Supporting Information

Zhang et al. 10.1073/pnas.1110552108

Supporting Information Corrected December 29, 2011

SI Materials and Methods

BAC Library Screening and Sequencing of the *Q/q* Locus. *Q/q*-containing BACs from the A and B genome of *Triticum turgidum* cv. LDN (1), and from *Aegilops tauschii* accession AL8/78 (2) were isolated by Southern hybridization using probe Tmap2 as previously described (3). One *T. turgidum* BAC of each type was sequenced by Myriad Genetics, and one *Ae. tauschii* BAC was sequenced at the Washington University Genome Sequencing Center. All other BACs were isolated using the PCR method with *Q/q*-specific primers (Table S4). *Q/q*-containing BACs from A, B, and D genomes of *Triticum aestivum* cvs. RE and CS were isolated from libraries described previously (4, 5). *Triticum urartu* (URA) and *Aegilops speltoides* (SPE) BACs where isolated from pooled libraries as described previously (6, 7). Shotgun sequencing of RE, CS, URA, and SPE BACs was performed at the Centre National de Sequencage as previously described (8).

Identification of Transposable Elements (TEs). TE annotation and prediction were essentially done as previously described (9). TEs were identified by BLASTn searches against two databases of repetitive elements: TREP (http://wheat.pw.usda.gov/ITMI/Repeats/index.shtml) (10) and Repbase (http://www.girinst.org/Repbase_Update.html) (11). Core domains (nucleic acid coordinates of known elements) were identified through BLASTn alignments against TREPn. Long-terminal repeats (LTRs) and limits were identified using BLASTn and CENSOR (12) alignments against Repbase and TREP databases. Putative polyproteins were identified by BLASTx alignments against TREPprot. No a priori cutoff was imposed for BLASTx and BLASTn. A complete reconstruction of nested insertions of TEs and split elements was conducted. TE prediction and classification followed the 80-80-80 rule recommended by the unified classification system for eukaryotic TEs: a query element belongs to the same family as a subject element referenced in Repbase or TREP if it is longer than 80 bp, and its sequence is at least 80% identical over at least 80% of their coding or internal domain, or in the terminal repeat regions, or both (13).

Novel TE families include elements that did not fit the 80-80-80 rule (13) showing weak or no similarity with the referenced TE families. Novel TEs were confirmed and analyzed by LTR_STRUC software (14), and/or BLASTx similarity searches against the NCBI nr database (http://www.ncbi.nlm.nih.gov/), as well as using the DOTTER program (15). When possible (i.e., for complete TEs), target-site duplications were identified to confirm an insertion by transposition. Each BAC sequence was also analyzed using the DOTTER program to identify or confirm direct repeats, LTRs, local duplications, and deletion events as well as MITEs. TEs were named and classified according to the previously suggested nomenclature (i.e., element name, BAC name, discovery appearance rank) and designated as complete, truncated, or degenerated as previously suggested (13) with two exceptions: the Sukkula large retrotransposon derivatives (LARDs) were considered to be *Gypsy*-like because their LTRs showed sequence BLASTn similarity to Erika (Gypsy-like) TEs. We maintained the *Athila* designation for those TEs that were initially referenced in TREP (http://wheat.pw.usda.gov/ITMI/Repeats/) to distinguish them from other Gypsy-like retrotransposons. They were designated Ga (*Gypsy* retrotransposons, which were initially designated as *Athila*) in the nomenclature and annotation files.

Annotation of Other Repetitive DNA. Short repeats were identified either as inverted repeats (by using EINVERTED with default parameters; http://bioweb.pasteur.fr/seqanal/interfaces/einverted.html) or tandem repeats (Tandem Repeat Finder, with default parameters; http://tandem.bu.edu/trf/trf.advanced.submit.html. Repeats longer than 100 bp were included in the annotation files.

Gene Structure Analysis. Gene prediction analysis was conducted for the 18.5% non-TE and nonrepeated DNA, using the FGENESH gene prediction program (with the Monocot matrix) as well as BLASTn, BLASTx, and tBLASTx alignments against dbEST (http://www.ncbi.nlm.nih.gov/), SwissProt (http://expasy.org/sprot/), and the rice genome databases.

Estimation of Retrotransposon Insertion Dates. Retrotransposons with both 5′ and 3′ LTRs and target-site duplications (TSDs) were considered to correspond to original insertions. The number of transitions and transversions in pairs of LTRs was calculated using MEGA3 software (16). Retrotransposon insertion dates were calculated using the substitution rate of 13×10^{-8} substitutions/site/year (17). SEs were estimated using the formula T 1/4 $K2P/2r$ (18). All retrotransposon insertion date estimates are presented in Table S1.

Primer Design and PCR-Based Analysis of Helitron-Insertion Haplotype Diversity. The Primer3 program (19) was used to design PCR primers (Table S4) on the basis of the sequence of gene–TE or TE-unassigned DNA junctions as previously described (20). Several pairs of primers were often used. PCR primers targeting internal TE sites were used in control amplifications. PCR amplification was carried out in a final volume of 10 μL with 200 nM of each dNTP, 500 nM each of forward and reverse primers, and 0.2 units of Taq polymerase (Perkin-Elmer) using the following "touchdown" procedure: 14 cycles of 30 s at 95 °C, 30 s at 72 °C minus 1 °C for each cycle, 30 s at 72 °C followed by 30 cycles of 30 s at 95 °C, 30 s at 55 °C, 30 s at 72 °C, and 1 cycle of 10 min at 72 °C.

cDNA Amplification and Sequencing. Total RNA was isolated from immature spikes of CS, LDN, and RE (with one-tenth of the length of the full-grown spike) using on-column digestion (RNAeasy Plant mini kit, Qiagen) following the manufacture's instructions. cDNA was prepared using TaqMan reverse transcription reagents with Oligo d(T)16 primer and MultiScribe Reverse Transcriptase (Applied Biosystems) as recommended by the manufacturer. *5Bq* and *5Dq* cDNAs were PCR amplified using primers based on the predicted LDN *5Bq* and *Ae. tauschii* *5Dq* cDNA sequences (Table S4). Gene-specific primers were designed from *5Bq* and *5Dq* cDNA sequences for rapid amplification of cDNA ends (RACE) using the BD SMART RACE cDNA amplification kit (BD Biosciences). The 5′ RACE and 3′ RACE ready cDNAs were prepared according to the manufacture's instructions and PCR amplified using primers shown in Table S4. Sequence alignments and amino acid sequence prediction were performed using CLUSTALW and SIXFRAME provided by San Diego Supercomputer Center (SDSC) biology workbench (http://workbench.sdsc.edu/).

