Title

Uncovering hidden variation in polyploid wheat.

Journal

Proceedings of the National Academy of Sciences of the United States of America, 114(6)

ISSN

0027-8424

Authors

Krasileva, Ksenia V
Vasquez-Gross, Hans A
Howell, Tyson
et al.

Publication Date

2017-02-01

DOI

10.1073/pnas.1619268114

Peer reviewed

PNAS PLUS

Uncovering hidden variation in polyploid wheat

Ksenia V. Krasileva[a,b,c], Hans A. Vasquez-Gross[a], Tyson Howell[a], Paul Bailey[c], Francine Paraiso[a], Leah Clissold[c], James Simmonds[d], Ricardo H. Ramirez-Gonzalez[c,d], Xiaodong Wang[a], Philippa Borrill[d], Christine Fosker[c], Sarah Ayling[c], Andrew L. Phillips[e], Cristobal Uauy[d,1,2], and Jorge Dubcovsky[a,f,1,2]

[a]Department of Plant Sciences, University of California, Davis, CA 95616; [b]The Sainsbury Laboratory, Norwich NR4 7UH, United Kingdom; [c]The Earlham Institute, Norwich NR4 7UG, United Kingdom; [d]John Innes Centre, Norwich NR4 7UH, United Kingdom; [e]Rothamsted Research, Harpenden AL5 2JQ, United Kingdom; and [f]Howard Hughes Medical Institute, Chevy Chase, MD 20815

Contributed by Jorge Dubcovsky, December 20, 2016 (sent for review November 22, 2016; reviewed by Beat Keller and Joachim Messing)

Comprehensive reverse genetic resources, which have been key to understanding gene function in diploid model organisms, are missing in many polyploid crops. Young polyploid species such as wheat, which was domesticated less than 10,000 y ago, have high levels of sequence identity among subgenomes that mask the effects of recessive alleles. Such redundancy reduces the probability of selection of favorable mutations during natural or human selection, but also allows wheat to tolerate high densities of induced mutations. Here we exploited this property to sequence and catalog more than 10 million mutations in the protein-coding regions of 2,735 mutant lines of tetraploid and hexaploid wheat. We detected, on average, 2,705 and 5,351 mutations per tetraploid and hexaploid line, respectively, which resulted in 35–40 mutations per kb in each population. With these mutation densities, we identified an average of 23–24 missense and truncation alleles per gene, with at least one truncation or deleterious missense mutation in more than 90% of the captured wheat genes per population. This public collection of mutant seed stocks and sequence data enables rapid identification of mutations in the different copies of the wheat genes, which can be combined to uncover previously hidden variation. Polyploidy is a central phenomenon in plant evolution, and many crop species have undergone recent genome duplication events. Therefore, the general strategy and methods developed herein can benefit other polyploid crops.

wheat | polyploidy | mutations | reverse genetics | exome capture

Since the dawn of agriculture, wheat has been a major dietary source of calories and protein for humans. The cultivated wheat species *Triticum turgidum* (tetraploid, AABB genome) and *Triticum aestivum* (hexaploid, AABBDD genome) originated via recent polyploidization events followed by domestication. *T. turgidum* originated less than 500,000 y ago from the hybridization of *Triticum urartu* (diploid, AA genome) and a now-extinct species related to *Aegilops speltoides* (diploid, SS similar to BB genome), whereas *T. aestivum* originated less than 10,000 y ago from the hybridization of tetraploid wheat with *Aegilops tauschii* (diploid, DD genome) (1).

As a result of the recent polyploidization, most genes in tetraploid and hexaploid wheat species are present in multiple functional copies, referred to as homeologs. These duplicated genes buffer the rapid natural changes occurring in the large and dynamic wheat genomes (1). As loss-of-function mutations in any single wheat homeolog are frequently masked by redundancy in other homeologs, this variation remains hidden from natural and human selection. This drawback becomes an advantage for the development of mutant populations, as redundancy confers tolerance to high densities of induced mutations (2). On average, mutation densities of ethyl methanesulfonate (EMS) mutant populations of hexaploid wheat (3–5) are as much as 10-fold higher than those of diploid barley (6). When mutations in individual homeologs have been identified, they can be combined to generate loss-of-function mutants and to overcome the masking effect of redundant homeologs.

Extensive utilization of the current wheat mutant populations has been limited by the need to physically access the DNAs of the mutant lines and by the time required for the mutant screens, which entail the development and optimization of genome-specific primers for each target gene. A pilot study using three Cadenza lines with known mutations in the *GA20ox* gene and a small capture array including 1,846 genes demonstrated that exome capture (7) followed by sequencing was a viable strategy to identify mutations in wheat (8). Whole-genome resequencing of mutant lines also has been used for species with small genomes (9), but is a very expensive alternative for the large genomes of tetraploid (12 Gb) and hexaploid (17 Gb) wheat (10).

In this study, we describe the development of a wheat exome capture platform and its use to sequence the coding regions of 2,735 mutant lines. We characterized the obtained mutations, organized them in a public database including more than 10 million mutations, identified deleterious alleles for ~90% of the captured wheat genes, and discuss potential applications.

Results

Development of a Wheat Exome Capture Design. In collaboration with NimbleGen, we developed an 84-Mb exome capture assay including overlapping probes covering 82,511 transcripts (*SI Appendix*, Method S1 and Table S1). We aligned these transcripts to

Significance

Pasta and bread wheat are polyploid species that carry multiple copies of each gene. Therefore, loss-of-function mutations in one gene copy are frequently masked by functional copies on other genomes. We sequenced the protein coding regions of 2,735 mutant lines and developed a public database including more than 10 million mutations. Researchers and breeders can search this database online, identify mutations in the different copies of their target gene, and request seeds to study gene function or improve wheat varieties. Mutations are being used to improve the nutritional value of wheat, increase the size of the wheat grains, and generate additional variability in flowering genes to improve wheat adaptation to new and changing environments.

Author contributions: C.U. and J.D. designed research; K.V.K., T.H., L.C., J.S., X.W., P. Borrill, and C.F. performed research; K.V.K., H.A.V.-G., T.H., P. Bailey, F.P., R.H.R.-G., S.A., A.L.P., C.U., and J.D. contributed new reagents/analytic tools; K.V.K., H.A.V.-G., T.H., P. Bailey, R.H.R.-G., C.U., and J.D. analyzed data; K.V.K., H.A.V.-G., T.H., C.U., and J.D. wrote the paper; A.L.P. contributed to the original idea and provided the Cadenza TILLING population; C.U. proposed the original idea, and codirected the project; and J.D. proposed the original idea, codirected the project, and coordinated international collaboration.

Reviewers: B.K., University of Zürich; and J.M., Rutgers University.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in NCBI BioProject (accession no. PRJNA258539) and European Read Archive (ENA study PRJEB11524). The variant calls are available at Plant Ensembl, plants.ensembl.org/Triticum_aestivum/Info/Index.

[1]C.U. and J.D. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: cristobal.uauy@jic.ac.uk or jdubcovsky@ucdavis.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1619268114/-/DCSupplemental.

the draft wheat genome (11) and identified 286,799 exons, which we padded with 30 bp of intronic sequence, when possible, to maintain coverage at splice sites. Given that the assay probes cross-hybridize efficiently at high levels of sequence identity (12), we did not include the most similar homeologs in the capture design. The coding regions of wheat homeologs average 97.2% identity [SD 1.8% (13)], which is similar to the genome divergence found in peanuts (14) and slightly greater than the divergence found in soybeans (15).

We multiplexed captures from eight tetraploid or four hexaploid wheat lines per Illumina lane and obtained >20 million 100-bp paired-end reads per sample by using methods described in *SI Appendix*, Method S2 and Table S2. To improve the proportion of mapped reads, we supplemented the reference wheat genome sequence (11) with de novo assemblies of unmapped reads from both wheat species (*SI Appendix*, Method S3 and Table S3). The de novo assemblies (*SI Appendix*, Table S3) resulted in an additional 40,975 contigs for Kronos (33.4 Mb) and 67,632 contigs for Cadenza (41.3 Mb). The expanded reference improved the proportion of mapped reads from 93% to 98% in Kronos and from 96% to 99% in Cadenza (*SI Appendix*, Table S3). The initial capture design (alpha) was tested in 42 "Kronos" and 49 "Cadenza" mutants, and probes with extreme capture efficiencies were rebalanced at NimbleGen, resulting in an adjusted design (beta) with more uniform coverage (*SI Appendix*, Fig. S1).

Identification and Characterization of Induced Mutations. We used the adjusted design to capture and sequence the coding regions of 1,535 EMS mutants from the tetraploid variety "Kronos" (5) and 1,200 EMS mutants from the hexaploid variety "Cadenza" (3). The development of these populations and the sequencing strategy are summarized in Fig. 1. Mutations were then identified using the Mutation and Polymorphism Survey (MAPS) bioinformatics pipeline (16), which was modified by additional filters described in *SI Appendix*, Figs. S2–S5. We modified several MAPS parameters to optimize the detection of mutations in polyploid wheat (*SI Appendix*, Method S4). Using these parameters, the MAPS pipeline identified 119.2 Mb and 162.4 Mb positions with adequate coverage and quality for mutation identification in tetraploid and hexaploid wheat, respectively. These positions, designated hereafter as "valid positions," were larger than the 84 Mb covered by the original design. This was an expected result, as the captured sequences included homeologs that were not in the original design. The 1.4:1 ratio between valid positions in hexaploid and tetraploid wheat is close to the 1.5:1 ratio expected between lines containing three and two genomes.

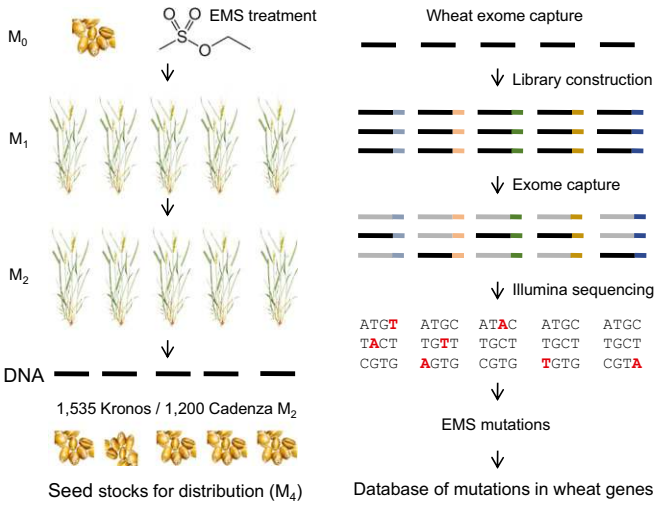


Fig. 1. Overview of the development of the sequenced mutant populations. M_0 seeds were mutagenized with EMS (resulting in M_1 plants), and a single M_2 plant was grown from each M_1 plant. Genomic DNAs were extracted from the M_2 plants, and M_3 seeds were obtained from the same plant. M_3 seeds were planted in the field to produce M_4 seed for distribution. Barcoded sequencing libraries were constructed, used for exome capture, and sequenced by using Illumina. Mutations were identified by using the MAPS pipeline and were deposited in public databases that can be searched online.

Real mutations are expected to be present in multiple reads, whereas sequencing errors tend to be independently distributed. Therefore, the minimum number of reads including a mutation that are required to call a mutation [minimum coverage (*MC*)] is a critical parameter to differentiate real mutations from sequencing errors. To select an *MC* threshold that maximizes the number of detected mutations while keeping a low error rate, we compared the coverage, the number of mutations, and the associated errors at different *MC* stringency levels (*SI Appendix*, Method S4 and Tables S4–S6). Based on these data, we selected a minimum coverage of five mutant reads for heterozygous (*HetMC5*) and three for homozygous (*HomMC3*) mutations as the optimal thresholds for mutant identification. The median coverage at mutation sites for Kronos and Cadenza was 21× (*SI Appendix*, Table S4).

By using the *HetMC5*/*HomMC3* threshold, we identified 4.15 million uniquely mapped EMS-type mutations (G to A and C to T) in tetraploid wheat (2,705 mutations per line) and 6.42 million in hexaploid wheat (5,351 mutations per line; Table 1). Dividing these numbers by the number of valid positions identified by MAPS, we estimated an average mutation density of 23 mutations per Mb per individual in Kronos (34.8 mutations per kb for 1,535 lines) and 33 mutations per Mb per individual in Cadenza (39.5 mutations per kb for 1,200 lines). The distribution of mutations along the wheat pseudomolecules paralleled the distribution of gene densities (Fig. 2*A*), suggesting a uniform distribution of mutations along the wheat coding regions.

At the *HetMC5*/*HomMC3* threshold, the nonmutagenized control lines showed a much lower number of polymorphisms than the mutant lines (0.5% in Kronos and 1.2% in Cadenza). On average, only 14 SNPs per plant were detected in the nonmutagenized Kronos, along with 63 in the nonmutagenized Cadenza controls (Fig. 2*B*), suggesting a low error rate. This was also supported by the low percentage of non–EMS-type mutations (all mutation types except G to A and C to T) observed in the mutagenized lines of Kronos (0.9%) and Cadenza (0.8%; Fig. 2*C* and Table 1). A third estimate of the error rate was obtained by calculating the number of non–EMS-type transitions (A to G and T to C) within the non-EMS SNPs. This number is very similar to the reciprocal EMS-type mutations (G to A and C to T; *SI Appendix*, Method S5), and can be used to estimate the number of potentially erroneous EMS-type mutations. By using this method, the predicted error rate among uniquely mapped EMS-type mutations was less than 0.2% in both Kronos and Cadenza (Table 1). The low error rate of our mutant identification pipeline was also validated experimentally by resequencing PCR products and dedicated SNP assays. Among 280 EMS-type mutations, 278 (99.3%) were confirmed across both populations (*Materials and Methods* and *SI Appendix*, Text S1 and Tables S7–S10).

At lower *MC* stringency levels, we observed higher numbers of mutations, but also an increase in the associated errors (Fig. 2*D*). At *HetMC3*/*HomMC2*, for example, we detected an additional 0.9 million EMS-type mutations in tetraploid wheat (*SI Appendix*, Table S5) and 1.7 million in hexaploid wheat (*SI Appendix*, Table S6) that were not previously detected at *HetMC5*/*HomMC3*. However, the estimated errors for heterozygous mutations at exact coverage of 3 (*HetC3*) increased to 5.6% for Kronos (*SI Appendix*, Table S5) and 10.0% for Cadenza (*SI Appendix*, Table S6). These additional mutations still have an acceptable probability of being correct and can be accessed by selecting the desired *MC* in the public database

Table 1. Characterization of mutations in tetraploid and hexaploid wheat (*HetMC5/HomMC3*)

Mutations and SNPs characteristics	Tetraploid Kronos	Hexaploid Cadenza
Uniquely mapped SNPs*	4,189,561	6,470,733
Heterozygous/homozygous ratio at M_2*	1.87	2.21
Uniquely mapped EMS-type mutations*	4,152,707	6,421,522
Average EMS-type mutations/line*	2,705	5,351
Average EMS-type mutations per kilobase (population)	34.8	39.5
EMS-type, %*	99.1	99.2
Non–EMS-type transitions*	7,323	10,569
Maximum error in uniquely mapped EMS-type, %*,†	0.18	0.16
RH SNPs	69,651	38,626
Heterozygous/homozygous ratio in RH	0.33	0.30
Average SNPs per megabase per individual in RH	592	441
EMS-type SNPs in RH	16,412	6,023
EMS-type in RH, %	23.6	15.6
Non–EMS-type transitions in RH	20,358	8,669
Multimap SNPs	321,511	955,074
Heterozygous/homozygous ratio in multimap	2.85	6.16
Multimap EMS-type mutations	315,537	933,515
EMS-type mutations in multimap SNPs, %	98.14	97.74
Non–EMS-type transitions in multimap	1,166	5,968
Maximum error in multimapped EMS-type, %	0.37	0.64
Gene models with at least one mutation (GM_1)‡	48,172	73,895
GM_1 with at least one truncation	28,604 (59%)	45,311 (61%)
GM_1 with at least one missense mutation	46,198 (96%)	69,543 (94%)
Average number of missense mutations per GM_1	21.4	22.6
GM_1 with truncation and/or deleterious missense§	43,787 (91%)	67,830 (92%)
No. of unique genes eliminated in large deletions	832	6,657

*Excluding RH and deletion regions.
†Estimated from the number of reciprocal A>G and T>C transitions among non–EMS-type mutations.
‡GM in Ensembl (a more detailed analysis of variant effect predictions is provided in *SI Appendix*, Text S3).
§Predicted deleterious missense mutations by SIFT (<0.05).

search tools from the project (dubcovskylab.ucdavis.edu/wheat_blast and www.wheat-tilling.com).

Residual Genetic Heterogeneity. The original breeder's seed stocks of Kronos and Cadenza that were mutagenized had small regions of residual genetic heterogeneity (RH) that were identified after sequencing based on their lower proportion of EMS-type mutations, higher mutation density, higher proportion of homozygous mutations, and presence in multiple individuals (Table 1, Fig. 3 *A* and *B*, and *SI Appendix*, Fig. S6). Using an index that combined those criteria (*SI Appendix*, Method S5 and Table S11), we identified 69,651 RH-SNPs in Kronos (1.7% of the total SNPs) and 38,626 RH-SNPs (0.6%) in Cadenza at the *HetMC5*/*HomMC3* threshold (Table 1 and *SI Appendix*, Table S12). These RH levels are consistent with seed obtained after pooling multiple F_7 plants for Kronos and multiple F_8 plants for Cadenza, which is normal breeding practice.

Mutations Present in Multiple Individuals. Even after removal of the RH regions, approximately 1.4 million EMS-type mutations shared by more than one individual were detected in Kronos and Cadenza. The frequency of these mutations decayed rapidly from two to six individuals (Fig. 3 *C* and *D*, red bars) and was very different from the frequency distribution in the RH regions (Fig. 3 *A* and *B*). The distribution of these EMS-type mutations approached a Poisson distribution (Fig. 3 *C* and *D*, light blue bars). However, the closest theoretical Poisson distribution was obtained by using only one fifth of the available G/C sites (*SI Appendix*, Method S6 and Tables S13 and S14). This result suggests that some G/C positions have a lower probability of being affected by the EMS mutagen.

We hypothesize that the EMS preference for certain sequences flanking the mutated sites (Fig. 3 *E* and *F* and *SI Appendix*, Method S7) can affect the probability of mutations in some G/C positions, as previously observed in rice (16). This hypothesis is also supported by the observation that sequence preferences in the region flanking EMS-type mutations were stronger in non-RH mutations shared by multiple individuals than in those present in only one individual (*SI Appendix*, Fig. S7). We do not rule out the possibility that differences in chromatin structure and DNA methylation may have also affected the probability of mutations at some G/C sites.

Reads Mapping to Multiple Locations. For some genes, we detected very few or no mutations. Characterization of these genes revealed that this was mainly caused by duplicated regions in the reference genome (e.g., highly similar homeologs or incorrect duplicated assemblies). Reads associated with multiple mapping (MM) locations were assigned low mapping quality values and were eliminated in the MAPS pipeline. To recover mutations in these locations, we developed a custom bioinformatics pipeline that assigned the MM reads to a single location, recorded alternative locations, modified the mapping quality score, and redirected the reads to MAPS (*SI Appendix*, Method S8 and Figs. S2 and S3). By using this pipeline, we recovered 16.6 million reads and identified an additional 1.25 million EMS-type mutations (Table 1). More MM high-quality mutations were observed in hexaploid (933,515) than in tetraploid wheat (315,537), which was expected based on the presence of an additional genome. We validated 22 of the 25 MM mutations tested by PCR and resequencing (*SI Appendix*, Method S8 and Table S15).

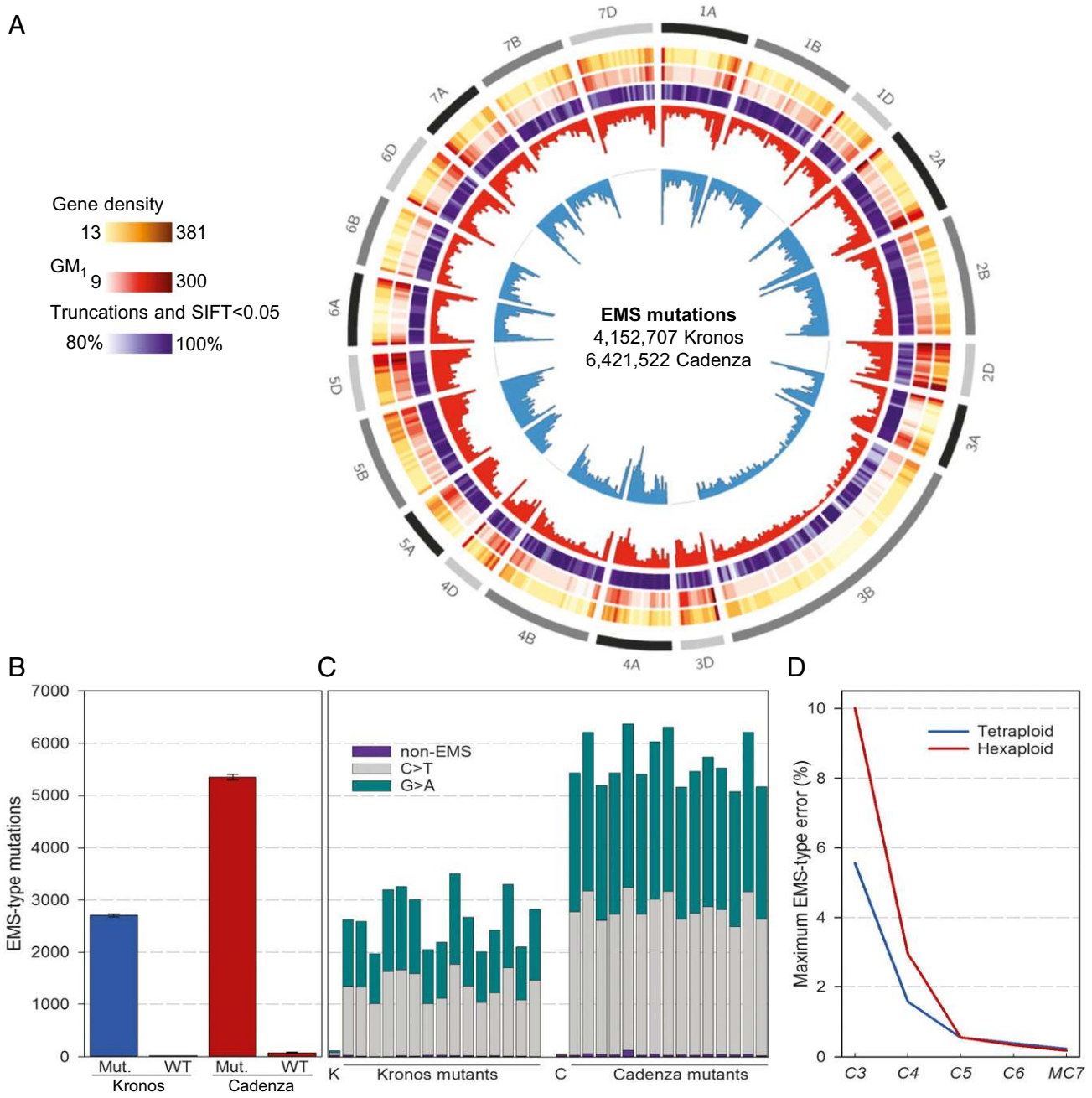


Fig. 2. Characterization of mutations in tetraploid Kronos and hexaploid Cadenza. (*A*) Genome-wide positions of identified mutations and their effects. The tracks from outside to inside represent gene density along wheat chromosomes (yellow-brown), number of gene models with at least one mutation (i.e., GM_1; red), percentage of GM_1 genes with at least one deleterious allele (truncation and/or missense mutation with SIFT score <0.05; purple), and total mutation densities for Cadenza (red) and Kronos (light blue). Each bin corresponds to a 10-Mb window. (*B*) Average number of SNPs in mutants and WT controls. (*C*) Distribution of mutations types in fifteen representative Kronos and Cadenza mutants ("K," nonmutagenized Kronos; "C," nonmutagenized Cadenza). Gray/teal indicates EMS-type mutations (C>T/G>A); violet indicates non–EMS-type mutations. (*D*) Estimated EMS-type error at different *HetMC* cutoffs (*C*, coverage; *MC*, minimum coverage).

