Supporting Information

Zhang et al. 10.1073/pnas.1319598110

SI Materials and Methods

Plant Materials. Seeds at selfed generation 2 (S_2) of four synthetic allotetraploid wheat lines and their respective parental genotypes were procured from Moshe Feldman (Weizmann Institute of Science). These lines were produced by intergeneric hybridization followed by colchicine-mediated genome doubling of the first filial generation (F_1) hybrids (1). We designated these four lines as AT1 [*Aegilops sharonensis* (accession TH02) (♀) × *Triticum monococcum* (accession TMB02) (♂), ($2n = 28$, genome $S^{sh}S^{sh}A^mA^m$)], AT2 [*Aegilops longissima* (accession TL05) (♀) × *Triticum urartu* (accession TMU06) (♂), ($2n = 28$, genome S^lS^lAA)], AT3 [*T. urartu* (accession TMU38) (♀) × *Aegilops tauschii* (accession TQ27) (♂), ($2n = 28$, genome AADD)], and AT4 [*Aegilops bicornis* (accession TB01) (♀) × *A. tauschii* (accession TQ27) (♂), ($2n = 28$, genome S^bS^bDD)]. All lines and their parental genotypes were propagated under strict selfing conditions in growth cabinets to obtain seeds of advanced generations. Detailed information on these synthetic lines and their diploid parental species was provided in Table S1.

Karyotyping by Sequential Genomic in Situ Hybridization and Fluorescence in Situ Hybridization. The protocols for genomic in situ hybridization (GISH) and fluorescence in situ hybridization (FISH) were followed essentially as described in Han et al. (2) and Kato et al. (3) with minor modifications (cf. ref. 4). Specifically, for GISH, genomic DNA was isolated from young leaves of all of the diploid parental genotypes of the four synthetic allotetraploid wheats using a modified cetyl trimethylammonium bromide (CTAB) method (5). These parental genotypes include *T. monococcum* (accession TMB02), *T. urartu* (accession TMU06 and TMU38), and *A. tauschii* (accession TQ27). Genomic DNA of TMB02, TMU06, TMU38, and TQ27 were labeled by nick translation with Chroma Tide Alexa Fluor 488–5-dUTP (Invitrogen; catalog no. C11397, green) and Texas Red-5-dCTP (Perkin-Elmer; catalog no. NEL 426, red), respectively. Genomic DNA of TH02, TL05, TQ27 and TB01 were used separately as blocker DNA. For FISH, repetitive DNA sequences pSc119.2 (6) and 45S ribosomal DNA (rDNA) were labeled with Alexa Fluor 488–5-dUTP (green); and pAs1 (7), 5S rDNA, and a microsatellite $(GAA)_n$ oligonucleotide were labeled with Texas Red-5-dCTP (red), and hybridized to the same set of slides used for GISH analysis. In general, sequential probing by pSc119.2 and pAS1, following GISH, was sufficient to enable identification of all of the parental chromosomes in each of the synthetic allotetraploid wheat lines. The three additional FISH probes, 5S rDNA, 45S rDNA, and $(GAA)_n$ oligonucleotide, were used independently to verify the chromosome designations.

Metaphase chromosome spreads were prepared according to the protocol described in Kato et al. (3). Slide denaturation, hybridization, and washing steps were carried out following manufacturer's instructions (Invitrogen; catalog no. C11397). Slides were examined with an Olympus BX61 fluorescence microscope and digitally photographed. The images were captured by using the Olympus IPP software package and visualized in the Photoshop Adobe System.