Transcript Analysis. Steady-state mRNA level in immature spikes collected from CS, CS del143 (*5AQ* deleted), CS 5BL-14 (*5Bq* deleted), CS 5DL-5 (*5Dq* deleted), and the *5AQ*/*5Dq* double deletion line (CS-5A/5Ddd; both *5AQ* and *5Dq* deleted) was measured by relative quantitative (RQ)-PCR with primers for *5AQ*, *5Bq* (*5Bq.1* and *5Bq.2*), and *5Dq* (*5Dq.1* and *5Dq.2*) using a 7500 Real-Time PCR system (Applied Biosystems). CS-del143 was generated by fast-neutron mutagenesis and contains an interstitial

deletion of ~2 Mb encompassing the *5AQ* locus (3). CS 5BL-14 and CS 5DL-5 are terminal chromosome deletions that lack the distal 25 and 24% of the chromosome 5B and 5D long arms, respectively (21). CS-5A/5Ddd was generated by crossing CS-del143 and CS 5DL-5. A CS–*Triticum dicoccoides* 5A disomic chromosome substitution line CS-DIC 5A (22) was also included in the analysis. In CS-DIC 5A, the chromosome substitution replaces *5AQ* with *5Aq*. An illustration of the genetic structure of these stocks is presented in Fig. S5. The wheat actin gene (Table S4) was used as an internal control for all RQ-PCR experiments. All reactions were done in quadruplicate and each experiment was repeated three times. Specificity of primers designed for *5AQ*, *5Bq*, and *5Dq* (Table S4) was verified using mRNA from the wheat deletion lines. A dissociation analysis was performed after each RQ-PCR assay to confirm that the amplification was specific. The amplification efficiency was tested for each gene using a previously described method (23).

Phylogenetic Analysis. DNA sequence alignments were created using ClustalX 2.0.12 (24) and further analyzed and edited, where necessary, using MacClade 4.08 (Sinauer Associates). Synonymous substitution rates and neighbor-joining phylogenetic trees were calculated using the Nei-Gojobori method with Jukes-Cantor correction for multiple substitutions as implemented in MEGA5 (16). All positions with gaps were excluded. CS *5Bq* sequence was excluded from the initial calculations because of the large deletion in the gene and was added later assuming one observed synonymous difference from RE *5Bq*. Divergence times of the *Q*/*q* genes were calculated using synonymous substitution rates and a molecular clock calibrated using 11.6 Mya for divergence between barley and wheat (17). Errors of the substitution rate calculations inherent in the analyzed sequence data sets (multiple alignments), and divergence times derived from them, were evaluated by the bootstrap method (500 repetitions) implemented by MEGA5. MEGA5 was used to perform molecular clock (third codon position) and Tajima's relative rate tests. Pairwise nucleotide substitution rates for noncoding sequences were calculated using Microsoft Excel after manual parsing sequence alignments with all gap positions excluded. These calculations used a 1-kb window sliding in 1-bp steps.

Morphology and Trait Analysis. The genetic stocks CS, CS-DIC 5A, CS-del143, CS 5BL-14, CS 5DL-5, and CS-5A/5Ddd were grown in a completely randomized design with eight replications to evaluate various morphological and agronomic traits. Plants were grown in a greenhouse at 22–25 °C with a 16-h photoperiod. Eight plants for each genetic stock were grown with one plant per pot. Spike emergence time was measured from planting date to the date that the first spike completely emerged from the boot. Plant height from the soil surface to the tip of the tallest spike was measured when plants were mature (completely senesced) immediately before harvest. At the same time, spike length and number of spikelets per spike were measured for three spikes of each plant. Number of tillers, number of seeds per plant, 1,000-kernel weight, and grain yield per plant were also measured. Number of seeds per spike was calculated by dividing the number of seeds per plant by the number of tillers for the same plant. Spike morphology (square/speltoid) was assessed by manual inspection as previously described (22). Glumes were also manually inspected for shape, thickness, and rigidity. Spikes were threshed manually to evaluate threshability and glume/rachis disarticulation.

1. Cenci A, et al. (2003) Construction and characterization of a half million clone BAC library of durum wheat (Triticum turgidum ssp. durum). *Theor Appl Genet* 107: 931–939.
2. Xu Z, et al. (2002) Construction and characterization of five large-insert BAC and BIBAC libraries of Aegilops tauschii, the diploid donor of the wheat D genome. *10th International Plant and Animal Genome Conference*, ed Heller SR (Scherago International, Jersey City, NJ), p 101.
3. Faris JD, Fellers JP, Brooks SA, Gill BS (2003) A bacterial artificial chromosome contig spanning the major domestication locus Q in wheat and identification of a candidate gene. *Genetics* 164:311–321.
4. Chalhoub B, Belcram H, Caboche M (2004) Efficient cloning of plant genomes into bacterial artificial chromosome (BAC) libraries with larger and more uniform insert size. *Plant Biotechnol J* 2:181–188.
5. Allouis S, et al. (2003) Construction and characterisation of a hexaploid wheat (Triticum aestivum L.) BAC library from the reference germplasm 'Chinese Spring'. *Cereal Res Commun* 31:331–338.
6. Isidore E, Scherrer B, Chalhoub B, Feuillet C, Keller B (2005) Ancient haplotypes resulting from extensive molecular rearrangements in the wheat A genome have been maintained in species of three different ploidy levels. *Genome Res* 15:526–536.
7. Salse J, et al. (2008) New insights into the origin of the B genome of hexaploid wheat: Evolutionary relationships at the SPA genomic region with the S genome of the diploid relative Aegilops speltoides. *BMC Genomics* 9:555.
8. Chantret N, et al. (2005) Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (Triticum and Aegilops). *Plant Cell* 17:1033–1045.
9. Charles M, et al. (2008) Dynamics and differential proliferation of transposable elements during the evolution of the B and A genomes of wheat. *Genetics* 180: 1071–1086.
10. Wicker T, Matthews DE, Keller B (2002) TREP: A database for Triticeae repetitive elements. *Trends Plant Sci* 7:561–562.
11. Jurka J (2000) Repbase update: A database and an electronic journal of repetitive elements. *Trends Genet* 16:418–420.
12. Phillips CM, et al. (2010) ACC2 gene polymorphisms, metabolic syndrome, and gene-nutrient interactions with dietary fat. *J Lipid Res* 51:3500–3507.
13. Wicker T, Yahiaoui N, Keller B (2007) Contrasting rates of evolution in Pm3 loci from three wheat species and rice. *Genetics* 177:1207–1216.
14. McCarthy EM, McDonald JF (2003) LTR_STRUC: A novel search and identification program for LTR retrotransposons. *Bioinformatics* 19:362–367.
15. Sonnhammer EL, Durbin R (1995) A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* 167:GC1–GC10.
16. Kumar S, Nei M, Dudley J, Tamura K (2008) MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform* 9:299–306.
17. Chalupska D, et al. (2008) Acc homoeoloci and the evolution of wheat genomes. *Proc Natl Acad Sci USA* 105:9691–9696.
18. Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16: 111–120.
19. Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 132:365–386.
20. Charles M, et al. (2009) Sixty million years in evolution of soft grain trait in grasses: Emergence of the softness locus in the common ancestor of Pooideae and Ehrhartoideae, after their divergence from Panicoideae. *Mol Biol Evol* 26:1651–1661.
21. Endo TR, Gill BS (1996) The deletion stocks of common wheat. *J Hered* 87:295–307.
22. Faris JD, Gill BS (2002) Genomic targeting and high-resolution mapping of the domestication gene Q in wheat. *Genome* 45:706–718.
23. Simons KJ, et al. (2006) Molecular characterization of the major wheat domestication gene Q. *Genetics* 172:547–555.
24. Chenna R, et al. (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 31:3497–3500.

