Science AAAS

Supplementary Materials for

Identification of Wheat Gene *Sr35* that Confers Resistance to Ug99 Stem Rust Race Group

Cyrille Saintenac, Wenjun Zhang, Andres Salcedo, Matthew N. Rouse, Harold N. Trick, Eduard Akhunov, Jorge Dubcovsky.

correspondence to: eakhunov@ksu.edu & jdubcovsky@ucdavis.edu

Supplementary Materials and Methods

Plant Materials and Stem Rust Assays

The *Sr35* high-density genetic map was generated from two populations derived from the cross between *T. monococcum ssp. monococcum* accession G2919 (= PI428170, donor of *Sr35*) and the susceptible *T. monococcum* ssp. *aegilopides* accession TA189 (= PI427796). The first population included 1,925 F_2 plants and the second population 725 BC_1F_1 lines from the backcrossing of TA189 as female to the F_1 hybrid. The first 269 BC_1F_1 were used to generate the map of *Sr35* published before (*8*). Stem rust seedling assays with races RKQQC (isolate 99KS76A) and QTHJC (isolate 75ND717C) were performed at Kansas State University (KSU) and at the USDA Cereal Disease Laboratory (CDL). Assays with races TTKSK (Ug99; isolate 04KEN156/04) and TRTTF (isolate 06YEM34-1) were performed at a biosafety level-3 facility (University of Minnesota).

Frozen urediniospores stored at -80 °C were heat-shocked for 6-10 min in a water bath at 42°C prior to use. Spores were suspended in Soltrol 170 light oil (Chevron Phillips Chemical Company, The Woodlands, TX). Inoculations were performed by air-spray using a compressor at the two-leaf stage (KSU) or one-leaf stage (CDL). Immediately after inoculation, plants were incubated in a dew chamber at 22 °C with 100% relative humidity in the dark for 16 h. Plants were then moved to a growth chamber (KSU) or greenhouse (CDL) running at 22 °C day and 18 °C night with 16 h of photoperiod. Disease Infection Types (ITs) were assessed 12-14 days after inoculation. Progeny-tests of the five critical recombinants are provided in Table S1.

Marker Development and Genotyping

Brachypodium genes colinear to the wheat *Sr35* region (Fig. 1A) were used to search the NCBI nucleotide databases and the 454 sequences of *T. aestivum* cv. Chinese Spring (*25*) using BLASTN. Wheat putative orthologues were re-sequenced in *T. monococcum* parental lines G2919 and TA189. The list of primers used for re-sequencing is provided in Table S3. PCR reactions were performed using AmpliTaq® Gold polymerase DNA (Applied Biosystems) and PCR products were purified with Exonuclease I and Shrimp alkaline phosphatase at 37 °C for 15 min. Sanger sequencing reactions were performed from purified PCR fragments using BigDye® Terminator v3.1 Cycle *Sequencing* Kit (Applied Biosystems). Reactions were processed on the ABI3730xl instrument. Alignments and detection of SNPs were performed using the Sequencher package (Gene Codes, Ann Arbor, MI, USA). SNP genotyping was done with KASP™ SNP genotyping assays (LGC Genomics Hoddesdon Herts, UK). Reactions were performed in 8 µl final volume and run in the StepOnePlus™ Real-Time PCR System (Applied Biosystems) according to the following protocol: 94 °C for 15 min, 35 cycles of 94 °C for 10 s and 60 °C for 30 s, and a reading at 35 °C for 30 s.

BAC Library Screening, Fingerprinting, and Construction of Physical Map

The BAC library of *T. monococcum* accession DV92 (*9*) was screened using PCR primers for proximal markers *SFGH* and *AK331487*. DNAs of 23 selected BACs were extracted using R.E.A.L. Prep 96 Plasmid Kit (Qiagen). Fingerprinting reactions were performed in a 3730 ABI instrument as described before (*26*). The resulting ABI files were processed with GenoProfiler and analyzed with FPC (FingerPrinted Contigs) (*27*). Contigs were built using a cut off value of $1e^{-45}$ and are presented in Fig. S1. Overlapping BACs 502I15, 64A22 and 245M16 covering the complete *Sr35* region were sequenced.

Filters for the BACs library of *T. urartu* accession G1812 (*28*) were screened using a probe derived from the AP2 domain-containing pseudogene (primers AP2_F1 and AP2_R2, Table S4) using hybridization protocols described before (*29*). The selected BAC clones were further tested with PCR primers CNLF and CNL9_R6 designed for a conserved region of the *CNL4* and *CNL9* genes. The single *T. urartu* BAC (288D18) that included both *AP2* and *CNL4/9* was sequenced.

BAC Sequencing and Annotation

BACs were sequenced using a combination of high coverage Roche 454 sequencing (~100X for the *T. monococcum* BACs and 124X for the *T. urartu* BAC) complemented with dideoxy Sanger sequencing of large sub-clones (6-8 kb) to resolve large repeats and to correct 454 sequencing errors in long homopolymer tracks. All gene sequences were confirmed by Sanger sequencing. Sequences were assembled using Newbler v. 2.6 (https://wiki.gacrc.uga.edu/wiki/Newbler#Version) and gaps were finished by Sanger sequencing of sub-clones spanning the gaps.

As a first step in the annotation process, repetitive elements were identified using the *Triticeae* Repeat Sequence Database (http://wheat.pw.usda.gov/ITMI/Repeats/). The non-

repetitive sequence was annotated using BLASTN searches in GenBank against the wheat EST collection at GrainGenes (http://www.graingenes.org/), the TIGR Wheat Genome Database (http://tigrblast.tigr.org/euk-blast/index.cgi?project=tae1) and gene models generated by comparison of diploid and tetraploid wheat transcriptomes (http://wheat.pw.usda.gov/GG2/WheatTranscriptome/) with the *T. aestivum* genomic sequence. Additional BLASTX searches were performed against the non-redundant GenBank plant proteins. Annotated sequences without gaps were submitted to GenBank using Sequin.

The orthologous *T. monococcum* (KC573058) and *T. urartu* (KC816724) sequences were compared using Artemis Comparison Tool (*30*) using a 92% identity threshold (Fig. 1E-F). Duplicated repetitive regions in non-colinear locations were removed for clarity.

Phylogenetic Analyses

The phylogenetic analysis in Fig. S2 included intact *CNL* genes from *T. monococcum* accession DV92 (*TmCNL1*, *TmCNL2*, *TmCNL4*, *TmCNL6*, and *TmCNL9*, GenBank KC573058) and *T. urartu* accession G1812 (*TuCNL-A*, *TuCNL-C*, *TuCNL-D*, GenBank KC816724). The *TmCNL* genes were then used to screen the *T. turgidum* cv. Kronos transcriptome (http://wheat.pw.usda.gov/GG2/WheatTranscriptome/), the genomic DNA sequences of *T. aestivum* cv. Chinese Spring chromosomes 3A and 3B (http://wheat-urgi.versailles.inra.fr/Seq-Repository), and the non-redundant database at NCBI (*Ta*AK331487) by BLASTN (>80% identity and E value < $1e^{-100}$). Complete *CNL* genes were identified in contigs Tt_k51_contig_23942 and Tt_k25_contig_23552 from *T. turgidum*, and in contigs TaIWGSC_chr3AL_k71_contig4437008 and TaIWGSC_chr3B_k71_contig_10707498 from *T. aestivum*. Genes from the colinear region from *Brachypodium distachyon* were described before (*8*).

Sequences of these *CNL* genes are relatively well conserved even in the LRR domain, which facilitated the alignment of both protein and coding DNA sequences. Sequences were aligned using Muscle (http://www.drive5.com/muscle/) as implemented in Mega v. 5 (*31*). Mega was used to construct the phylogenetic trees using the neighbor-joining method. The Poisson and Maximum Composite Likelihood models were used to calculate pair-wise distances in protein and nucleotide sequence alignments, respectively. Bootstrap confidence values were calculated based on 1000 iterations.

Screening and Characterization of EMS Mutants

G2919 seeds were soaked in water for 8 h and treated in ethyl-methanesulfonate (EMS) solution for 16 h at room temperature under shacking at 100 rpm. Seeds were rinsed thoroughly with tap water for 4 h. Two different doses of EMS (0.15 and 0.2 %) were used to generate the population. EMS-treated M_1 plants were grown under greenhouse conditions (22/18 °C day/night temperatures) and spikes were covered to avoid cross-pollination. Eight seeds per M_2 family were planted for 651 families, and less than five seeds were planted for the other 436 families. At the two-leaf stage, plants were inoculated with *Pgt* race RKQQC (virulent on *Sr21* and avirulent on *Sr35*). Phenotyping was done twelve days after inoculation. Sixteen M_3 plants from M_2-

families *cnl9*^{1296} and *cnl9*^{1120} identified as susceptible were inoculated with RKQQC to confirm homozygocity of the mutations.

