores (Cadenza/Avalon specific scores) for these iSelect SNP positions were extracted for each of the Yr7/yr7 datasets and plotted in Fig 5E. Fig 5E shows that the vast majority of Yr7-linked iSelect SNP positions on chromosome 2B (216 or 87.8%), fall within the peak region 633,430,001–734,910,001bp that we identified using mapping-by-sequencing. We noted that the 5 SNPs that we previously found in our candidate genetically defined interval 'Block 6' (Fig 3), all showed high Yr7 Cadenza specific ( $>1E+36$ ) and yr7 Avalon specific homozygosity scores ( $>1E+16$ ) but also low Yr7 Avalon specific ( $<6$ ) and yr7 Cadenza specific homozygosity scores ( $<1$ ) that are characteristic of the Yr7 locus. Furthermore, the 5 SNPs were found in the 10.3 Mbp interval 683,029,160–693,314,916bp on chromosome 2B.

**Table 2. Candidate genes.** Annotated disease resistance genes within the candidate intervals of 633,430,001–663,920,001bp, 681,790,001–693,030,001bp and 702,660,001–721,190,001bp on RefSeqv1 chromosome 2B.

Gene name	Start	End	Annotation	Further information
TraesCS2B01G486100	683043955	683047878	NBS-LRR disease resistance protein-like protein	<a href="https://plants.ensembl.org/Triticum_aestivum/Gene/Summary?g=TraesCS2B02G486100;r=2B:683042399–683049870;t=TraesCS2B02G486100.1;db=core">https://plants.ensembl.org/Triticum_aestivum/Gene/Summary?g=TraesCS2B02G486100;r=2B:683042399–683049870;t=TraesCS2B02G486100.1;db=core</a>
TraesCS2B01G486200	683053644	683059018	NBS-LRR disease resistance protein-like protein	<a href="https://plants.ensembl.org/Triticum_aestivum/Gene/Summary?g=TraesCS2B02G486200;r=2B:683053513–683062553;t=TraesCS2B02G486200.1;db=core">https://plants.ensembl.org/Triticum_aestivum/Gene/Summary?g=TraesCS2B02G486200;r=2B:683053513–683062553;t=TraesCS2B02G486200.1;db=core</a>
TraesCS2B01G486300	683066277	683072620	NBS-LRR disease resistance protein-like protein	<a href="https://plants.ensembl.org/Triticum_aestivum/Gene/Summary?g=TraesCS2B02G486300;r=2B:683065941–683075708;t=TraesCS2B02G486300.1;db=core">https://plants.ensembl.org/Triticum_aestivum/Gene/Summary?g=TraesCS2B02G486300;r=2B:683065941–683075708;t=TraesCS2B02G486300.1;db=core</a>
TraesCS2B01G486400	683127582	683132111	NBS-LRR disease resistance protein-like protein	<a href="https://plants.ensembl.org/Triticum_aestivum/Gene/Summary?g=TraesCS2B02G486400;r=2B:683127549–683135331;t=TraesCS2B02G486400.1;db=core">https://plants.ensembl.org/Triticum_aestivum/Gene/Summary?g=TraesCS2B02G486400;r=2B:683127549–683135331;t=TraesCS2B02G486400.1;db=core</a>
TraesCS2B01G486700	683159224	683163025	NBS-LRR disease resistance protein-like protein	<a href="https://plants.ensembl.org/Triticum_aestivum/Gene/Summary?g=TraesCS2B02G486700;r=2B:683159256–683166341;t=TraesCS2B02G486700.1;db=core">https://plants.ensembl.org/Triticum_aestivum/Gene/Summary?g=TraesCS2B02G486700;r=2B:683159256–683166341;t=TraesCS2B02G486700.1;db=core</a>
TraesCS2B01G487700	683752480	683755002	NBS-LRR disease resistance protein-like protein	<a href="https://plants.ensembl.org/Triticum_aestivum/Gene/Summary?g=TraesCS2B02G487700;r=2B:683752480–683755002;t=TraesCS2B02G487700.1;db=core">https://plants.ensembl.org/Triticum_aestivum/Gene/Summary?g=TraesCS2B02G487700;r=2B:683752480–683755002;t=TraesCS2B02G487700.1;db=core</a>
TraesCS2B01G488000	685265721	685270904	NBS-LRR disease resistance protein-like protein	<a href="https://plants.ensembl.org/Triticum_aestivum/Gene/Summary?g=TraesCS2B02G488000;r=2B:685265721–685270904;t=TraesCS2B02G488000.1;db=core">https://plants.ensembl.org/Triticum_aestivum/Gene/Summary?g=TraesCS2B02G488000;r=2B:685265721–685270904;t=TraesCS2B02G488000.1;db=core</a>
TraesCS2B01G488400	685741277	685746619	NBS-LRR disease resistance protein-like protein	<a href="https://plants.ensembl.org/Triticum_aestivum/Gene/Summary?g=TraesCS2B02G488400;r=2B:685741193–685746862;t=TraesCS2B02G488400.1;db=core">https://plants.ensembl.org/Triticum_aestivum/Gene/Summary?g=TraesCS2B02G488400;r=2B:685741193–685746862;t=TraesCS2B02G488400.1;db=core</a>
TraesCS2B01G489400	686809999	686815417	NBS-LRR disease resistance protein-like protein	<a href="https://plants.ensembl.org/Triticum_aestivum/Gene/Summary?g=TraesCS2B02G489400;r=2B:686809725–686815511;t=TraesCS2B02G489400.1;db=core">https://plants.ensembl.org/Triticum_aestivum/Gene/Summary?g=TraesCS2B02G489400;r=2B:686809725–686815511;t=TraesCS2B02G489400.1;db=core</a>
TraesCS2B01G507100	703972543	703975392	NBS-LRR disease resistance protein-like protein	<a href="https://plants.ensembl.org/Triticum_aestivum/Gene/Summary?g=TraesCS2B02G507100;r=2B:703972469–703976356;t=TraesCS2B02G507100.1;db=core">https://plants.ensembl.org/Triticum_aestivum/Gene/Summary?g=TraesCS2B02G507100;r=2B:703972469–703976356;t=TraesCS2B02G507100.1;db=core</a>