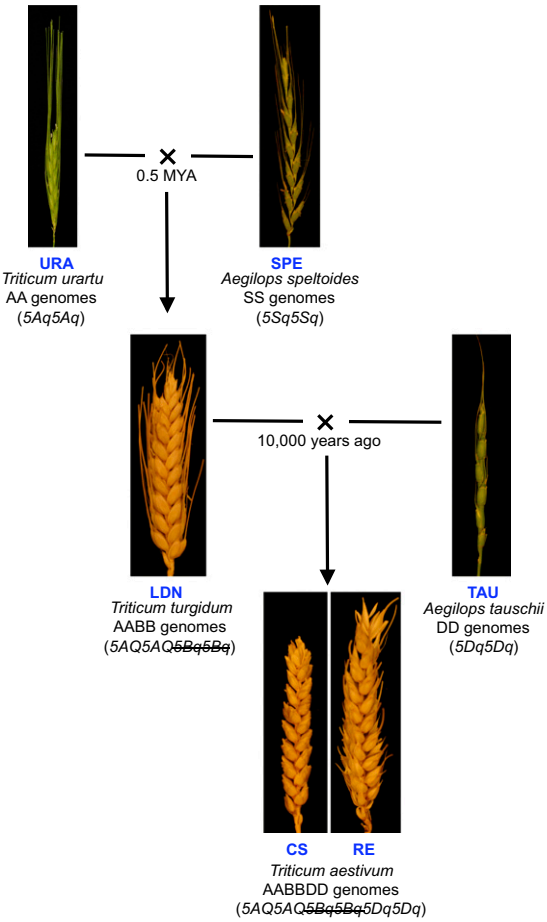

Fig. S1. The amphiploidization events involved in the evolution of common bread wheat illustrated with photographs of spikes of species/cultivars used for BAC sequence analysis of the *Q/q* loci. Designation of accessions used in this research is indicated in blue. Taxonomical names, genome constitutions, and *Q* loci genotypes are shown in parentheses with strikethrough indicating a pseudogene.