Small and Large Deletions. In addition to EMS-type mutations, MAPS identified 616 and 1,268 small (<20 bp) deletions in Kronos and Cadenza, respectively (at *HetMC5*/*HomMC3*, excluding RH regions; *SI Appendix*, Table S16). We selected 15 of these predicted deletions for validation by Sanger resequencing (*SI Appendix*, Table S17 and Fig. S8) and were able to validate 14 of them (93.3%; *SI Appendix*, Table S17).

By using a dedicated bioinformatics pipeline (*SI Appendix*, Method S9 and Fig. S4), we detected a total of 870 large homozygous deletions covering five or more exons in Kronos and 7,971 in Cadenza (*SI Appendix*, Tables S18 and Figs. S9–S11). Overall, the large deletion events were mainly independent (94% of deletions were restricted to one or two individuals; *SI Appendix*, Text S2 and Fig. S11), and were confined to relatively small physical intervals comprising a median size of four to eight scaffolds in Kronos and Cadenza, respectively (*SI Appendix*, Fig. S10). The tetraploid Kronos population had fewer numbers of lines with homozygous deletions compared with the hexaploid Cadenza population, and those deletions were, on average, smaller and confined to fewer individuals. This is most likely a consequence of the higher EMS dosage used in the mutagenesis of the hexaploid

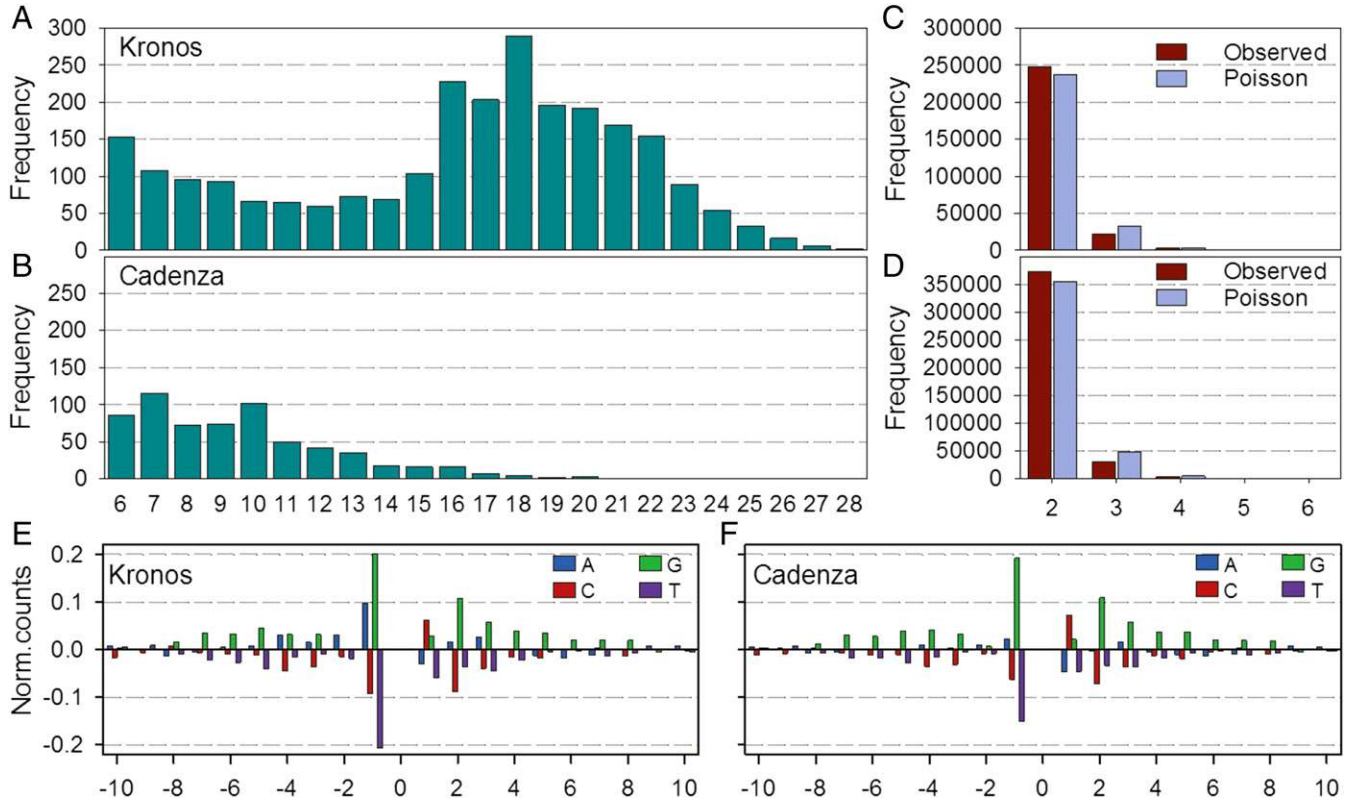


Fig. 3. EMS mutations present in multiple individuals. EMS sequence preference and RH: (*A*, *C*, and *E*) Tetraploid Kronos and (*B*, *D*, and *F*) hexaploid Cadenza. (*A* and *B*) Mutations shared by multiple individuals in RH regions. (*C* and *D*) Observed (red) and closest Poisson distribution (light blue) of mutations present in non-RH regions of multiple individuals. (*A*–*D*) The *x*-axis indicates the number of individuals sharing the same mutation. (*E* and *F*) Sequence preference in regions flanking EMS-type mutations (*SI Appendix*, Fig. S7). The *x*-axis indicates the number of nucleotides upstream (negative) and downstream (positive) from the mutated site.

lines, which is also reflected in the higher average mutation density in Cadenza relative to Kronos. For validation, we selected 11 homozygous large deletions and were able to confirm all of them (*SI Appendix*, Method S10 and Table S19).

Effect of Induced Mutations on Gene Models. We analyzed the effect of EMS-type mutations on gene models with at least one mutation (GM_1; Table 1 and *SI Appendix*, Method S11, Text S3, and Tables S20 and S21). In tetraploid wheat, 59% of GM_1 genes contained at least one truncation (premature stop or splice site) mutation and 96% at least one missense mutation (average, 21.4). In hexaploid wheat, 61% of GM_1 genes contained at least one truncation and 94% contained at least one missense mutation (average, 22.6; Table 1 and *SI Appendix*, Table S21). By using the "sorting intolerant from tolerant" (SIFT) algorithm (17), we found that more than 85% of GM_1 genes across both populations had at least one deleterious missense mutation (SIFT < 0.05). Results combining the SIFT and truncation analyses suggest that our database includes high-quality uniquely mapped mutations that eliminate or reduce function for more than 90% of the captured wheat genes (Table 1 and Fig. 2*A*). As an example of the high frequency of mutations in these populations, we show the presence of truncations or deleterious missense mutations in most of the genes from the wheat starch biosynthesis (Fig. 4*A* and *SI Appendix*, Method S12, Table S22, and Fig. S12) and flowering pathways (Fig. 4*B* and *SI Appendix*, Method S12, Table S23, and Fig. S12).

Mutations in the *Starch Branching Enzyme* genes have been already used to develop pasta and wheat germplasm with increased levels of resistant starch (18, 19), a dietary fiber associated with beneficial effects on human health (20–22). Mutations in wheat flowering genes have been used to dissect the wheat flowering pathway and to modulate wheat flowering time (23–28). For four of these genes, the effects of mutations in individual homeologs were negligible compared with those of the null mutations affecting all homeologs (Fig. 4 *C*–*F*). These results illustrate the limited effects of individual recessive mutations in polyploid wheat.

Access to Mutations, Seed Stocks, and SNP Markers. The EMS-type mutations detected in the Kronos and Cadenza populations at different stringency levels are accessible in public databases and can be visualized using a JBrowse graphic interface (*SI Appendix*, Text S4). Once the desired mutations are identified, the corresponding M_4 seeds can be requested from the University of California, Davis (dubcovskylab.ucdavis.edu/wheat-tilling), and the UK Germplasm Resources Unit (https://www.seedstor.ac.uk/shopping-cart-tilling.php).

In addition, predesigned "Kompetitive Allele Specific PCR" (KASP) primers are available to validate the mutations and to select them for downstream research and breeding applications. In total, we designed 2,771,688 KASP assays for the Kronos population and 3,872,892 assays for Cadenza, the majority of which are genome-specific or -semispecific (72.9% Kronos and 82.8% Cadenza). These primers are provided as part of the output from the public databases.

Discussion

Advantages and Limitations of Sequenced Mutant Populations. The exome-sequenced tetraploid and hexaploid mutant populations can be used for complementary purposes. The tetraploid mutant population is best suited for basic research projects because it allows quicker generation of complete null mutants. This can be achieved through a single cross between A and B genome mutant

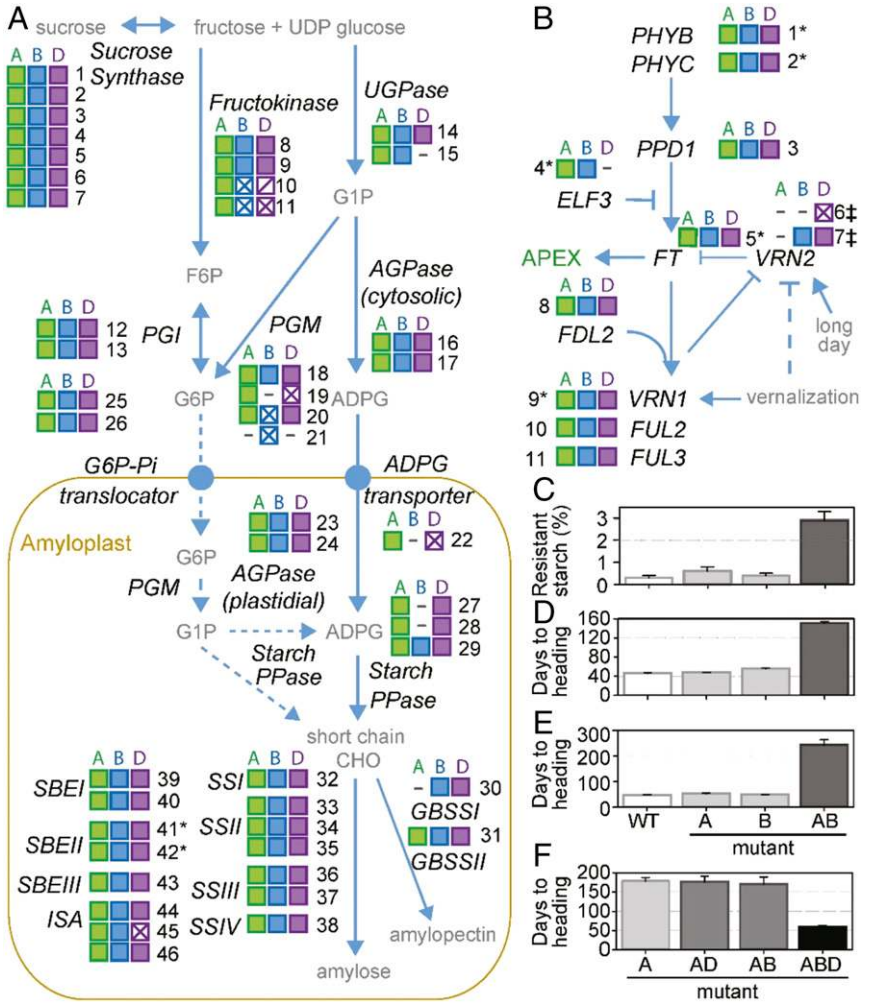


Fig. 4. Predicted effect of mutant alleles on wheat starch biosynthesis and flowering pathways genes. (*A*) Wheat starch biosynthesis and (*B*) flowering pathways. Squares represent individual genes and are colored according to genomes ("A," green; "B," blue; "D," violet). Filled squares represent genes with at least one truncation mutation from the mutant database, squares with crossed diagonals indicate deleterious missense mutations, and single diagonals denote tolerated missense mutations as predicted by SIFT (17). Dashed lines indicate homeologs absent or nonfunctional in the reference. Details for each gene are presented in *SI Appendix*, Tables S22 and S23. (*C–F*) Effect of loss-of-function mutations for (*C*) *STARCH BRANCHING ENZYME IIa/b* (18), (*D*) *PHYTOCHROME C* (25), (*E*) *PHYTOCHROME B* (27), and (*F*) *VERNALIZATION 2* (26). Asterisk indicates published characterization of mutant phenotype. Double dagger indicates published characterization of mutant phenotype based on natural mutations.

lines, followed by self-pollination of the F_1 plant, to generate a segregating F_2 population. Homozygous F_2 double mutants can then be selected by using predesigned KASP assays to generate complete null mutants in the third generation. For breeding applications, however, the hexaploid population is most relevant, as bread wheat represents more than 95% of the wheat grown globally (29). In hexaploid wheat, an additional cross is required to combine mutations across all three genomes. Both populations are in genetic backgrounds with a spring growth habit and have relatively short generation times (10–16 wk depending on growing conditions).

The high mutation rates of these populations are valuable for the identification of novel alleles, including truncation mutations. However, they can also have a negative impact in the characterization of a specific mutation. Strategies to account for this high level of background mutation include the use of multiple independent mutants, backcrossing to reduce mutation load, and selecting for isogenic sibling lines that share background mutations. For mutations with subtle phenotypic effects or that require field phenotyping, it is advisable to backcross the mutant lines to the nonmutagenized parent for at least two generations before combining homeologous mutations (30, 31). Sibling lines can then be selected for homozygous WT or null-mutant alleles. Given that isogenic sibling lines share many of the same background mutations, comparisons between them provide a better assessment of the mutation effect than comparisons vs. the nonmutagenized control. Conversely, for highly penetrant traits, the comparison of complete null and WT F_2 sibling lines without backcrossing may suffice for the analysis.

Recent developments in genome-editing technologies offer the potential to induce mutant alleles in targeted regions of the wheat genome (32–34). Simultaneous homozygous deletions in all three homeologs have been achieved by using transient expression of the Cas9 endonuclease and short guide RNAs complementary to the three homeologs (34). However, homozygous triple mutants in the first T_0 generation were rare (68 plants from 7,680 bombarded embryos) and only ~0.5% (41 of 7,680) corresponded to transgene-free triple homozygous mutants (34). This suggests that, for most genome editing experiments, crossing will still be required to combine single or double mutant lines or to segregate out the DNA construct. Genome editing has the advantage of lower levels of background mutations, but the disadvantage of the high upfront

expense in construct design and transformation costs. By contrast, the sequenced mutant populations provide researchers with instant access to the mutant alleles with a simple online search and inexpensive seed request. Efficient wheat transformation is still limited to a small number of large research institutions, so the public sequenced mutant populations have the potential to democratize access to reverse genetic resources in wheat.

It is still not clear if genome-edited crops will be regulated as nontransgenic in all countries, which may impose constrains for globally traded crops such as wheat. There are still no commercially available transgenic wheat varieties, whereas EMS mutations have been used in agriculture for almost a century and are not under any regulation. We therefore predict that the mutant populations developed in this work will be very valuable and highly complementary to editing approaches in the future.

Uncovering Hidden Recessive Variation. The results for the flowering repressor *VRN2* mutants (Fig. 4*F*) are particularly illustrative of the limited phenotypic effect of recessive mutations in polyploid wheat. To date, no polyploid wheat variety has been described with a spring growth habit associated with recessive *vrn2* alleles (26). This is not caused by a lack of effect in polyploid wheat, as loss-of-function mutations at all three *VRN2* homeologs in hexaploid wheat results in a spring growth habit (26). It is also not caused by limited selection pressure, as more than 10 independent dominant mutations for spring growth habit have been described for the meristem identity gene *VRN1* (26). Finally, it is unlikely that the lack of spring types associated with *vrn-2* is caused by a low adaptive value, as most accessions of cultivated diploid wheat *Triticum monococcum* (35) and a large number of diploid barley varieties (36, 37) have a spring growth habit associated with recessive *vrn-2* mutations. Based on the previous evidence, we hypothesize that recessive mutations at the *VRN2* locus in polyploid wheat have remained hidden from selection for more than 8,000 y by the redundancy conferred by multiple homeologs.

Given the recent origin of the polyploid wheat species, many potentially useful induced and natural mutations are likely masked by functional redundancy among homeologs. The >10,000,000 sequenced mutations identified in wheat coding regions in the present study facilitate the identification of loss-of-function mutations in different homeologs and generates a large number of alleles. These mutations can be combined to study gene function and to reveal previously hidden phenotypic variation. Likewise, the effects of candidate genes from diploid grass species can now be studied directly in wheat, as recently shown for the wheat *TaGW2-A1* mutants with increased grain size identified in the tetraploid population (31). The strategy and methods developed herein can be also applied to other young polyploid crops with closely related genomes.

In summary, the mutant populations sequenced in the present study represent an invaluable resource for wheat functional genetics and provide a powerful tool to uncover variation previously hidden to human and natural selection in a central crop species for global food security.

Materials and Methods

Exome Capture Design. The wheat exome capture designs used in the present study were developed in collaboration with NimbleGen (Roche) and are publicly available to order from Roche catalog numbers 140228_Wheat_Dubcovsky_D18_REZ_HX1 (tetraploid wheat) and 140430_Wheat_TGAC_D14_REZ_HX1(hexaploid wheat). The sequences in this design comprise protein-coding transcript data from *T. turgidum* and *T. aestivum* transcriptome studies, wheat ESTs, wheat sequences homologous to barley gene models (38) not present in the previous wheat datasets, and hand-annotated sequences that were crowd-sourced from the wheat research community (*SI Appendix*, Table S1). A detailed description of the methods used for the development of the exome capture is presented in *SI Appendix*, Method S1.

A total of 82,511 protein coding sequences passed through all filters into the final design (*SI Appendix*, Table S1). Exon prediction was performed by aligning transcripts to the Chinese Spring Survey (CSS) sequences using the exonerate program (39) as described previously (40). Individual exons were padded with 30 bp from the introns on each side of the exons to increase capture efficiency at exon/intron borders. For *T. aestivum* sequences, all exon and padded sequences were derived from the genomic assembly. For *T. turgidum* sequences, original sequences were retained for the exons and only padding was supplemented from the *T. aestivum* genome. In total, we included 219,383 padded and 67,416 unpadded exons in the design, covering a total of 84 Mb (*SI Appendix*, Table S1). The exome capture design is available for BLAST and can be downloaded at dubcovskylab.ucdavis.edu/wheat-tilling and www.wheat-tilling.com.

Sample Preparation. A detailed description of the methods used for genomic DNA extraction and shearing, library construction and barcoding, capture hybridization and DNA recovery, amplification. and sequencing are presented in *SI Appendix*, Method S2 and Table S2.

Data Processing and Mapping Rates. Illumina 100-bp paired-end reads were preprocessed to trim 3′ adapter sequences and low quality. Trimmed reads were aligned to genome scaffolds of the CSS sequence (AB genomes for Kronos and ABD genomes for Cadenza) using bwa (41). For Cadenza, CSS scaffolds for chromosome 3B were replaced with the 3B pseudomolecule assembly (42). Alignments were sorted by using samtools (41), and duplicate reads were removed with Picard tools. Additional information is provided in *SI Appendix*, Method S3.

De Novo Assembly of Unmapped Reads. To increase the proportion of mapped reads, we supplemented the CSS reference with a de novo assembly of unmapped reads from Kronos and Cadenza. Supplementary de novo assemblies for tetraploid and hexaploid wheat were constructed separately by using 43,073,616 unmapped reads from 14 Kronos samples and 56,988,370 unmapped reads from 10 Cadenza samples (*SI Appendix*, Method S3 and Table S3).

MAPS Parameter Optimization. To identify EMS-induced mutations, we used the MAPS software that was previously tested on rice and eight wheat mutant lines (16) (*SI Appendix*, Method S4 and Fig. S2). In each MAPS run, we processed batches of 24 samples for tetraploid wheat and 24 or 32 samples for hexaploid wheat. Only SNPs detected in a single sample of the batch are reported by the MAPS pipeline. This removes varietal SNPs between the CS reference and Kronos/Cadenza and is also critical in polyploid species to eliminate polymorphisms among homeologs, which are present across all samples.

In wheat, EMS generates almost exclusively G to A and C to T changes, which are referred in the present study as EMS-type mutations. Therefore, the proportion of non–EMS-type SNPs can be used as a first approximation to the error rate. We also estimated the error rate by comparing the number of SNPs detected in the nonmutagenized WT line (no mutations expected) with the average number of mutations detected in the EMS-treated plants. By using these error estimates, we empirically adjusted several parameters in the MAPS pipeline to minimize the detection of false mutations without losing too much sensitivity (*SI Appendix*, Method S4).

Validation of EMS Mutations. We validated EMS mutations by examining their status in M_4 plants by using genome-specific KASP assays designed with PolyMarker (43) and through direct Sanger sequencing (*SI Appendix*, Tables S8–S10). The main objectives were to confirm the presence of the mutation in the M_4 progeny seed deposited in the public repositories and classify the M_2 as homozygous or heterozygous in the M_4 progeny seed.

Correction for Heterozygous/Homozygous Mutation Classification. In a single M_2 individual, the frequency of WT and mutant alleles is expected to be close to 50% for a heterozygous mutation. However, MAPS classifies mutations as heterozygous even when a single WT read is present at low frequency. A single mismapped read can result in a homozygous mutation being misclassified as heterozygous and in the overestimation of the ratio between heterozygous to homozygous mutations. This problem is exacerbated in polyploid species with similar genomes. To correct these errors, we introduced a filter in the bioinformatics pipeline that reclassified heterozygous mutations as homozygous when the frequency of the WT allele was less than 15% of reads (*SI Appendix*, Fig. S5).

Calculation of Mutation Density and Coverage. To estimate the mutation density (number of mutations per kb of captured sequence) across the population, we divided the total number of uniquely mapped mutations identified at

HetMC5/*HomMC3* (excluding RH) by the average number of positions used by MAPS to identify mutations (119.2 Mb for Kronos and 162.4 Mb for Cadenza). To calculate these last two numbers, we first obtained the number of bases covered by at least one read at quality higher than 20 in at least $N - 4$ samples from the MAPS batch from the intermediate "MAPS-assay" file (*SI Appendix*, Fig. S2). We then determined the number of bases covered by at least four reads in each individual, and used the average across all mutants in the population to estimate the average number of positions used by MAPS to identify mutations.

The coverage values presented in *SI Appendix*, Table S4, are based on mutant positions. These positions are selected by MAPS to have a minimum coverage of three reads and to be present in a high proportion of the samples in the same run, and therefore their coverage can be higher than the average from the complete population. However, the coverage values obtained for Kronos by using the previous method (26.6) was almost identical to the value obtained for 89.2 million positions independently of mutations in the comparison of the α-design and β-design in Kronos (28.8; *SI Appendix*, Fig. S1). This result suggests that coverage values estimated from mutant positions are not very different from the ones in the overall population.

EMS Sequence Preference. Sequence preference in sites adjacent to EMS mutation sites were calculated by using the method described previously (16). Briefly, we measured nucleotide frequencies in 20-bp regions flanking the mutated G nucleotides and compared it to regions flanking a nonmutagenized G located 40–50 bp upstream and downstream of each mutation site. Sequence preference at each position was expressed as the difference between the percent frequency in the base flanking the mutated G and the corresponding frequency in the control sites (Fig. 3 *E* and *F* and *SI Appendix*, Fig. S7).

Reads Mapped to Multiple Locations (i.e., Multimapped Reads). The bioinformatics pipeline used to detect MM is described in *SI Appendix*, Fig. S3, and the methods used to select the primary location, visualize the alternative locations, and validate the MM mutations are described in *SI Appendix*, Method S8.

Identification and Validation of Large Deletions. To identify and characterize homozygous deletions in our mutant populations, we developed a custom bioinformatics pipeline that examines relative coverage of exons within and across mutant lines (*SI Appendix*, Method S9 and Fig. S4). The methods used to validate these large mutations are described in *SI Appendix*, Method S10).

Variant Effect Prediction. Mutation effects on gene function were predicted on the final SNP files (*HetMC5*/*HomMC3*) without RH using the Variant Effect Predictor program (44) from Ensembl tools release 78 in offline mode (*SI Appendix*, Method S11).