<https://doi.org/10.1371/journal.pone.0231157.t002>

### Combining gene enrichment and mapping-by-sequencing for the Yr7 and yr7 DH bulks allows identification of candidate causal SNPs in resistance gene analogues

Once we had established that the combination of gene enrichment, mapping-by-sequencing and the homozygosity haplotyping algorithm, was specifically and robustly detecting an interval which overlapped the genetically defined interval, the number and nature of Yr7-linked polymorphisms could be examined in more detail. The candidate window encompassed 589 high confidence genes, 10 of which are annotated as disease resistance associated and are therefore candidate genes for Yr7 resistance (Table 2). The 10 genes are mainly clustered together in the region 683,043,955–686,815,417bp (~4Mbp). In a previous study a candidate Yr7 gene was suggested and its gene sequence reported [6]. We used BLAST to align this Yr7 gene sequence to the Chinese Spring RefSeqv1.1 where its top hit was to the homolog TraesCS2B01G488000 (alignment length 433bp, score 1630, e-value <0.01 and identity 81%). This gene is also seen in the peak interval that was defined here and is one of only 10 of the gene candidates that we report in Table 2. There are 9 homozygous Cadenza specific alleles in the Yr7 bulk in this gene TraesCS2B01G488000 (in a minimum of 70% of the sequencing reads), of which, 4 are predicted to be non-synonymous resulting in codon changes (Table 3). These are potential markers for the Yr7 gene against a Chinese Spring background. Interestingly, there were no homozygous Avalon specific alleles in the yr7 bulk in this gene, although



Table 3. Annotation of Cadenza specific alleles that are homozygous in the Yr7 bulk.

Position	Ref allele	Alt allele	Concordance	Region	Gene	SNP status	Ref codon	SNP codon	Sample
685269082	A	G	0.75	CDS	TraesCS2B01G488000	Non-Syn	E	G	Yr7
685269083	A	C	0.70	CDS	TraesCS2B01G488000	Syn	E	E	Yr7
685269107	G	C	0.72	CDS	TraesCS2B01G488000	Syn	L	L	Yr7
685269174	G	A	0.89	CDS	TraesCS2B01G488000	Non-Syn	D	A	Yr7
685269205	T	C	0.91	CDS	TraesCS2B01G488000	Non-Syn	I	T	Yr7
685269217	C	T	0.82	CDS	TraesCS2B01G488000	Non-Syn	A	V	Yr7
685269221	G	A	0.83	CDS	TraesCS2B01G488000	Syn	L	L	Yr7
685269242	A	G	0.89	CDS	TraesCS2B01G488000	Syn	L	L	Yr7
685269266	A	G	0.91	CDS	TraesCS2B01G488000	Syn	L	L	Yr7

<https://doi.org/10.1371/journal.pone.0231157.t003>

there was sufficient sequencing coverage in this region to define SNPs. Therefore, it appears that Avalon matches Chinese Spring closely in this region.