Re-sequencing of the four candidate genes in the two mutants (primers provided in Table S3) revealed mutations only in *CNL9*. Mutant *cnl9*^{1296}showed a G to A mutation characteristic of the EMS treatment, but mutant *cnl9*^{1120} showed three closely linked mutations (positions 854, 856 and 858) that are not the expected result of EMS mutagenesis. To test the possibility that these changes were the result of seed contamination or accidental cross-pollination during the development of the mutant population, we estimated the level of SNPs between the wild type genotype G2919 and the two mutants. We also determined the relative proportion of homozygous and heterozygous mutations and compared them with the expected ratios (Table S6). All these studies were performed using genotyping-by-sequencing (GBS) following a protocol described before (*11*).

Briefly, complexity-reduced genomic libraries were prepared for each of the three genotypes using restriction enzymes *Pst*I and *Mse*I and a set of barcoded adapters with sticky ends complimentary to the 3' overhangs of the two restriction enzymes. The digested and heat-inactivated DNA samples (400 ng) were ligated with adapters using QIAquick PCR Purification Kit (Qiagen) and pooled. Libraries were amplified and purified using Qiagen columns. After estimating the average DNA-fragment length using Bioanalyzer (~240 bp), each library was diluted to 10 nM concentration and sequenced on Illumina MiSeq (2 x150 bp read run). A custom Perl pipeline was used to process and analyze complexity-reduced genomic sequence data. Reads were quality trimmed from each side and barcode sequences were used to assign reads to individual genotypes. Sequence reads were clustered using CD-HIT (*32*) using a 100% similarity threshold. One representative read from each G2919 cluster was used as a reference for mapping mutant sequence data using *bowtie* (*33*) permitting 3 mismatches in the first 28 bases. Variant discovery and genotype calling was performed using SAMTools software (*34*). Variable sites showing more than 2 states were removed, and only sites that had total coverage of at least 10 reads and at least 2 reads per SNP allele were used for genotype calling. Results of the GBS analysis are summarized in Table S6 and Supplementary text.

Prediction of Sites under Positive Selection

To test if the CNL9 amino acid positions 854, 856, and 858 associated with susceptibility in *cnl9*^{1120} and in PI428167-2 were under positive selection we used the method implemented in the HyPhy program (*35*). We aligned the coding sequences of 14 wheat and *Brachypodium* CNL genes described in the Supplemental Materials section "Phylogenetic analyses", and screened 943 codons for evidence of selection using a multi-partition fixed effects likelihood approach (*36*). The method corrects for recombination by performing selection tests within segments of a gene located between recombination breakpoints. This approach reduces the proportion of false-positives among identified sites by taking into account the effect of recombination on phylogeny (*37*). The MG94xHKY85 evolutionary model was fit to data and used to estimate the rate of synonymous (*dS*) and non-synonymous (*dN*) mutations at each codon along the branches of the phylogenetic tree. The likelihood ratio (LRT) test was used to compare the maximum-likelihood

estimates of *dN* and *dS* rates with the null model assuming *dN* = *dS*. A site was considered positively selected if the LRT *P*-value was significant and *dN* > *dS*. The codons subjected to positive selection are listed in Table S7.

Production and Characterization of Transgenic Plants

The hexaploid wheat cultivars Fielder and Bobwhite (susceptible to stem rust races RKQQC and Ug99) were transformed with a genomic construct including *Sr35* with its native promoter. The construct used for transformation included 2,462 bp upstream of the start codon, the complete coding region and 2,615 bp downstream of the stop codon (including the three 3'UTR introns and exons). The genomic DNA fragment was amplified using primers CNL9_cloning_F1 and CNL9_cloning_R1 and an annealing temperature of 65°C (Table S4). The amplified fragment was purified from an agarose gel with the QIAEX II gel extraction kit (Qiagen). A-overhang was added with one unit of NEB *Taq* polymerase at 72°C for 20 min and the fragments were subcloned into the pCR® II-TOPO® vector (Invitrogen). DNA from this plasmid was sequenced by Sanger to validate the construct. Plasmids were used for biolistic co-transformation with plasmid pAHC20 containing the *bar* gene under the control of the maize ubiquitin promoter (*38*) using protocol described before (*39*).

Briefly, the premature seeds were surface sterilized with 20% sodium hypochlorite and 0.02% TWEEN-20. Immature embryos were then excised on CM4 media to initiate the formation of somatic embryo. Somatic embryos proliferated in CM4+ osmoticum (0.2 M mannitol, 0.2 M sorbitol) were co-bombarded with pAHC20 and *Sr35* plasmids at 1:1 ratio by using the particle inflow gun. Wheat calli were placed on CM4 medium containing 5 mg/L glufosinate. Cultures were kept in the medium of 10 mg/L glufosinate for 10-15 weeks. The growing embryogenic tissues were transferred to shoot production medium (MSP) with 5 mg/L glufosinate selection until green shoots were observed. The cultures were then re-transferred to elongation and rooting medium (MSE) containing 5 mg/L glufosinate but not 2,4-D for 2-3 weeks.

To examine the expression of the *bar* resistance gene, freshly prepared 0.2% (v/v) solution of herbicide Liberty ® was applied on the second/third youngest leaf of the T_0 plants. Positive lines with resistant green leaves were selected for further PCR screening analysis. Each plant was tested for the presence of *bar* and *Sr35* genes using a set of gene-specific primers (Table S4). RNA was isolated from infected leaves and *Sr35* expression was estimated by qRT-PCR. Transgenic plants at the tillering stage were inoculated with *Sr35*-avirulent races RKQQC and TTKSK and with *Sr35*-virulent race QTHJC to confirm race specificity.

We used four pairs of PCR primers (CNL9_F4/CNL9_R4, Table S3 and M13/CNL9_R2, CNL9_F1/R1 and CNL_F19 and T7, Table S4) to screen the T_0 plants for the presence of the transgene, and selected three putative transgenics in Fielder (#962, #1163, and #1123) and one in Bobwhite (#1007). T_1 plants from the four putative transgenic events were genotyped with the same primers described above and phenotyped with races TTKSK (Ug99), RKQQC and QTHJC. Only the #1123 T_1 lines segregated for the transgene and for resistance to Ug99 (Fig. 2C). *CNL9*

transcript levels in the resistant transgenic and in the non-transgenic control plants are presented in Fig. S4.

Progeny tests were performed for three T_1 plants showing susceptible, resistant, and intermediate phenotypes. Fifteen to nineteen T_2 plants per each T_1 plant were inoculated with TTKSK (Ug99), RKQQC and QTHJC as described before. Quantification of fungal DNA was performed following published protocols (*40*). Briefly, genomic DNA was isolated from infected primary leaves collected for all genotypes simultaneously. DNA was quantified by SYBRGreen® Super Mix (BioRad, Hercules CA) and was diluted to 3 ng/µl. The wheat *Phytochelatin synthase* gene (BJ274652) (*41*) was used as internal control. The number of *Pgt-Actin* DNA molecules relative to the number of *T. aestivum* (*Ta*)*Phytochelatin synthase* DNA molecules was calculated by the formula $2^{(C_T\ Ta\text{-}Phytochelatin\ synthase}$ - $C_T\ Pgt\text{-}Actin\ )}$, where C_T is the threshold cycle.

5' and 3' RACE

Rapid amplification of cDNA ends (RACE) was performed using G2919 cDNA obtained before stem rust inoculation as described above and the FirstChoice® RLM-RACE Kit (Invitrogen). The outer PCR of the 5' RACE reaction was performed using the 5' outer primer provided in the kit combined with the specific primer CNL9-5'race_R1 (Table S4). The inner PCR of the 5' RACE was carried out using the 5' inner primer provided in the kit and the specific primer CNL9-5'race_R20 (Table S4). The outer PCR of the 3' RACE reaction was performed using the outer 3' primer provided by the kit and the specific primer CNL9-3'race_F17 (Table S4). Finally, the inner PCR of the 3' RACE used the 3' inner primer provided by the kit and the specific primer CNL9-3'race_F21 (Table S4).