Access to and Visualization of Mutations. The raw reads for the tetraploid and hexaploid projects are available from National Center for Biotechnology Information BioProject PRJNA258539 and European Read Archive ENA (European Nucleotide Archive) study PRJEB11524, respectively. In addition, we deposited the uniquely mapped EMS-type mutations at *HetMC5*/*HomMC3* (excluding RH) in EnsemblPlants. These platforms are searchable through string searches or BLAST queries.

Access to Mutant Seed Stocks and SNP Markers. For the Kronos and Cadenza Targeting Induced Local Lesions in Genomes (TILLING) populations, M_3 seed was collected from the individual M_2 plants used for DNA extraction and exome sequencing. For initial seed bulking, ~30 M_3 siblings were grown in the field as single rows and all M_4 seed was harvested and bulked for each mutant line. For lines with low yields (<60 g), an additional set of M_3 siblings was grown in the glasshouse or field to increase seed quantity. Additional backups of the complete tetraploid mutant population have been deposited in Centro Internacional de Mejoramiento de Maiz y Trigo (Mexico), Shandong University, the University of Saskatchewan, the quarantine repository in Australia, the Cereal Disease Laboratory, and Washington State University. Likewise, backups of the complete hexaploid mutant population have been deposited at Rothamsted Research, National Institute of Agricultural Botany, the French National Institute for Agricultural Research, and University College Dublin.

To generate KASP assays for the *HetMC5*/*HomMC3* EMS-type mutations in the database, we ran the PolyMarker pipeline (43). For the allele-specific primers, fluorophore-compatible tails need to be added to the 5′ end before oligo synthesis (45).

Code Availability. All code is available through the wheat TILLING project GitHub page: https://github.com/DubcovskyLab/wheat_tilling_pub.

ACKNOWLEDGMENTS. The authors thank Dr. Marcelo Soria for help with the EST analysis; Dr. Wenjun Zhang for mutation validation; Dr. Luca Comai for valuable suggestions for the analysis of TILLING datasets; Dr. Robert King for assistance with the deletion analyses; Meric Lieberman for Mutation and Polymorphism Survey (MAPS) advice; Dr. Martin Trick for developing initial stages of the www.wheat-tilling.com database; Drs. Mike Ambrose, Adrian Turner, and Richard Horler for developing the TILLING seed database at SeedStor; the John Innes Centre (JIC) field trials and horticultural team for plant husbandry; members of the Platforms and Pipelines Group for the next-generation sequencing and library construction; and the Norwich BioScience Institutes Computing infrastructure for Science group through the JIC and Earlham Institute clusters. This work was supported by the Howard Hughes Medical Institute (J.D.); Gordon and Betty Moore Foundation Grant GBMF3031 (to J.D.); NRI Competitive Grants 2011-68002-30029 and 2017-67007-25939 from the US Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA; to J.D.); UK Biotechnology and Biological Sciences Research Council (BBSRC) Grants BB/J003557/1, BB/J003913/1, and BB/J003743/1 (to C.U., K.V.K., A.P., S.A., and L.C.); BBSRC and Institute Strategic Programme Grant at The Earlham Institute BB/J004669/1; BBSRC National Capability in Genomics at The Earlham Institute Grant BB/J010375/1; BBSRC Future Leader Fellowship BB/M014045/1 (to P. Borrill); USDA NIFA postdoctoral fellowship 2012-67012-19811 (to K.V.K.), a Norwich Research Park PhD Studentship (to R.H.R.-G.), and an The Earlham Institute Funding and Maintenance Grant (to R.H.R.-G.). The following institutions contributed funding to sequence 100 tetraploid mutant lines each: Shandong University, University of Saskatchewan, Commonwealth Scientific and Industrial Research Organisation, USDA–Agricultural Research Service Cereal Disease Laboratory, and Washington State University.

1. Dubcovsky J, Dvorak J (2007) Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science* 316(5833):1862–1866.
2. Tsai H, et al. (2013) Production of a high-efficiency TILLING population through polyploidization. *Plant Physiol* 161(4):1604–1614.
3. Rakszegi M, et al. (2010) Diversity of agronomic and morphological traits in a mutant population of bread wheat studied in the Healthgrain program. *Euphytica* 174: 409–421.
4. Slade AJ, Knauf VC (2005) TILLING moves beyond functional genomics into crop improvement. *Transgenic Res* 14(2):109–115.
5. Uauy C, et al. (2009) A modified TILLING approach to detect induced mutations in tetraploid and hexaploid wheat. *BMC Plant Biol* 9:115.
6. Wang TL, Uauy C, Robson F, Till B (2012) TILLING *in extremis*. *Plant Biotechnol J* 10(7):761–772.
7. Mamanova L, et al. (2010) Target-enrichment strategies for next-generation sequencing. *Nat Methods* 7(2):111–118.
8. King R, et al. (2015) Mutation scanning in wheat by exon capture and next-generation sequencing. *PLoS One* 10(9):e0137549.
9. Jiao Y, et al. (2016) A sorghum mutant resource as an efficient platform for gene discovery in grasses. *Plant Cell* 28(7):1551–1562.
10. Bennett MD, Smith JB (1976) Nuclear DNA amounts in angiosperms. *Philos Trans R Soc Lond B Biol Sci* 274(933):227–274.
11. Mayer KFX, et al. (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* 345(6194):1251788.
12. Jordan KW, et al. (2015) A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. *Genome Biol* 16:48.
13. Schreiber AW, et al. (2012) Transcriptome-scale homoeolog-specific transcript assemblies of bread wheat. *BMC Genomics* 13:492.
14. Guo Y, Abernathy B, Zeng Y, Ozias-Akins P (2015) TILLING by sequencing to identify induced mutations in stress resistance genes of peanut (*Arachis hypogaea*). *BMC Genomics* 16:157.
15. Cannon SB, Shoemaker RC (2012) Evolutionary and comparative analyses of the soybean genome. *Breed Sci* 61(5):437–444.
16. Henry IM, et al. (2014) Efficient genome-wide detection and cataloging of EMS-induced mutations using exome capture and next-generation sequencing. *Plant Cell* 26(4):1382–1397.
17. Ng PC, Henikoff S (2003) SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 31(13):3812–3814.
18. Hazard B, Zhang X, Naemeh M, Dubcovsky J (2014) Registration of durum wheat germplasm lines with combined mutations in *SBEIIa* and *SBEIIb* genes conferring increased amylose and resistant starch. *J Plant Regist* 8(3):334–338.
19. Schönhofen A, Hazard B, Zhang X, Dubcovsky J (2016) Registration of common wheat germplasm with mutations in *SBEII* genes conferring increased grain amylose and resistant starch content. *J Plant Regist* 10(2):200–205.
20. Amini A, Khalili L, Keshtiban AK, Homayouni A (2016) Resistant starch as a bioactive compound in colorectal cancer prevention. *Bioactive Foods in Health Promotion*, eds Watson RR, Preedy VR (Academic, Cambridge, UK), pp 773–780.
21. Robertson MD, Bickerton AS, Dennis AL, Vidal H, Frayn KN (2005) Insulin-sensitizing effects of dietary resistant starch and effects on skeletal muscle and adipose tissue metabolism. *Am J Clin Nutr* 82(3):559–567.
22. Wong THT, Louie JCY (2016) The relationship between resistant starch and glycemic control: A review on current evidence and possible mechanisms. *Starke* 68:1–9.
23. Alvarez MA, Tranquilli G, Lewis S, Kippes N, Dubcovsky J (2016) Genetic and physical mapping of the earliness per se locus *Eps-A^{m}1* in *Triticum monococcum* identifies *EARLY FLOWERING 3* (*ELF3*) as a candidate gene. *Funct Integr Genomics* 16(4):365–382.

24. Chen A, Dubcovsky J (2012) Wheat TILLING mutants show that the vernalization gene *VRN1* down-regulates the flowering repressor *VRN2* in leaves but is not essential for flowering. *PLoS Genet* 8(12):e1003134.
25. Chen A, et al. (2014) PHYTOCHROME C plays a major role in the acceleration of wheat flowering under long-day photoperiod. *Proc Natl Acad Sci USA* 111(28):10037–10044.
26. Kippes N, Chen A, Zhang X, Lukaszewski AJ, Dubcovsky J (2016) Development and characterization of a spring hexaploid wheat line with no functional *VRN2* genes. *Theor Appl Genet* 129(7):1417–1428.
27. Pearce S, Kippes N, Chen A, Debernardi JM, Dubcovsky J (2016) RNA-seq studies using wheat *PHYTOCHROME B* and *PHYTOCHROME C* mutants reveal shared and specific functions in the regulation of flowering and shade-avoidance pathways. *BMC Plant Biol* 16(1):141.
28. Lv B, et al. (2014) Characterization of *FLOWERING LOCUS T1* (*FT1*) gene in *Brachypodium* and wheat. *PLoS One* 9(4):e94171.
29. Taylor RD, Koo WW (2015) 2015 Outlook of the U.S. and World Wheat Industries, 2015-2024. Agribusiness & Applied Economics. Agribusiness & Applied Economics Report 738 (North Dakota State Univ, Fargo, ND), p 23.
30. Hazard B, et al. (2015) Mutations in durum wheat *SBEII* genes affect grain yield components, quality, and fermentation responses in rats. *Crop Sci* 55(6):2813–2825.
31. Simmonds J, et al. (2016) A splice acceptor site mutation in *TaGW2-A1* increases thousand grain weight in tetraploid and hexaploid wheat through wider and longer grains. *Theor Appl Genet* 129(6):1099–1112.
32. Shan Q, Wang Y, Li J, Gao C (2014) Genome editing in rice and wheat using the CRISPR/Cas system. *Nat Protoc* 9(10):2395–2410.
33. Wang W, Akhunova A, Chao S, Akhunov E (2016) Optimizing multiplex CRISPR/Cas9-based genome editing for wheat. *bioRxiv*, 10.1101/051342.
34. Zhang Y, et al. (2016) Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. *Nat Commun* 7:12617.
35. Yan L, et al. (2004) The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. *Science* 303(5664):1640–1644.
36. Karsai I, et al. (2005) The *Vrn-H2* locus is a major determinant of flowering time in a facultative x winter growth habit barley (*Hordeum vulgare* L.) mapping population. *Theor Appl Genet* 110(8):1458–1466.
37. Szűcs P, et al. (2007) Validation of the *VRN-H2*/*VRN-H1* epistatic model in barley reveals that intron length variation in *VRN-H1* may account for a continuum of vernalization sensitivity. *Mol Genet Genomics* 277(3):249–261.
38. Mayer KFX, et al. (2012) A physical, genetic and functional sequence assembly of the barley genome. *Nature* 491(7426):711–716.
39. Slater GSC, Birney E (2005) Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6:31.
40. Krasileva KV, et al. (2013) Separating homeologs by phasing in the tetraploid wheat transcriptome. *Genome Biol* 14(6):R66.
41. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
42. Choulet F, et al. (2014) Structural and functional partitioning of bread wheat chromosome 3B. *Science* 345(6194):1249721.
43. Ramirez-Gonzalez RH, Uauy C, Caccamo M (2015) PolyMarker: A fast polyploid primer design pipeline. *Bioinformatics* 31(12):2038–2039.
44. McLaren W, et al. (2016) TheEnsembl Variant Effect Predictor. *Genome Biol* 17(1):122.
45. Ramirez-Gonzalez RH, et al. (2015) RNA-Seq bulked segregant analysis enables the identification of high-resolution genetic markers for breeding in hexaploid wheat. *Plant Biotechnol J* 13(5):613–624.

Supplementary Information Appendix

Uncovering hidden variation in the young polyploid wheat genomes

K.V. Krasileva, H. Vasquez-Gross, T. Howell, P. Bailey, F. Paraiso, L. Clissold, J. Simmonds, R. H. Ramirez-Gonzalez, X. Wang, P. Borrill C. Fosker, S. Ayling, A. Phillips, C. Uauy, J. Dubcovsky

* correspondence to: jdubcovsky@ucdavis.edu and Cristobal.Uauy@jic.ac.uk

This file includes:

SI Appendix, Methods S1 to S12

SI Appendix, Text S1 to S4

SI Appendix, Figures S1 to S12

SI Appendix, Tables S1 to S23

SI Appendix, References

SI Appendix, Methods

SI Appendix, Method S1. Exome capture design

First we obtained 56,831 protein-coding sequences previously annotated in *T. turgidum* cv. Kronos transcriptome (*SI Appendix*, Table S1) (1). The *T. aestivum* transcriptome assemblies from cultivars Kukri (2) and Chinese Spring (3) were combined with the Kronos transcriptome using CD-HIT-EST clustering (94% identity cutoff) (4), and only protein coding sequences annotated by *findorf* were retained (1). We then used the CD-HIT-EST-2D program to add sequences from four additional datasets (*SI Appendix*, Table S1): i) full-length cDNAs from the RIKEN Plant Science Center Japan (5), ii) 30,497 contigs assembled from senescing leaves of hexaploid cv. Bobwhite (6) and annotated with *findorf*, iii) wheat proteins from NCBI that were not present in the *T. turgidum* predicted proteins, and iv) wheat EST sequences available from NCBI (as of Oct 2012). For the last data set (iv), the sequences were passed through the *SeqTrim* pipeline (7) to remove poly-A, poly-T tails, and chimeric reads, and then assembled with the TIGR Gene Indices clustering tools (TGICL) (8) after masking vector contaminants, transposons and repeats using *cross_match*, UniVec (NCBI) and TREP databases (9). *Blastx* with e-value cutoff $1e^{-5}$ against the *Viridiplantae* section of the non-redundant (nr) nucleotide collection of GenBank was used to select ESTs with protein coding-potential.

In addition, we identified 2,002 protein-coding sequences from the barley genome project (10) that were not present in our wheat datasets at >85% identity cutoff and added the corresponding wheat homologs to the dataset. After removal of transposons, the remaining 1,798 sequences were used to search for wheat exons in the Chinese Spring chromosome survey (CSS) sequence (11). Matching sequences were retained as wheat exons and added to the capture design. Finally, we set up a BLAST database and invited wheat researchers to submit sequences not present in our study. Based on these analyses, we added 123 hand-curated sequences to the dataset.

Sequences from all sources were combined and the final set was passed through CD-HIT-EST clustering (99% identity cutoff) to remove any residual redundancy. The dataset was further curated by eliminating any contaminants from human and *E. coli* DNA, wheat plastids and ribosomal sequences using BLAT (12), and other contaminants (e.g. DNA from wheat pathogens) using taxonomy-based searches as described previously (1). The filtered non-

redundant contigs were analyzed with *findorf* to identify coding regions and to remove potential pseudogenes. Transposons were removed based on similarity to the TREP database with BLAST (*blastn*, $1e^{-10}$) and transposon-associated Pfam domains with HMMER (*hmmscan*, $1e^{-3}$). To prevent the elimination of important repetitive gene families such as disease resistance genes (R-genes) and gliadins during the removal of repetitive sequences, we manually curated 451 NLR resistance genes (R-genes) and 189 gliadins and included them in the design without passing through any masking filters. Sequences containing large runs of N's were split using the *seqtk cutN* program (https://github.com/lh3/seqtk).

SI Appendix, Method S2. Sample preparation

Genomic DNA was extracted from leaves of individual M_2 adult plants (Fig. 1). For the *T. turgidum* samples, DNA was extracted following a large-scale extraction protocol that includes an initial step of nuclei purification, followed by proteinase K and phenol-chloroform purification and normalization to a final concentration of 200 ng/µL in low-EDTA TE buffer (0.1 mM EDTA, 10 mM Tris-HCl, pH 8.0) (13). For the *T. aestivum* samples, DNA was extracted using MagAttract DNA Blood M96 Kit (Qiagen) following the instructions provided by the supplier. The freeze-dried material was lysed with ammonium acetate and precipitated on to Agencourt Genfind v2 (Beckman Coulter_ A41497) magnetic beads and washed several times. The purified *T. aestivum* DNA was eluted into low-EDTA TE using the Beckman FXp robotic system and the samples were normalised to 20 ng/µL on a Beckman NX platform. Purified DNA samples were sheared on a Covaris E220 instrument using settings specific for each species (*SI Appendix*, Table S2). Genomic DNA libraries for both species were constructed with a Sciclone G3 robotics (PerkinElmer) using the High-Throughput Library Preparation kits from KAPA Biosystems, Inc. (Wilmington, MA, USA, catalog number KK8234) following the Maestro KAPA HTP protocol with dual-SPRI bead size selection (https://www.kapabiosystems.com/assets/KAPA-HTP-LPK_Sciclone-User-Guide.pdf). The tetraploid libraries were barcoded using NEXTflex-96™ oligos 1-48 (Bioo Scientific, Austin, TX, USA, catalogue number 514105), whereas the hexaploid libraries used Roche adapters – set "A" (1-12) (Roche, catalogue number 07141530001). During the library preparation step, samples were amplified by PCR, using five amplification cycles for *T. turgidum* and six for *T.*

aestivum. The products were purified with Agencourt AMPure beads (Beckman Coulter, A63881) on the Sciclone G3 platform and were eluted in ultrapure DNAse/RNAse free distilled water.

In preparation for the capture, DNA libraries were pooled together and blocking oligos for Illumina adapters and repetitive DNA sequences (developer reagent) were added to minimize non-specific binding and improve the number of reads on target (*SI Appendix*, Table S2). The DNA mixture was dried using a speed vacuum centrifuge. Capture hybridization and washes were done according to recommended protocols from Roche SeqCapEZ User Guide 4.2 v7 or automated on the Sciclone machine (Perkin Elmer). The DNA pellet was dissolved in 7.5 µL of hybridization Solution 5 and 3 µL of Solution 6 (Roche, catalogue number 5634253001), denatured at 95 ºC for 10 minutes and hybridized to the SEQCAP EZ probes (140228_Wheat_Dubcovsky_D18_REZ_HX1 for *T. turgidum* and 140430_Wheat_TGAC_D14_REZ_HX1 for *T. aestivum*) for 70 h at 47 ºC in a thermocycler (lid temperature set to 57 ºC). The hybridization reaction was recovered using Dynabeads M-270 streptavidin beads (Invitrogen, 653-06) and washed according to the manufacturer's protocol (Roche, catalogue number 5634253001).

The captured DNA was amplified for ten cycles for *T. turgidum* and seven cycles for *T. aestivum* using KAPA Readymix Amplification kit (KAPA Biosystems, Inc., catalogue number KK2612) and purified in 1.8 x volume of Agencourt AMPure beads (Beckman Coulter, catalogue number A63881). Captured DNA was eluted in 30 µL of ultrapure water and quantified using QUBIT Systems equipment. The fold enrichment of the targeted exons was estimated by qRT-PCR using primers for two wheat housekeeping marker genes (Nuclear-encoded Rubisco, Ta_cDNA_5.1 Fw ATCGGATTCGACAACATGC; Ta_cDNA_5.1 Rev ATATGGCCTGTCGTGAGTGA; and Malate Dehydrogenase Ta_cDNA_51.1 Fw AAAGGCGTCAAGATGGAGTT Ta_cDNA_51.1 Rev GGAATCCACCAACCATAACC).

Each tetraploid wheat capture (8-plex pool) was sequenced in one lane of Illumina HiSeq2000 (1/8 of a lane per sample). For 16 Kronos samples that had fewer than 20 million read-mates, an additional round of Illumina sequencing was performed including the 16 lines in one Illumina lane, and the two sources of reads were combined *in silico* for each mutant line. Captures from Cadenza include an additional genome (D genome) and have a higher proportion of duplicated

reads (due to an additional PCR cycle during library construction), so a smaller number of lines were pooled per Illumina lane. The hexaploid wheat 4-plex pools were run on one lane of Illumina HiSeq2500, and the 8-plex pools were run on two lanes of Illumina HiSeq2500 (total ¼ of a sequencing lane per sample). All hexaploid wheat samples had more than 20 million read-mates.

SI Appendix, Method S3. *De novo* assembly

Data processing and mapping rates: The 3' adapter sequences and low quality bases of the Illumina 100-bp paired-end reads were trimmed using the *scythe* (https://github.com/vsbuffalo/scythe) and *sickle* programs (https://github.com/najoshi/sickle). Trimmed reads were aligned to the A and B genome scaffolds of the CSS sequence for Kronos and to the A, B and D genome scaffolds for Cadenza (hexaploid) data using *bwa aln* and *bwa sampe* programs (14). In the case of the Cadenza samples, CSS scaffolds for chromosome 3B were replaced with the new 3B pseudomolecule assembly (http://plants.ensembl.org/Triticum_aestivum/Info/Index) supplemented by CSS 3B scaffolds which were absent in the new 3B pseudomolecule. Alignments were sorted using *samtools* (15) and duplicate reads were marked and removed with Picard tools *rmdup* (http://broadinstitute.github.io/picard/). Mapping statistics were calculated with *samtools view*.

To increase the proportion of mapped reads, we supplemented the CSS reference with a *de novo* assembly of unmapped reads from Kronos and Cadenza. We expected this additional sequence to include variety-specific genes or genes currently absent in the reference assembly. This is particularly important for capturing rapidly evolving NLRs resistance genes that are unique to the mutagenized genotypes. We hypothesized that by combining the sequences of multiple independent captures into the *de novo* assembly, we would dilute the noise (different non-targeted sequences included in the individual captures) and increase the signal (targeted sequences present in the capture), enhancing the signal to noise ratio.

De novo assembly of unmapped reads. Unmapped reads were extracted from the bam files using *samtools view* (15) with the 0x0004 bitwise flag and converted to fastq files using Bedtools v2.17.0 *bamtofastq* (16). Reads were assembled with *MaSuRCA* software, chosen for its high performance on the pine genome (17). We compared assembles at k-mers 31, 51 and 63 on the

Kronos dataset and evaluated the results by N50, length of the assembled region, number of reads mapped, number of reads mapped in pairs, number of reads mapped above Q30. The k-mer 63 assembly performed best based on all metrics and was chosen for both Kronos and Cadenza final assemblies. The *de novo* assemblies (*SI Appendix*, Table S3) were added to the references as unknown chromosome UCW_Kronos_ChrU for Kronos (40,975 contigs, 33.4 Mb) and TGAC_Cadenza_U for Cadenza (67,632 contigs, 41.3 Mb).

SI Appendix, Method S4. of MAPS parameter optimization

From the alignment of the reads to the improved references (CSS survey sequence + *de novo* assemblies, *SI Appendix*, Method S3), we called SNPs using default *mpileup* parameters and mapping quality higher than 20 (*SI Appendix*, Fig. S2). We then used the MAPS pipeline to select bases in the reference covered by at least one read at quality higher than 20 in a minimum number of samples. This number is determined by the *MinLib* parameter, which was set equal to the total number of samples in the batch minus four. For example, we used *MinLib* = 20 for batches of 24 samples and *MinLib* = 28 for batches of 32 samples. This number was selected to ensure that at least half of the lines in each capture including eight individuals had a minimum coverage of one read at quality higher than 20. This threshold showed a low number of false positives and was adopted for the complete project.

An additional MAPS parameter that is critical to differentiate real mutations from sequencing errors is the minimum number of reads carrying the mutation (minimum coverage, henceforth, *MC*) required to call a mutation. This threshold is established independently for homozygous and heterozygous using the parameters *HomMC* and *HetMC*, respectively. Unless indicated otherwise, all the numbers presented in this study were calculated at *HomMC*=3 (homozygous mutation present in all reads from an individual and detected at least three times) and *HetMC*=5 (heterozygous mutation detected in at least five reads). Statistics for different *HomMC*/*HetMC* combinations and their corresponding estimated errors are provided in *SI Appendix*, Tables S5 and S6 for Kronos and Cadenza, respectively. When the mutations detected at lower thresholds were analyzed separately from the rest of the mutations, the estimated error rate was higher than at *HetMC5/HomMC3* but still lower than 10.0 %. Although it is safer to use mutations identified at high stringency levels (e.g. *HetMC5/HomMC3*), there is still a good probability to find a

mutation detected at a lower threshold (>90%). At *HetMC3/HomMC2*, the number of detected EMS-type mutations increased to 5,085,379 in Kronos and 8,083,066 mutations in Cadenza (total ~13 million mutations, *SI Appendix*, Tables S5-S6).