The 5 SNPs that we previously found in our candidate genetically defined interval ‘Block 6’ that showed the highest-log(P) scores in the GWA scan (Fig 2), were found in the 10Mbp interval 683,029,160–693,314,916bp on chromosome 2B. This interval encompasses our candidate gene homolog TraesCS2B01G488000 (Table 2) that is located at 685,265,721–685,270,904bp. Therefore, gene capture from the same Yr7 and yr7 bulks combined with this mapping-by-sequencing mutant identification pipeline accurately reproduced and refined the results from the genetic mapping analysis to an interval containing 10 genes and ultimately a strong candidate gene. It uses a sliding window scoring algorithm to score regions of interest by prioritizing long homozygous parental haplotypes to smooth out noise in the homozygosity signal.

## Conclusions

### Mapping of Yr7

A strong body of evidence links yellow rust resistance present in the cultivars ‘Lee’ and ‘Thatcher’ (used as differential hosts) and their derivatives with the Yr7 genetic locus. Yr7 was named by Macer in 1963 [29]. Pathotyping studies led to the postulation of Yr7 in French varieties ‘Camp Remy’ and UK varieties ‘Tonic’, ‘Cadenza’ and ‘Brock’ [3]. Yr7 has been tentatively mapped by association with QTL in several previous studies. [30] interpreted a QTL (QYR1, gwm501-gwm47) interval on chromosome2B observed in their Camp Remy x Michigan Amber RIL population as Yr7. Furthermore, the Yr7 gene was mapped to the long arm of chromosome 2B [7] and recently, a candidate gene validated by isolating several independent susceptible mutants from the ‘Cadenza’ background [6]. Here we show the progression from genetic mapping to mapping-by-sequencing and the overlap of the defined intervals by each approach culminating with interval refinement and identification of a logical candidate gene using mapping-by-sequencing.

In this work, we are able to re-map Yr7 with great precision, identifying a set of only 10 gene candidates that matched both positional and functional annotation criteria, based only on our own results. Comparison of this list with the published identity of Yr7 by [6] which emerged while this work was in progress showed that in fact, our shortlist of 10 NBS-LRRs contained the ‘Chinese Spring’ homolog of Yr7, thus giving retrospective validation to our method. We highlight the benefit of combining our mapping-by-sequencing approach with the near complete wheat reference genome sequence (RefSeqv1) in refining our candidate interval. Furthermore, this was possible even though the reference sequence is based on the variety Chinese Spring while our trait was found in the variety Cadenza. The identification of



Yr7 by [6] exploited the availability of a saturated mutant population in the resistant background 'Cadenza' to isolate multiple loss-of-function alleles and targeted enrichment sequencing of disease resistance genes (MutRenSeq). This highly targeted sequencing approach and dominant loss-of-function screen will not always be available. The novelty of this study lies in our ability to fine-map directly in hexaploid wheat using whole exome capture sequencing data and a hybrid GWAS/mapping-by-sequencing approach that could be applied to any trait, rather than just in application to disease resistance or NLR genes.

**A modified homozygous haplotyping algorithm and mapping-by-sequencing pipeline for hexaploid species.** Here, we have taken yellow rust resistant and susceptible bulked doubled-haploid populations, performed gene enrichment and, using mapping-by-sequencing, identified a 60Mbp region on chromosome 2B that contains the Yr7 locus. The candidate window encompassed 589 high confidence genes, including 10 disease resistance associated genes that included the closest Chinese Spring homolog (TraesCS2B01G488000) of the candidate Yr7 gene reported previously [6].

This analysis has taken the principles demonstrated by SHOREmap [31] and applied them to polyploid wheat using a homozygote-scoring algorithm that highlights longer homozygous haplotypes shared between the mutant parental line and the bulked dataset. This extends a proof of concept approach combining genic enrichment and a sliding window mapping-by-synteny analysis using a pseudo-genome. This was firstly carried out in the diploid wheat species *T. monococcum* identifying a region on chromosome 3 that was likely to contain the *Eps-3A<sup>m</sup>* deletion that results in early flowering [4] and later in hexaploid wheat to map the Yr6 yellow rust resistance locus [23]. Here, this sliding window mapping-by-sequencing analysis, implementing wheat chromosomal pseudomolecules directly, has been successfully applied to define an interval of interest in a polyploid species.