Assay of Homolog-Specific Copy-Number Variations by Pyrosequencing. The protocol was followed as reported in Mochida et al. (8). The 17 and 31 protein-coding genes (Dataset S2) that were confirmed by gene-specific PCR amplification to exist in both diploid parental species of the synthetic allotetraploid wheat lines AT1 and AT2, respectively, were used for the analysis. The corresponding EST sequences of these genes were downloaded from the National Center for Biotechnology Information website (www.ncbi.nlm.nih.gov). The SeqMan program (Lasergene) was used to identify the parental subgenome-specific SNPs [popularly known as "homeologous sequence variants" (HSVs) (cf. ref. 9)], which enabled classification of the EST sequences into three clusters where each corresponds with a subgenome of hexaploid wheat (BBAADD). Gene-specific PCR amplification primers and pyrosequencing primers were designed for each of these genes using Soft Assay Design software and verified on the pyrosequencing system, PyroMarkQ96 ID (Qiagen) (see Dataset S2 for primer details). The gene-specific primers were designed from the fully conserved sequences between the two homeolog-specific EST-clusters and were planed such that they flanked the HSVs targeted by the pyrosequencing primers. For pyrosequencing, either forward or the reverse PCR amplification primers were labeled with biotin. Primers for pyrosequencing were designed to anneal to the regions immediately upstream of the SNP sites on biotinylated strands. PCR amplifications were carried out using HotStar *Taq* DNA Polymerase following manufacturers' instructions (Qiagen). The thermal cycling program used was as follows: 95 °C for 5 min, followed by 50 cycles at 95 °C for 30 s, 56–68 °C for 30 s and 72 °C for another 30 s, and a final extension for 5 min at 72 °C. PCR products were analyzed by electrophoresis on 2% (weight/vol) agarose gel. We analyzed 24 individual plants each of synthetic allotetraploid wheat lines, AT1 and AT2, which were selected arbitrarily from two (S_4 and S_5) or three (S_6, S_7, and S_8) generations of AT1 and AT2, respectively (Table S2). Results for all 408 data points (17 genes × 24 plants) of AT1 and 744 data points (31 genes × 24 plants) of AT2 were obtained (Dataset S3). To determine the rate of experimental errors (e.g., DNA quantification or pipetting errors during mixing of DNA samples) a set of 10 independent 1:1 DNA mixes of parental genotypes of AT1 and AT2 were subjected to the same pyrosequencing analysis.

Profiling of Seed Setting. Greenhouse-grown plants of two synthetic allotetraploid wheat lines, AT2 (S^lS^lAA) and AT3 (AADD), were used to score rate of seed setting which was calculated by diving the number of filled seeds by the total number of kernels of all spikes of a given plant (cf. ref. 4).

Data Collection and Statistics. First, arbitrarily chosen seeds for each of the four synthetic allotetraploid wheat lines across the available selfed generations were karyotyped (Table S1). The genome and chromosome biases for chromosome loss or gain and the chromosome number range based on karyotyping of a total of 92, 96, 205, and 96 plants of the four lines respectively (Table S1) were calculated and presented in Fig. 2. Statistical significance was calculated using the chi-squared test, i.e., prop. test with *R* (version 2.15.0). Next, the same plants of the four synthetic wheat lines were also used to analyze chromosome structural stability. The genome and chromosome biases for structural chromosome changes were calculated and presented in Fig. 3. Statistical significance was calculated (prop. test) with *R* (version 2.15.0). Two lines, AT1 ($S^{sh}S^{sh}A^mA^m$) and AT2 (S^lS^lAA), were analogous in genome constitution to the natural tetraploid wheats, and both were found to have a localized, persistent genomic change of the pSc119.2 DNA repeat. So we chose 24 plants from each of these two lines and used 17 and 31 gene primers to see if genomic changes in the coding genes in these two lines existed. The data are presented in Fig. 4 and Fig. S7. To determine the relationship between karyotype instability and

seed setting, we chose 44 and 140 plants of the AT2 and AT3 lines, respectively. These two lines represent the line (AT2) showing karyotype-stability and one containing massive aneuploidy and rearrangements (AT3). These two lines were chosen also because they had sufficient plants karyotyped for this purpose. The data are presented in Fig. S8.