In addition to *HetMC*, the MAPS pipeline uses the *HetMinPer* parameter to reduce the probability of calling sequencing errors as heterozygous mutations in regions of high coverage. *HetMinPer* determines the minimum percent of mutant reads required for calling a heterozygous mutation. This parameter was set at 20% in diploid rice (18) but was adjusted in this study to 15% for tetraploid Kronos and to 10% for hexaploid Cadenza to account for the differences in ploidy level. In polyploid wheat, reads from different homoeologs can map to the same reference if one of the homoeologs is absent in the reference.

SI Appendix, Method S5. Residual heterogeneity (RH)

The seeds used to generate the Kronos and Cadenza TILLING populations were obtained from active breeding programs. Usually, wheat breeders self-pollinate lines for 6-10 generations before pooling the seeds of multiple plants to produce the final commercial seed stock. Depending on the number of generations of self-pollination before pooling multiple plants, different levels of residual genetic heterogeneity (henceforth "RH") are expected from the naturally occurring polymorphisms between the parental lines of the varieties. If the same RH region is present in more than one of the lines analyzed within the same MAPS run, the SNPs are not reported by the program. However, if the RH region is present in only one line in the run, the SNPs are reported as mutations by MAPS (even though they were not induced by EMS mutagenesis). It is important to identify RH regions because they affect the estimation of several mutation parameters and also because they can complicate the validation of mutations within these regions.

Criteria to identify RH regions. Four characteristics were used to differentiate the RH regions from regions carrying real EMS-induced mutations:

i) The RH-SNPs are more likely to be present in multiple individuals in the population, since different inbred plants are pooled and then self-pollinated to generate the commercial seed. Among the identified RH-SNPs, the mode of the distributions of SNP shared by different

numbers of individuals was 18 lines in Kronos and 10 in Cadenza (Figs. 3A-B). By contrast, the mode for the non-RH region was 1 line (99% of the mutations were found in a single individual).
ii) The RH-SNPs are expected to show a higher percent of non-EMS-type mutations than the regions containing only EMS-induced mutations. Among the identified RH-SNPs, the percentage of non-EMS-type mutations (76.4% in Kronos and 84.4% in Cadenza) was more than 75-fold higher than the percentage detected in the non-RH regions (<1% in both populations, Table 1).
iii) SNPs from RH regions are expected to have a higher proportion of homozygous mutations due to the multiple generations of self-pollination during seed increases. The SNPs in the identified RH regions showed a six- to seven-fold lower ratio of heterozygous to homozygous mutations (Kronos = 0.33 and Cadenza = 0.30) than those in the non-RH regions (Kronos = 1.87 and Cadenza = 2.21).
iv) RH regions are expected to have a higher average SNP density than the regions including only induced mutations. Among the identified RH-SNPs, SNP densities per individual line were 25.7-fold higher than in the non-RH regions in Kronos and 13.4-fold higher in Cadenza (*SI Appendix*, Fig. S6).

Bioinformatics pipeline to identify RH regions. To identify these RH regions, we developed a custom pipeline (https://github.com/DubcovskyLab/wheat_tilling_pub), which uses the output files generated from the MAPS pipeline. The first step of the pipeline breaks large scaffolds in the reference into 10 kb bins to avoid calling all mutations on a large scaffold as RH if it has a small RH region. This was particularly important for the large 3B pseudomolecule. Next, for each bin in each individual the pipeline calculates a score based on the criteria described in the previous section and in *SI Appendix*, Table S11. Intervals with a score of 12.5 or higher are tagged as RH regions in the database and users are warned in the JBrowse viewer if they are in a RH region. Using this bioinformatics pipeline, we identified 69,651 SNPs in RH regions in the tetraploid population (1.7%), and 38,626 SNPs in RH regions in the hexaploid population (0.6%) at *HetMC5/HomMC3* (Table 1).

EMS-type and reciprocal transitions in RH Regions. Within the identified RH regions of tetraploid wheat, we identified similar numbers of EMS-type G>A and C>T mutations (16,412) and reciprocal A>G and T>C transitions (20,358) at *HetMC5/HomMC3*. At the same stringency, we also detected similar numbers in Cadenza (EMS-type 6,023 and reciprocal transitions 8,669). We took advantage of this similarity to use the number of A>G and T>C transitions within the

non-EMS SNPs as an estimate of the maximum number of non EMS-induced G>A and C>T SNPs that could have been incorrectly included as EMS-type mutations in the non-RH regions.

Real EMS-type induced mutations within RH regions. Real EMS-type induced mutations are also present within the RH regions and could be tentatively identified by their presence in single lines (see blue arrows in *SI Appendix*, Fig. S6). However, the relatively high SNP density in the RH regions increases the probability that a linked SNP rather than the induced mutation caused the distinctive phenotype found in the mutant line. Two different strategies can be used to avoid this problem depending on the status of the mutation. For homozygous mutations, the phenotype of the line with the putative EMS-type mutation can be compared the phenotypes of other lines carrying the same RH region(s). If only the plants carrying the putative EMS-induced mutation show the phenotype, this would suggest that the phenotype is not caused by the SNPs present in the linked RH region. For heterozygous mutations, sibling lines with and without the EMS-induced mutation can be compared.

SI Appendix, Method S6. Estimation of the proportion of "accessible" G residues

The large number of mutations detected in multiple individuals provided a unique opportunity to estimate the probability that the G/C sites present in the sequenced region would be affected by the EMS mutagen. The duplicated EMS-induced mutations followed an approximate Poisson distribution with a maximum at 2 individuals and a rapid decay as the number of lines including the same mutation increased (Fig. 3C-D, red bars). To estimate the "proportion of accessible G residues", we first estimated the total number of G/C sites in the captured sequence using the percent G/C content in our capture design (average 46.8%). The probability of mutation was calculated by dividing the total number of observed EMS-type mutations (Table 1) by the number of predicted G/C sites. We then estimated the proportion of these G/C sites that would have generated a Poisson distribution most similar to the observed data (Fig. 3C-D, light blue bars). A reduced number of "accessible" G/C sites results in a higher probability of mutation, and higher predicted Poisson frequencies. We found this optimum similarity when the Poisson distribution was calculated using only 17.9 % of the G sites in the captured sequence from Kronos and 20.7 % of the sites in the captured sequence from Cadenza (*SI Appendix*, Tables S13 and S14). Although this is just an approximation, these numbers suggest that a large proportion

of G residues in the coding regions have a very small probability of being modified by the EMS mutagen.

SI Appendix, Method S7. EMS sequence preference

To estimate EMS sequence preference, we followed the method described before for rice (18). In both Kronos and Cadenza, we observed relatively high frequency of C bases at position +1 downstream of the mutated G, and of G at position -1 and +2 relative to the mutated G. A negative bias for T was also observed in both populations 1 bp upstream of the mutagenized site, a profile that is very similar to what was described before for rice (18). A weaker preference for G was also observed for up to 8 bp upstream or downstream of the mutagenized site (Fig. 3E-F and *SI Appendix*, Fig. S7).

To test if mutations present in two or more lines (581,992 EMS-type mutations in Kronos and 858,444 in Cadenza) have a stronger EMS sequence preference than mutations present in only one line, we analyzed both groups of mutations separately. The mutations present in two or more lines showed stronger sequence preferences at positions -1, +1 and +2 in both populations. These results suggest that G residues flanked by sequences similar to the favored EMS preference profile have higher probabilities of being affected by the mutagen and therefore a higher chance of occurring in multiple individuals.

As a consequence of this EMS sequence preference, the potential number and distribution of mutations in a particular gene is determined by its nucleotide sequence (e.g. G/C content and their sequence context).

SI Appendix, Method S8. Reads mapped to multiple locations

Analyses of regions included in the capture design but that showed no mutations revealed the presence of highly similar scaffolds in the reference (e.g. recently duplicated paralogs, and artificially duplicated scaffolds generated during the assembly of the reference sequence). Reads that mapped to these regions were assigned to multiple mapping locations and, as a result, received very low mapping quality scores. These reads, designated hereafter as “multi-mapping

reads" or simply "MM", fell below the selected mapping quality threshold of 20, creating blind spots with few or no mapped reads. To recover the mutations from these regions, we created a separate bioinformatics pipeline, outlined in *SI Appendix*, Fig. S3. Briefly, reads with a *BWA* (14) mapping quality of less than 20 that had more than one but fewer than eleven mapping locations were extracted for each mutant. A "best" mapping location was chosen from among all potential mapping locations, while keeping a record of all alternative mapping locations.

The following criteria were applied sequentially to each possible mapping location until only a single mapping location was selected. First we selected the location with the lowest edit distance (number of deletions, insertions, and substitutions needed to transform the reference sequence into the read sequence) from the *BWA* "NM" flag (14). If there were locations with identical edit distances, we selected the position with the highest number of alignment matches to avoid favoring indels given the same edit distance. If the previous two parameters were identical for multiple locations, we selected the location in the longest scaffold. The majority of reads could be assigned to a location using the three criteria above, but in the few cases where reads mapped equally well to two scaffolds of the same length, the scaffold that occured last alphanumerically was chosen. When reads mapped to multiple locations in a single scaffold, the highest bp mapping position on a scaffold was chosen.

Once a best mapping location was determined, the BAM/SAM line was updated to reflect the new mapping location and the mapping quality was changed to 255 (unknown) so that it would pass the MAPS mapping quality threshold of 20 (multi-map-corrector-V1.6.py available on https://github.com/DubcovskyLab/wheat_tilling_pub). MM reads recovered from hexaploid wheat were processed in batches of 24-32 individuals using the same *MinLibs* threshold as the main pipeline. To accelerate the mapping process, MM reads from tetraploid wheat were processed with the MAPS pipeline in larger batches of 72-164 individuals with *MinLibs* set to 83% of the number of individuals in the batch, to match the proportion of 20/24 used in the main pipeline. Using this pipeline, we identified 448,152 EMS-type mutations in Kronos and 1,427,823 in Cadenza at *HetMC3/ HomMC2*, leading to a total of 14.9 million mutations detected at this stringency level.

To help users identify multi-mapped mutations, a red bar is displayed on JBrowse when multi-mapped mutations can be found on a different scaffold(s). Alternative scaffolds are listed with

the corresponding hyperlinks. When multiple mapping locations are due to artificial duplications of the reference, the real location in the genome will be unique and validation will be simple. However, when alternate MM locations are caused by very similar paralogous or homoeologous sequences, the user will need to determine experimentally which of the alternative locations has the mutation.

We selected 25 MM mutations for validation (*SI Appendix*, Table S15). The validation strategy consisted of genome specific PCR amplification across the most likely multi-mapped region from six M_4 plants and two wild-type as controls followed by Sanger sequencing of the PCR products. We confirmed 22 MM mutations and in 20 of them we also confirmed the expected segregation pattern based on the M_2 classification as heterozygous or homozygous. In two cases, a homozygous mutation based on the M_2 classification was found segregating in the M_4 progeny. For three Cadenza assays, we could not identify the putative mutation in six M_4 plants, leading to an overall validation rate of 88% (22/25 assays).

We also observed the complementary situation to the multi-mapped reads: some duplicated regions with high levels of sequence identity in Kronos and Cadenza were represented by a single scaffold in the reference sequence. *PPD-B1* in Kronos and *ZCCT-B2* genes are examples of recently duplicated genes in Kronos represented by a single scaffold in the current CSS reference. These duplicated regions can be identified by three distinctive characteristics. First, almost all the mutations in these regions are expected to be classified as heterozygous because wild-type reads from the alternative copy are always present. For example, in the duplicated *PPD-B1* gene from Kronos all 189 mutations detected for this gene were classified as heterozygous. By contrast, the heterozygous to homozygous ratio for the non-duplicated *PPD-A1* homoeolog was normal (73/42). Second, a higher ratio of wild-type to mutant reads coverage is expected in the heterozygous mutations, also due to the additional wild-type sequences. As an example, the wild-type/mutant reads ratio was 2.99 for *PPD-B1* and 1.25 for *PPD-A1* (close to the 1.2 population average). Finally, the average mutation density in the duplicated regions is expected to be roughly twice as high as in non-duplicated regions since both copies are captured and both can be mutagenized. We found 189 mutations for the duplicated *PPD-B1* gene and 115 for the non-duplicated *PPD-A1* in Kronos, confirming the previous expectation.

SI Appendix, Method S9. Detection of large deletions

To identify and characterize homozygous deletions in our mutant populations, we developed a custom bioinformatics pipeline (available at https://github.com/homonecloco/bio.tilling) that examines relative coverage of exons within and across mutant lines (*SI Appendix,* Fig. S4). First, we calculated the raw coverage of each exon based on the IWGSC2 annotation. We used *bedtools* (16) to count the total number of reads that overlap a specified exon. These coverage values were then normalized in a two-step process, first accounting for the variation in coverage of each exon within an individual mutant and second to account for variation in each exon across the population. This two-step normalization allows direct comparison of coverages across individuals and exons.

First, this coverage was normalized by dividing it by exon length and total number of mapped Illumina reads per individual mutant line to account for differences in the size of exons and total number of mapped reads per line. The result was multiplied by 10^9 to avoid small decimal numbers. Exons with coverage values of 0 across all mutant lines were removed to avoid extreme values. The relative coverage of each exon *i* in mutant line *j* is given by the formula:

$$RelativeCoverage_{i,j} = \frac{ExonCoverage_{i,j} \times 10^9}{ExonLength_i \times TotalReadsSample_j}$$

where *ExonCoverage* is the total number of reads that overlap with a specified exon and $TotalReadsSample_j$ are the total number of mapped Illumina reads in mutant line *j*.

Second, the $RelativeCoverage_{i,j}$ values were normalized across the population for each individual exon. A normalized exon coverage matrix ($XNORM_{ij}$) was calculated by dividing the $RelativeCoverage_{i,j}$ for a particular exon and individual by the average coverage of that exon across the complete population:

$$XNorm_{i,j} = \frac{RelativeCoverage_{i,j}}{mean(\text{RELATIVECOVERAGE}_i)}$$

Given this two-step normalization process, a distribution of normalized exon coverages across mutant lines with mean 1 and standard deviation $sd(\text{XNORM}_i)$ was obtained for each exon. A similar distribution with mean 1 and standard deviation $sd(\text{XNORM}_j)$ was obtained for each mutant line in the Kronos and Cadenza populations. Exons and mutant lines that were too variable (e.g. normalized standard deviation ≥ 0.3) were removed from the analysis and the two-

step normalization process was repeated. The 1,535 Kronos and 1,200 Cadenza M_2 mutant lines were analyzed using the methods described above and a total of 1,494 Kronos and 1,011 Cadenza mutant lines passed an initial quality control filter in which the sample had *sd*(XNorm$_j$) < 0.3 (*SI Appendix*, Tables S18 and S19). In addition, a Kronos wild-type sample and 25 Cadenza wild-type samples were processed alongside the mutants as controls.

Based on the $XNorm_{i,j}$ value, each exon was classified into two exclusive categories: exons with coverage within 3 standard deviations of the normalized mean (*No3SigmaDel*) and exons with coverage below 3 standard deviations of the normalized mean (*3SigmaDel*). This allowed us to identify individual exons in a given mutant line with unusually low coverage as we expect over 99.7% of the coverage to be within ± 3 standard deviations. These categories were calculated as:

- *No3SigmaDel$_j$*: $XNorm_{i,j} > 1 - (3 \times sd(\text{XNORM}_i))$
- *3SigmaDel$_j$*: $XNorm_{i,j} \leq 1 - (3 \times sd(\text{XNORM}_i))$

Within the *3SigmaDel* category, the subset of exons with less than 10% of the XNORM$_i$ coverage were considered as homozygous deleted exons (*HomDel$_j$*). This classification was performed independently for all exons in the Kronos and Cadenza populations.

We hypothesized that large deletions should extend across multiple adjacent exons. Therefore, we examined CSS scaffolds (*sc*) to identify those in which multiple exons within the scaffold were classified as putative homozygous deletions. Each scaffold was scored based on the proportion of exons classified as *HomDel* compared to the total number of valid exons in the scaffold (exons with *sd*(XNORM$_i$) ≤ 0.3):

$$ScaffoldScore_{sc} = \frac{count(HomDel_{sc})}{count(No3SigmaDel_{sc} + 3SigmaDel_{sc})}$$

We selected scaffolds with at least 5 valid exons to ensure that we had at least 5 independent estimates of deleted exons across each scaffold. Those scaffolds with at least 5 valid exons in which more than 75% of the exons were classified as *HomDel$_j$* ($ScaffoldScore_{sc} > 0.75$) were considered homozygous deletions. Where possible, the genetic position of the IWGSC scaffold was determined using the POPSEQ genetic map (19).

SI Appendix, Method S10. Validation of large deletions

To validate the homozygous deletions detected in the M_2 mutants, we used KASP assays (*SI Appendix*, Fig. S9). We chose eleven deleted scaffolds across five mutants lines (three Kronos and two Cadenza) which represented seven independent deletion events based on the predicted chromosome position of the scaffolds (*SI Appendix*, Table S19). The deleted scaffolds had $ScaffoldScore_{sc}$ ranging from 0.78 to 1.00 (*SI Appendix*, Method S9) and all were unique to the specific mutant lines (deletions per scaffold equals 1). The deletion events included single (e.g. Kronos1017) and multi-scaffold deletions events (e.g. Kronos376; 64 scaffolds deleted on chromosome 3B). We validated the predicted homozygous deletions by assessing their status in the M_4 progeny.

We developed two types of KASP assays to perform this validation: flanking the deletion and within deleted scaffolds.

Flanking the deletion: We used EMS SNPs identified in scaffolds surrounding the deleted region based on the POPSEQ genetic map. These KASP assays were designed as described in Material and Methods and targeted specific EMS mutations present in the mutant lines. These assays were used to confirm that the scaffolds surrounding the deleted region were present in the M_4 progeny and that the mutations segregated as expected.

Within the deleted scaffolds: Homozygous deleted scaffolds are expected to be absent in all M_4 progeny. We first identified homoeologous variants within the target deleted region: for example, for a D-genome scaffold deletion in Cadenza we identified homoeologous variants between the D-genome and the A/B genomes. Using this information, one KASP primer was designed to incorporate the homoeologue-specific variant in the 3' end of the primer, the second specific primer was designed to amplify the other two alternative genomes and the third common primer is non-homoeolog specific. Following the example above, one KASP primer would target the D genome variant, the alternative KASP primer would amplify the A and B genomes and the third common primer would amplify all three. The expected result of this assay would be a "heterozygous" cluster for a wild-type plant since both the D and the A/B genome primers would amplify. In the case of a deletion which is missing the D-genome, the assay should be "homozygous" for the A/B primer since the D-genome specific KASP primer would fail to

amplify the deleted gDNA. A schematic of this is shown in *SI Appendix*, Fig. S9. An analogous strategy was used to validate the large deletions in tetraploid Kronos.

We screened 10-12 M_4 plants for each of the 11 target homozygous deletions using 4-5 KASP assays flanking the deletion and 1-3 assays within the deleted scaffolds. For all predicted homozygous mutations, we obtained results consistent with the presence of a homozygous deletion in the original M_2 plant. In all cases, the KASP assays on either side of the homozygous deletion yielded the expected result and segregation pattern, except in one case where a predicted heterozygous mutation was identified as homozygous in all lines. Likewise, for each of the independent homozygous deletions between one and four independent KASP assays yielded the "homozygous" clusters as detailed in *SI Appendix*, Fig. S9. The actual result for the deletion assay on IWGSC_CSS_1DL_scaff_2208937 is shown as an example (*SI Appendix*, Fig. S9, right panel).

SI Appendix, Method S11. Variant Effect Prediction (VEP)

The *T. aestivum* Variant Effect Predictor (VEP) cache file containing the annotation data for use with the VEP was downloaded from Ensembl (ftp://ftp.ensemblgenomes.org/pub/). Release 30 of the cache file was used to obtain all mutation effects and SIFT scores for the missense mutation (20). Mutation effects on gene function were predicted using the Variant Effect Predictor (VEP) program (21) from Ensembl tools release 78 in offline mode.

To estimate the number of genes disrupted by stop/splice or missense mutations, we extracted all effects (http://www.ensembl.org/info/genome/variation/predicted_data.html) predicted by VEP and counted them for each gene using a dedicated script (calcMutationEffectStatsFromVCF.py, https://github.com/DubcovskyLab/wheat_tilling_pub). The number of mutation effects was also calculated for each gene. If a mutation affected more than one gene due to overlapping gene models, we counted both effects. If the same mutation occurred in more than one mutant line, we counted it multiple times in *SI Appendix*, Table S20, which summarizes the effects predicted by VEP. These duplicated mutations were then reported as affecting a single gene model in *SI Appendix*, Table S21.

The quality of variant effect prediction depends on the quality of the predicted gene models. Since wheat genome annotation is still in its initial stages some of the mutations identified by VEP as 'intergenic' might be in genes that are incompletely or not annotated yet. Since the wheat genome annotation is still in its initial stages, we advise users to manually examine the gene models available for their target genes. In the absence of appropriate gene models, mutations can be manually annotated as we did for the genes in *SI Appendix*, Tables S22-23 and outlined in http://www.wheat-training.com/tilling-mutant-resources/.

SI Appendix, Method S12. Annotation of starch biosynthesis and flowering genes

The rice starch biosynthetic genes (22) were converted from MSU rice gene nomenclature to RAP nomenclature using RAP-DB ID Converter (http://rapdb.dna.affrc.go.jp/tools/converter). The wheat orthologues of the RAP nomenclature rice genes were identified using gene trees available at *EnsemblPlants* based on the IWGSC gene models. The orthologous relationship between wheat and rice genes was confirmed using reciprocal BLAST on *EnsemblPlants*, and by checking that the percentage identity between the rice and wheat genes was > 75 %. In cases where gene duplication occurred between wheat and rice, all wheat orthologues were retained. Gene names within multigene families were assigned by comparison to cDNAs available at NCBI and literature search (references in *SI Appendix*, Table S22 footnote).

When *EnsemblPlants* gene models were absent or incomplete for the genes in the flowering pathway, we obtained gene structures from previous publications and indicated the corresponding GenBank numbers in *SI Appendix*, Table S23. For incomplete gene models in both pathways, a BLAST search was used to recover missing exons from the CSS scaffolds and the mutation effects were manually annotated. For those mutations we re-examined their effects using ParseSNP (http://blocks.fhcrc.org/~proweb/input/) which reproduces the VEP and SIFT outputs.

SI Appendix, Text

SI Appendix, Text S1. Validation of uniquely mapped SNPs

Kronos: A total of 80 mutations were assayed across 8 independent M_4 mutant families (10 EMS mutations per family). Sixteen M_4 plants were tested per family in addition to six wild-type Kronos DNA samples and two no-template controls. Of the 80 designed KASP assays, 71 (88.8%) produced valid clusters that could be classified. Of these, we confirmed the expected mutation in 70 (98.6%), whereas only a single mutation in Kronos4346 could not be confirmed with the KASP assay (1.4%, *SI Appendix*, Table S8). This last mutation was heterozygous in the M_2 and may have been lost by genetic drift during seeds increases to M_4. The other mutations in Kronos4346 confirmed that the M_4 seed was correct.

All confirmed mutations, except one, were correctly classified as heterozygous or homozygous. The only exception was one SNP classified as heterozygous based on M_2 sequencing data (Kronos3288: 8 wild-type reads and 15 mutant reads), but found to be homozygous in the tested M_4 seeds. A possible explanation for this difference is fixation of the mutation by genetic drift. The correctly predicted homozygous lines included three mutations that were corrected by the bioinformatics filter applied after MAPS (*SI Appendix*, Table S8). This filter converts heterozygous to homozygous mutations when the frequency of the minor allele is less than 15% of reads (*SI Appendix*, Fig**.** S5).