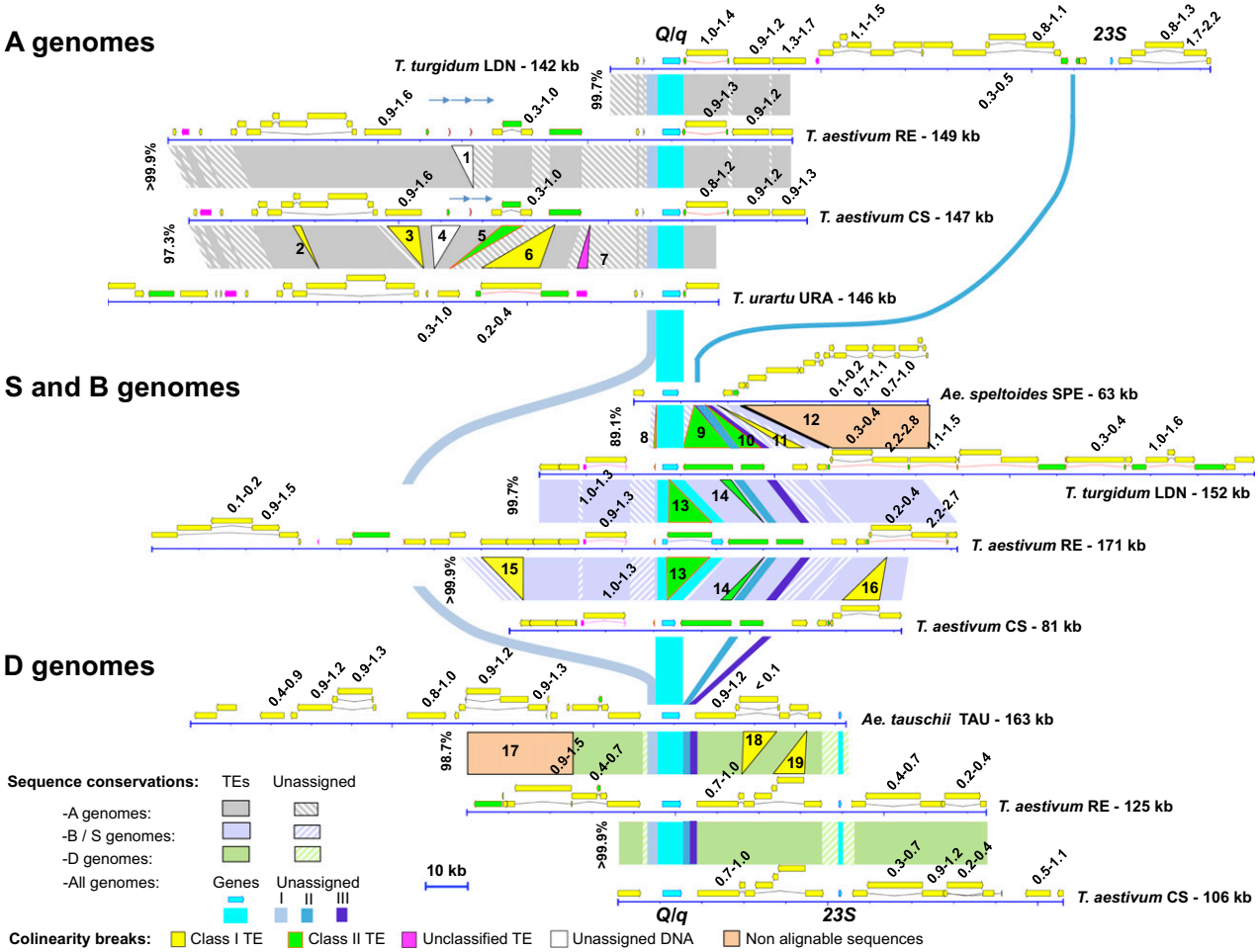

Fig. S2. Composition of the *Q/q* loci in wheat. Vertical numbers show an overall % identity of the overlapping sequences marked in gray, violet, and green. Diagonal numbers show TE insertion date estimates in million years ago (Mya). See also Table S1. Nineteen large intragenome insertions and deletions and breaks in genome colinearity marked with numbers are as follows: (1) 5,101–5,105 bp sequence repeated in tandem (blue arrows in the figure: an additional repetition in RE); (2) complete solo-LTR RLC *Angela* (insertion in polyploids); (3) complete *gypsy* RLG Fatima1 (recombined to solo-LTR in URA); (4) unassigned segment with complete DTM Argo (insertion in A genomes of polyploids); (5) newly identified complete CACTA (insertion in A genomes of polyploids); (6) complete *copia* RLC Maximus (insertion in URA); (7) unclassified TE (insertion in URA); (8) partial class II TE (insertion in all B genomes); (9) class II TE (insertion in all B genomes); (10) class II TE (insertion in all B genomes); (11) class I TE (insertion in B genome of RE); (12) break in colinearity, nonalignable (NA) sequences between B and S genomes; (13) Insertion of a complete helitron transposon in the B genome of RE; (14) class II TE (partial deletion in RE); (15) Tandem duplication of four copies of *copia* retrotransposon *Angela* in RE, through a mechanism of unequal homologous recombination; (16) complete *copia* retrotransposon RLC *Angela* (insertion in RE and CS); (17) break of colinearity, nonalignable genomic segments of TEs completely different when comparing the D genome haplotype sequence of the diploid TAU to that of the hexaploid wheat; (18) complete *copia* retrotransposon RLC *Valerie* (insertion in TAU); and (19) partial *copia* retrotransposon RLC Angela and complete *gypsy* retrotransposon RLG Ifis (insertion in RE and CS).