1. Ozkan H, Levy AA, Feldman M (2001) Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. *Plant Cell* 13(8):1735–1747.
2. Han F, Liu B, Fedak G, Liu Z (2004) Genomic constitution and variation in five partial amphiploids of wheat—Thinopyrum intermedium as revealed by GISH, multicolor GISH and seed storage protein analysis. *Theor Appl Genet* 109(5):1070–1076.
3. Kato A, Lamb JC, Birchler JA (2004) Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. *Proc Natl Acad Sci USA* 101(37):13554–13559.
4. Zhang H, et al. (2013) Persistent whole-chromosome aneuploidy is generally associated with nascent allohexaploid wheat. *Proc Natl Acad Sci USA* 110(9):3447–3452.
5. Kidwell KK, Osborn TS (1992) Simple plant DNA isolation procedures. Plant genomes: Methods for Genetic and Physical Mapping, eds Beckmann JS, Osborn TC (Kluwer, The Netherlands), pp 1–13.
6. McIntyre CL, Pereira S, Moran LB, Appels R (1990) New Secale cereale (rye) DNA derivatives for the detection of rye chromosome segments in wheat. *Genome* 33(5): 635–640.
7. Rayburn AL, Gill B (1986) Molecular identification of the D-genome chromosomes of wheat. *J Hered* 77(4):253–255.
8. Mochida K, Yamazaki Y, Ogihara Y (2003) Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. *Mol Genet Genomics* 270(5):371–377.
9. Rustgi S, Bandopadhyay R, Balyan HS, Gupta PK (2009) EST-SNPs in bread wheat: Discovery, validation, genotyping and haplotype structure. *Czech J Genet Plant Breed* 45(3):106–116.

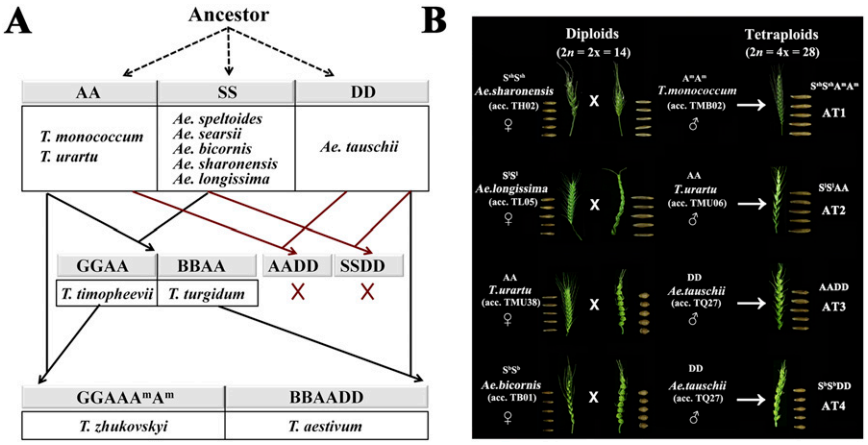

Fig. S1. The evolutionary trajectories of natural tetra- and hexaploid wheats and the four synthetic allotetraploid wheats used in this study. (*A*) diagrammatic representation of the evolutionary history of allopolyploid wheats via sequential allotetra- and allohexaploidization events. Maroon arrows and Xs represent conceivably occurred allotetraploidization events that failed to establish as species. (*B*) Typical spike and seed morphologies of the four synthetic allotetraploid wheats used in this study.