Within the defined candidate gene, we were able to select homozygous alleles at positions that were conserved with the parental lines. This revealed a small subset of 9 Cadenza specific alleles in the Yr7 bulk, of which, 4 are predicted to be non-synonymous and are potential markers for the Yr7 gene. The identification of a small genetic region containing a candidate gene of interest in a hexaploid demonstrates the power of this analysis and its broad applicability.

### Combining QTL, GWAS and mapping-by-synteny approaches to pinpoint candidate causative polymorphisms

Each of the approaches taken has contributed something to the discovery of potential causative SNPs underlying the Yr7 locus. The line-by-line SNP genotyping of the biparental population provided an ordered genome scan that treated homoeologues as distinct. The GWAS approach gave further context as to the population history of Yr7 deployment and the historical recombination evident in the variety panel provided more genetic discrimination than many hundreds more AxC lines could possibly have provided, a fact that was particularly important given the unfavorable physical: genetic distance ratio found in the Yr7 interval. The nature of this particular YR panel i.e. the fact that it is composed purely of commercial UK winter wheat varieties and important historic founders provides breeders with knowledge of those lineages in which Yr7 is found and those where it is not, allowing them to pick the most appropriate Yr7 donor and flanking markers to suit the objectives of the cross and the constraints of their marker assisted selection schemes. Finally, the mapping-by-sequencing defined a list of only 10 genes of which one was the candidate gene itself and also generated additional markers for this gene. We noted that the SNPs that we found in our candidate genetically defined interval closely overlapped the interval defined by mapping-by-sequencing.

## Perspectives for gene cloning in hexaploid wheat

Although the technique of mapping is by now an established method in the genetic toolbox of species which boast a fully assembled genome, polyploid large-genome crop species whose genomes have continued to evade gold standard sequencing and are expensive to re-sequence are still problematic to work with. [32] showed the application of transcriptome re-sequencing from model bulks in a tetraploid species, which offered one route to the rapid saturation of genetic intervals defined by bulked segregants provided the target genes were expressed to an adequate level. [4] solved the issue of expression variation inherent in the transcriptome re-sequencing approach by employing exome re-sequencing for gene mapping in hexaploid wheat. In this work, we go a step further by successfully deploying a mapping-by-sequencing approach in a hexaploid using a near complete reference genome sequence. Although the homozygous haplotype scoring map (Fig 4) was initially created blind to knowledge of bulk composition or expected genomic location, it succeeded in remapping the expected interval and in saturating the interval with parent-specific polymorphisms without significant discovery bias and in a manner independent of expression level of the underlying gene. The peak detection sensitivity will depend on the validity of the consensus gene order across all homoeologous genomes and the gene representation in the exome capture. However, evidently this is not an issue for the mapping of Yr7.

Short lists of candidate genes or polymorphisms produced by the hybrid mapping-by-sequencing/GWAS approach demonstrated here can be subjected to validation of their function in wheat. Current methods of choice include *in silico* TILLING (if the wild type background possesses the trait of interest) [33] or gene editing of multiple homeoalleles (which can in principle be applied in any genetic background) [34].

## Supporting information

**S1 Table. Creation of Yr7 and yr7 bulks.**

(XLSX)

**S2 Table. Segregation of the infection type at seedling stage in AxC population against isolate 08/21.**

(XLSX)

**S3 Table. Avalon x Cadenza (AxC) Doubled Haploid population—Yellow rust seedling test scores.**

(XLSX)

**S4 Table. YR Panel of wheat varieties—Yellow rust seedling test scores.**

(XLSX)

## Acknowledgments

The authors would like to thank BBSRC and the John Oldacre Foundation for funding and Amelia Hubbard (UKCPVS) for provision of pathotyped yellow rust isolates.

## Author Contributions

**Conceptualization:** Anthony Hall, Donal M. O'Sullivan.

**Data curation:** Pauline Bansept-Basler.

**Formal analysis:** Laura-Jayne Gardiner, Pauline Bansept-Basler, Mohamed El-Soda.

**Funding acquisition:** Anthony Hall, Donal M. O'Sullivan.

**Investigation:** Laura-Jayne Gardiner, Pauline Bansept-Basler, Mohamed El-Soda.

**Methodology:** Laura-Jayne Gardiner, Pauline Bansept-Basler.

**Project administration:** Donal M. O'Sullivan.

**Resources:** Anthony Hall.

**Supervision:** Anthony Hall, Donal M. O'Sullivan.

**Writing – original draft:** Laura-Jayne Gardiner, Pauline Bansept-Basler, Donal M. O'Sullivan.

**Writing – review & editing:** Laura-Jayne Gardiner, Pauline Bansept-Basler, Mohamed El-Soda, Anthony Hall, Donal M. O'Sullivan.