PCR products from the 5' and 3' RACE reactions were purified with the QIAquick PCR purification kit (Qiagen). Fragments were sub-cloned into the pCR® II-TOPO® vector (Invitrogen) and transformed into EC100 electro-competent *E. coli*. A total of 48 colonies were PCR amplified using M13 primers. PCR fragments were sequenced by Sanger method on an ABI 3730xl, and sequences were aligned to *T. monococcum* genomic sequence KC573058 to identify the transcriptional initiation and termination sites for *CNL9*.

Expression Analysis

Transcript levels of the four candidate genes (*APGG1, CNL4*, *CNL6*, and *CNL9*) were determined by quantitative reverse transcription polymerase chain reaction (qRT-PCR) (Fig. 2D), using the wheat *Phytochelatin synthase* gene (BJ274652) as internal control (*41*). Transcript levels were expressed as the ratio between the initial numbers of molecules in the target and the internal control (fold-*Phytochelatin synthase* levels) using the $2^{-\Delta C_T}$ method as described before (*42*). Primer efficiencies were evaluated for the different primers combinations using serial five-fold dilutions of G2919 cDNA (1/5, 1/25, 1/125, 1/625) and only primers with efficiencies higher than 95% were used for the qRT-PCR experiments.

Leaves from six different plants of *T. monococcum* G2919 (*Sr35*-resistant) were collected six days after inoculation with *Sr35*-virulent race RKQQC. Total RNA was extracted using Trizol protocol (Invitrogen) following manufacturer's instructions. RNA quality was estimated using Bioanalyser (Agilent), quantified with Nanodrop-1000 (Thermo Scientific), and treated with DNAse I (Invitrogen Inc, Carlsbad, CA, USA). The first strand cDNA was synthesized from 5 µg of total RNA with Superscript III and anchored oligo $dT_{(20)}$ (1h at 50°C). Quantitative PCR were carried out using a Biorad iCycler IQ Real Time PCR using 2X IQTM SYBR® Green Super Mix reagent (BioRad, Hercules, CA). The following thermal profile was applied to all reactions: 95 °C for 5 m; 40 cycles of 95 °C for 30 s, 58 °C for 30 s, 72 °C for 40 s. At the end of this profile the PCR samples were subjected to a dissociation analysis to confirm specificity of amplification. Three independent reactions were performed for each biological replicate. All primers sequences are described in the Table S5.

Alternative Splicing

RT-PCR of G2919 cDNA samples with *CNL9* specific primer CNL9_F5 and primer CNL9_R15 (Table S5) revealed the presence of two alternative splicing forms. Both amplification products were cloned in the pCR® II-TOPO® vector (Invitrogen) and sequenced using the Sanger method in an ABI 3730xl. The main isoform showed correct splicing of all four introns, but the second isoform retained the third intron (2^{nd} intron of the 3' UTR) (Fig. S6A). Retention of the same intron was observed in transcripts from *T. turgidum* ssp. *durum* (Table S9).

To determine the relative transcript levels of the two alternative splice variants, we developed *CNL9* isoform-specific PCR primers CNL9Main_rv_2 and CNL9_iso_intron_rv_2 (Fig. S6B and Table S5). Primer specificities were determined using DNAs from *T. monococcum* BAC clones carrying a single CNL gene (Fig. S6B). These primers were used to amplify the two isoforms from leaves of six plants of *T. monococcum* accession G2919 collected before (0 h) and after inoculation with RKQQC (24 h, 48 h, 96 h and 144 h). A mock inoculated control was analyzed in parallel (Fig. 2E).

Supplementary Text

Genotyping-by-sequencing

In total, we obtained 6,074,021 sites covered by at least 10 reads in each line. We identified 65 SNPs between *cnl9*^{1120} and G2919 out of which 37 were homozygous and 28 heterozygous. There were a total of 55 SNPs between *cnl9*^{1296} and G2919 of which 24 were homozygous mutations and 31 were heterozygous sites. The overall mutation density in *cnl9*^{1120} and *cnl9*^{1296} was 1 mutation per 93 kb and 110 kb, respectively. These values were very similar to the EMS mutation density previously reported for a *T. monococcum* TILLING population (1 mutation per 92 kb) (*43*).

The ratios between homozygous and heterozygous plants were also as expected for M_2 and M_3 mutants. The re-sequenced M_2 mutant *cnl9*^{1296} showed a ratio of 24 homozygous to 31 heterozygous, which was not significantly different from the expected ratio of 1 homozygous to 2 heterozygous (Table S6, $\chi^2 = 1.4$, $P = 0.24$). The re-sequenced M_3 mutant *cnl9*^{1120} showed a ratio of 37 homozygous to 28 heterozygous mutations, which was not significantly different from the expected ratio of 3 homozygous to 2 heterozygous mutations (Table S6, $\chi^2 = 0.13$, $P = 0.72$).

The total number of mutations detected in *cnl9*^{1120} and *cnl9*^{1296} was not significantly different between mutants ($\chi^2 = 0.83$; $P = 0.3613$), and is 300-fold lower than the number of polymorphisms expected between different *T. monococcum* accessions (*44*). These results indicate that it is very unlikely that *cnl9*^{1120} mutant resulted from seed contamination or accidental cross-pollination.

Supplementary Figures

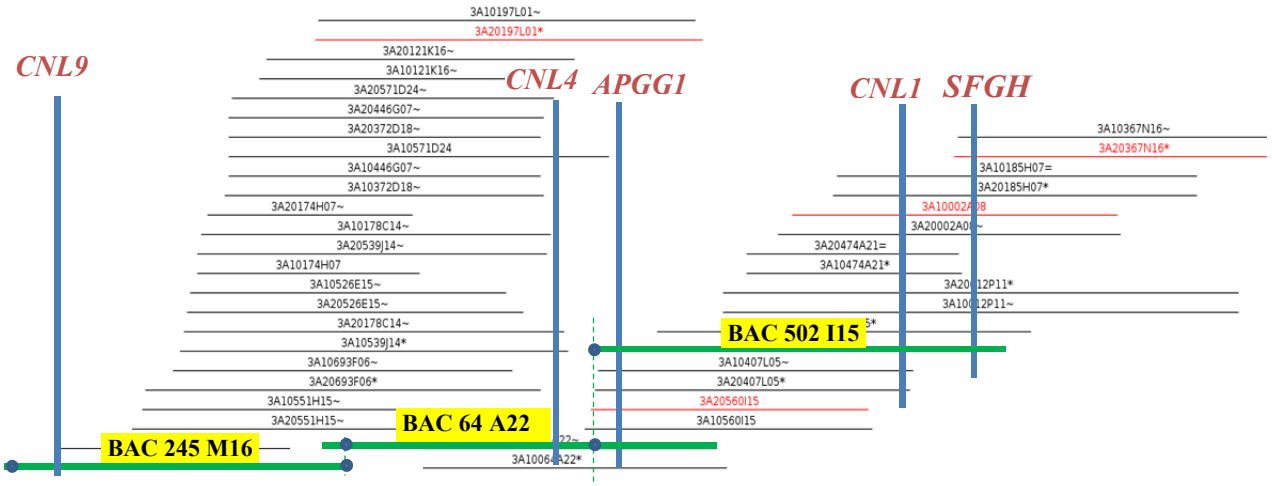

Fig. S1.

Fingerprints of 23 *T. monococcum* BACs selected with *Sr35* proximal markers *SFGH* and *AK331487* and assembled with FPC. AK331487 is a *CNL* gene and the selected primers amplified multiple members of the *CNL* family present in this contig. Two independent DNA samples were fingerprinted from each BAC. The FPC parameters used for the assembly included: overlap=15, Max=50, from end=0 and Min shared bands=14. The BACs selected for sequencing are indicated in green. The approximate locations of genes *CNL9*, *CNL4, APGG1*, *CNL1*, and *SFGH* are indicated by vertical blue lines (the figure is oriented with the proximal region to the right). Blue dots represent BAC end markers used to validate the overlap between the BACs selected for sequencing.