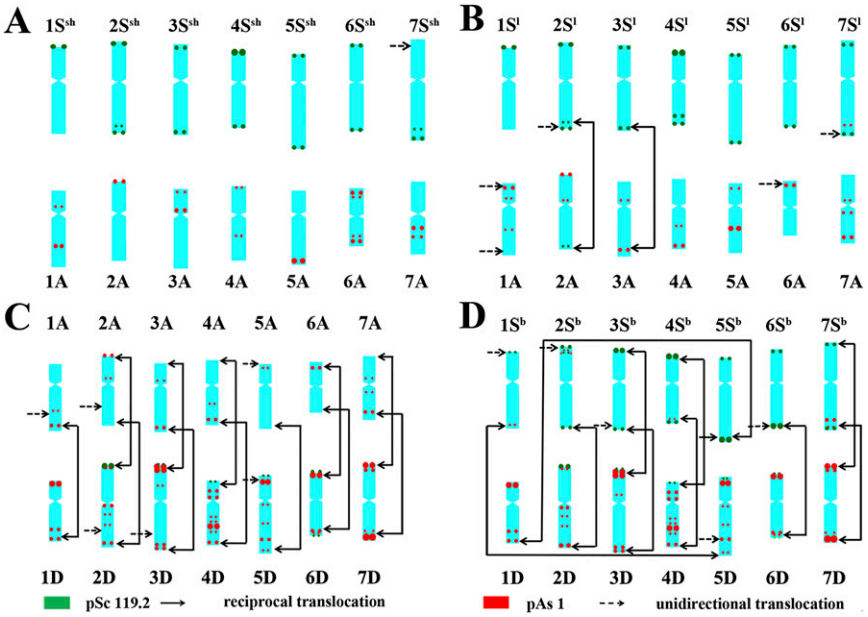

Fig. S2. Ideograms of four synthetic allotetraploid wheat lines (AT1, AT2, AT3, and AT4), based on sequential FISH with two repetitive DNA probes, pSc119.2 (green dots) and pAs1 (red dots), showing structural changes and intersubgenomic rearrangements. Confidence in chromosome designations comes from GISH results, and each designation was further confirmed by the results of an independent FISH experiment using three additional probes: 5S rDNA, 45S rDNA, and $(GAA)_n$ (Fig. S3). Reciprocal intersubgenomic translocations and unidirectional homeologous transfer of DNA segments are marked by solid and dashed arrows, respectively. *A*, *B*, *C*, and *D*, respectively, are AT1 (genome $S^{sh}S^{sh}A^mA^m$), AT2 (genome S^lS^lAA), AT3 (genome AADD), and AT4 (genome S^bS^bDD).

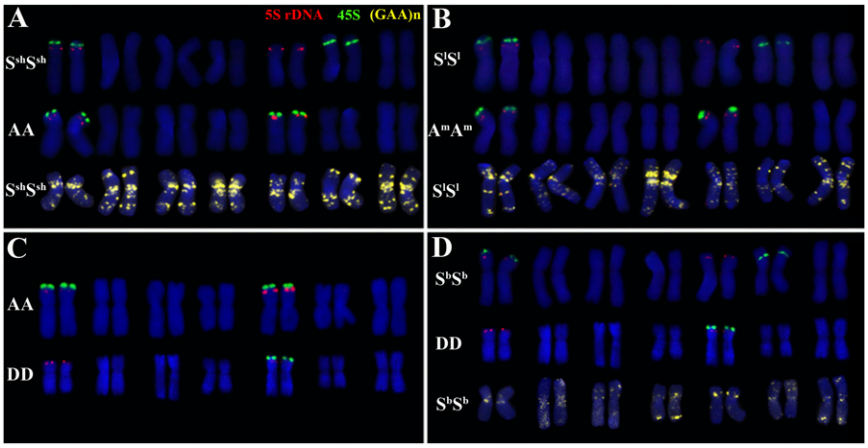

Fig. S3. Independent validation of chromosome identifications by sequential FISH with three additional DNA repeats, 5S rDNA, 45S rDNA, and a microsatellite $(GAA)_n$ oligonucleotide. (*A*) AT1 (genome $S^{sh}S^{sh}A^mA^m$): first, when probed with 5S rDNA (red signals), chromosomes $1S^{sh}$, $5S^{sh}$, 1A, and 5A were identified; second, when probed with 45S rDNA (green signals), chromosomes $1S^{sh}$, $6S^{sh}$, 1A, and 5A were identified; third, when probed with $(GAA)_n$ oligonucleotide (yellow signals), all seven chromosomes of the S^{sh} genome were distinguished. (*B*) AT2 (S^lS^lAA): first, when probed with 5S rDNA (red signals), chromosomes $1S^l$, $5S^l$, $1A^m$, and $5A^m$ were identified; second, when probed with 45S rDNA (green signals), chromosomes $1S^l$, $6S^l$, $1A^m$, and $5A^m$ were identified; third, when probed with $(GAA)_n$ oligonucleotide (yellow signals), all seven chromosomes of the S^l genome were distinguished. (*C*) AT3 (AADD): first, when probed with 5S rDNA (red signals), chromosomes 1A, 5A, 1D, and 5D were identified; second, when probed with 45S rDNA (green signals), chromosomes 1A, 5A, and 5D were identified. (*D*) AT4 (S^bS^bDD): first, when probed with 5S rDNA (red signals), chromosomes $1S^b$, $5S^b$, 1D, and 5D were identified; second, when probed with 45S rDNA (green signals), chromosomes $1S^b$, $6S^b$, and 5D were identified; third, when probed with $(GAA)_n$ oligonucleotide (yellow signals), all seven chromosomes of the S^b genome were distinguished.