## References

1. Beddow JM, Pardey PG, Chai Y, Hurley TM, Kriticos DJ, Braun H-J, et al. Research investment implications of shifts in the global geography of wheat stripe rust. *Nature Plants*. 2015; 1: 15132. <https://doi.org/10.1038/nplants.2015.132> PMID: 27251389
2. Ma J, Wingen L, Orford S, Fenwick P, Wang J, Griffiths S. Using the UK reference population Avalon × Cadenza as a platform to compare breeding strategies in elite Western European bread wheat. *Molecular Breeding*. 2015; 35: 1–18.
3. Singh RP, Singh PK, Rutkoski J, Hodson DP, He X, Jørgensen LN, et al. Disease impact on wheat yield potential and prospects of genetic control. *Annual Review of Phytopathology*. 2016; 54: 303–322. <https://doi.org/10.1146/annurev-phyto-080615-095835> PMID: 27296137
4. Gardiner L-J, Bansept-Basler P, Olohan L, Joynson R, Brenchley R, Hall N, et al. Mapping-by-sequencing in complex polyploid genomes using genic sequence capture: a case study to map yellow rust resistance in hexaploid wheat. *The Plant Journal*. 2016; 87: 403–419. <https://doi.org/10.1111/tpj.13204> PMID: 27144898
5. Johnson R. Past, present and future opportunities in breeding for disease resistance, with examples from wheat. *Euphytica*. 1992; 63: 3–22.
6. Marchal C, Zhang J, Zhang P, Fenwick P, Steuernagel B, Adamski NM, et al. BED-domain-containing immune receptors confer diverse resistance spectra to yellow rust. *Nature Plants*. 2018; 4: 662–668. <https://doi.org/10.1038/s41477-018-0236-4> PMID: 30150615
7. Zhang P, McIntosh RA, Hoxha S, Dong C. Wheat stripe rust resistance genes Yr5 and Yr7 are allelic. *Theor Appl Genet*. 2009; 120: 25–29. <https://doi.org/10.1007/s00122-009-1156-5> PMID: 19763533
8. McGrann GRD, Smith PH, Burt C, Mateos GR, Chama TN, MacCormack R, et al. Genomic and genetic analysis of the wheat race-specific yellow rust resistance gene Yr5. *Journal of Plant Science and Molecular Breeding*. 2014; 3: 2.
9. Chen X, Coram T, Huang X, Wag Mand Dolezal A. Understanding Molecular Mechanisms of Durable and Non-durable Resistance to Stripe Rust in Wheat Using a Transcriptomics Approach. *Curr Genomics*. 2013; 14(2): 111–126 <https://doi.org/10.2174/1389202911314020004> PMID: 24082821
10. Agenbag GM, Pretorius ZA, Boyd LA, Bender CM, MacCormack R, Prins R. High-resolution mapping and new marker development for adult plant stripe rust resistance QTL in the wheat cultivar Karioga. *Molecular Breeding*. 2014; 34: 2005–2020.
11. Zegeye H, Rasheed A, Makdis F, Badebo A, Ogbonnaya FC. Genome-wide association mapping for seedling and adult plant resistance to stripe rust in synthetic hexaploid wheat. *PLoS One*. 2014; 9: e105593. <https://doi.org/10.1371/journal.pone.0105593> PMID: 25153126
12. Mackay I, Powell W. Methods for linkage disequilibrium mapping in crops. *Trends in Plant Science*. 2007; 12: 57–63. <https://doi.org/10.1016/j.tplants.2006.12.001> PMID: 17224302
13. Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, Nordborg M, et al. Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLoS Genetics*. 2010; 6: e1000940. <https://doi.org/10.1371/journal.pgen.1000940> PMID: 20463887
14. Varshney RK, Terauchi R, McCouch SR. Harvesting the Promising Fruits of Genomics: Applying Genome Sequencing Technologies to Crop Breeding. *Plos Biology*. 2014; 12: e1001883. <https://doi.org/10.1371/journal.pbio.1001883> PMID: 24914810