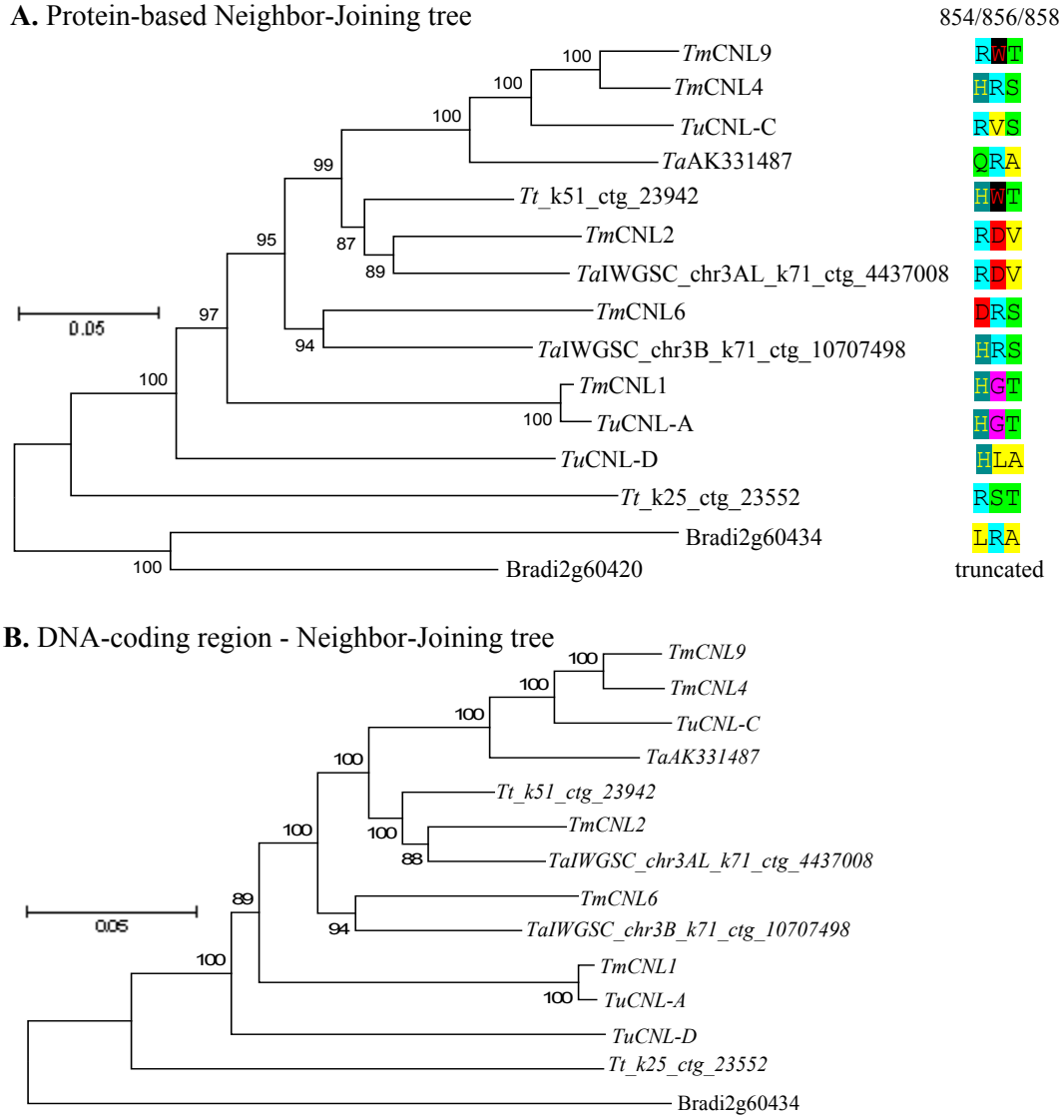

Fig. S2.

Neighbor-Joining tree of complete (**A**) CNL like proteins, and (**B**) DNA coding regions. Sequences within this group are conserved and easy to align even in the LRR region. For sequences sources see Supplemental Material and Methods section Phylogenetic analyses. Sequences were aligned with Muscle as implemented in Mega v. 5 (*31*). Mega was also used to generate the phylogenetic trees using the pair-wise deletion method. Bootstrap confidence values based on 1000 iterations are indicated in the nodes. The letters to the right of the protein-based tree indicate the three amino-acids at positions 854, 856 and 858 affected by the putative conversion event in mutant *cnl9*^{1120} and accession PI428167-2 that resulted in susceptibility to Ug99. Note that all Ug99-resistance accessions of *T. monococcum* (*CNL9*) have a distinctive combination of amino acids at these positions (Table S2). Positions 856 and 858 showed evidence of positive selection (see Table S7).

A

>TmCNL9 *T. monococcum* DV92: sites predicted to be under positive selection (Table S7)

MEIAMGAIGSLLPKLGELLIGEITLEKKVRKGIESLITELKLMQAVLSKVSKVPADQLDEGVKIWAGNVKELSYQMEDIVDAFMVR
VGDGGESTNPKNRVKKILKKVKKLFKNGKDLHRISAALEEVVLQAKQLAELRQRYEQEMRDTSANTSVDPRMMALYTDVTELVGIE
ETRDKLINMLTEGDDWSKHPLKTISIVGFGGLGKTTLAKAAYDKIKVQFDCGAFVSVSRNPEMKKVLKDILYGLDKVKYENIHNAA
RDEKYLIDDIIEFLNDKRYLIVIDDIWNEKAWELIKCAFSKKSPGSRLITTTRNVSVSEACCSSEDDIYRMEPLSNDVSRTLFCKR
IFSQEEGCPQELLKVSEEILKKCGGVPLAIITIASLLANKGHIKAKDEWYALLSSIGHGLTKNRSLEQMKKILLFSYYDLPSYLKP
CLLYLSIFPEDREIRRARLIWRWISEGFVYSEKQDISLYELGDSYFNELVNRSMIQPIGIDDEGKVKACRVHDMVLDLICSLSSEE
NFVTILDDPRRKMPNSESKVRRLSIQNSKIDVDTTRMEHMRSVTVFSDNVVGKVLDISRFKVLRVLDLEGCHVSDVGYVGNLLHLR
YLGLKGTHVKDLPMEIGKLQFLLTLDLRGTKIEVLPWSVVQLRRLMCLYVDYGMKLPSGIGNLTFLEVLDDLGLSDVDLDFVKELG
RLTKLRVLRLDFHGFDQSMGKALEESISNMYKLDSLDVFVNRGLINCLSEHWVPPPRLCRLAFPSKRSWFKTLPSWINPSSLPLLS
YLDITLFEVRSEDIQLLGTLPALVYLEIWNYSVFEEAHEVEAPVLSSGAALFPCATECRFIGIGAVPSMFPQGAAPRLKRLWFTLFP
AKWISSENIGLGMRHLPSLQRVVVDVISEGASREEADEAEAALRAAAEDHPNRPILDIW*

B

Distribution of the 25 sites predicted to be under positive selection (Table S7)

	Length	Observed	Expected [1]	Chi-square test
Outside domains	82	1	2.23	
Coiled-coil	116	0	3.16	
Nucleotide binding	352	6	9.58	*P*= 0.0037[2]
Leucine-rich repeat	369	18	10.04	
N-proximal-half	184	4	9	
C-distal-half	184	14	9	*P*= 0.0184

[1] Based on relative length of the region

[2] The coiled-coil and outside regions were merged for the Chi-square test

Fig. S3.

Distribution of amino acids showing evidence of positive selection in CNL9. (**A**) Protein sequence. Pink highlight: coiled-coil domain, green highlight: nucleotide binding domain, light-blue highlight: leucine rich repeat, red letters with yellow highlight: amino acids showing significant evidence of positive selection (Table S7). The first exon is underlined and the second exon is not. The squares around position 854, 856 and 858 indicate the three amino acids differences detected in the Ug99 susceptible mutant *cnl9*^{1120}. (**B**) Distribution of amino acids showing evidence of positive selection. Note that the distribution along the protein is not proportional to the length of the domains (*P*=0.0037). Within the leucine-rich repeat the sites with evidence of positive selection are concentrated in the C-terminal half (*P*=0.0047) of the domain. Several studies have shown that the C-terminal half of the leucine-rich domain is frequently involved in pathogen recognition and that is frequently subject to diversifying selection (*15, 45*).

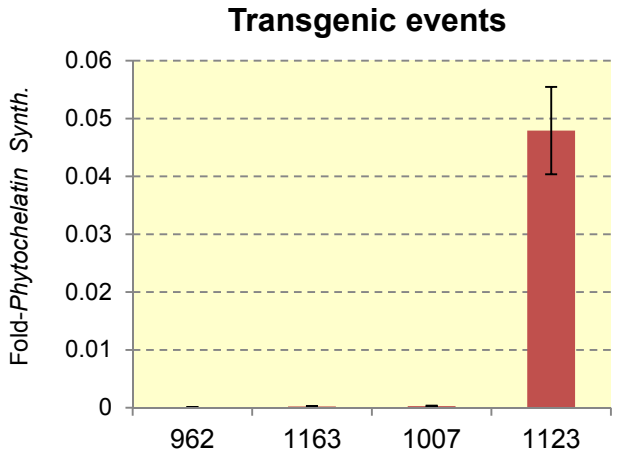

Fig. S4.