In addition, we validated 62 mutations (from 59 independent M_4 Kronos families) by direct sequencing of genes currently being studied in our laboratories (*SI Appendix*, Table S9). We confirmed the presence of the mutation in 61 of the 62 amplicons (98.4%), with a single mutation in Kronos910 that could not be confirmed. To test if this was due to a planting error, we re-sequenced specific mutations from each of the 24 lines that were sown in the same row as Kronos910 in the field. We discovered a planting shift that affected six lines including Kronos910. After the IDs of these six lines were corrected we were able to validate the Kronos910 mutation. We also confirmed the segregation for 60 mutations based on the prediction by the MAPS pipeline and the heterozygous-to-homozygous correction (*SI Appendix*, Table S9). A single mutation identified as heterozygous in the original M_2 DNA was found to be homozygous in the tested M_4 seeds (Kronos3634: 8 wild-type reads and 7 mutant reads). This

line seemed like a true M_2 heterozygote based on coverage and, as indicated above for Kronos4346, the difference may be the result of genetic drift. In summary, using both the KASP assays and direct sequencing, and accounting for the Kronos910 planting error, 132 out of 133 mutations were confirmed (99.25%), of which 130 (98.48%) segregated as predicted by the MAPS pipeline and the heterozygous-to-homozygous correction in the M_4 families. The 0.75% error found for the Kronos population is not very different from the 0.2% estimated error (Table 1). None of the methods used to estimate error account for the loss of heterozygous mutations by genetic drift or outcrossing with other mutants during the two generations between the sequenced M_2 data and the M_4 seeds used for distribution. Genetic drift and outcrossing are expected to be higher in lines with pollen-sterility problems where M_4 seeds were obtained from few plants and the probability of outcrossing is higher.

Cadenza: A total of 172 mutations were assayed across 19 independent M_4 mutant families (between 8 and 10 EMS mutations per family). Twelve M_4 plants were tested per family in addition to four wild-type Cadenza DNA samples, seven random M_4 mutant lines and one no-template control. In total, 147 out of 172 designed KASP assays (85.5%) produced valid clusters. Of these, we confirmed 146 expected mutations (99.32%), with a single false positive in Cadenza0548 being identified (0.68%, *SI Appendix*, Table S10). This mutation was heterozygous in the M_2 and could have been lost due to genetic drift.

Among the 146 confirmed mutations 139 segregated as expected (95.21%), including 7 mutations that were corrected by the heterozygous-to-homozygous filter. Two homozygous mutations originally classified as heterozygous may be explained by genetic drift (more frequent in lines with few available M_2 seeds). In addition, five mutations (four in Cadenza1538 and one in Cadenza1551) were scored as heterozygous despite being identified as homozygous in the M_2 line (*SI Appendix*, Table S10). In Cadenza1551, the only exception, was one mutation originally identified as heterozygous in M_2 but corrected by our pipeline to homozygous. Since the other four homozygous M_2 mutations in Cadenza1551 were validated as homozygous, the single exception is likely an over-correction. However, in Cadenza1538, four out of the five homozygous M_2 mutations were heterozygous in the M_4 validation. Outcrossing with surrounding mutants provides a simple explanation for these exceptions. If this explanation is correct, it will indicate a rate of outcrossing of 1 in 19 individuals (5.3%, *SI Appendix*, Table

S10), which is within the range previously reported for outcrossing in common wheat (23). Outcrossing can also explain some of the lost mutations.

SI Appendix, Text S2. Characterization of Large Deletions

In both populations, the majority of the lines did not have scaffolds with evidence of homozygous deletions ($ScaffoldScore_{sc} > 0.75$), and none were detected in the wild-type samples. In tetraploid Kronos, 115 lines (7.7%) had at least one scaffold with five or more exons that was classified as a homozygous deletion; 27 of these consisted of lines with a single homozygous deletion, and the majority of lines (87; 75.7%) had 10 or fewer scaffolds deleted (*SI Appendix*, Table S18, Fig. S10A, red line). In hexaploid Cadenza, 293 lines (29%) had at least one homozygous deletion, with 165 (56.3%) of them having 10 of fewer scaffolds deleted (*SI Appendix*, Table S18, Fig. S10B, red line). For those lines carrying at least one deletion in Kronos and Cadenza, the median number of deleted scaffolds was 4 and 8 scaffolds, respectively.

In both populations, the majority of the deletions within an individual were restricted to a single chromosome arm. For example, among the 88 Kronos mutants with at least 2 scaffolds deleted, 79 (90%) had deletions restricted to a single chromosome arm. The physically defined nature of the mutations was further supported by the POPSEQ genetic positions of the deleted scaffolds, which in the majority of cases mapped to the same or adjacent genetic bins. Similar to Kronos, the majority of the Cadenza mutant lines with two or more scaffolds deleted (237 lines) were restricted to either a single chromosome arm (162 lines; 68%) or two chromosome arms (53 lines; 22%). The co-localization of homozygous deletions based on the chromosome arm assignment and the POPSEQ genetic position, and the fact that the scaffolds were assessed independently for their deletion status, suggested that the bioinformatics pipeline was effective at identifying deletions with a low false-positive rate. This was further confirmed by the validation of 11 homozygous deletions across the Kronos and Cadenza populations (*SI Appendix*, Method S10; Table S19).

In Cadenza, seventeen lines had over 100 scaffolds deleted (*SI Appendix*, Fig. S10B). Among them, one was homozygous for a complete chromosome deletion, eleven for complete arm deletions, and two for deletions including most of the sequences from a chromosome arm. The

observed frequency of nullisomics (0.08%) is four-fold higher than the predicted frequency of nullisomics in non-mutagenized wheat populations (0.02%). This last value was estimated by multiplying the frequency of monosomics in stable non-mutagenized wheat varieties (0.69%) by the frequency of nullisomics (3%) in the progeny of wheat monosomic plants (24). The frequency of complete arm deletions (0.92%) is also higher than expected from the misdivision of monosomics in non-mutagenized populations (0.07%). The later value was estimated by multiplying the frequency of monosomics in non-mutagenized populations (0.69%) (24) by an estimate of the maximum average frequency of telocentrics (10%) in the progeny of wheat monosomics (25). Taken together, these observations suggest that the EMS treatment increased the frequency of aneuploids and large deletions in the M_2 plants.

We also assessed the frequency at which specific scaffolds were deleted across each population (*SI Appendix,* Table S18). Overall, 5% of the scaffolds with 5 or more exons had evidence of at least one homozygous deletion in Kronos (785/15,629 scaffolds) whereas a larger proportion (28.3%) of scaffolds were deleted in at least one mutant individual in Cadenza (5,433/19,191 scaffolds). Most scaffolds were deleted in a single mutant or were shared between two mutant lines (97% Kronos, 94% Cadenza).

Scaffolds that contain homozygous deletions are of interest because they are likely to lead to complete loss of gene function. We therefore examined the number of gene transcripts that were affected in the 785 and 5,433 unique scaffolds deleted in the Kronos and Cadenza populations based on *Ensembl* release 30. In Kronos, we identified 832 (1.7%) gene models that were deleted in at least one line, whereas in Cadenza we identified 6,657 (9.0%) gene models deleted (including those within complete chromosome and chromosome arm deletions). A total of 348 gene models were deleted in both populations. The low frequency of deleted gene models suggests that these populations will not be adequate to identify deletions including tightly linked duplicated genes, which are difficult to tackle by point mutations. Dedicated wheat radiation mutant populations are likely a better option for this objective.

.

SI Appendix, Text S3. Variant Effect Prediction

Using currently available gene models, we were able to assign effects to >50% of all mutations corresponding to 48,172 genes in tetraploid Kronos and 73,895 in hexaploid Cadenza (*SI*

Appendix, Tables S20 and S21). We also summarized the total number of genes that possessed at least one mutation that resulted in a truncation (gain of a premature stop codon or a change to the splice donor or acceptor sites) or a missense mutation (*SI Appendix*, Table S21). In these summary calculations, we did not include 'upstream_gene_variant' and 'downstream_gene_variant' effects as these can belong to other unannotated genes. In total, 96% of tetraploid and 94% of hexaploid genes that had at least one mutation included a missense allele. On average, we annotated 1.58 and 1.81 truncations (stop codons or mutations in splice sites) per gene model and 21 and 23 missense mutations per gene model in Kronos and Cadenza, respectively (*SI Appendix*, Table S21). Detailed information about variant effect predictions for each gene is available in the project websites http://www.wheat-tilling.com and http://dubcovskylab.ucdavis.edu/wheat-tilling under file names GeneAnnotationTableSummary_Cadenza_main_set for Cadenza and GeneAnnotationTableSummary_Kronos_main_set for Kronos.

We also implemented the sorting intolerant from tolerant (SIFT) analysis within VEP to predict the effect of missense mutations on protein function for the 48,172 Kronos and 73,895 Cadenza genes. We identified missense alleles predicted to be deleterious (SIFT score < 0.05) for 40,913 (85%) Kronos and 66,734 (90%) Cadenza genes (*SI Appendix*, Table S21). Combining the SIFT and truncation analyses, these results revealed that a total of 43,787 (91%) Kronos and 67,830 (92%) Cadenza genes had at least one mutation leading to a truncation and/or a deleterious allele as predicted by SIFT results (< 0.05).

SI Appendix, Text S4. EMS mutant database and JBrowse graphic interface

Wheat EMS mutant database. The SQL schema of this database includes over thirty tables joined by unique IDs to query distinct parts of the mutation results, which are described in detail in (https://github.com/homonecloco/bioruby-wheat-db). These tables include assemblies, biotypes, chromosomes, mutations, effects, genes, markers, mutations, primers, multi-map mutations, scaffolds, species, SNPs, and others to organize the data. When querying only specific datatypes, the schema design uses separate tables to enhance performance. The current schema also allows for flexible storage of multiple line types. The combination of these tables powers the

BLAST results table, line search page, and downloadable flat-file generation. The EMS mutant database at www.wheat-tilling.com allows users to query the database in three different ways:

1. IWGSC scaffold name: this refers to the IWGSC gDNA scaffold to which a mutation is mapped, e.g. IWGSC_CSS_1BS_scaff_3451992. The scaffold names are in the same format as the IWGSC scaffolds on *EnsemblPlants*. This is different from the name of an IWGSC scaffold on the URGI BLAST server, which consists of a longer identifier e.g. IWGSC_chr1BS_ab_k71_contigs_longerthan_200_3451992. Note that both name formats have the same numerical identifier at the end of the name (here 3451992).
2. IWGSC gene model: this refers to the IWGSC gene model to which a mutation is mapped, e.g. Traes_7BS_C9F4BC10E. This nomenclature is consistent with that of *EnsemblPlants*. The database also allows searches with the transcript name (e.g. Traes_7BS_C9F4BC10E.1)
3. Mutant line identifier: Each mutant in the Kronos and Cadenza population has a unique 4-digit identifier following the genotype name. Hence mutant number 3091 of the Kronos population is called Kronos3091 and mutant number 624 in the Cadenza population is called Cadenza0624. This feature allows users to query for all the mutations in a given mutant line.

Other features of the database include:

- The database can be searched simultaneously for mutations in the Kronos and Cadenza populations, or a single database can be chosen and searched independently.
- The search results can be reported in HTML format or can be downloaded as an Excel file. Results contain hyperlinks to the specified IWGSC scaffold sequence in FASTA format and to the gene page of *EnsemblPlants*.
- Multiple scaffold names, gene names, or line identifiers can be queried.
- The BLAST search includes only those scaffolds and genes for which mutations were identified. This search also incorporates the *de novo* assemblies of the Kronos and Cadenza reads that did not map to the CSS reference sequence (*SI Appendix*, Method S3).

The results from the database query are formatted with several headers that provide information regarding the IWGSC scaffold, mutant line, mutation position, zygosity, predicted effect on protein sequence, SIFT score, and KASP primer for SNP validation. The UK website used

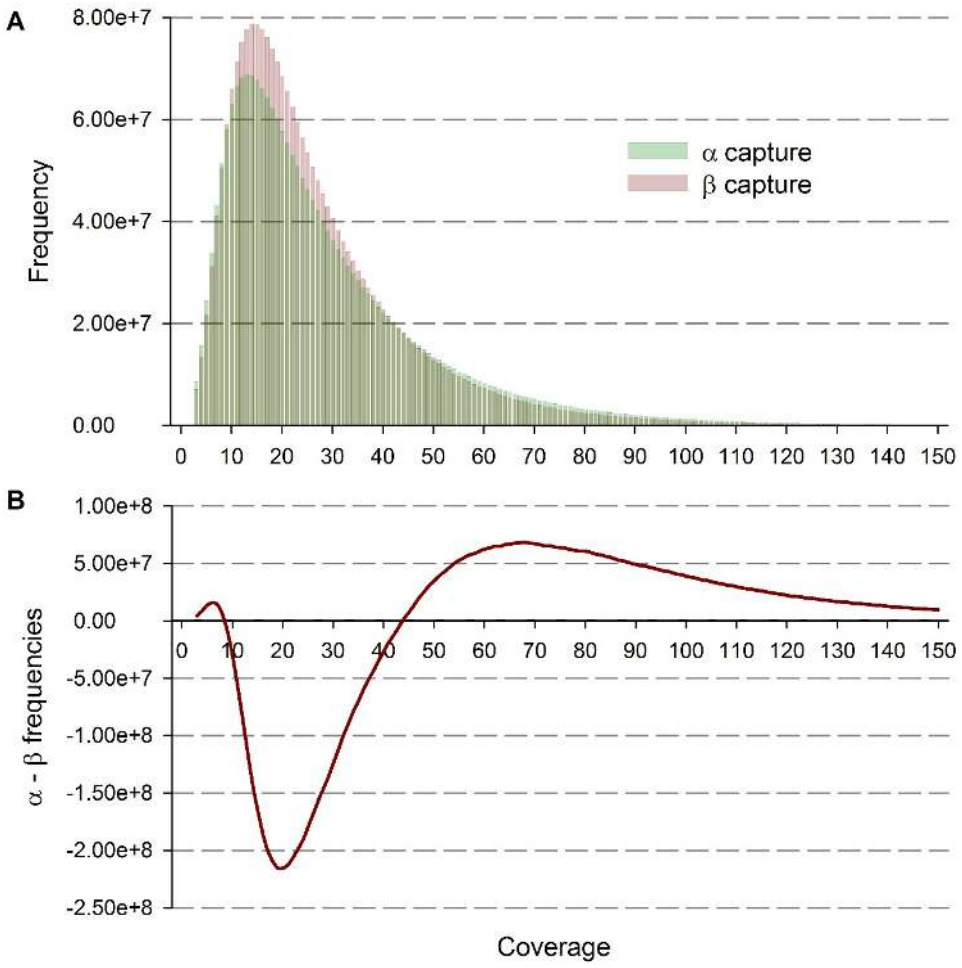
SequenceServer (26) and BioRuby (27, 28) to power the BLAST search and data processing. A detailed explanation of each header is outlined in the www.wheat-tilling.com and http://dubcovskylab.ucdavis.edu/wheat-tilling websites.

JBrowse graphic interface. For our online web BLAST, we used the Viroblast package (https://els.comotion.uw.edu/express_license_technologies/viroblast) and for visualization, we used JBrowse (http://jbrowse.org/). BLAST results from the project page (http://dubcovskylab.ucdavis.edu/wheat-tilling) are presented in a table where users are given their top BLAST hits, the number of mutations and confidence interval for the mutations on the scaffold hit, and a visualization link to the JBrowse installation. This required editing the Viroblast source code to enable these functionalities and implementing relational database queries to our PostGreSQL 9.3 backend. The SQL schema of this database is identical to the one for the Cadenza population and powers the BLAST results table, line search page, and downloadable flat-file generation.

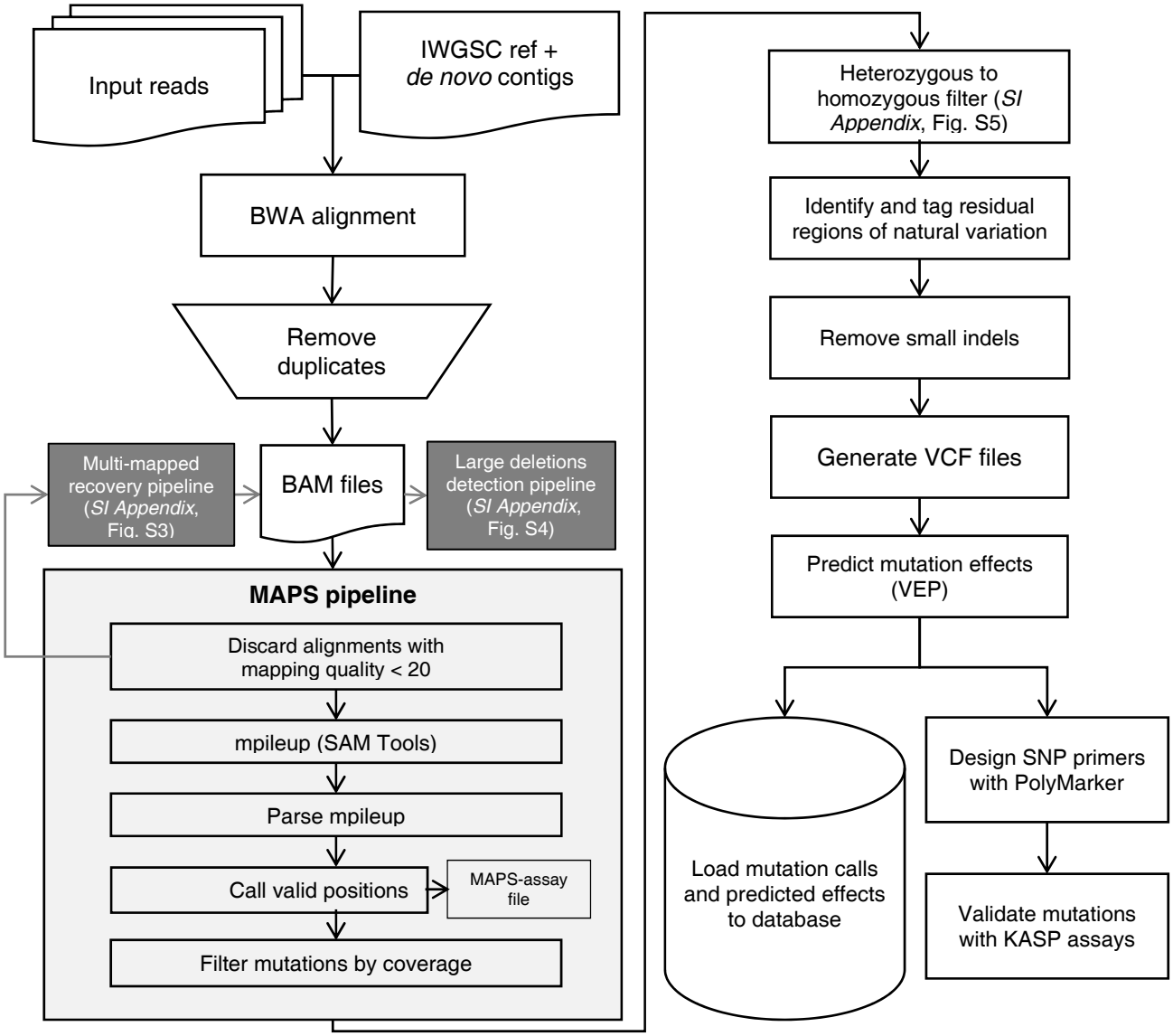
JBrowse (29) implements javascript and HTML5 to rapidly display genomics data on a web browser. Standard setup scripts were run on the reference files to create JBrowse data sets for each wheat chromosomal arm. The mutation information in VCF format was used to visualize the mutant data. Additional 'INFO' fields were used in VCF format to display 'VEP SNP Effect', 'hethom ratio', and 'Seed Stock Availability' when a user clicks on a particular mutation. Furthermore, JBrowse functionality was extended by adding javascript code to color code mutation effects based on severity (red = truncations, violet = missense, green = synonymous, and blue = non-coding regions). The additional code also allows additional options when right-clicking on a mutation in the browser. These additional right-click options allow users to download the mutation data in TSV format or to go directly to the Seed Order form to request a seed. JBrowse javascript changes can be viewed in the project github page (https://github.com/DubcovskyLab/wheat_tilling_pub) under the jbrowse_config directory. A detailed explanation of the different tracks and options is provided in the http://dubcovskylab.ucdavis.edu/wheat-tilling website.

SI Appendix, Figures

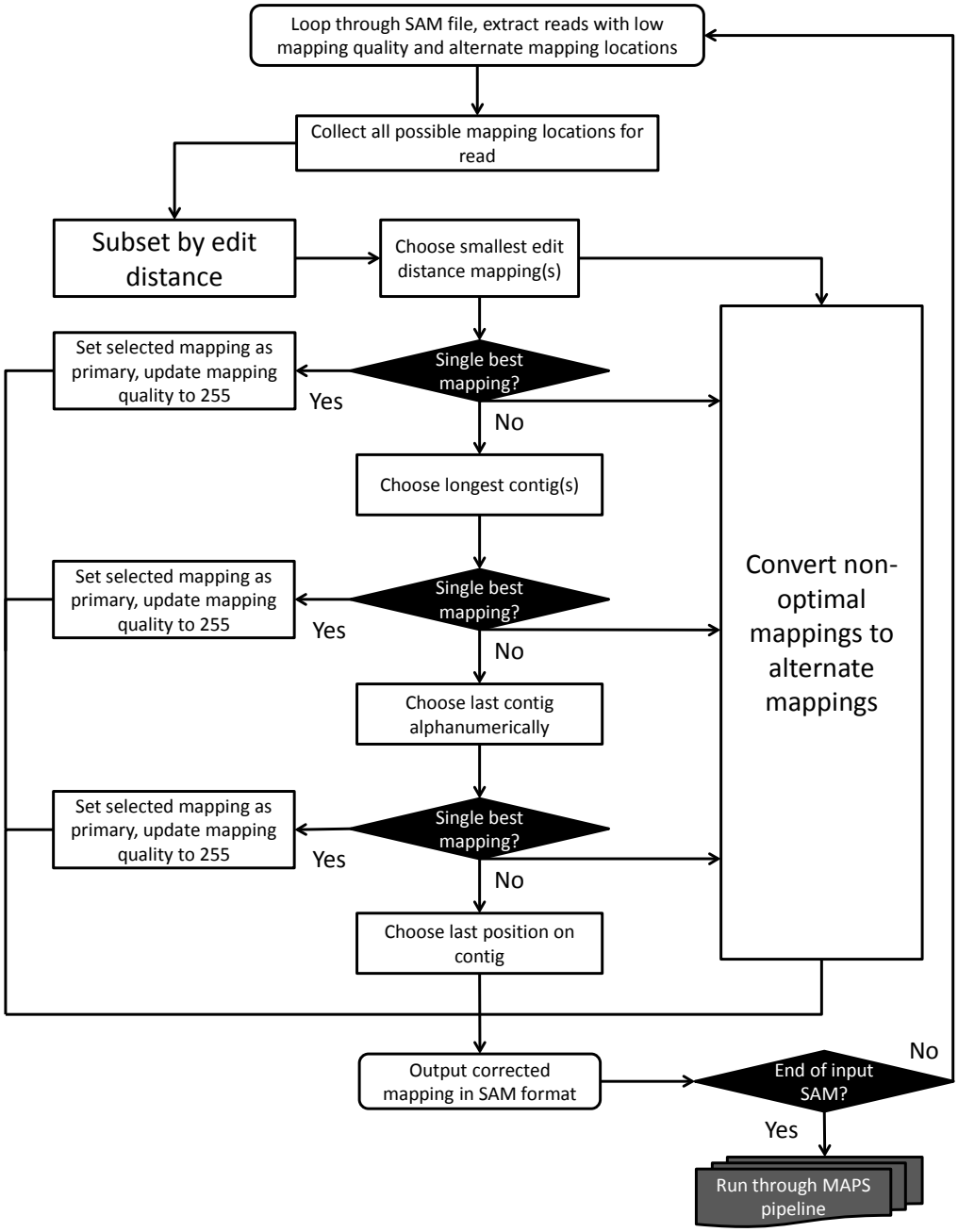
***SI Appendix*, Figure S1**. **Comparison between wheat alpha and beta exome capture assays**. (**A**) Read coverage distribution from 24 tetraploid lines captured with the α-design and 24 captured with the β-design (see *SI Appendix*, Method S1). From the BAM files, we identified a total of 89.2 million positions per individual with coverage ≥ 3 in all 48 lines from both designs, and adjusted the distributions to an identical total number of reads. Note the higher frequency of positions with very high coverage in the α-design. (**B**) Differences between the frequencies in the α- and the β-design. Positive numbers indicate higher values in the α-design and negative numbers indicate higher values in the β-design. The β-design showed relatively higher frequencies in the central coverages (9 x to 43 x), which resulted in a smaller standard deviation (3.8) than in the α-design (4.0). Based on the more homogeneous coverage of the β-design capture we used this design for the rest of the project.