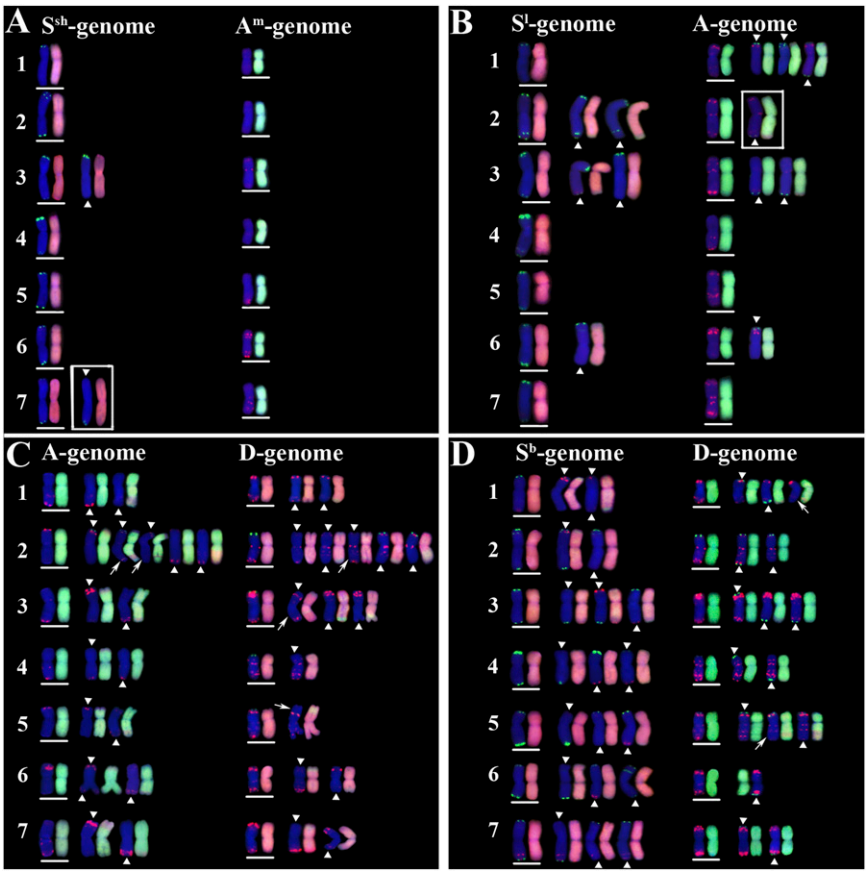

Fig. S4. Examples of restructured chromosomes in the four sets of synthetic allotetraploid wheats based on sequential FISH/GISH-based karyotyping. All wild-type parental chromosomes (underlined) are included for comparisons. Arrows and arrowheads denote terminal and interstitial intersubgenomic rearrangements, respectively, which can be either reciprocal translocation or unidirectional homeologous segment transfer. The localized persistent genomic changes (see *Results*, second section, for details) detected by FISH with pSc119.2 (green signal), which occurred only in AT1 ($S^{sh}S^{sh}A^mA^m$) and AT2 (S^lS^lAA), are boxed *A*, *B*, *C*, and *D* are FISH (*Left*) and GISH (*Right*) images of AT1, AT2, AT3 and AT4, respectively.