Transcript levels of *CNL9* in four putative T_0 transgenic plants. Quantitative RT-PCR using primers NL9_F22 and CNL9Main_rv_2 (Table S5) showed *CNL9* expression only in event #1123. In the T_1 progeny of the four putative T_0 transgenic plants, the presence of the transgene was confirmed only in #1123, which suggests that either the other three T_0s were false positives or the transgene was lost. The *CNL9* transcript levels in the positive #1123 transgenic hexaploid plants were roughly 9-fold lower than the transcript levels observed in *T. monococcum* G2919 in Fig. 2D. In the T_1 and T_2 progeny of #1123, the plants carrying the transgene were resistant to Ug99 (Fig. 2C, Fig. S5, and Table S8).

.

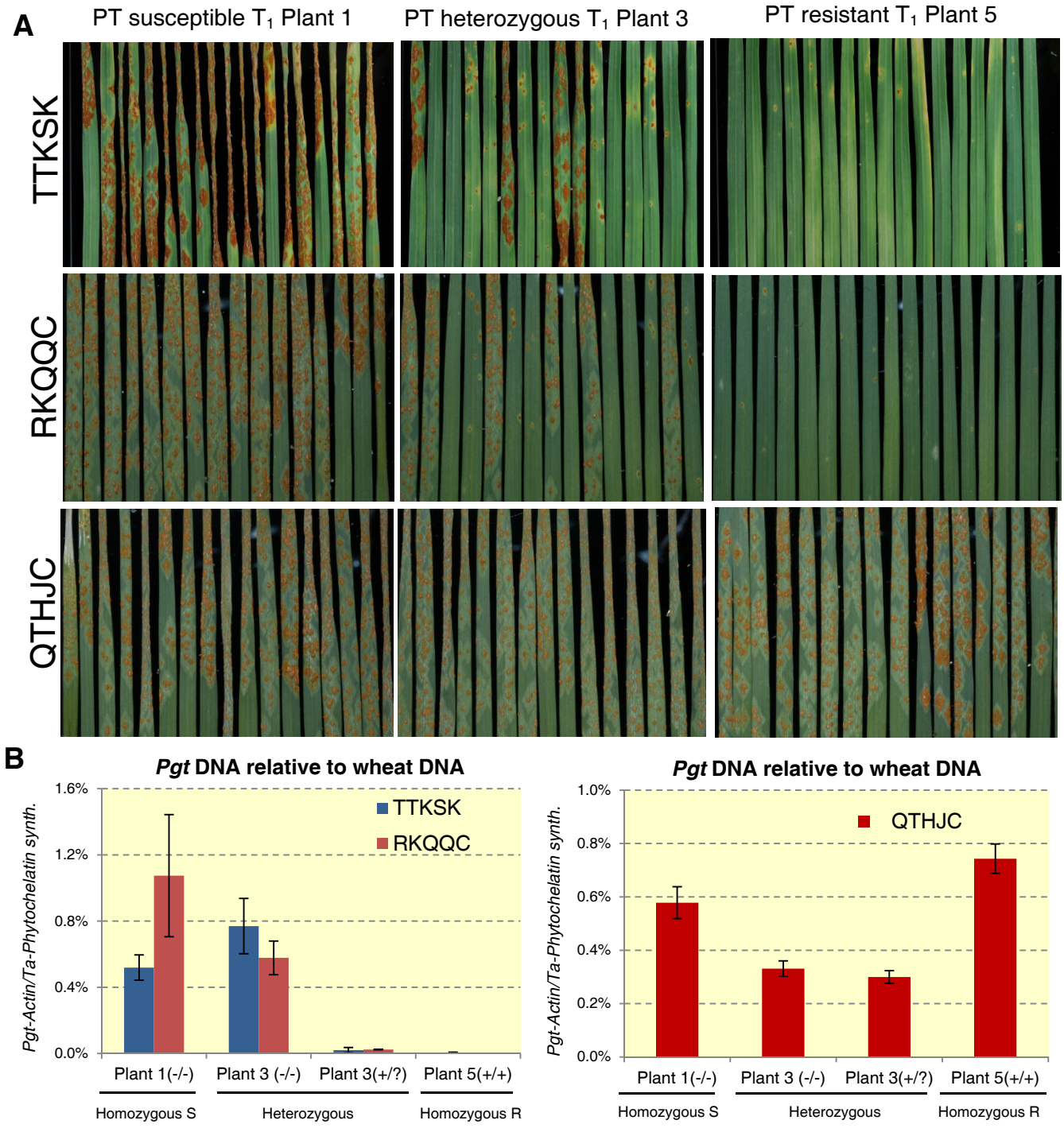

Fig. S5.

Progeny tests of *CNL9* transgenic hexaploid wheat event #1123. Progeny of T_1 plants 1, 3 and 5 were inoculated with two races avirulent on *Sr35* (TTKSK and RKQQC) and one virulent to *Sr35* (QTHJC). (**A**) Column 1: progeny of T_1 plant 1 homozygous for the absence of *CNL9* (-/-). Column 2: progeny of T_1 plant 3 heterozygous for *CNL9* (+/-). Column 3: progeny of T_1 plant 5 homozygous for the presence of *CNL9* (+/+). (**B**) Quantitative PCR of genomic DNA of *Pgt* races TTKSK, RKQQC, and QTHJC relative to wheat DNA for the same plants described in A. Since we only have a dominant marker for the transgene, "+/?"= +/+ or +/- in T_2 plants.

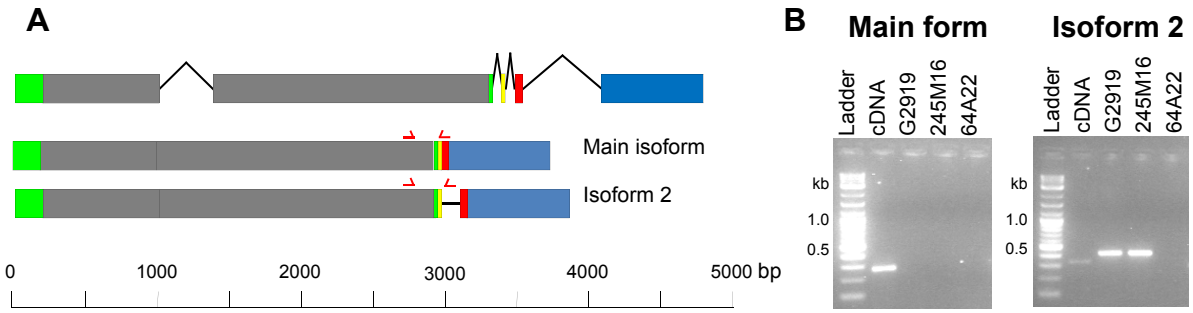

Fig. S6.

Alternative splicing forms of *Sr35*. (**A**) The *CNL9* gene has two transcribed exons (in gray) and three exons and three introns in the 3' UTR. The gene encodes two isoforms that differ in the absence (main form) or presence (isoform 2) of the third intron (2nd intron of the 3' UTR). The positions of the isoform-specific primers are indicated with red half arrows. (**B**) Validation of isoform-specific primers for RT-PCR. PCR was performed using cDNAs from G2919 seedlings, G2919 genomic DNA, and DNAs from BAC clones 245M16 (*CNL9*) and 64A22 (*CNL4*). Left panel: primers NL9_F22 and NL9Main_rv_2 are specific for the main CNL9 isoform and only amplify from cDNA samples (273 bp product). Right panel: Isoform 2 specific primers NL9_F22 and NL9_iso_intron_rv_2 amplify products of 302 bp from cDNA and 390 bp from genomic DNA due to the presence of an additional intron. Note the specificity of these primers for the BAC containing *CNL9*. Primers are listed in Table S5.