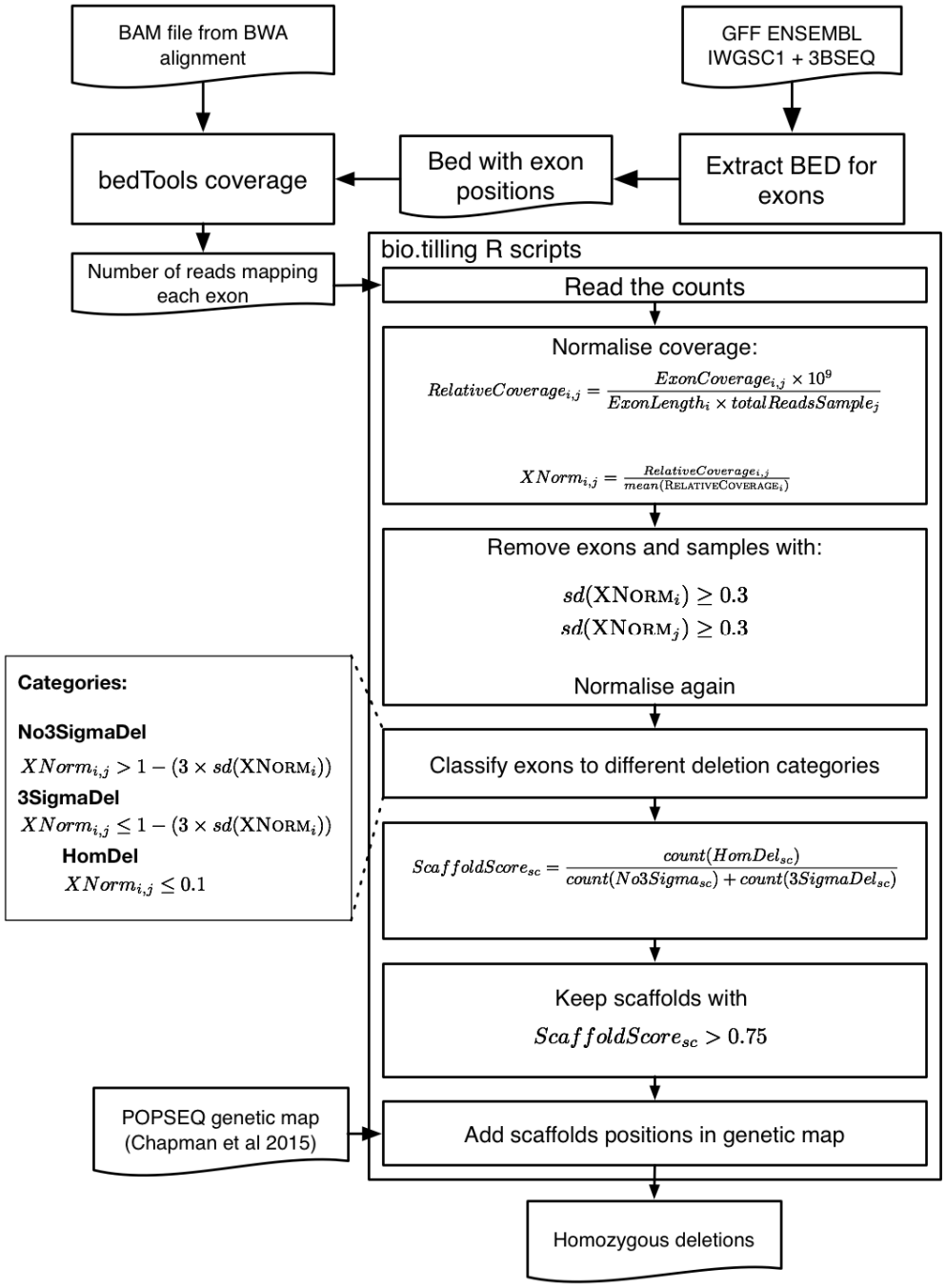
***SI Appendix*, Figure S2. Mutation calling pipeline.** Illumina 100-bp paired-end reads were aligned with *bwa* to the Chinese Spring wheat genome reference supplemented with *de novo* assembled contigs (*SI Appendix*, Method S3). Duplicate reads were removed using Picard tools. BAM files were generated and used for the MAPS pipeline in batches of 24-32 lines (18) to identify mutations. Heterozygous mutations with a low coverage of the wild-type allele (<15%) were converted to homozygous (*SI Appendix*, Fig. S5) and VCF files were generated. The effects of the identified mutations were predicted by *Ensembl* Variant Effect Prediction (VEP) (21), and were loaded into a PostGreSQL relational database. A subset of the mutations was selected for experimental validation. The "Multi-mapped reads recovery pipeline" and the "Large deletions detection pipeline" are described in *SI Appendix*, Figs. S3 and S4, respectively.



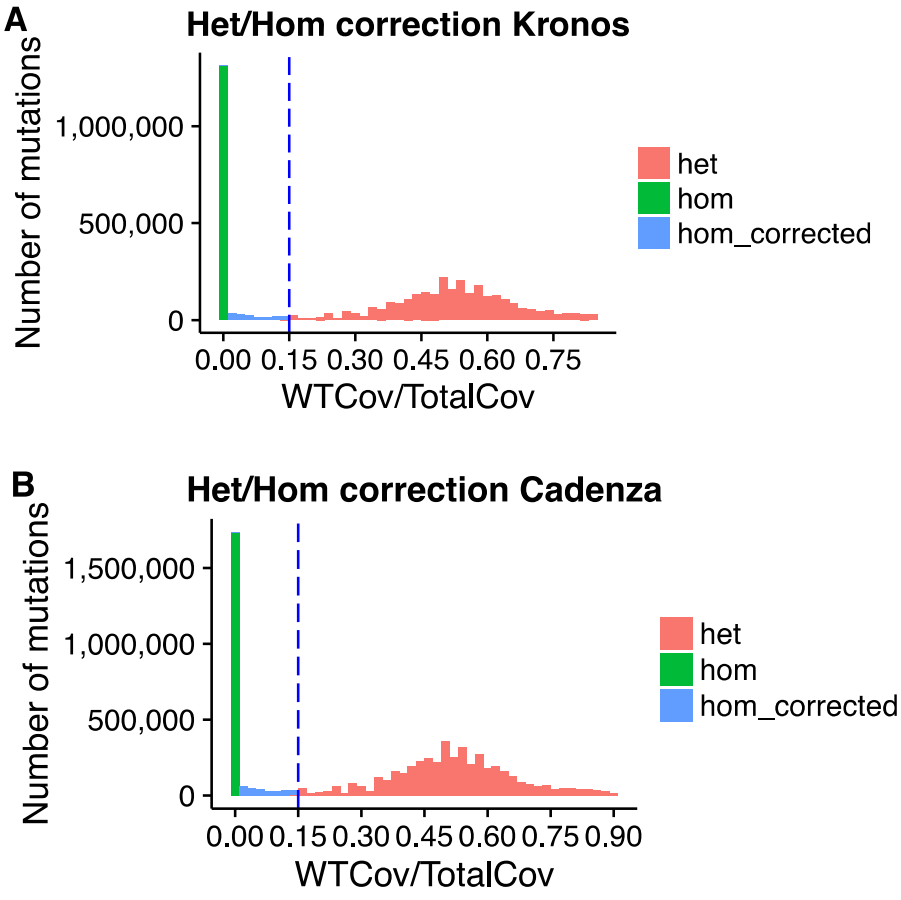
***SI Appendix*, Figure S3. Multi-mapped reads recovery pipeline:** Reads that map to multiple locations are assigned low mapping quality values and are excluded by MAPS. To recover these mutations, we generated the pipeline described below.



***SI Appendix*, Figure S4. Large deletions detection pipeline**. We determined homozygous deletions in contigs with at least five exons (*SI Appendix*, Method S9). Briefly, the pipeline calculates the relative coverage of individual exons within a mutant line and then normalizes the coverage values across all mutants in each mutant population independently. Scaffolds in which more than 75% of the exons had less than 3 standard deviations and less than 10% of the normalized mean coverage were considered homozygous deletions ($ScaffoldScore_{sc} > 0.75$).



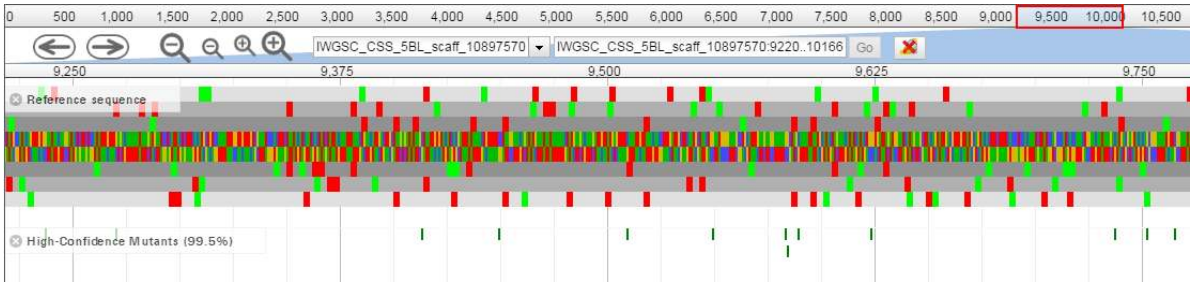
***SI Appendix*, Figure S5. Heterozygous to homozygous filter.** The MAPS pipeline classifies a mutation as homozygous only if all the reads are mutant (green bar). Therefore, incorrectly mapped homoeologous reads (or errors) result in the misclassification of homozygous mutations as heterozygous (blue bars). Plotting the fraction of wild-type reads (WTCov) out of total reads (TotalCov) on the X-axis and their frequency on the Y-axis shows two distributions for sites classified as heterozygous by MAPS. The main distribution fits the expected Normal distribution centred at 50% WTCov/TotalCov reads (pink). However, more mutations than those expected based on this distribution are observed close to the WTCov/TotalCov = 0% value corresponding to the homozygous class. Based on this distribution we selected a threshold of 15% (blue dashed line), and converted heterozygous classifications into homozygous classifications when the proportion of wild-type reads was less than 15%. After this correction, 150,208 Kronos and 285,653 Cadenza uniquely mapped mutations were reclassified from heterozygous to homozygous. After this correction, the ratio of heterozygous to homozygous mutations dropped from 2.20 to 1.87 in Kronos, and from 2.74 to 2.21 in Cadenza (at *HetMC5/HomMC3*).



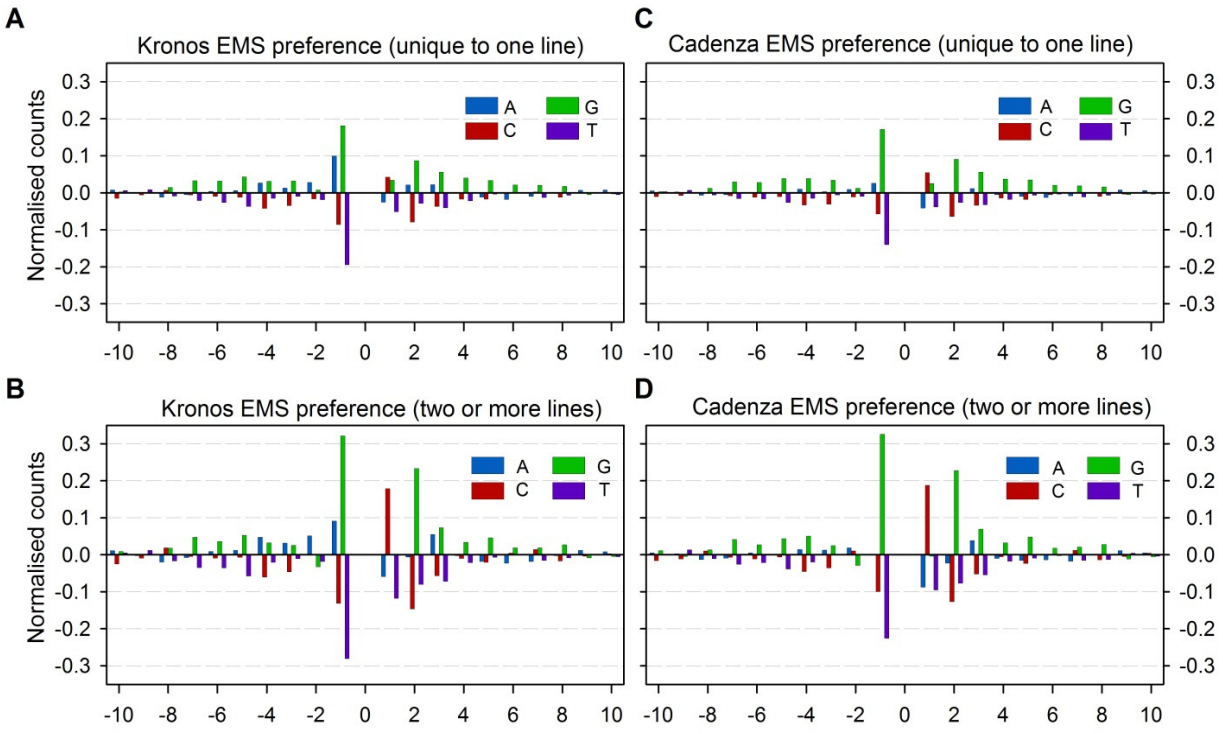
***SI Appendix*, Figure S6. JBrowse view of the distribution of mutations in residual heterogeneity (RH) and non RH regions in tetraploid wheat.** Small green lines represent mutations in a single individual. (**A**) RH-region: Red arrows indicate the presence of RH mutations which are visible as mutations in a common position across multiple individuals. Blue arrows point to some examples of putative induced mutations in the RH region since they are only present in a single individual. (**B**) Non-RH region. Note the absence of mutations mapped in multiple individuals.

A

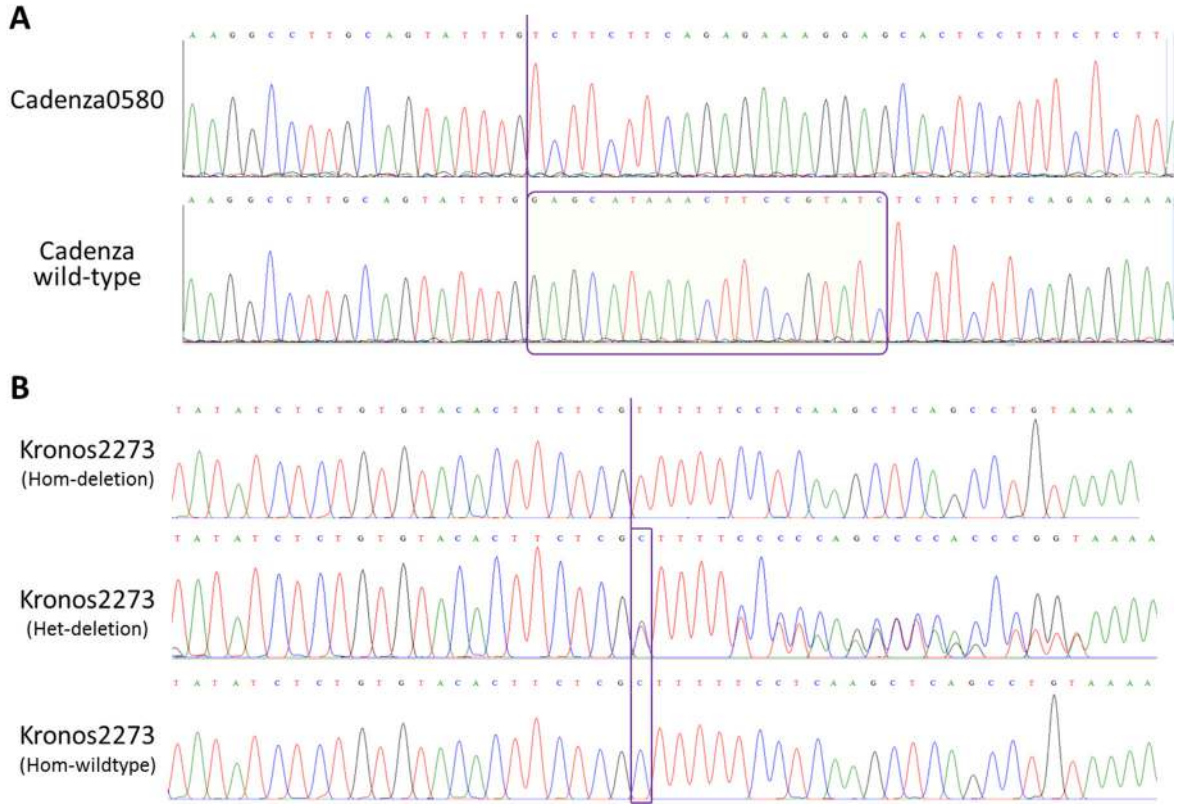
B



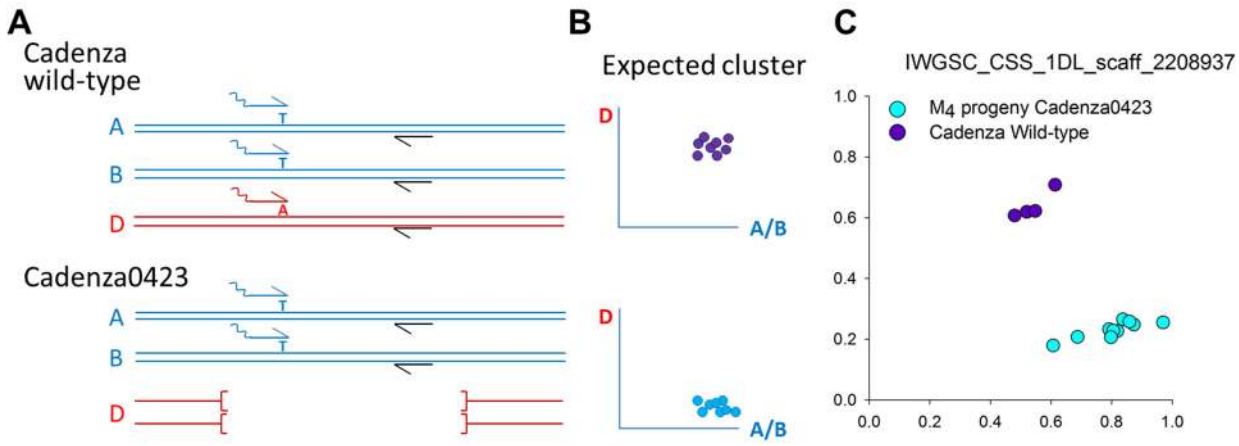
***SI Appendix*, Figure S7. Observed EMS preference normalized against randomly chosen surrounding sequences.** We examined sequences surrounding a mutated G at all mutated sites in Kronos (**A, B**) and Cadenza (**C, D**). The X-axis indicates the position of the base relative to the mutated G. The Y-axis shows the counts of nucleotide frequencies normalized against randomly chosen sites upstream or downstream of mutations (18). Sequence preferences are described in detail in *SI Appendix*, Method S7. Note that preference effects were stronger in mutations observed in more than one individual (**B, D**) than in those observed in only one individual (**A, C**).



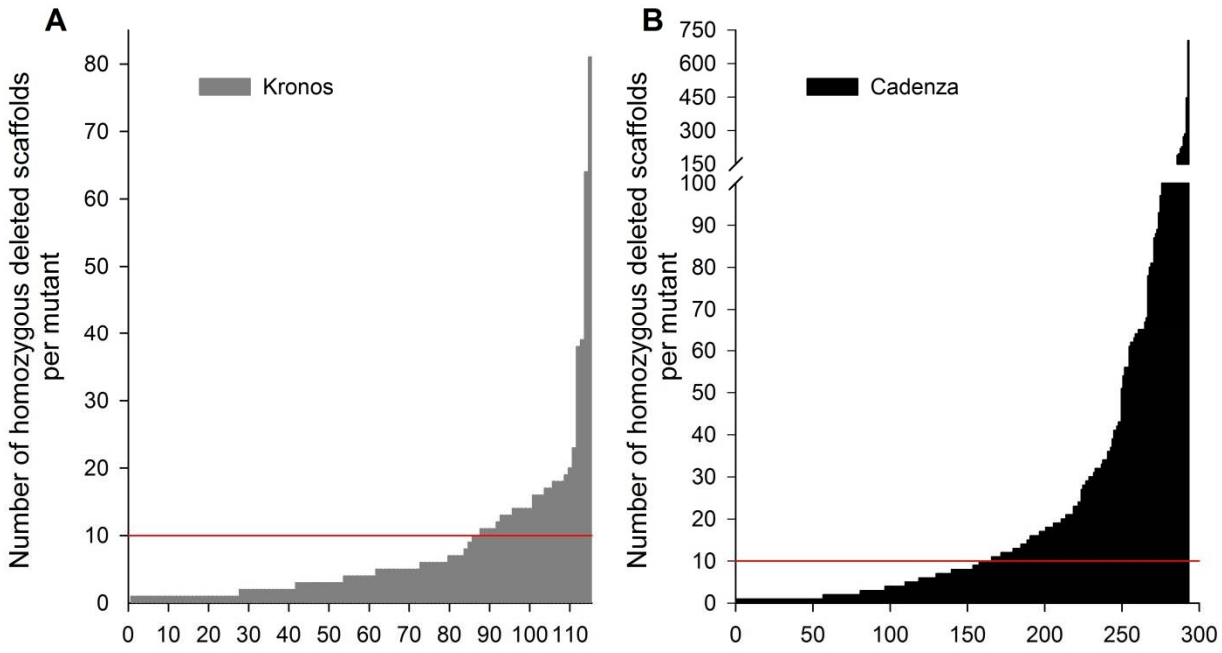
***SI Appendix*, Figure S8. Chromatograms of the validation of small deletions in the Kronos and Cadenza mutant populations.** (**A**) Cadenza0580 carries a homozygous 19-bp deletion in IWGSC_CSS_4AL_scaff_7167665 compared to wild-type Cadenza (purple-frame box). (**B**) Three chromatogram traces of Kronos2273 M_4 siblings segregating for the presence of a homozygous 1-bp deletion in IWGSC_CSS_1AL_scaff_3977540. The top panel shows a M_4 homozygous deletion mutant, the middle panel a heterozygous individual with a mixed trace from the 1-bp deletion onwards, and the bottom trace a homozygous wild-type M_4 sequence with the expected cytosine residue.



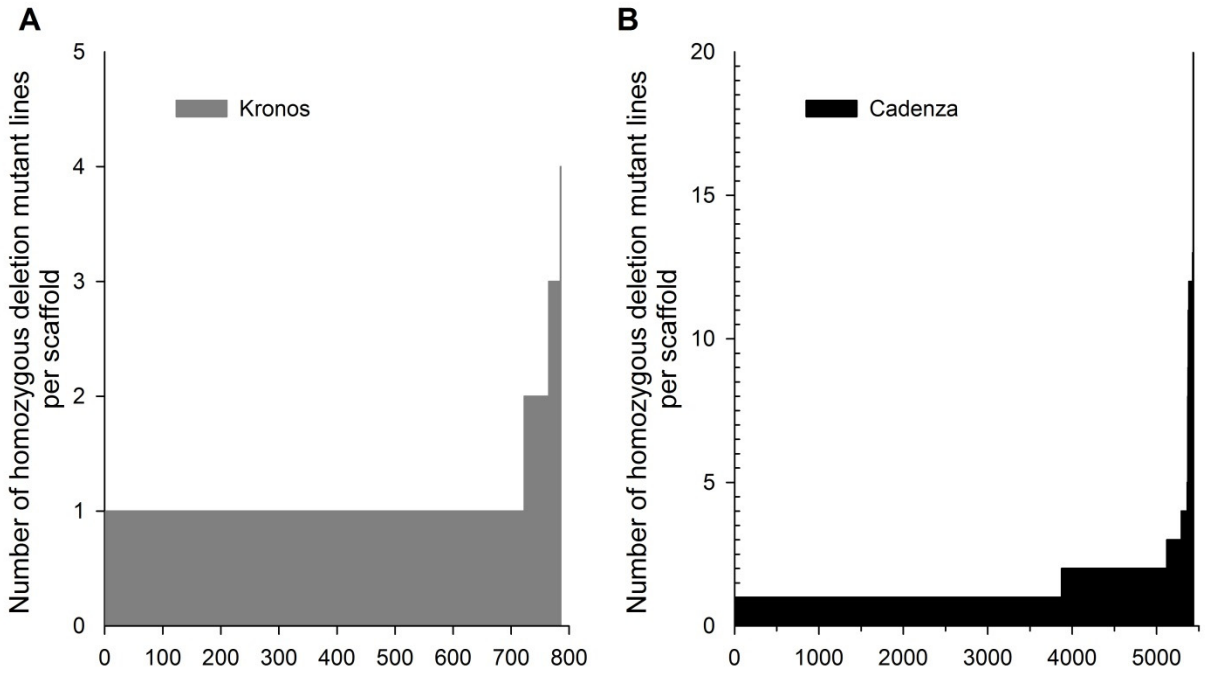
***SI Appendix*, Figure S9. Strategy to validate homozygous large deletions in EMS mutants.** (**A**) Example of a D-genome deletion assessed in Cadenza0423 M_4 plants. The KASP assay is designed to amplify the D-genome variant (red, A variant, red arrow) and the A/B genome variant (blue, T variant, blue arrow). The assay includes a common reverse primer (black arrow). (**B**) Wild-type Cadenza is expected to produce a "heterozygous" cluster (purple, top) which incorporates both the D (red) and the A/B genome primers (blue). The homozygous deletion mutants lack the D-genome (missing region in square brackets). Therefore, the assay should only incorporate the A/B primer leading to a "homozygous" blue cluster (bottom). (**C**) Actual results of the KASP assay for the deletion assay of IWGSC_CSS_1DL_scaff_2208937 on four wild-type Cadenza lines (purple) and 10 M_4 progeny plants of Cadenza0423 (*SI Appendix*, Table S19).