15. Takagi H, Abe A, Yoshida K, Kosugi S, Natsume S, Mitsuoka C, et al. QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *The Plant Journal*. 2013; 74: 174–183. <https://doi.org/10.1111/tpj.12105> PMID: 23289725
16. Allen AM, Barker GLA, Wilkinson P, Burridge A, Winfield M, Coghill J, et al. Discovery and development of exome-based, co-dominant single nucleotide polymorphism markers in hexaploid wheat (*Triticum aestivum* L.). *Plant Biotechnology Journal*. 2013; 11: 279–295. <https://doi.org/10.1111/pbi.12009> PMID: 23279710
17. Hubbard A, Lewis CM, Yoshida K, Ramirez-Gonzalez RH, de Vallavielle-Pope C, Thomas J, et al. Field pathogenomics reveals the emergence of a diverse wheat yellow rust population. *Genome Biology*. 2015; 16:23. <https://doi.org/10.1186/s13059-015-0590-8> PMID: 25723868
18. McNeal FHUSDoA. *A uniform system for recording and processing cereal research data*. [ Beltsville, Md.]: Agricultural Research Service, United States Department of Agriculture. 1971
19. Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, et al. Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnology Journal*. 2014; 12:787–796. <https://doi.org/10.1111/pbi.12183> PMID: 24646323
20. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity*. 2005; 95(3):221–7. <https://doi.org/10.1038/sj.hdy.6800717> PMID: 16077740
21. Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, et al. Efficient control of population structure in model organism association mapping. *Genetics*. 2008; 178: 1709–1723. <https://doi.org/10.1534/genetics.107.080101> PMID: 18385116
22. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, et al.: genome association and prediction integrated tool. *Bioinformatics*. 2012; 28: 2397–2399. <https://doi.org/10.1093/bioinformatics/bts444> PMID: 22796960
23. Gardiner L-J, Gawroński P, Olohan L, Schnurbusch T, Hall N, Hall A. Using genic sequence capture in combination with a syntenic pseudo genome to map a deletion mutant in a wheat species. *The Plant Journal*. 2014; 80: 895–904. <https://doi.org/10.1111/tpj.12660> PMID: 25205592
24. IWGSC (The International Wheat Genome Sequencing Consortium) et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*. 2018; 361: 6403.
25. Chapman J, Mascher M, Buluc A, Barry K, Georganas E et al. A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. *Genome Biology*. 2015; 16: 26. <https://doi.org/10.1186/s13059-015-0582-8> PMID: 25637298
26. Li H and Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25: 1754. <https://doi.org/10.1093/bioinformatics/btp324> PMID: 19451168
27. Li H, Handsaker B, Wysoker A, 1000 Genome Project Data Processing Subgroup et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25: 2078–9. <https://doi.org/10.1093/bioinformatics/btp352> PMID: 19505943
28. Koboldt D, Zhang Q, Larson D, Shen D, McLellan M, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research*. 2012; 22: 568–576. <https://doi.org/10.1101/gr.129684.111> PMID: 22300766
29. Macer RCF. The formal and monosomic genetic analysis of stripe rust (*Puccinia striiformis*) resistance in wheat. *Second International Wheat Genetics Symposium*, Vol. 2. Lund, Sweden: Hereditas; 1963. pp. 127–142.
30. Boukhatem N, Baret PV, Mingeot D, Jacquemin JM. Quantitative trait loci for resistance against Yellow rust in two wheat-derived recombinant inbred line populations. *Theoretical and Applied Genetics*. 2002; 104: 111–118. <https://doi.org/10.1007/s001220200013> PMID: 12579435
31. Schneeberger K, Ossowski S, Lanz C, Juul T, Petersen AH, Nielsen K et al. SHOREmap: simultaneous mapping and mutant identification by deep sequencing. *Nature Methods*. 2009; 6: 550–551 <https://doi.org/10.1038/nmeth0809-550> PMID: 19644454
32. Trick M, Adamski N, Mugford S, Jiang C, Febrer M and Uauy C. Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. *BMC Plant Biology*. 2012; 12: 14. <https://doi.org/10.1186/1471-2229-12-14> PMID: 22280551
33. Uncovering hidden variation in polyploid wheat. Krasileva KV, Vasquez-Gross HA, Howell T, Bailey P, Paraiso F, Clissold L, et al. *Proceedings of the National Academy of Sciences*. 2017; 114:E913–E921.
34. Wang Y, Cheng X, Shan Q, Zhang Y, Liu J, Gao C, et al. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nature Biotechnology* 2014; 32:947–951. <https://doi.org/10.1038/nbt.2969> PMID: 25038773