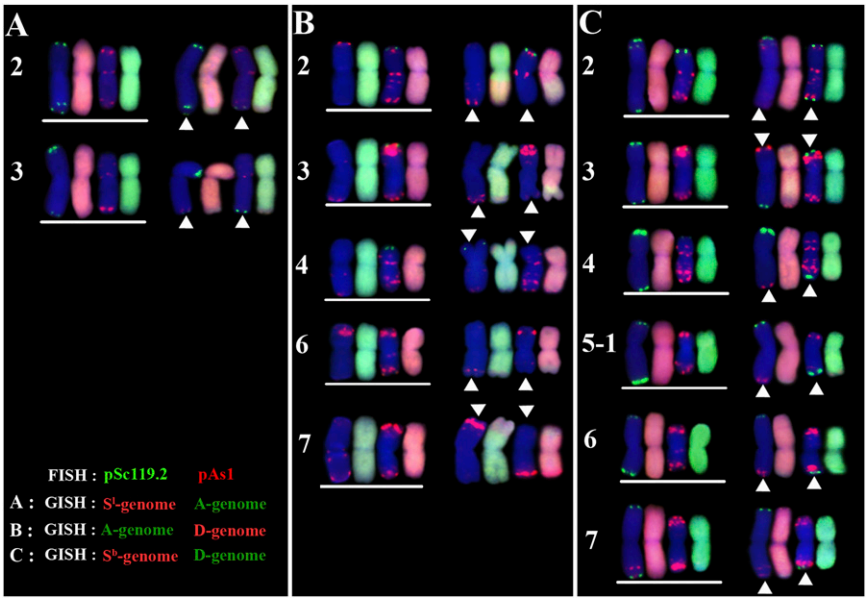

Fig. S5. Interhomeologous chromosome translocations (arrowheads) identified by sequential FISH/GISH-based karyotyping in three synthetic allotetraploid wheats, (*A*) AT2 (genome S^lS^lAA), (*B*) AT3 (genome AADD), and (*C*) AT4 (genome S^bS^bDD). The corresponding wild-type parental chromosomes are underlined.

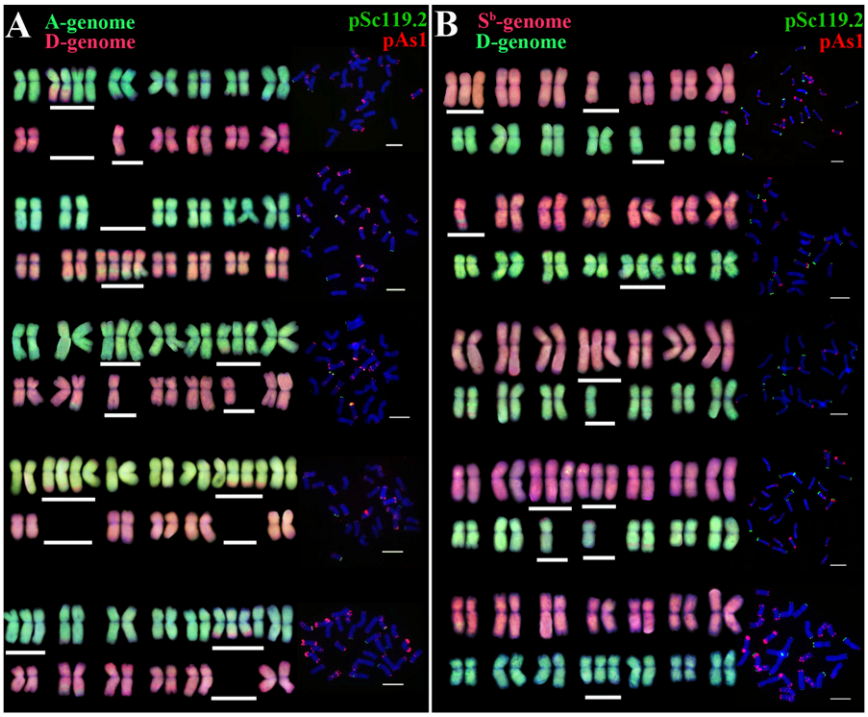

Fig. S6. Examples of aneuploid karyotypes of synthetic allotetraploid wheats AT3 and AT4 based on sequential FISH and GISH of the same chromosome spreads. The aneuploid chromosomes in each GISH karyotype are underlined. The karyotypes of AT3 (*A*) include (each pair of underlined chromosomes from top to bottom): 2A/2D tetrasomy/nullisomy + 3D monosomy ($2n = 27$), 3A/3D nullisomy/tetrasomy ($2n = 28$), 3A/3D + 6A/6D double-trisomy/monosomy ($2n = 28$), 2A/2D + 6A/6D double-tetrasomy/nullisomy ($2n = 28$), and 6A/6D tetrasomy/nullisomy + 1A trisomy ($2n = 29$). The karyotypes of AT4 (*B*) include (each pair of underlined chromosomes from top to bottom): $1S^b$/5D trisomy/monosomy + $4S^b$ monosomy ($2n = 27$), $1S^b$/5D monosomy/trisomy ($2n = 28$), $4S^b$/4D trisomy/monosomy ($2n = 28$), $3S^b$/3D + $4S^b$/4D double-trisomy/monosomy ($2n = 28$), and 4D trisomy ($2n = 29$). (Scale bars, 10 μm.)