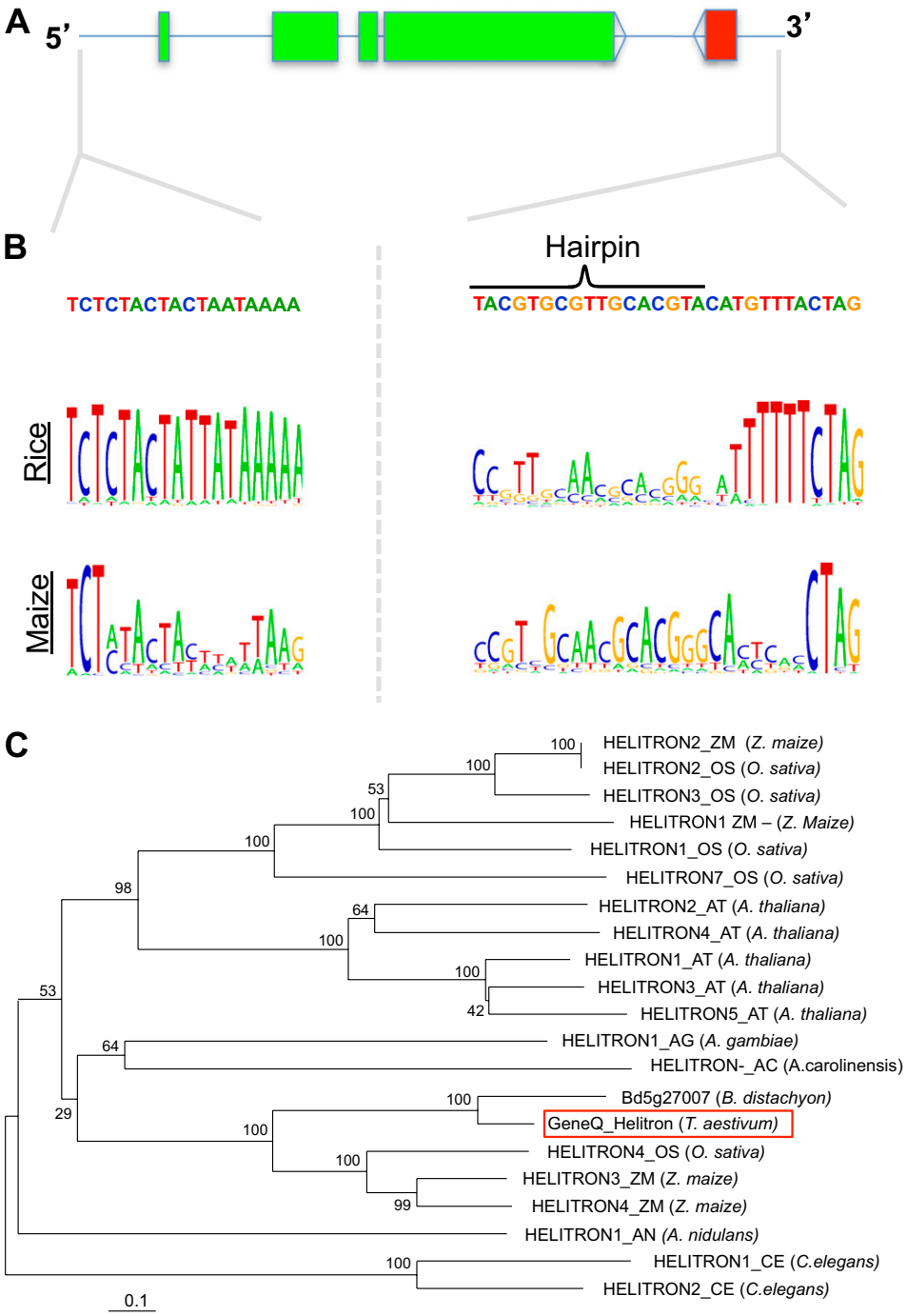

Fig. S3. Analysis of the 9,535-bp helitron inserted in opposite orientation in the *5Bq* gene of *T. aestivum* cv. RE. (*A*) Intron/exon structure of the helitron, predicting a putative gene of four exons (shown in green), where the putative RepHel protein is of 1,405 amino acids. A gene encoding a 128-amino-acid fragment of a glyoxalate reductase-like protein interrupted by a frameshift is also present in the helitron (shown in red) inserted in the same orientation as *q*. (*B*) Comparison of the 5′- and 3′-end sequences of the identified helitron shows a typical AT insertion site, the 5′ TC, the 3′ CTAG, and the sequences leading to the hairpin structure formation at 3′ ends, show high conservation with rice and maize helitron consensus sequences. (*C*) Bootstrapped neighbor-joining tree of predicted RepHel proteins encoded by helitrons from various species [found in repbase (http://www.girinst.org/repbase/update/index.html)]. Reference helitron sequences and species are indicated. The protein encoded by the *T. aestivum* cv. RE *5Bq* helitron is shown in the red box. It is 76% similar to RepHel encoded by Heltiron4 Os of rice (*AC105746.1*) and 87% similar to a putative *Brachypodium distachyon* helicase (*Bradi5g20077*). The insertion site, presence of the RepHel-like gene, and fragments of other genes are characteristic features of helitrons (1, 2).

1. Du C, Fefelova N, Caronna J, He L, Dooner HK (2009) The polychromatic Helitron landscape of the maize genome. *Proc Natl Acad Sci USA* 106:19916–19921.
2. Yang L, Bennetzen JL (2009) Distribution, diversity, evolution, and survival of Helitrons in the maize genome. *Proc Natl Acad Sci USA* 106:19922–19927.

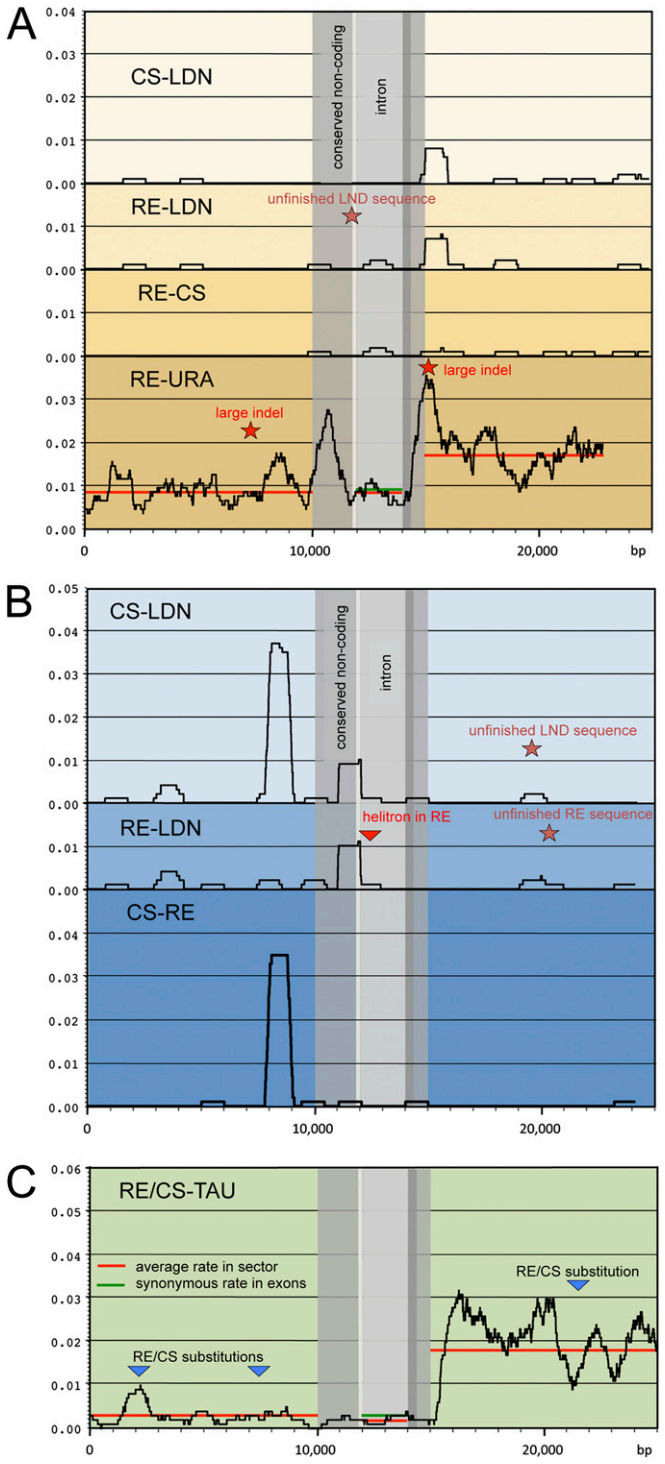

Fig. S4. Pairwise comparisons of the noncoding parts of the *Q*/*q* genes and the neighboring intergenic regions. Nucleotide substitution rates in a 1-kb window sliding by 1 bp in *5AQ*/*q*, *5Bq*, and *5Dq* for different pairs of wheats are shown in *A*, *B*, and *C*, respectively. Introns, light gray; UTRs, dark gray; and conserved sequences at the 5′ and 3′ ends of the *Q*/*q* gene, medium gray. Red lines indicate average substitution rates for the marked region (intergenic or intron) and green lines indicate synonymous nucleotide substitution rates for the *Q*/*q* coding sequence.