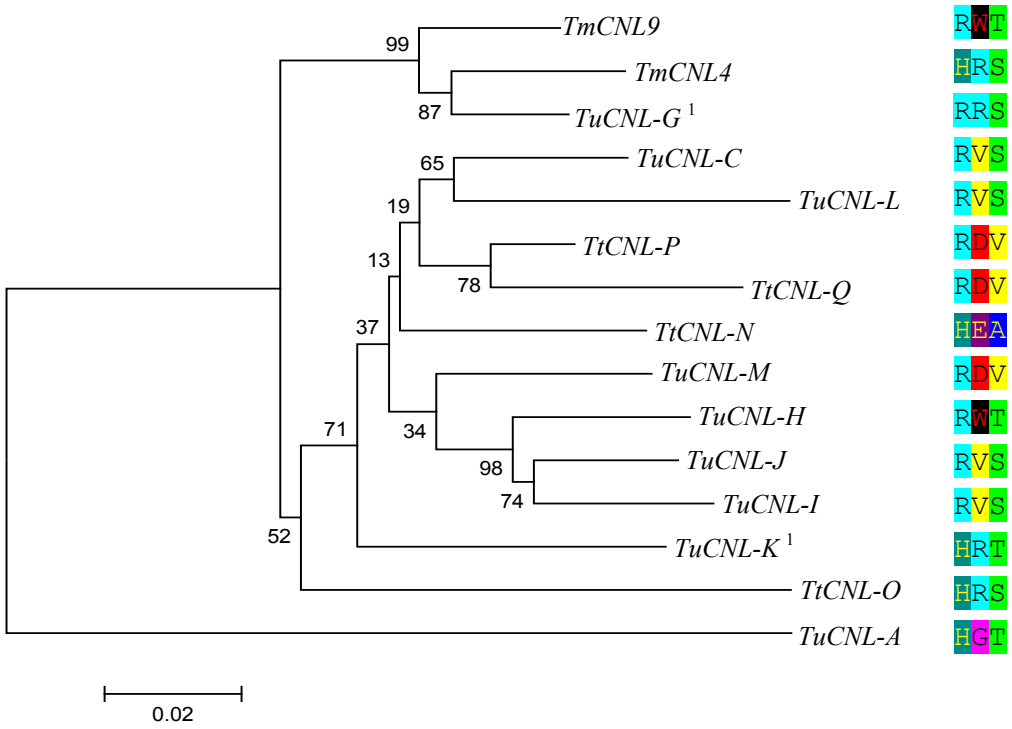

Fig. S7.

Neighbor-Joining tree of the distal region of *CNL-C* like genes identified in a collection of 41 *T. urartu* accessions and 12 *T. turgidum* ssp. *dicoccoides* accessions (Table S10). The distal region of exon 2 and the first 158 bp of the 3'UTR were amplified with primers CNL_F3 and CNL9_R6 (Table S3), which are conserved between *TmCNL4*, *TmCNL9* and *TuCNL-C* genes. The tree was generated using Mega version 5.1 (*31*). Bootstrap confidence values based on 1000 iterations are indicated in the nodes. *TuCNL-A* was included as an out-group. The letters to the right of the tree indicate the three predicted amino-acids affected by the putative conversion event in *cnl9*^{1120} and PI428167-2 (positions 854, 856 and 858), which resulted in susceptibility to Ug99. Amino acid position 856 and 858 showed evidence of selection (Table S7).

[1] Sequences from two independent genomic regions (BE406908 and DMC1) show that accessions classified as *T. urartu* haplotypes *TuCNL-G* and *TuCNL-K* in the NSGC, are more similar to *T. monococcum* accessions DV92, G3116, G2919, T189 (99.9 to 100% identical) than to the other five *T. urartu* haplotypes (96.7 to 99.0 identical). These results suggest that these accessions are misclassified.

Table S1

Phenotypic characterization of the five critical recombinant lines with *Pgt* race TTKSK (Ug99). The *Sr35* candidate region (213 kb) is delimited by markers 502-I15 $_{F2R2}$ and pAP2. Primers are listed in Table S4.

Recombinant #	Proximal marker	Distal marker	Ug99 Segregation
F_2TA189-1826	502-I5$_{F2R2}$ (SR[1])	APGG1map (RR)	0 S & 21 R (RR)
F_2TA189-53	pCNL7 (SS)	pAP2 (SR)	17 S & 0 R (SS)
BC_1F_1TA189-768	pCNL7 (SS)	pAP2 (SR)	13 S & 0 R (SS)
F_2TA189-1231	pAP2 (RR)	245M16R-end (SR)	0 S & 18 R (RR)
F_2TA189-1836	pAP2 (SR)	245M16R-end (SS)	3 S & 13 R (SR)

[1] SS= homozygous susceptible, RR= homozygous resistant, SR= Heterozygous.

Table S2.

Natural variation in *Sr35* candidate genes among a collection of 49 *T. monococcum* accessions previously characterized for their *Sr35*-mediated resistance to stem rust (*6*). Primers used for re-sequencing are described in Table S3.

Haplo-type.	Acc. No.	RKQQC[a]	*APGG1*	*CNL4*	*CNL6*	*CNL9*	Accessions
R1	9	R	+[b]	+D[c]	+	+	DV92, PI119422-1[d], PI167591-1, PI167611-1, PI264935, PI277137, PI428161, PI428164, PI428175
R2	15	R	+	+G	+	+	G2919 (=PI428170), PI119422-2, PI191381-1, PI191383, PI272560, PI306543, PI343181, PI355534, PI355536, PI428152-1, PI428166-1, PI428167-1, PI428169, PI538721, PI538722-1
S1	1	S	+	+G	+	+3 SNPs	PI428167-2
S2	2	S	+	+G	+	+3 SNPs/ 242bp del[e]	PI355537, PI538722-2
S3	2	S	+	+G	+	Del.	PI119422-3, PI3555526
S4	7	S	Del.	+G	Del.	Del.	PI191381-2, PI272557, PI272558, PI277138, PI306542, PI352475, PI355543
S5	3	S	Del.	Del.	+	Del.	TA416, PI355523, PI542473
S6	10	S	Del.	Del.	Del.	Del.	G3116, TA189, PI167526, PI167591-2, PI167611-2, PI221416, PI277140, PI355528, PI428152-2, PI428166-2

[a] *Pgt* race RKQQC is virulent on *T. monococcum Sr21* resistance gene but avirulent on *Sr35* and therefore can be used to detect the presence of *Sr35* independently of *Sr21*.

[b] "+" indicates presence of the gene and no-sequence modifications compared to G2919 unless followed by a difference.

[c] In *CNL4* exon 2, we detected two haplotypes characterized by a 6 bp indel and 4 SNPs within a 16 bp segment. The "D" haplotype (found first in DV92) codes for the amino acids sequence "EWAHQL", whereas the "G" haplotype (found first in G2919) codes for the amino acids "DWVP".

[d] Seven accessions have two haplotypes and one (PI119422) three haplotypes, which are indicated by a -1, -2 and -3 after their respective PI identification numbers.

[e] Two of the three accessions carrying the three SNPs in *CNL9* also have a 242 bp deletion in the same gene, which likely occurred after the proposed conversion event that generated the 3 SNPs.

Table S3.

Primers used for re-sequencing genes in the *Sr35* region in different germplasm.