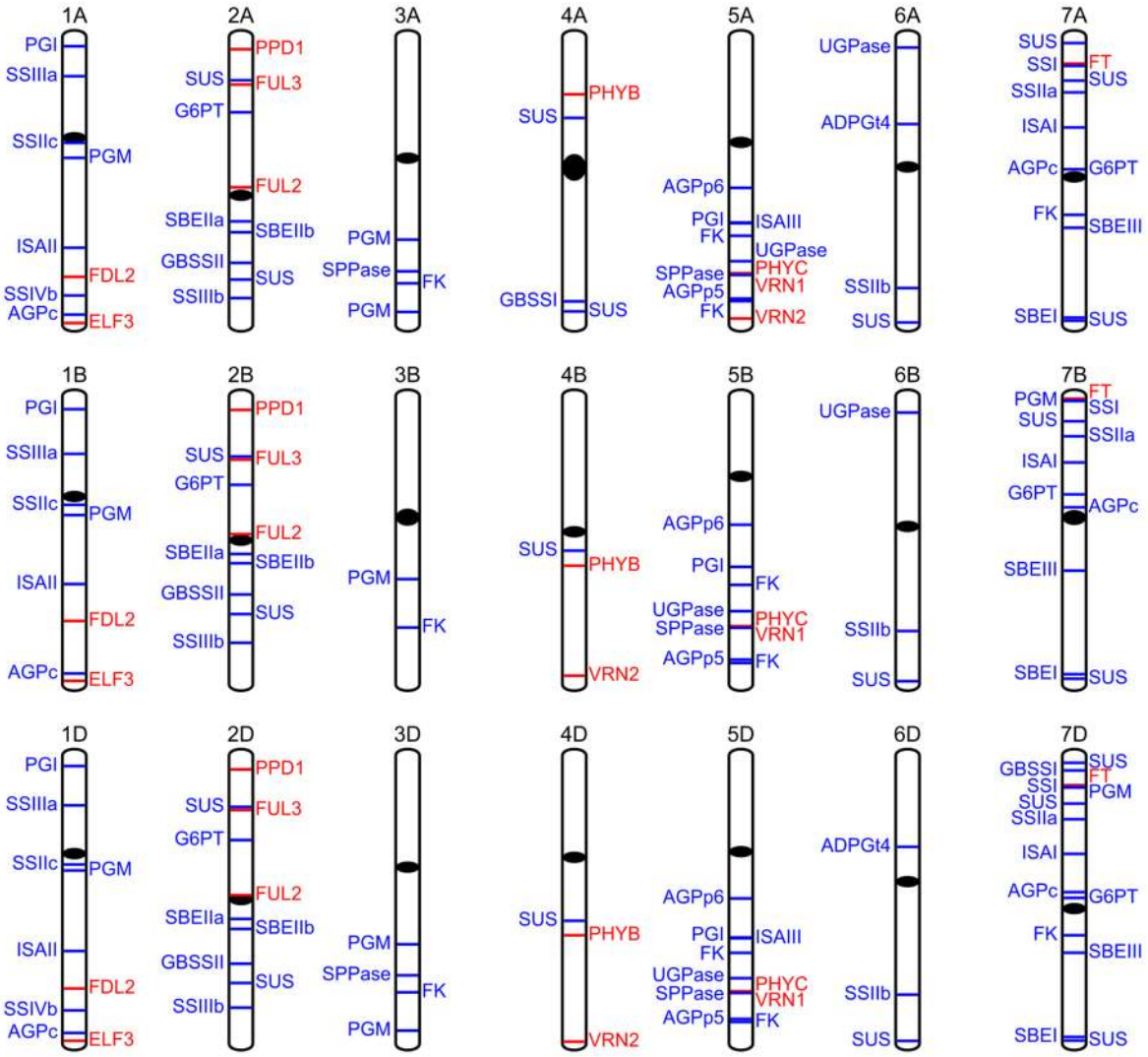
***SI Appendix*, Figure S10**. **Number of homozygous deleted scaffolds per mutant.** (**A**) Kronos and (**B**) Cadenza mutant populations. Mutant lines with at least one homozygous deletion are ordered on the X-axis based on the number of homozygous deleted scaffolds within each line and are assigned numbers 1 to 115 for Kronos and 1 to 293 for Cadenza). The red line indicates lines with between 1 and 10 homozygous deleted scaffolds. In Cadenza, the Y-axis includes a break to better represent the majority of the mutant lines; seventeen Cadenza lines have over 100 homozygous deleted scaffolds. The 1,379 Kronos and 718 Cadenza mutant lines which do not have predicted homozygous deletions are not included in this figure.



***SI Appendix*, Figure S11. Number of mutant lines which carry a homozygous deletion for a given scaffold.** (**A**) Kronos and (**B**) Cadenza mutant populations. Scaffolds deleted in at least one mutant individual are ordered on the X-axis based on the number of occurrences within the populations and are assigned numbers 1 to 785 for Kronos and 1 to 5,433 for Cadenza. The scaffolds which are not deleted in the Kronos (14,844) and Cadenza (13,758) populations are not shown.



***SI Appendix*, Figure S12**. Chromosome location of genes from Tables S22 and S23 based on their position on the IWGSC WGA v0.4 assembly.



Abbreviations: ADPGt4 (ADP-glucose transporter 4), AGPc (ADP-glucose pyrophosphorylase (cytosol)), AGPp5 (ADP-glucose pyrophosphorylase 5 (plastid), AGPp6 (ADP-glucose pyrophosphorylase 6 (plastid)), ELF3 (Early flowering 3), FDL2 (FD-like 2), FK (Fructokinase), FT (Flowering locus T), FUL2 (Fruitfull-like 2), FUL3 (Fruitfull-like 3), G6PT (G6P-Pi translocator), GBSSI (Granule-bound starch synthase I), GBSSII (Granule-bound starch synthase II), ISAI (Isoamylase I), ISAII (Isoamylase II), ISAIII (Isoamylase III), PGI (Glucose-6-phosphate isomerase), PGM (Phosphoglucomutase), PHYB (Phytochrome B), PHYC (Phytochrome C), PPD1 (Photoperiod 1), SBEI (Starch branching enzyme I), SBEIIa (Starch branching enzyme IIa), SBEIIb (Starch branching enzyme IIb), SBEIII (Starch branching enzyme III),SPPase (Starch PPase), SSI (Starch synthase I), SSIIa (Starch synthase IIa), SSIIb (Starch synthase IIb), SSIIc (Starch synthase IIc), SSIIIa (Starch synthase IIa), SSIIIb (Starch synthase IIIb), SSIVb (Starch synthase isoform IV), SUS (Sucrose synthase), UGPase (UDP-glucose pyrophosphorylase), VRN1 (Vernalization 1), VRN2 (Vernalization 2).

SI Appendix, Tables

***SI Appendix*, Table S1.** Wheat exome capture design[1].

Category	No. of sequences
1. T. turgidum 'Kronos' transcripts	56,831
2. T. aestivum transcripts, complementary	23,759
3. H. vulgare transcripts matching *T. aestivum* genome	1,798
4. Genes contributed by the community	123
Non-redundant protein coding transcripts (<95% identical)	82,511
Number of padded exons	219,383
Number of unpadded exons	67,416
Total number of exons	286,799
Total number of bases in capture	84,038,015

[1] Publicly available from Roche catalogue numbers 140228_Wheat_Dubcovsky_D18_REZ_HX1 (tetraploid wheat) and 140430_Wheat_TGAC_D14_REZ_HX1 (hexaploid wheat).

***SI Appendix*, Table S2**. Library preparation and capture setup.

	T. turgidum	***T. aestivum***
CovarisE220 settings		
Duty cycle	20%	30%
Intensity	175 W	450 W
Cycles per burst	200	200
Time	90 s	115 s
Library preparation		
PCR cycles	5	6
Capture set-up		
Multiplexing per capture	8 samples	4 (or 8) samples
DNA input per sample	0.15 µg	0.35 (or 0.15) µg
Combined DNA input	1.2 µg	1.4 µg
Developer reagent [1]	12 µL	14 µL
Universal adapter blocker [2]	2.4 µL	2.8 µL
Barcode-specific blockers	0.5 µL of 250 µM stock (NEXTflex™ INV-HE Index Oligos) [2]	1 µL of 1,000 µM (TS-HE SeqCap HE Indices) [2]

[1] Roche, 6684335001.

[2] Tetraploid: Bioo Scientific, NEXTflex™ DNA Barcode Blockers 514134. Hexaploid: Roche, 06777287001.

***SI Appendix*, Table S3**. Kronos and Cadenza references with supplementary *de novo* assemblies (ChrU).

	Kronos	Cadenza
Assembly statistics ChrU		
Total length (bp)	33,388,548	41,301,548
Number of sequences	40,975	67,632
N25	1466	994
N50	935	646
N75	602	360
% GC	54.06%	54.97%
Mapping statistics ChrU		
Total number of input reads	43,073,616	56,988,370
Mapped (all)	33,964,726	30,159,927
Mapped (proper pairs)	33,168,732	28,681,670
Mapping quality 20	32,494,205	28,477,375
kmer (word size)	63	63
Length cutoff (bp)	300	0
Degenerate contig length cut-off	>200	>300
Reference statistics		
Total length (bp) without ChrU	7,426,889,742	10,332,975,726
Total number of sequences without ChrU	7,348,894	10,232,593
Source of 3B sequences	CSS14	CSS14+3B38
Total length (bp) plus ChrU	7,460,278,290	10,393,511,112
Total number of sequences plus ChrU	7,389,869	10,302,538
Mapping statistics for improved reference with ChrU		
Total number of test samples	24	86
Mean number of input reads	63,464,697	87,835,717
Mean number of mapped reads without ChrU	58,922,353	84,619,330
Mean number of mapped reads plus ChrU	62,358,434	86,687,447
% mapped reads to ref. without ChrU	93%	96%
% mapped reads to ref. plus ChrU	98%	99%

***SI Appendix*, Table S4.** Calculation of read coverage (before RH and deletion removal).

	Kronos	**Cadenza**
HetMC5/HomMC3		
Avg. coverage at mutation sites	26.60 X	29.01 X
Median coverage at mutation sites [1]	20.57 X	20.97 X
Standard deviation avg. coverage among individuals	± 5.54	± 6.74
Average standard deviation across all mutations	± 20.56	± 23.60
HetMC3/HomMC2		
Avg. coverage at mutation sites (*HetMC3/HomMC2*)	22.02 X	24.41 X
Median coverage at mutation sites [1]	16.43 X	17.1 X
Standard deviation avg. coverage among individuals	± 5.03	± 5.74
Average standard deviation across all mutations	± 19.40	± 21.78

[1] The lower values of the median relative to the means reflect distributions skewed to the right (higher coverage) as shown in *SI Appendix*, Fig. S1.

***SI Appendix*, Table S5.** Uniquely mapped EMS-type mutations at different stringency levels in 1,535 mutagenized tetraploid Kronos lines (excluding RH regions).

Coverage	# SNPs	Het/ Hom	EMS SNP	Avg. EMS SNP / line	%EMS	non-EMS transitions	%EMS error
HetMC3/HomMC2 [1]	5,525,228	2.46	5,085,379	3,313	92.04	35,707	0.70
HetMC4/HomMC3	4,601,287	2.15	4,507,550	2,937	97.96	12,525	0.28
HetMC5/HomMC3 [2]	**4,189,561**	**1.87**	**4,152,707**	**2,705**	**99.12**	**7,323**	**0.18**
HetMC6/HomMC4	3,771,030	1.78	3,745,578	2,440	99.33	5,885	0.16
HetC3	809,943		453,092	295	55.94	25,257	5.57
HetC4	421,189		362,657	236	86.10	5,741	1.58
HetC5	320,269		308,163	201	96.22	1,708	0.55
HetC6	263,162		258,349	168	98.17	997	0.39
HetMC7	2,150,372		2,129,733	1,387	99.04	4,888	0.23
HomC2	148,310		148,306	97	(100) [3]	Excluded [3]	
HomC3	107,289		107,289	70	(100) [3]	Excluded [3]	
HomC4	92,952		92,952	61	(100) [3]	Excluded [3]	
HomMC5	1,264,544		1,264,544	824	(100) [3]	Excluded [3]	

[1] *MC*= minimum coverage (e.g. *HetMC3* indicates mutations detected as heterozygous with a mean coverage of 3 or more reads. *C*= coverage at the exact level.

[2] Default stringency level *HetMC5/HomMC3* used in the main text is indicated in bold.

[3] Since most sequencing errors are heterozygous, all homozygous non EMS-type mutations were assumed to be RH and were removed by the RH pipeline, resulting in 100% EMS-type mutations.

***SI Appendix*, Table S6**. Uniquely mapped EMS-type mutations at different stringency levels in 1,200 EMS mutagenized Cadenza lines (excluding RH regions).

Coverage	# SNPs	Het/ Hom	EMS SNP	Avg. EMS SNP/line	%EMS	non-EMS transition	%EMS error
HetMC3/HomMC2 [1]	8,599,721	2.85	8,083,066	6,736	93.99	108,261	1.34
HetMC4/HomMC3	7,203,110	2.58	7,054,109	5,878	97.93	29,019	0.41
HetMC5/HomMC3 [2]	**6,470,733**	**2.21**	**6,421,522**	**5,351**	**99.24**	**10,569**	**0.16**
HetMC6/HomMC4	5,798,403	2.11	5,760,826	4,801	99.35	7,873	0.14
HetC3	1,197,870		821,320	684	68.57	82,177	10.01
HetC4	742,007		640,395	534	86.31	18,950	2.96
HetC5	521,816		509,428	425	97.63	2,865	0.56
HetC6	426,012		419,110	349	98.38	1,445	0.34
HetMC7	3,510,063		3,479,388	2,900	99.13	6,428	0.18
HomC2	23,491		231,489	193	(100) [3]	Excluded [3]	
HomC3	155,755		155,755	130	(100) [3]	Excluded [3]	
HomC4	131,238		131,238	109	(100) [3]	Excluded [3]	
HomMC5	1,731,090		1,731,090	1,443	(100) [3]	Excluded [3]	

[1] *MC*= minimum coverage (e.g. HetMC3 indicates mutations detected as heterozygous with a mean coverage of 3 or more reads. *C*= coverage at the exact level.

[2] Default stringency level *HetMC5/HomMC3* used in the main text is indicated in bold.

[3] Since most sequencing errors are heterozygous, all homozygous non EMS-type mutations were assumed to be RH and were removed by the RH pipeline, resulting in 100% EMS-type mutations.

***SI Appendix*, Table S7:** Summary of EMS mutations validated using KASP assays and Sanger sequencing in Kronos and Cadenza M_4 families.

	Kronos		Cadenza	
	No.	**%**	**No.**	**%**
Independent M_4 families tested	67		19	
Valid KASP assays and Sanger sequence	133		147	
False positives	1 [1]	0.7%	1	0.7%
Mutations confirmed	132	99.2%	146	99.3%
Expected segregation (MAPS with correction[2])	130	98.5%	139	95.2%
HOM mutation originally classified as HET	2	1.5%	2	1.4%
HET mutation originally classified as HOM[2]	0	0.0%	5	3.4%

[1] This is after correcting for the mutation which initially failed to validate in Kronos910. We confirmed that the mutation was absent due to a planting error and that the mutation was real when the ID of the M_4 seeds was corrected.

[2] This includes correction from heterozygous-to-homozygous (*SI Appendix*, Fig S5).

***SI Appendix*, Table S8**: Detailed information of mutations validated by KASP assays in Kronos M_4 families.

IWGSC contig	Line	Pos.	WT	Mut.	Val.	Pred.	Obs M_4	Primer 1 (Kronos)	Primer 2 (mutant)	Common Primer
IWGSC_CSS_1AS_scaff_3284790	Kronos3085	7449	G	A	Y	Het	Het	ccacaccttgagcctcgc	ccacaccttgagcctcgt	gtgattttgccaggggaga
IWGSC_CSS_1BL_scaff_3897513	Kronos3085	1515	C	T	Y	Het	Het	gcttccactgggtcctgc	gcttccactgggtcctgt	acaaggactgcttcagagac
IWGSC_CSS_2AL_scaff_6434745	Kronos3085	3424	C	T	Y	Het	Het	cctcggttttgcaaatttctatgc	cctcggttttgcaaatttctatgt	ggcaatggcataacaacagata
IWGSC_CSS_3AS_scaff_3408995	Kronos3085	732	C	T	Y	Het	Het	aggccatttcgaattccgc	aggccatttcgaattccgt	ggtgttatccagaacctgagtg
IWGSC_CSS_3B_scaff_10708748	Kronos3085	2675	G	A	Y	Het	Het	gttgcatgcttcacccagg	gttgcatgcttcacccaga	gtaacaatctgagttcgtagcac
IWGSC_CSS_4AL_scaff_7132733	Kronos3085	1799	C	T	Y	Hom	Hom	cacccgtgagtgaccctc	cacccgtgagtgaccctt	accgcctagaaagaaagcttc
IWGSC_CSS_5AS_scaff_1534693	Kronos3085	4605	C	T	Y	Het	Het	cagcttcctggccctcatc	cagcttcctggccctcatt	gtacctcacgagtcatgagag
IWGSC_CSS_6AS_scaff_4361911	Kronos3085	8857	G	A	Y	Het	Het	tcacgaaagacgacttcaacctcc	tcacgaaagacgacttcaacctct	catgaggtgctgcatctccatca
IWGSC_CSS_6BS_scaff_3008326	Kronos3085	1528	G	A	Y	Het	Het	ccatgttgtactggtggtgc	ccatgttgtactggtggtgt	ggaagcatggcaagtgca
IWGSC_CSS_7AS_scaff_4214385	Kronos3085	27835	C	T	Y	Hom	Hom	cgtaccttcgttgggaaagg	cgtaccttcgttgggaaaga	ctcttggtcagctgtataagact
IWGSC_CSS_1AL_scaff_3929964	Kronos3191	1336	C	T	Y	Het	Het	tttcggccatacctgacatc	tttcggccatacctgacatt	attgcctccagttcttgcag
IWGSC_CSS_1BL_scaff_3899789	Kronos3191	7925	C	T	Y	Het	Het	actctcactggcagcagc	actctcactggcagcagt	caacgtggtgcccatcgta
IWGSC_CSS_2AL_scaff_6426728	Kronos3191	1481	G	A	Y	Hom	Hom	gaaactgccgcagctcgc	gaaactgccgcagctcgt	ccagcagctcgtgagaaa
IWGSC_CSS_2BL_scaff_7960273	Kronos3191	690	C	T	Y	Hom	Hom	gccattcatccttaggcgc	gccattcatccttaggcgt	acatgcaattgctgatgactg
IWGSC_CSS_3AS_scaff_3286603	Kronos3191	2975	G	A	Y	Hom[1]	Hom	ccgtgtggtttgttgtggg	ccgtgtggtttgttgtgga	gaaaggaacgtgtcatgcag
IWGSC_CSS_5AL_scaff_2694249	Kronos3191	2399	C	T	Y	Het	Het	gccttccagatagagccgc	gccttccagatagagccgt	cgccacatcgacattcctg
IWGSC_CSS_5BL_scaff_10923577	Kronos3191	3713	C	T	Y	Het	Het	gtggattgcctgagcttgc	gtggattgcctgagcttgt	tggtggccttcttgggac
IWGSC_CSS_6AL_scaff_5823017	Kronos3191	13225	C	T	Y	Hom	Hom	ccctttcgagcctctggag	ccctttcgagcctctggaa	ttcgagaaggcccatcga
IWGSC_CSS_6BS_scaff_2955394	Kronos3191	1622	C	T	Y	Hom[1]	Hom	gtggagatgaaggtctagcaag	gtggagatgaaggtctagcaaa	gatactcgtgcaatgggtgt
IWGSC_CSS_7BL_scaff_6739382	Kronos3191	12261	G	A	Y	Hom	Hom	gagacaagctttgaattgctcc	gagacaagctttgaattgctct	cgagtgaccttcatttcccg
IWGSC_CSS_1AS_scaff_3276389	Kronos3288	9720	C	T	Y	Hom	Hom	accagcaggaccaatgtctc	accagcaggaccaatgtctt	atgatgcaacctcagccat
IWGSC_CSS_2AL_scaff_6367515	Kronos3288	6976	G	A	Y	Het	Het	caggtcgagtgtctccgg	caggtcgagtgtctccga	ggggtgatctggaagggc
IWGSC_CSS_2AL_scaff_6422019	Kronos3288	4523	G	A	Y	Het	Het	cgctaggtccctgcatagg	cgctaggtccctgcataga	acgcacgctaagccgtac
IWGSC_CSS_3AL_scaff_4284850	Kronos3288	7901	C	T	Y	Hom	Hom	tggctttggacaacatcgg	tggctttggacaacatcga	tgtcagcatcgacagccag
IWGSC_CSS_3B_scaff_10436253	Kronos3288	3228	G	A	---[2]	Het	---[2]	aggctggtgaaatgagtggg	aggctggtgaaatgagtgga	tctccttcacagacctggg
IWGSC_CSS_4AS_scaff_5962359	Kronos3288	13049	G	A	Y	Het	**Hom**	ccatcaagaagtacgagttcgac	ccatcaagaagtacgagttcgat	accatgcccagcttgtca
IWGSC_CSS_5AL_scaff_2751724	Kronos3288	7179	C	T	---[2]	Het	---[2]	ctggaaaagggactccgcc	ctggaaaagggactccgct	acaactgggtcgtgggga