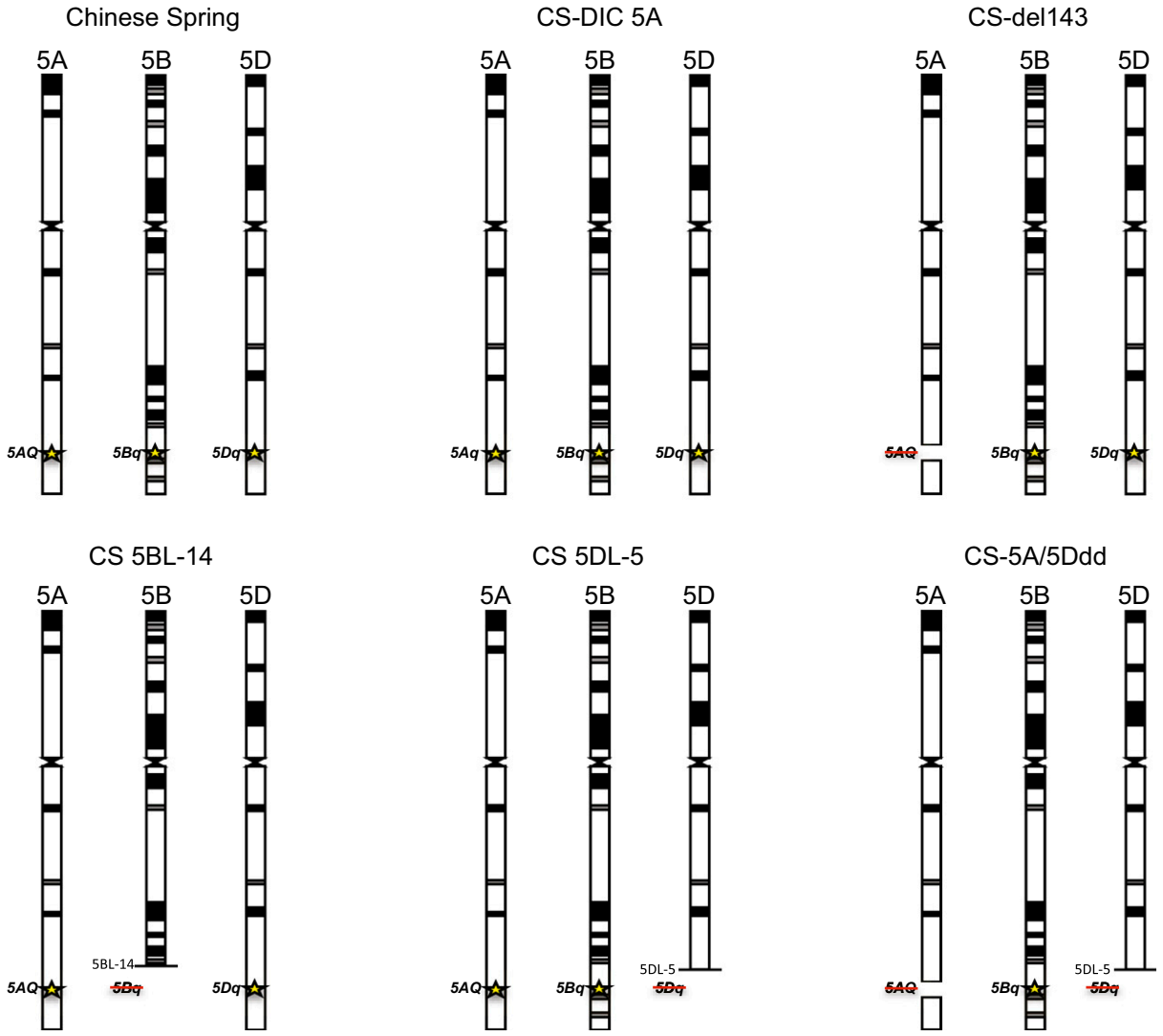

Fig. S5. Homoeologous group 5 chromosome constitutions for the genetic stocks used for *Q/q* expression and phenotypic analysis. Locations of homoeologous *Q/q* loci on the long arms of group 5 chromosomes are indicated by yellow stars. Absence of a *Q/q* locus is indicated by a red strike through. Black and gray regions along the chromosomes indicate dark and light cytological C bands, respectively.

Table S1. Divergence and insertion dates estimated for LTR retrotransposons with both left and right LTRs, and target site duplications identified

TE name	Insertion date (Mya)	Divergence	Std error
A_ContigQRG1_Ale1.fasta	0.66	0.0172	0.0084
A_ContigQRG1_Fatima1.fasta	1.25	0.0326	0.0083
A_ContigQRG1_Melina1.fasta	1.07	0.0279	0.0048
A_ContigQRG1_Wis1.fasta	1.06	0.0276	0.004
A_ContigQURA_Ale1.fasta	0.66	0.0172	0.0083
A_ContigQURA_Claudia1.fasta	0.28	0.0073	0.0022
A_CS12224M17Q_Ale1.fasta	0.66	0.0172	0.0085
A_CS12224M17Q_Fatima1.fasta	1.25	0.0326	0.0083
A_CS12224M17Q_Melina.fasta	1.01	0.0263	0.0047
A_CS12224M17Q_Sabrina2.fasta	1.12	0.029	0.0046
A_CS12224M17Q_Wis1.fasta	1.06	0.0276	0.0039
A_LDN376H15_Angela1.fasta	0.36	0.0093	0.0025
A_LDN376H15_Angela2.fasta	1.95	0.0507	0.0055
A_LDN376H15_Fatima1.fasta	1.06	0.0275	0.0072
A_LDN376H15_Melina1.fasta	1.23	0.032	0.0053
A_LDN376H15_Sabrina1.fasta	1.49	0.0387	0.005
A_LDN376H15_Sabrina2.fasta	1.30	0.0338	0.005
A_LDN376H15_Sabrina3.fasta	0.93	0.0241	0.0039
A_LDN376H15_Wis1.fasta	1.02	0.0264	0.0039
B_1255P23QRG4_Angela1.fasta	1.11	0.0288	0.0046
B_1255P23QRG4_Derami1.fasta	2.42	0.0629	0.0067
B_1255P23QRG4_Jeli1.fasta	1.20	0.0312	0.0075
B_1255P23QRG4_Wis1.fasta	0.15	0.004	0.0016
B_1255P23QRG4_Wis2.fasta	0.33	0.0086	0.0022
B_CS102N4QF_Angela1.fasta	1.14	0.0296	0.0045
B_LDN1004P5_Angela1.fasta	1.11	0.0289	0.0043
B_LDN1004P5_Barbara1.fasta	1.28	0.0334	0.0043
B_LDN1004P5_Derami1.fasta	2.52	0.0654	0.0073
B_LDN1004P5_Fatima3.fasta	1.31	0.034	0.0079
B_LDN1004P5_Laura1.fasta	0.36	0.0094	0.0015
B_LDN1004P5_Wis1.fasta	0.33	0.0086	0.0022
D_AeT20P19_Angela2.fasta	1.08	0.0282	0.0043
D_AeT20P19_Angela3.fasta	1.02	0.0265	0.0037
D_AeT20P19_Barbara1.fasta	1.02	0.0266	0.0042
D_AeT20P19_Derami1.fasta	1.09	0.0283	0.0042
D_AeT20P19_Maximus1.fasta	1.01	0.0262	0.0041
D_AeT20P19_NewQ04.fasta	0.67	0.0174	0.0068
D_AeT20P19_Valerie1.fasta	0.88	0.023	0.0034
D_AeT20P19_Valerie2.fasta	0.03	0.0009	0.0006
D_ContigQRG5_Ifis1.fasta	1.17	0.0305	0.0077
D_ContigQRG5_Ifis2.fasta	0.55	0.0142	0.005
D_ContigQRG5_Maximus1.fasta	0.87	0.0225	0.0037
D_ContigQRG5_Sabrina1.fasta	0.53	0.0139	0.0031
D_ContigQRG5_Wis1.fasta	0.34	0.0088	0.0022
D_CS201F2QGD_Angela3.fasta	1.04	0.0271	0.0043
D_CS201F2QGD_Ifis1.fasta	0.52	0.0136	0.005
D_CS201F2QGD_Maximus1.fasta	0.84	0.0218	0.0038
D_CS201F2QGD_NewQ04.fasta	0.78	0.0204	0.0076
D_CS201F2QGD_Wis1.fasta	0.32	0.0082	0.0022
S_SH120_23G16_Quinta2.fasta	0.16	0.0041	0.0021
S_SH120_23G16_Quinta3.fasta	0.88	0.0228	0.0047
S_SH120_23G16_Quinta4.fasta	0.83	0.0217	0.0046