Forward	Sequence	Reverse	Sequence
SFGH_F1	TGCCCCATGACCTTCTCCCT	SFGH _R1	CCAACACCAAAATCCCAACTATCTG
SFGH _F2	AGGCAGATAGTTGGGATTTTGGTGT	SFGH _R2	AAGGGCAGTTTATTGGGTTAGCA
SFGH _F3	GGGCACAACTAAATCAGAATGGGAG	SFGH _R3	TGCACGCCTCCTCGAAGTTG
AK331487 _1F	GGCCCTCTCCTCCCGAAGCT	AK331487 _1R	CGAAACCGAGACAAAAGCACCAC
AK331487 _2F	CCAAAGCAGCATACGACAAGATCA	AK331487 _2R	GGCTTCCGGGGCTCTTCTTG
AK331487 _3F	AAGCCGACTAATCACGACAACCC	AK331487 _3R	CGGCAACCTTCTACATTTCCTTCA
AK332451_F1	GCGCTTGCTAGGTGTGGATCA	AK332451_R1	CCCCAATGATCACAGCACCCACT
AK332451_F2	GGGTCTTCTTACAAACTCTGGAGGA	AK332451_R2	GGCACCATCACCAAAAATTTCTG
AK332451_F3	TGGTGATGGTGCCGAAAGTCTGT	AK332451_R3	CCACCAGCCTCCGAGTCTTCAGA
BM138354_F1	GGCATGGTCCACAGATCAGCA	BM138354_R1	TCCTGGTCCTCGGTGAAGCTC
BM138354_F2	AGTCGGACGGGAAGGTGCTCT	BM138354_R2	GACGACAGGAACCCCCACCAG
BM138354_F3	AGTCGGACGGGAAGGTGCTCT	BM138354_R3	TCCGCTCAAATGTCTCCCACC
AK331482_1F	CGATGGCGAAGCAAGGTGTT	AK331482_1R	GCTGGACGTTGGCCTTGACC
AK331482_2F	GGACAGCGACGGCCAGAAGA	AK331482_2R	TTCTCTGTCGGGTCCCCTGTCTT
CJ656351_1F	CCCCGCTGGATGACGTGAAA	CJ656351_1R	CAGGTTGAGCTATTAGGAGTTTGTC
CJ656351_2F	TGCAGAAGAAATTGGCGTGGACAT	CJ656351_2R	TCTGATTGCTGAATAGAACCCTAACG
CJ656351_3F	TTGCAGAAGAAATTGGCGTGGAC	CJ656351_3R	TCCGATTCATGGCAGATTCCCTTAC
CNL4_F5	TGACCTGACGTGATTAGAAGC	CNL4_R2	CACACACATAATTGAGATCGTCA
CNL4_F6	TGGGAGGGTAATACAAGTCTC	CNL9_R4	ATACCCAACATCGGAGACATG
CNL9_F4	AGAGAGCAAGGTTCGCAGG	CNL4_R3	TGAACCAACTTTCATCTGACTC
CNL4_F3	CTCATCAATTGCCTGAGCCAAG	CNL9_R6	GTATCTAGCGAACCTCAATCG
CNL6_F2	TCCTCTAGGACACGATTACTC	CNL6_R2	TGGGTAAAACGTCTGTGCAAC
CNL6_F3	CCTAAGTGGACGAGCTCAAC	CNL6_R4	CAAGCCTCTACCTTGTCTTCC
CNL6_F4	ACAGAAGATCAATGTTTGGAGG	CNL6_R3	GCAGTACTTCTATCTTAGTCCT
CNL6_F5	GATACTGATGTTGAGGACCTAC	CNL6_R6	GACCTGACTGAAGTACACATG
CNL9_F1	GGACAACGAGAGAACACACCA	CNL9_R3	GCACATAATTGAGATCGTCG
CNL9_F3	TCTCAATTATGTGCGTGTGAGT	CNL9_R4	ATACCCAACATCGGAGACATG
CNL9_F4	AGAGAGCAAGGTTCGCAGG	CNL9_R5	GAACCAACTTCGTTTTGACGG
CNL9_F5	CTCATCAACTGCTTGAGCGAAC	CNL9_R6	GTATCTAGCGAACCTCAATCG
CNL_F3	ATCTGACCAAGCTCAGGGTG	CNL9_R6	GTATCTAGCGAACCTCAATCG
APGG1_F1	CTCCTCATTTAGGACGCCTTTG	APGG1_R4	CTCGTTCTGCATCCTCACCAG
APGG1_F3	GCTCTCCTCATTGAGCTGGC	APGG1_R3	GGTGGGAGTACTGCTAACTG
APGG1_F4	GGATTGGTGGAGTTTCGCTTG	APGG1_R6	CCAGTTGGATTTTCAATGTGCA

Table S4.

Primers used for library screening, high-density map of critical recombination events (Table S1), 5' and 3' RACE and transgenic production and screening.

Forward	Sequence	Reverse	Sequence	Function
SFGH_BAC_F1	CATTTGCTCCAATTGCTAACCC	SFGH_BAC_R2	GCCAGAAAAGGGATGCTACACT	*Tm* BAC lib.
AK331487_3F	AAGCCGACTAATCACGACAACCC	AK331487_3R	CGGCAACCTTCTACATTTCCTTCA	*Tm* BAC lib.
AP2_F1	TCGGGCGGAAGCGGAGAT	AP2_R2	CGTCGCACCCTCCTCCTCCT	*Tu* BAC lib.
CNLF	CCAACCGTCCCATCCTTGA	CNL9_R6	GTATCTAGCGAACCTCAATCG	*Tu* BAC lib.
APGG1map_F	CTCCATTCAGAAGGTTCGTCAGTT	APGG1map_R	GACGCCTTTGTGCAATTCCTT	Fine-mapping[1]
502I-15_F2	GCAAGTCACTTTACTCGTACC	502-I15_R2	TGAGCTCGGTGATAGAGTTAC	Fine-mapping
pCNL7_F	CCAGGCAGATTAACCAGCAAGAA	pCNL7_R	CAGCGATGCCACAAGTGAGGTT	Fine-mapping
pAP2_LF1	GTAACTTCGGAGTGGAGTTCTC	pAP2_LR1	GGAGAGTTCCATTGGCTGTTC	Fine-mapping
245M16-end_F2	CCAGCTTCTCATTTGACGCCA	245M16-end_R2	CGGTGAGATCAGCTTATTACG	Fine-mapping
CNL9-5'R_R1	GAGAAAGGGCATCGAATCTCTC	CNL9-5'R_R20	GGCTATCGGCTCTCTCCTCCC	5'RACE
CNL9-3'race_F17	CTCCCTTCATCGATTGTCAGTCTCT	CNL9-3'race_F21	TGCAGCCGAGATCAGATGAGAA	3'RACE
CNL9_F1	GGACAACGAGAGAACACACCA	CNL9_R1	GAGAGATTCGATGCCCTTTCTC	Transgenic
CNL9_F4	AGAGAGCAAGGTTCGCAGG	CNL9_R4	ATACCCAACATCGGAGACATG	Transgenic
M13_F	CACACAGGAAACAGCTATGAC	CNL9_R2	CGTCTCTCTGCCATTCCTCCT	Transgenic
CNL9_F19	GCGGGTCGTCTATGCGTGTC	T7	AATACGACTCACTATAGGG	Transgenic
BARabF	CCTGCCTTCATACGCTATTTATTTGC	BARabR	CTTCAGCAGGTGGGTGTAGAGCGT	Transgenic
CNL9_cloning _F1	TGCCTATGCTCTGCGCTCGA	CNL9_cloning _R1	CCCCGTCGTTTAAGCACTTGA	Transgenic

[1] Markers derived from the BAC sequences to delimit the *Sr35* region (Table S1)

Table S5.

Primers used in the expression studies

Forward	Sequence	Reverse	Sequence	Function
NL9_F5	CTCATCAATTGCTTGAGCGAAC	NL9_R15	ATGGGTGGTGGACAGAATGAAGT	Alt.splic.
NL4_qpcrF1	TTGGCTTTCGAGTCAGATG	NL4_qpcrR1	CGAACATAGAAGAATAATCCAGT	qRT-PCR
APGG1_qpcrF3	GCCAGTGCATTCGCCATCAG	APGG1_qpcrR3	AATGGCTTTTGGTGCTGCTTTC	qRT-PCR
Phytochelatin_F1	GTATGTCCTCCTACCTCACGAAGT	Phytochelatin_R1	CGCTGCTGCGATAATCTGCT	Int. control
NL6_qpcrF1	TGATGAGGAAGACAAGGTAGA	NL6_qpcrR1	CTTCCAGTGACATCCAACA	qRT-PCR
NL9_qpcrF2	GGTGCTCCGGCTTGACTTCC	NL9_qprR2	GGCACCCAATGTTCGCTCAA	qRT-PCR
NL9_F22	GGCTTAAACGCCTTTGGTTCA	NL9Main_rv_2*	CGGCTGCCTGAAACTCAATC	Main isoform
NL9_F22	GGCTTAAACGCCTTTGGTTCA	NL9_iso_intron_rv_2**	GAAAACACTTTGTACAAAATCGTGG	Isoform 2

* Specific to the main isoform of *CNL9*.

** Specific to the *CNL9* isoform that retained the third intron (=2nd intron of the 3' UTR).

Table S6.

Genotyping-by-sequencing of *T. monococcum* accession G2919 and mutant lines *cnl9*^{1296} and *cnl9*^{1120}. Expected and observed ratios of homozygous to heterozygous mutations were compared using Chi-Square tests.

Genotype	**Reads passed quality filter**	**Reads mapped to the reference**	**Generation**	**Observed ratio Hom:Het**	**Expected ratio Hom:Het**	χ^2 ***P***
G2919	3,621,929	3,610,138		-	-	
cnl9^{1296}	2,017,687	1,866,001	M_2	24:31	1:2	0.24
cnl9^{1120}	2,610,205	2,358,048	M_3	37:28	3:2	0.72

Table S7.

Predicted sites under positive selection based on HyPhy program (*46*) (*P* threshold =0.1). The coordinates in column one are based on the CNL9 protein. A site was considered positively selected if *dN* > *dS* and the LRT *P*-value was significant. *dS*= synonymous mutations, *dN*= non-synonymous mutations, LRT= likelihood ratio test. The location of these sites in the predicted protein is presented in Fig. S3A and the statistical analyses of their distribution in Fig. S3B. The accessions used in the analysis are described in the Supplemental Material and Methods section "Phylogenetic analyses".