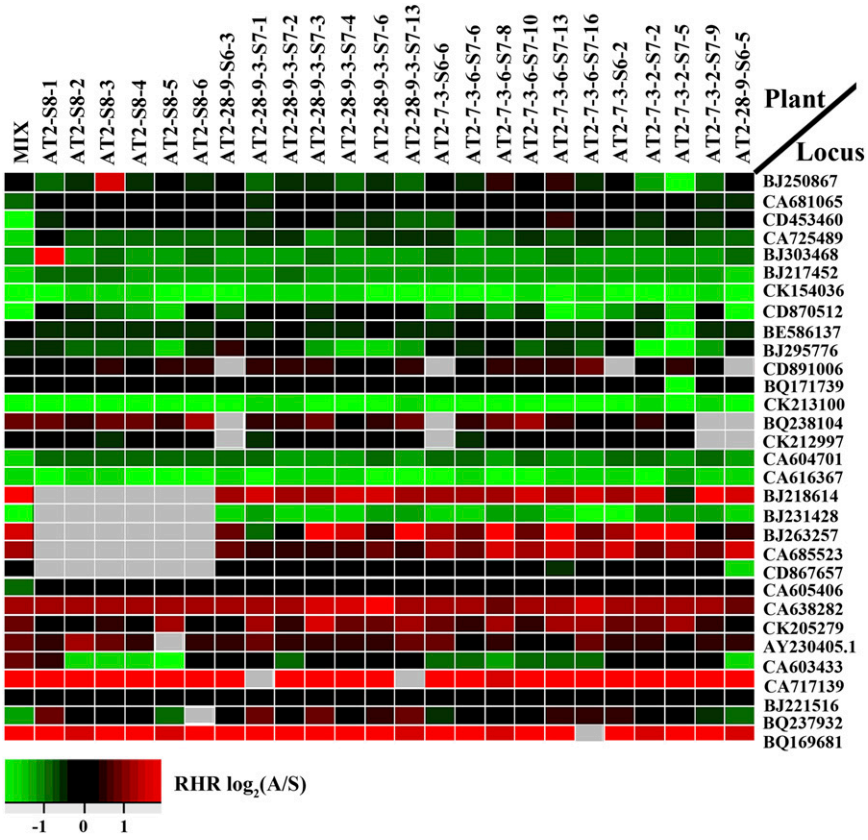

Fig. S7. Copy-number variations (CNVs) recorded in 24 individual AT2 (genome S^lS^lAA) plants for a set of 31 genes (Dataset S2) using locus-specific pyrosequencing. The relative homolog ratio in each of the 744 data points (31 genes × 24 plants) was calculated based on three biological replications, and its deviation from that of parental 1:1 DNA mixes was tested for statistical significance (for details, see *SI Materials and Methods*). Gray indicates missing data due to insufficient DNA.