Insertion dates were calculated using the substitution rate of 13 x 10-9 substitutions/site/year. SEs were estimated using the formula *T* 1/4 *K2P/2r* (1).

1. Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16:111–120.

Table S2. Helitron is found only in hexaploid wheat cvs Renan and Mironovskaya 808

	Helitron	+	–	+	–
Haplotype	Primers	21F / 21R	22F / 22R	32F / 32R	32F / 32bR
44 - AABBDD-*T. aestivum* cv. Arminda			1		1
45 - AABBDD-*T. aestivum* cv. Chinese Spring			1		1
46 - AABBDD- *T. aestivum* ssp. *macha* 1793MA86112			1		1
47 - AABBDD- *T. aestivum* ssp. *compactum rufulum* 71VREB68368			1		1
48 - AABBDD- *T. aestivum* ssp. *compactum crebicum* 72VREB68369			1		1
49 - AABBDD- *T. aestivum* ssp. *compactum erinaceum* 75VREB68372			1		1
50 - AABBDD- *T. aestivum* ssp. *spaerococcum tumidum perciv globosum* REB77050			1		1
51 - AABBDD-*T. aestivum* cv. Recital			1		1
52 - AABBDD-*T. aestivum* cv. Renan		1		1	
52 - AABBDD-*T. aestivum* cv. Mironovskaya 808		1		1	
53 - AABBDD-*T. aestivum* cv. Soisson			1		1
54 - AABBDD-*T. aestivum* ssp. *spelta* SP81300			1		1
55 - AABBDD- *T. aestivum* ssp. *spelta* SP86054			1		1
56 - AABBDD- *T. aestivum* cv. Vilmorin			1		1
13 - AABBDD- *T. aestivum* cv. Chopin			1		1
5 - AABB-*T. turgidum* ssp. *dicoccoides* DIC 118			1		1
6 - AABB- *T. turgidum* ssp. *dicoccoides* DIC 119			1		1
7 - AABB- *T. turgidum* ssp. *dicoccoides* DIC 120			1		1
8 - AABB- *T. turgidum* ssp. *dicoccoides* DIC 132			1		1
9 - AABB- *T. turgidum* ssp. *dicoccoides* DIC 196			1		1
10 - AABB- *T. turgidum* ssp. *dicoccoides* DIC 44			1		
11 - AABB- *T. turgidum* ssp. *dicoccoides* DIC 46			1		
12 - AABB- *T. turgidum* ssp. *dicoccoides* DIC 54			1		
13 - AABB- *T. turgidum* ssp. *dicoccoides* DIC 55			1		1
14 - AABB- *T. turgidum* ssp. *dicoccoides* DIC 87			1		
15 - AABB-*T. turgidum* ssp. *dicoccum* DIC 298			1		1
16 - AABB-T. *turgidum* ssp. *dicoccum* DIM-05			1		1
17 - AABB-*T. turgidum* ssp. *dicoccum* DIM-15			1		1
18 - AABB-*T. turgidum* ssp. *dicoccum* DIM-78			1		1
19 - AABB-*T. turgidum* ssp. *durum* HTTH01			1		1
20 - AABB-*T. turgidum* ssp. *dicoccum* TTC01			1		1
21 - AABB-*T. turgidum* ssp. *dicoccum* TTC02			1		1
22 - AABB- *T. turgidum* ssp. *dicoccoides* TTD01			1		1
23 - AABB- *T. turgidum* ssp. *dicoccoides* TTD03			1		1
24 - AABB- *T. turgidum* ssp. *dicoccoides* TTD06			1		1
25 - AABB- *T. turgidum* ssp. *dicoccoides* TTD09			1		1
26 - AABB- *T. turgidum* ssp. *dicoccoides* TTD121			1		1
27 - AABB- *T. turgidum* ssp. *dicoccoides* TTD125			1		1
28 - AABB- *T. turgidum* ssp. *dicoccoides* TTD15			1		1
29 - AABB- *T. turgidum* ssp. *dicoccoides* TTD20			1		1
30 - AABB- *T. turgidum* ssp. *dicoccoides* TTD23			1		1
32 - AABB- *T. turgidum* ssp. *dicoccoides* TTD48			1		1
33 - AABB- *T. turgidum* ssp. *dicoccoides* TTD58			1		
34 - AABB- *T. turgidum* ssp. *dicoccoides* TTD72			1		
35 - AABB- *T. turgidum* ssp. *dicoccoides* TTD86			1		
36 - AABB- *T. turgidum* ssp. *dicoccoides* TTD88			1		
37 - AABB-T *T. turgidum* ssp. *dicoccoides* TTD95			1		
38 - AABB-*T. turgidum* ssp. *polinicum* TTP03			1		
39 - AABB-*T. turgidum* ssp. *durum* TTR04			1		
40 - AABB-*T. turgidum* ssp. *durum* TTR05			1		
41 - AABB-T.*turgidum* ssp *durum* TTR16			1		1
42 - AABB-T. *turgidum* ssp *durum* TTR19			1		1
43 - AABB-T.*turgidum* ssp.*turgidum* TTT02			1		1
11 - AABB- *T. turgidum* ssp. *durum* cv. Langdon			1		1
12 - AABB- *T. turgidum* ssp. *durum* cv. Capelli			1		1
1 - AA-*T. monococcum* TMB02			1		
2 - AA-*T. urartu* TMU38			1		
1 - AA - *T. beoticum*-no_ref			1		
2 - AA- *T. beoticum*-no_ref			1		
3 - AA- *T. beoticum*-no_ref			1		

Table S2 Cont.