Position CNL9	***dN***	***dS***	**LRT**	***P*-value**
151	3.34	0	5.78	0.016
224	2.22	0	4.23	0.040
249	6.82	0	6.57	0.010
252	1.83	5.00E-09	4.43	0.035
323	3.35	5.00E-09	3.96	0.047
385	3.03	5.00E-09	4.71	0.030
492	6.10	5.00E-09	4.81	0.028
608	2.33	5.00E-09	4.46	0.035
655	4.43	5.00E-09	4.92	0.027
677	2.08	5.00E-09	4.41	0.036
702	3.31	5.00E-09	5.30	0.021
728	3.63	0.5564	4.07	0.044
730	5.55	5.00E-09	7.27	0.007
750	2.03	5.00E-09	4.18	0.041
752	10.99	0	8.54	0.003
787	1.69	5.00E-09	3.95	0.047
799	5.37	5.00E-09	6.98	0.008
803	14.72	0.959814	4.43	0.035
808	2.62	5.00E-09	4.91	0.027
810	3.60	5.00E-09	4.55	0.033
811	4.12	5.00E-09	6.70	0.010
856[1]	14.08	0	3.30	0.069
858[1]	3.62	0.5554	4.30	0.038
883	6.27	5.00E-09	5.89	0.015
886	2.63	5.00E-09	5.23	0.022

[1] These positions are included in the three linked SNPs associated with susceptibility to Ug99.

Table S8.

Segregation of the T_1 plants from transgenic of event #1123. Only the 3+ scores are susceptible (all others indicate different levels of resistance). The scale used for infection types (ITs) is included as a footnote. The "," separating scores indicates separate infection types on a single plant.

Plant No.	***CNL9*** presence	Infection Type [1] **Ug99=TTKSK** (*Sr35*-avirulent)
1[2]	-	3+
2	+	3-c LIF
3[2]	+	3-c LIF
4	+	0, ;
5[2]	+	0, ;
6	+	0, ;, 1
7	+	0, ;, 1
8	+	1, 3-
9	+	1, 3-
10	+	1, 3-
11	+	0
12	-	3+
13	+	0
14	+	1, 3- LIF
15	+	1, 3- LIF
		QTHJC (*Sr35*-virulent)
16	-	3+
17	+	3+
18	+	3+
19	+	3+
20	+	3+

[1] Infection types (ITs) scale (0 to 4) (*47*). IT="0" indicates immune reaction: no uredia or other visible symptoms. IT= ';' indicates a nearly immune reaction: no uredia, but visible hypersensitive flecks. IT= '1' indicates a very resistant reaction: small, round uredia were observed immediately surrounded by necrotic or chlorotic plant tissue. IT= '3' indicates a moderately susceptible reaction: medium, elongated uredia observed often associated with limited chlorotic plant tissue. When multiple infection types were observed on the same leaf, all infection types were recorded in order, starting with the most common IT. Signs '+' or '-' indicate larger or smaller size uredia within each infection type. 'c' indicates substantial plant chlorosis. 'LIF' indicates low infection frequency.

[2] Progeny tests shown in Fig. S5

Table S9.

Structure of the 3'UTR in *CNL* related genes from the *Sr35* cluster in *Triticum* species and in the colinear region of *Brachypodium distachyon*. Only genes with available transcripts and genomic sequences are included. The last column indicates the number of introns in the 3'UTR and the detection of alternative splicing variants. *T. urartu* (*Tu*) and *T. turgidum* ssp. *durum* cv. Kronos (*Tt*) transcripts from http://wheat.pw.usda.gov/GG2/WheatTranscriptome/, *T. aestivum* (*Ta*) genomic sequences from http://wheat-urgi.versailles.inra.fr/Seq-Repository and *Brachypodium distachyon* (*Bd*) transcripts and genomic sequences from http://plants.ensembl.org/Brachypodium_distachyon.

Closest *TmCNL*	**Genomic**	**Transcripts**	**No. of introns**
TmCNL2	*Tu*BAC KC816724 *pCNL-B*	Tu_k41_ctg_14030	3
TmCNL4	*Tu*BAC KC816724 *CNL-C*	Tu_k41_ctg_14031	3
TmCNL4[1]	*Ta*IWGSC_ctg_4437008	Tt_k51_ctg_23942/Tt_k45_ctg_23639	3[2]
~ all	*Ta*IWGSC_ctg_4392395 / *Ta*_UK_ctg14718[3]	Tt_k25_ctg_23552	3
~ all	*Bd*_Chr2g_57,751,622-57,755,669	Bradi2g60420.1	2
~ all	*Bd*_Chr2g_57,761,415-57,765,543	Bradi2g60434.1	2

[1] At the protein level *Ta*_IWGSC_contig_4437008 and Tt_k51_contig_23942 are more similar to TmCNL2, but their 3'UTRs are more similar to *TmCNL4*.

[2] Alternative splicing of the 2nd intron in the 3'UTR, which is retained in the Tt_k51_ctg_23942 transcript but not in the Tt_k51_ctg_23942 transcript.

[3] Two partial *Ta* contigs were required to cover the complete transcript.

Table S10.

Natural variation in the distal region of *CNL* genes (700 bp of second exon and first 158 bp of the 3'UTR amplified using primers CNL_F3 and CNL9_R6, Table S3) in 41 accessions of *T. urartu* (*Tu*) and 19 accessions of *T. turgidum ssp. dicoccoides* (*Tt*). These primers are conserved between *TuCNL-C, TmCNL4* and *TmCNL9*. The "3aa" column indicates the three critical amino acids encoded by these sequences at positions 854, 856 and 858 (replacement of RWT at these positions by HRS in mutant *cnl9*[1120] resulted in susceptibility to Ug99). The phylogenetic relationships based on these partial *CNL* haplotypes are presented in Fig. S7.

Gene	GenBank	No.	3aa	Accession number
TuCNL-C	KC816724	19	RVS	PI428184, PI428193, PI428199, PI428202, PI428203, PI428204, PI428207, PI428216, PI428235, PI428240, PI487268, PI538720, PI538726, G1780[2], G1789, G1791, G1802, G1810, G1812.
TuCNL-G [3]	KC876118	1	RRS	PI428217 (=G1841)
TuCNL-H	KC876119	1	RWT	PI227669 (=G1545)
TuCNL-I	KC876120	3	RVS	PI428224, PI428226, G1903
TuCNL-J	KC876121	13	RVS	PI428188, PI428191, PI428201, PI428212, PI428213, PI428223, PI428238, PI428239, PI428244, PI428245, PI538728, PI538729, PI538730
TuCNL-K [3]	KC876117	2	HRT	PI503319, PI538724
TuCNL-L	KC876116	1	RVS	PI428186
TuCNL-M	KC876115	1	RDV	PI487270
TtCNL-N	KF113357	1	HEA	PI560872
TtCNL-O	KF113356	6	HRS	PI428117, PI538697, PI428107, PI503315, PI428111, PI503313
TtCNL-P	KF113355	3	RDV	PI538699, PI428123, PI428113
TtCNL-Q	KF113354	2	RDV	PI487264, 27-37[4]
TtPseudogene	---	7	---	PI352324, PI428015, PI428141, PI470984, PI503314, PI538681, PI560874

[1] PI and CItr germplasm correspond to Germplasm Resources Information Network (GRIN) numbers.

[2] "G" numbers correspond to accessions collected by B. L. Johnson from the Department of Plant Sciences, University of California, Riverside, (G1780 to G1812 were all collected in Turkey, 40 to 45 km west of Kiziltepe).

[3] Sequences from two independent genomic regions (BE406908 and DMC1) show that accessions classified as *T. urartu* haplotypes *TuCNL-G* and *TuCNL-K* in the NSGC, are more similar to *T. monococcum* accessions DV92, G3116, G2919, T189 (99.9 to 100% identical) than to other five *T. urartu* haplotypes (96.7 to 99.0 identical). These results suggest that these accessions are misclassified.

[4] The 'Location-Genotype' identification number "27-37" is from a *Triticum turgidum* ssp. *dicoccoides* population deposited at the University of Haifa wheat germplasm collection.

Author Contributions

CS and WZ performed most of the experimental work with support from AS and HT. MR was responsible for all disease screenings with Ug99, TRTTF and QTHJC and the corresponding figures. CS, WZ, MR and EA contributed sections to the first draft that was assembled by JD. All authors contributed to the revision of the manuscript. EA and JD proposed the idea, wrote the grants and contributed equally to the direction of the project.