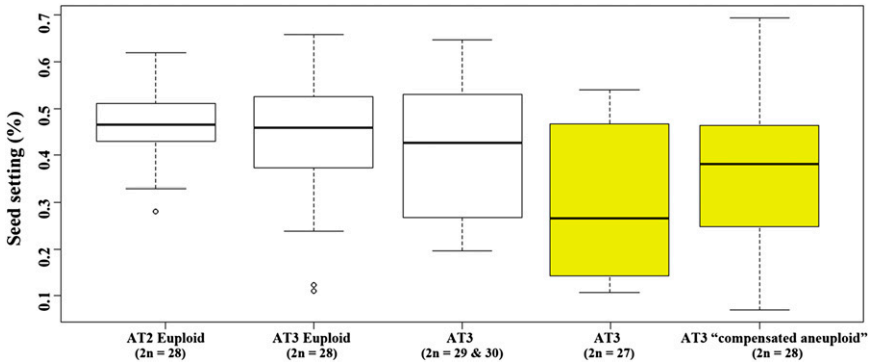

Fig. S8. Effect of aneuploidy (and associated structural chromosome changes) on seed setting, displayed as range of seeds produced by different karyotypes in box-plots. The synthetic allotetraploid wheat AT2 (euploid, $2n = 28$) is included as a reference for comparison. The following plant groups are compared: selected euploid individuals of AT3 ($2n = 28$), selected aneuploid individuals with a chromosome number of $2n = 29$ or 30, selected aneuploid individuals with a chromosome number of $2n = 27$, and selected "compensated aneuploid" individuals with a chromosome number of $2n = 28$. Same-color boxes indicate statistical insignificance in pairwise comparisons ($P > 0.05$, pairwise *t* test), while boxes with different colors (white vs. yellow) indicate statistical significance ($P < 0.05$, pairwise *t* test).

Table S1. The four sets of synthetic allotetraploid wheats used in this study

Designation	Cross combination ($2n = 4x = 28$) (♀ × ♂)	Genome constitution	No. of plants	Selfed generation
AT1	*A. sharonensis* (TH02) × *T. monococcum* ssp. *aegilopoilds* (TMB02)	$S^{sh}S^{sh}A^mA^m$	92	S_4-S_5
AT2	*A. longissima* (TL05) × *T. urartu* (TMU06)	S^lS^lAA	96	S_5-S_8
AT3	*T. urartu* (TMU38) × *A. tauschii* (TQ27)	AADD	205	S_3-S_6
AT4	*A. bicornis* (TB01) × *A. tauschii* (TQ27)	S^bS^bDD	96	S_3-S_5

Seeds at generation S_2 were procured from Moshe Feldman (The Weizmann Institute of Science).

Table S2. AT1 and AT2 plants used for locus-specific pyrosequencing-based assay to study CNVs for the protein-coding genes

AT1			AT2	
No. of plants	Plant designation	Chromosome no. ($2n = 4x$)	Plant designation	Chromosome no. ($2n = 4x$)
1	AT1-S5-111	28	AT2-S8-1	28
2	AT1-S5-112	28	AT2-S8-2	28
3	AT1-S5-115	28	AT2-S8-3	28
4	AT1-S5-116	28	AT2-S8-4	28
5	AT1-S5-117	28	AT2-S8-5	28
6	AT1-1-11-1-S5-8	28	AT2-S8-6	28
7	AT1-1-11-2-S5-1	28	AT2-28-9-S6-3	28
8	AT1-1-11-3-S5-7	28	AT2-28-9-3-S7-1	28
9	AT1-1-11-3-S5-8	28	AT2-28-9-3-S7-2	28
10	AT1-1-11-3-S5-14	28	AT2-28-9-3-S7-3	28
11	AT1-1-11-4-S5-2	28	AT2-28-9-3-S7-4	28
12	AT1-1-11-4-S5-13	28	AT2-28-9-3-S7-6	28
13	AT1-1-10-S4-2	28	AT2-28-9-3-S7-13	28
14	AT1-1-10-S4-3	28	AT2-7-3-S6-6	28
15	AT1-1-10-S4-4	28	AT2-7-3-6-S7-6	28
16	AT1-1-10-S4-5	28	AT2-7-3-6-S7-8	28
17	AT1-1-10-S4-6	28	AT2-7-3-6-S7-10	28
18	AT1-1-81-S4-1	28	AT2-7-3-6-S7-13	28
19	AT1-1-81-S4-2	28	AT2-7-3-6-S7-16	28
20	AT1-1-13-S4-3	28	AT2-7-3-S6-2	28
21	AT1-1-13-S4-4	28	AT2-7-3-2-S7-2	28
22	AT1-1-13-S4-5	28	AT2-7-3-2-S7-5	28
23	AT1-1-13-S4-6	28	AT2-7-3-2-S7-9	28
24	AT1-1-13-S4-7	28	AT2-28-9-S6-5	28

Other Supporting Information Files

Dataset S1 (XLSX)
Dataset S2 (XLSX)
Dataset S3 (XLSX)