	Helitron	+	–	+	–
Haplotype	Primers	21F / 21R	22F / 22R	32F / 32R	32F / 32bR
6 - AA- *T. urartu*-no_ref			1		
7 - AA- *T. beoticum*-no_ref			1		
8 - AA- *T. beoticum*-no_ref			1		
9 - AA- *T. urartu*-no_ref			1		
70 - AAGG-*T. timopheevii* ssp. *armeniacum* TIA02			1		
71 - AAGG-*T. timopheevii* ssp. *armeniacum* TIA06			1		
72 - AAGG- *T. timopheevii* ssp. armeniacum TIA28			1		
73 - AAGG- *T. timopheevii* ssp. armeniacum TIA50			1		
74 - AAGG- *T. timopheevii* ssp. timopheevii TIM01			1		
75 - AAGG- *T. timopheevii* ssp.*timopheevii viticulosum* TIM05			1		
77 - AAGG- *T. timopheevii* ssp.*timopheevii* zhuk (k38555) TIM07			1		
78 - AAGG- *T. timopheevii* TIM09			1		
79 - AAGG- *T. timopheevii* ssp. *typicum* TR103			1		
80 - AAGG-T. *araraticum*TR94			1		
81 - AAGG-*T. araraticum*TR95			1		
82 - AAGG-*T. araraticum*TR96			1		
83 - AAGG-*T. araraticum*TR99			1		
14 - AAGG- T *T. timopheevii* ssp. *timopheevii* var typica-no_ref			1		
15 - AAGG- *T. timopheevii* ssp. *araticum* var Schulze-motel-no_ref			1		
16 - AAGG- *T. timopheevii* ssp. *militinae* -no_ref			1		
18 - AAGG- *T. timopheevii* ssp. *timopheevii* -no_ref			1		
84 - DD-*Ae. kotschyi* var. *palaestina* AEG 506_1			1		

Empty cells, no PCR amplification; cells showing number 1, PCR amplification confirming helitron insertion.

Table S3. Traits of genetic stocks with various combinations of *Q/q* homoeoalleles and chromosomal deletions including the *Q/q* locus

				Yield components				
Wheat line *Q/q* genotype (AABBDD)	Spike length (mm)	Spike emergence (days)	Plant height (cm)	Tillers/ plant	Spikelets/ spike	Seeds/ spike	Seed weight (mg)	Grain yield (g/plant)
CS 5DL-5 (*QQ* ϕϕ –)	67 (63–72)	114 (112–117)	102 (96–108)	11.4 (9.2–13.6)	20 (18.4–21.2)	12 (6–19)	20 (18–23)	3.0 [3.0] (1.2–4.7)
CS (*QQ* ϕϕ *qq*)	72 (69–75)	80 (78–82)	118 (116–120)	10.3 (9.3–11.2)	25 (24.6–25.5)	36 (30–42)	31 (28–33)	11 [11] (9.6–13.0)
CS 5BL-14 (*QQ* – *qq*)	62 (57–67)	76 (72–79)	98 (88–107)	8.6 (7.1–10.2)	20 (18.4–21.8)	9 (4–14)	18 (16–19)	1.6 [1.4] (0.5–2.7)
CS-5A/5Ddd (– ϕϕ –)	81 (78–83)	113 (110–115)	110 (105–115)	12.1 (9.7–14.6)	17 (16.1–17.9)	6 (5–7)	18 (17–20)	1.4 [1.3] (1.0–1.7)
CS-del143 (– ϕϕ *qq*)	101 (98–105)	73 (71–75)	123 (119–127)	10.8 (9.7–11.8)	25 (24.6–26.1)	37 (32–42)	29 (28–30)	12 [12] (9.8–13.3)
CS-DIC 5A (*qq* ϕϕ *qq*)	110 (107–114)	68 (66–69)	129 (122–134)	9.9 (8.2–11.5)	28 (27.2–28.3)	47 (43–50)	29 (27–31)	13 [13] (11.6–14.7)

ϕ, pseudogene; –, chromosomal deletion including the *Q/q* locus; 95% confidence intervals are shown in parentheses; grain yield calculated on the basis of the yield components is shown in brackets.

Table S4. PCR primers

Primer	Sequence (5′ -> 3′)
Helitron haplotype diversity	
21F	CTGATGATGATGCGTGATTT
21R	TACTGTATATGGCATCTGATTAGAG
22F	AGGAGCACACGCAAGTAGTACCTCT
22R	CTACAAAGGGCATGATCGAAC
32F	GCAGGTAATCATCTAAGCTACTATAAGC
32R	GGATTGGAGCGAGTGTAAGATTTC
32bR	CAGCTTGTGAATTTCTTTTCAACGTAATCTTGTA
BAC library screening	
AP2start (FW)	ATGGTGCTGGATCTCAATGTGGAGTCGCCGGCGGA
AP2.12R (RW)	CAAAGTACCTTGCAGCTTCAACTTCGCTGTCAA
AP2.13R (RW)	CTCTTGGGATCGTGCGCGGTGGGTTGCGACATC
5Bq cDNA amplification	
5BqF2	TGCTTAACTCCGCGGACGCCGGCGGCTT
5BqR2	GTGCGCGGTAGGTTGCGACATCCGA
5Dq cDNA amplification	
5Dq1bF	GGATGATGACGGGGCAGCTGGCA
5Dq3aR	TGGGGAAGCACGACGGCGGCTCAGGGGGCCTTGG
5Bq RACE	
5Bq5′race1	CATATATTATCCGGCGCACTGACTACC
5Bq3′race1	GGTAGTCAGTGCGCCGGATAATATATG
5Dq RACE	
5Dq5′race3	CGCCACCCGCTGCGCCATCACG
5Dq3′race1	CCCCTGAGCCGCCGTCGTGCTTCCCCA
5AQ/q RQ-PCR	
RQAQ/q.F1	GGATCTGCGGATGTCGCAACCC
RQAQ/qR2	ATGCACAGGCCACTGGGAC
5Bq RQ-PCR	
RQBSF1	CAGTGCGCCGGATAATATATG
RQBSR1	ATGTGCACGAACTCCTCCTTG
5Dq RQ-PCR	
5D3′race1	CCCCTGAGCCGCCGTCGTGCTTCCCCA
Dq3′UTR.R2	CAGCCAGTCACACTCACACATGGCCT
Wheat Actin RQ-PCR	
Actin.F	ATGGAAGCTGCTGGAATCCAT
Actin.R	CCTTGCTCATACGGTCAGCAATAC
5Bq 2-bp deletion analysis	
5Bq2bp.del.F1	CAGCCCTCTCATCTCCCTACTCA
5Bq2bp.del.R3	AATGGATTCAAACGAGGCCTGA