IWGSC contig	Line	Pos.	WT	Mut.	Val.	Pred.	Obs M_4	Primer 1 (Kronos)	Primer 2 (mutant)	Common Primer
IWGSC_CSS_6AL_scaff_5778773	Kronos3288	6853	G	A	Y	Het	Het	gagtgaccttcccgtctttc	gagtgaccttcccgtctttt	ggagaacagctactcggct
IWGSC_CSS_6AS_scaff_4392100	Kronos3288	3434	C	T	Y	Het	Het	atggaagcacaggtgaccg	atggaagcacaggtgacca	ggaagcgaaagtgaacaaaca
IWGSC_CSS_7BL_scaff_6744240	Kronos3288	9772	G	A	Y	Het	Het	agctgttcttctcctacttcaag	agctgttcttctcctacttcaaa	caggtcgttcttgagctcc
IWGSC_CSS_1AL_scaff_3887185	Kronos3413	9708	C	T	Y	Hom	Hom	gcacgcctttatcgaggtaaag	gcacgcctttatcgaggtaaaa	agaaacagcagagcgcaa
IWGSC_CSS_2AL_scaff_6379082	Kronos3413	4307	G	A	---[2]	Het	---[2]	ggaaaacggcgtcaaaggg	ggaaaacggcgtcaaagga	tcagtgtgccagagagcc
IWGSC_CSS_2BS_scaff_3381362	Kronos3413	5160	C	T	Y	Hom[1]	Hom	caacttctgggctgtagtgtg	caacttctgggctgtagtgta	tgagaattctgacgcaaaagac
IWGSC_CSS_3AS_scaff_3296605	Kronos3413	6154	G	A	Y	Het	Het	ctggtcacgggctctagc	ctggtcacgggctctagt	cagcactgagagacatggac
IWGSC_CSS_3B_scaff_10693516	Kronos3413	12632	C	T	Y	Het	Het	ctaggcttggacaaacaggc	ctaggcttggacaaacaggt	agcttgcatctatgggcatt
IWGSC_CSS_5AS_scaff_1547699	Kronos3413	2686	G	A	Y	Het	Het	gctacaaccttcaccaatcgc	gctacaaccttcaccaatcgt	gacggctttgaagtgtcatc
IWGSC_CSS_5BL_scaff_10856077	Kronos3413	5853	G	A	Y	Het	Het	agagcttcaccccatgctc	agagcttcaccccatgctt	acgcacatttaatagctgaagc
IWGSC_CSS_6AL_scaff_5750718	Kronos3413	11046	G	A	Y	Hom	Hom	cacgcttcccgacttcttatag	cacgcttcccgacttcttataa	agacgatgtgatcaggattcag
IWGSC_CSS_7AL_scaff_4433177	Kronos3413	3511	C	T	Y	Het	Het	gatgctccgtcaggctgg	gatgctccgtcaggctga	cactactggacaagctcttgg
IWGSC_CSS_7BL_scaff_6742567	Kronos3413	667	C	T	Y	Het	Het	gttgcttgcgtggcagac	gttgcttgcgtggcagat	cattttgcaccgtgtgtctg
IWGSC_CSS_1AL_scaff_3976389	Kronos3935	10941	C	T	Y	Hom	Hom	ggtgaggagatcggcgatg	ggtgaggagatcggcgata	cagtcatctacatgagaggtcag
IWGSC_CSS_1BL_scaff_3873362	Kronos3935	1392	G	A	Y	Het	Het	cagatctgaagcctagcacatg	cagatctgaagcctagcacata	actaccagaatcagcacaaaaac
IWGSC_CSS_2BL_scaff_7882382	Kronos3935	2721	C	T	Y	Het	Het	gcaagctaagatgtaccgtagc	gcaagctaagatgtaccgtagt	gccacagtaggagaaagactt
IWGSC_CSS_3AL_scaff_4242376	Kronos3935	2410	C	T	Y	Het	Het	agaacccaaaacccgtacttag	agaacccaaaacccgtacttaa	gtagggtccatcctaaagcttg
IWGSC_CSS_3B_scaff_10485067	Kronos3935	3349	C	T	Y	Hom	Hom	gcttgagcaactactccaactg	gcttgagcaactactccaacta	gcaatttcctttatccgcagt
IWGSC_CSS_4AS_scaff_5984153	Kronos3935	6006	G	A	Y	Het	Het	agcaggtctggccaagttg	agcaggtctggccaagtta	cgaatgtatgagtaggcgct
IWGSC_CSS_4BL_scaff_7019402	Kronos3935	9081	C	T	Y	Het	Het	tgcaatcatgtagtgagctgg	tgcaatcatgtagtgagctga	agcatgatccctagaaccatac
IWGSC_CSS_5BL_scaff_10842786	Kronos3935	3304	G	A	Y	Het	Het	tggttcccgaagcctgaac	tggttcccgaagcctgaat	cgcatacttgaaacatgagcac
IWGSC_CSS_6BS_scaff_3045205	Kronos3935	2293	G	A	Y	Het	Het	aaggaccaagcccaaactctcg	aaggaccaagcccaaactctca	agtgatcaagcccaatgtcgca
IWGSC_CSS_7AL_scaff_4555249	Kronos3935	4487	C	T	Y	Het	Het	cagtgctcgagatggcgc	cagtgctcgagatggcgt	ccttgcaaccctcctgatt
IWGSC_CSS_1AL_scaff_3890367	Kronos4240	1639	G	A	---[2]	Het	---[2]	ttccaggtgtggtctgcac	ttccaggtgtggtctgcat	aagtctggtagtgatgatggtg
IWGSC_CSS_1BL_scaff_3918498	Kronos4240	6096	G	A	Y	Het	Het	ttgcatgccccaagaagag	ttgcatgccccaagaagaa	tgggcgaactggtaatgtgg
IWGSC_CSS_2BS_scaff_5131713	Kronos4240	5900	G	A	Y	Het	Het	cctttatcgaggaaagagacacc	cctttatcgaggaaagagacact	caccattgtagggttccttttttc
IWGSC_CSS_3B_scaff_10667202	Kronos4240	5470	G	A	---[2]	Het	---[2]	gactatcaggcggaggatgg	gactatcaggcggaggatga	tgtcatcctgctcctgcg
IWGSC_CSS_4BL_scaff_7037371	Kronos4240	5967	G	A	---[2]	Het	---[2]	ccatatacaagctggtggtcatg	ccatatacaagctggtggtcata	tgacgcacgttctgccaa
IWGSC_CSS_5AL_scaff_2769540	Kronos4240	9626	C	T	Y	Het	Het	tgcagtgtgggaaacggag	tgcagtgtgggaaacggaa	catgagtgagatcttcctgct

IWGSC contig	Line	Pos.	WT	Mut.	Val.	Pred.	Obs M_4	Primer 1 (Kronos)	Primer 2 (mutant)	Common Primer
IWGSC_CSS_5BL_scaff_10871091	Kronos4240	7062	G	A	Y	Het	Het	gccaaggaaccataacctgc	gccaaggaaccataacctgt	ggactcttggcaaccgga
IWGSC_CSS_6AL_scaff_5800333	Kronos4240	2360	G	A	Y	Het	Het	cgacaggattgtgagcgc	cgacaggattgtgagcgt	tcagatgctgcaagattcatct
IWGSC_CSS_7AS_scaff_4208434	Kronos4240	666	C	T	---[2]	Het	---[2]	aggtgtgttggttgggtgg	aggtgtgttggttgggtga	ccacaccatcatgcatagca
IWGSC_CSS_7BL_scaff_6716931	Kronos4240	2613	G	A	Y	Het	Het	ggtgggtatttgcttggtgag	ggtgggtatttgcttggtgaa	tggtggactcgacagtgta
IWGSC_CSS_2AL_scaff_6382649	Kronos4346	3051	G	A	**N**	Het	**wt**	gctcgatgtacatgttcacctc	gctcgatgtacatgttcacctt	tggtttctcccctcctctc
IWGSC_CSS_2BL_scaff_8029221	Kronos4346	2860	G	A	Y	Het	Het	tgcttccgctcttgctcc	tgcttccgctcttgctct	atttgcattcgatcgggcc
IWGSC_CSS_3AS_scaff_3289508	Kronos4346	1008	C	T	---[2]	Het	---[2]	gaagctgcgatggctcttg	gaagctgcgatggctctta	ccatctgatcagagacgctttt
IWGSC_CSS_3B_scaff_10460714	Kronos4346	14359	C	T	Y	Hom	Hom	ctaccttgccatgcgacatg	ctaccttgccatgcgacata	agcaccccagtctttgacg
IWGSC_CSS_4AS_scaff_5989735	Kronos4346	6404	G	A	Y	Hom	Hom	acgcatgctaacatcagcc	acgcatgctaacatcagct	actcaagataccaccgcacg
IWGSC_CSS_5BL_scaff_7648030	Kronos4346	6893	C	T	Y	Het	Het	tacccttcctactggcagg	tacccttcctactggcaga	ttttcagaggaacacaggtatca
IWGSC_CSS_6AL_scaff_5755840	Kronos4346	778	C	T	Y	Het	Het	atcgagtaagctgtcaccgc	atcgagtaagctgtcaccgt	acctgcatgtcacatccac
IWGSC_CSS_6BS_scaff_2972151	Kronos4346	7876	G	A	Y	Hom	Hom	gcagcaatgtcactgtttgg	gcagcaatgtcactgtttga	gcttggactgggcatttatg
IWGSC_CSS_7AL_scaff_4542983	Kronos4346	18700	G	A	Y	Het	Het	gcagggctaccggatacc	gcagggctaccggatact	catctgccggttaaacatgc
IWGSC_CSS_7BS_scaff_3098098	Kronos4346	5183	C	T	Y	Het	Het	gcgatatggtacttgcaatgag	gcgatatggtacttgcaatgaa	ttacattgcttatagtttgccgg
IWGSC_CSS_1AS_scaff_3259804	Kronos4485	219	C	T	Y	Het	Het	gtcggcacaaccccttgc	gtcggcacaaccccttgt	gcttctttaaggagggcga
IWGSC_CSS_2AL_scaff_6315418	Kronos4485	10490	G	A	Y	Hom	Hom	gcccctctcaaccttctcagc	gcccctctcaaccttctcagt	ttcagacgctcgaggaatttccc
IWGSC_CSS_2BS_scaff_5181092	Kronos4485	3742	G	A	Y	Het	Het	tggccagcacacctgcag	tggccagcacacctgcaa	tggacgatgagtgatggaaat
IWGSC_CSS_3B_scaff_10425015	Kronos4485	2372	C	T	Y	Het	Het	gctactgaagttggctcgg	gctactgaagttggctcga	cttcacatccttgggggttc
IWGSC_CSS_3B_scaff_10775915	Kronos4485	4701	C	T	Y	Het	Het	ccaagggctgcagagagg	ccaagggctgcagagaga	agacctcacgatgtcctcc
IWGSC_CSS_5AL_scaff_2754304	Kronos4485	2301	G	A	Y	Het	Het	taaccctgccatcgcccg	taaccctgccatcgccca	cattggccagccatgact
IWGSC_CSS_5BL_scaff_10919959	Kronos4485	1867	C	T	Y	Hom	Hom	gatgccctttgtggagaagg	gatgccctttgtggagaaga	tcttgttcccgaaacatgtca
IWGSC_CSS_6AL_scaff_5784069	Kronos4485	10805	G	A	---[2]	Het	---[2]	cgctatatgcattcctcggc	cgctatatgcattcctcggt	ctccagatgatgctcactgg
IWGSC_CSS_7AS_scaff_4245431	Kronos4485	3402	G	A	Y	Hom	Hom	aaggcgcctggtgtttcc	aaggcgcctggtgtttct	agtaagtggaacagctaagatcat
IWGSC_CSS_7BL_scaff_6667357	Kronos4485	641	C	T	Y	Het	Het	gatcagctgctcattcgagg	gatcagctgctcattcgaga	ttccctgtcaattgatgccc

[1] Heterozygous corrected to homozygous by bioinformatics filter applied after MAPS (*SI Appendix*, Figs. S2 and S5).

[2] KASP assays that failed to produce valid cluster.

***SI Appendix*, Table S9**: Detailed information of mutations validated by direct sequencing in Kronos M_4 families.

IWGSC contig	Line	Pos.	WT	Mut.	Val.	Pred.	Obs M_4	Left Primer	Right Primer
IWGSC_CSS_1AS_scaff_3287964	Kronos4254	3430	G	A	Y	Hom[1]	Hom	gataggaattagactgattc	acagtatggcttcttttcagc
IWGSC_CSS_1AS_scaff_3287964	Kronos2088	1305	G	A	Y	Het	Het	atgcatagatcaggaagaatg	catcaatgggcagaatccaagctcaggga
IWGSC_CSS_1BS_scaff_3463122	Kronos2536	14395	G	A	Y	Hom	Hom	ctcttcggcaagatctgtga	tcggccggtacagttagac
IWGSC_CSS_1BS_scaff_3463128	Kronos3720	14552	C	T	Y	Hom[1]	Hom	gccggtggccaagaagagc	cacatcacaattatgtactcctgc
IWGSC_CSS_1BS_scaff_3463128	Kronos2531	14332	G	A	Y	Het	Het	actcctgcttaattaaatttgtaccctgca	atatcaccaactgattgtttc
IWGSC_CSS_1BS_scaff_3473017	Kronos1051	4776	G	A	Y	Het	Het	tgcagaccaacaggctaaga	atgtgtccagcgctttacct
IWGSC_CSS_2AL_scaff_5133291	Kronos2233	785	G	A	Y	Het	Het	ccgggggagctccaagtaca	agctaggaggagagcgatggtggtaaata
IWGSC_CSS_2BL_scaff_2901645	Kronos3634	1002	G	A	Y	Het	**Hom**	cagcatctctctgccggcac	tcggtatcgaacagccccaagtagaggtac
IWGSC_CSS_2BL_scaff_7943322	Kronos2209	7138	C	T	Y	Hom	Hom	ggtaacatccaaacaagtcttatcacctagct	acatgttctttattatacctttatagaca
IWGSC_CSS_2BL_scaff_8085378	Kronos2991	17069	G	A	Y	Het	Het	ccattgaccatcaactagatca	tgaaccagcattcctgata
IWGSC_CSS_2BS_scaff_5169641	Kronos2489	4370	C	T	Y	Hom	Hom	catttatagctcaagttttagtggtg	gcactacagcagggcctctt
IWGSC_CSS_2BS_scaff_5176399	Kronos2594	2045	G	A	Y	Het	Het	gcctccgccagcgcttctgccagcagt	catgaagaaaatcgatcctaaaagg
IWGSC_CSS_2BS_scaff_5176399	Kronos2091	1915	C	T	Y	Het	Het	gctcgtgaacgttcccgtggtc	gctgaggtcggcgccgcacctctcaggct
IWGSC_CSS_2BS_scaff_5186635	Kronos2620	2234	C	T	Y	Hom	Hom	tgggaaagcatgcatattga	gctgatgtgacgggaaaact
IWGSC_CSS_3AL_scaff_4257087	Kronos656	180	G	A	Y	Hom	Hom	ctcccctggtatgtctgcat	acatggtgatgtcaaactgctaa
IWGSC_CSS_3AL_scaff_4263764	Kronos692	1370	C	T	Y	Hom	Hom	gtgaactgatgtcttcaaacaagggtcg	gaggcggatttctgcggcata
IWGSC_CSS_3AL_scaff_4342642	Kronos463	4737	G	A	Y	Hom	Hom	aattccatggagagtctgccacataactag	gaaccatttaatatgtgctcctta
IWGSC_CSS_3AL_scaff_4387054	Kronos3179	1039	G	A	Y	Het	Het	ttatctatatgtaattcgttcc	tagggaaagcaaatacaacagtcacctgt
IWGSC_CSS_3AS_scaff_3324175	Kronos2019	6301	C	T	Y	Hom[1]	Hom	tggcatcctggaagctcttt	attggggtgagatctgagca
IWGSC_CSS_3AS_scaff_3324175	Kronos0401	2737	G	A	Y	Het	**Hom**	tgcttgatgttggttgtggt	gaaacccatttgagcgagag
IWGSC_CSS_3B_scaff_10430895	Kronos445	11936	C	T	Y	Het	Het	actaccttggatggctttttcttgcgtcga	aaacaaagtagacataattatttag
IWGSC_CSS_3B_scaff_10600478	Kronos2215	1060	G	A	Y	Hom	Hom	atttccctcctctccctgtaattctgca	acgcatgatctggtgtagaatg
IWGSC_CSS_3B_scaff_10614985	Kronos2205	2968	G	A	Y	Hom	Hom	tggcatcctggaagctcttt	atagggctgtgatctgagca
IWGSC_CSS_3B_scaff_10614985	Kronos2241	4509	C	T	Y	Hom	Hom	tgctcagatcacagccctat	caaatatggccttgtcagca
IWGSC_CSS_3B_scaff_10645141	Kronos234	1914	C	T	Y	Hom	Hom	gcgtgaaactgacttcaaggat	gctcggcccattccagaaatgtaaa
IWGSC_CSS_4AL_scaff_6409814	Kronos776	672	G	A	Y	Hom	Hom	ttggaatcaatggatcaatgagt	ggtgcgcaggggaagcagacttgaatt
IWGSC_CSS_4AL_scaff_7152393	Kronos2713	11229	G	A	Y	Hom	Hom	ggattcaggcacataatacactttctgc	gacagtcgcattattcatctactgatcatcg

IWGSC contig	Line	Pos.	WT	Mut.	Val.	Pred.	Obs M_4	Seq. primer 1	Seq. primer 2
IWGSC_CSS_4AL_scaff_7152393	Kronos2170	11400	C	T	Y	Het	Het	ggattcaggcacataatacactttctgc	gacagtcgcattattcatctactgatcatcg
IWGSC_CSS_4BS_scaff_4913659	Kronos934	5188	C	T	Y	Het	Het	gactcctcctgcagcaccta	aacccggcgttgccgagg
IWGSC_CSS_5BL_scaff_10820313	Kronos 4369	5760	C	T	Y	Het	Het	gcttccagtcagacctcgtc	aattgaccattgatcgcaca
IWGSC_CSS_5AL_scaff_2205676	Kronos463	3435	C	T	Y	Het	Het	agctaattgtgtactcactttatgcatgca	ttgttagctgtgagttaaactcgttag
IWGSC_CSS_5AL_scaff_2205676	Kronos2338	1655	C	T	Y	Het	Het	cacaaagtaggtttaattgatctg	tcagtcaaaattgtgcaatcaccc
IWGSC_CSS_5BL_scaff_10886394	Kronos1440	7121	G	A	Y	Hom	Hom	actgcactctctcccagctc	aatcaatggctggctaatgg
IWGSC_CSS_5BL_scaff_10886394	Kronos2135	7181	C	T	Y	Het	Het	actgcactctctcccagctc	aatcaatggctggctaatgg
IWGSC_CSS_5BL_scaff_10886394	Kronos2084	7161	G	A	Y	Het	Het	actgcactctctcccagctc	aatcaatggctggctaatgg
IWGSC_CSS_5BL_scaff_10904272	Kronos3946	7511	C	T	Y	Het	Het	cttctccgctccaaggccagagccaagcgg	tcatctgatctccttcgtgc
IWGSC_CSS_5BL_scaff_10913218	Kronos334	9754	C	T	Y	Het	Het	gccccttccggcgtgagggac	gggtggtgttcatgatggct
IWGSC_CSS_5BS_scaff_2237861	Kronos344	5635	G	A	Y	Hom	Hom	gcacagggaagccatacaa	tggcactaaacatttccgct
IWGSC_CSS_6AS_scaff_4388419	Kronos2467	2552	C	T	Y	Hom[1]	Hom	tagccgagtttgaccagaaaaagcgaaggt	gctgagttggtgtgaccata
IWGSC_CSS_6AS_scaff_4388419	Kronos4671	2778	G	A	Y	Het	Het	tgaagctcggcaagagaacc	gctgagttggtgtgaccata
IWGSC_CSS_6AS_scaff_4388419	Kronos2513	1864	G	A	Y	Het	Het	catttcactctagattcctcaaaat	agcctgaccgaataggaaatttggctggt
IWGSC_CSS_6AS_scaff_4392667	Kronos2942	9224	G	A	Y	Hom	Hom	tctgctcaattgttttcgatcga	aggatcaggtaatccacaagaactagtcga
IWGSC_CSS_6AS_scaff_4392667	Kronos563	7151	C	T	Y	Het	Het	aggaaataagagaagtatcaatcaatcga	ccacctaggaaagaagatact
IWGSC_CSS_6AS_scaff_4392667	Kronos263	5879	C	T	Y	Het	Het	ggcgaggcccacaccaagtt	gagcgtcaagaaacacccaccggctgcagt
IWGSC_CSS_6BL_scaff_4207030	Kronos4353	595	G	A	Y	Hom	Hom	tgcgcccatctgtctctgcatctacctgca	aaacctctatatacagtatgttat
IWGSC_CSS_6BL_scaff_4385559	Kronos278	1922	C	T	Y	Het	Het	tcaagttcaccttgcttctgatt	cggctgggagtgcgagacgagaagcgttt
IWGSC_CSS_6BS_scaff_2928775	Kronos422	1360	C	T	Y	Hom[1]	Hom	acattttgtgctaaaacttttgtat	aaataggaaattaggctgttgccttggat
IWGSC_CSS_6BS_scaff_2928775	Kronos1071	1238	C	T	Y	Het	Het	ttcaaagttactcacacaaaagctccatg	gtttggaaattcattcgtaagcg
IWGSC_CSS_6BS_scaff_527202	Kronos679	3929	C	T	Y	Het	Het	taataaatgcttttgccattgatgctgca	tgggaaaggccttccatgacta
IWGSC_CSS_7AL_scaff_4536617	Kronos849	7318	G	A	Y	Hom	Hom	cccgaacagttagtattgcatac	gcatccctaagtggtcttttt
IWGSC_CSS_7AS_scaff_4058401	Kronos3953	199	G	A	Y	Hom	Hom	gaactgacgaggttcggtctccccgctgca	aaacgggcagcgtggatttgt
IWGSC_CSS_7AS_scaff_4176378	Kronos3078	790	G	A	Y	Het	Het	tgatggttgtgtggattagaatca	ctggtgcgcaggggaagcagacttggaat
IWGSC_CSS_7AS_scaff_4253310	Kronos1107	7739	G	A	Y	Hom	Hom	tcgatctacactaggaagaaggaag	gtgggccatgggtagg
IWGSC_CSS_7AS_scaff_4253310	Kronos910[2]	7914	G	A	(N)	Het[2]	(wt)[2]	gcagttgttgacagggattagcgtc	cagctacggagtccgctccttct
IWGSC_CSS_7AS_scaff_4253310	Kronos1279	7741	G	A	Y	Hom	Hom	tcgatctacactaggaagaaggaag	gtgggccatgggtagg

IWGSC contig	Line	Pos.	WT	Mut.	Val.	Pred.	Obs M_4	Left Primer	Right Primer
IWGSC_CSS_7AS_scaff_4253310	Kronos508	7872	G	A	Y	Het	Het	tcgatctacactaggaagaaggaag	gtgggccatgggtagg
IWGSC_CSS_7AS_scaff_4253310	Kronos944	7826	G	A	Y	Het	Het	tcgatctacactaggaagaaggaag	gtgggccatgggtagg
UCW_Kronos_U_deg7180000269681	Kronos3946	1040	G	A	Y	Het	Het	gtggtttgtccgatggttga	tgataatgtgggcgaagtga
UCW_Kronos_U_deg7180000269681	Kronos2992	567	G	A	Y	Het	Het	gctggcctttgtccgtctgattcctctag	tatccagtgttaatcacgagc
UCW_Kronos_U_jcf7180000426752	Kronos579	747	C	T	Y	Hom	Hom	gagtccaagtactgcgacagg	gggtttatctggctctgagaag
UCW_Kronos_U_jcf7180000438968	Kronos2350	2000	C	T	Y	Het	Het	gcatgcagatgaagcacagt	gcagaaaaggtcctggaaga
UCW_Kronos_U_jcf7180000447250	Kronos2381	995	C	T	Y	Het	Het	tgcagaccaacaggctaaga	atgtatgtgtccggtgctca

[1] Heterozygous corrected to homozygous by bioinformatics filter applied after MAPS (*SI Appendix*, Figs. S2 and S5).

[2] We later confirmed that the mutation was absent due to a planting error, and the mutation was validated when the ID of the M_4 seeds was corrected.

***SI Appendix*, Table S10**: Detailed information of mutations validated by KASP assays in Cadenza M_4 families.

IWGSC contig	Line	Pos.	WT	Mut.	Val.	Pred.	Obs M_4	Primer 1 (Cadenza)	Primer 2 (mutant)	Common Primer
IWGSC_CSS_3B_scaff_10445294	Cadenza1772	6019	C	T	Y	Het	Het	caggatagtgggactgtcaaag	caggatagtgggactgtcaaaa	ggagacggctgtggacatt
IWGSC_CSS_3DL_scaff_6955403	Cadenza1772	2418	C	T	Y	Hom[1]	Hom	tcagcggattgtcgggatg	tcagcggattgtcgggata	tgtccatgaatcttgtccacg
IWGSC_CSS_4AL_scaff_7106846	Cadenza1772	11277	G	A	Y	Hom	Hom	tgggatccatgcctacactg	tgggatccatgcctacacta	gatggtggatttgccgcta
IWGSC_CSS_4AS_scaff_5991335	Cadenza1772	15710	G	A	Y	Hom	Hom	ctggccctgcgctgctac	ctggccctgcgctgctat	gtggaagttcagaaggaccag
IWGSC_CSS_4BS_scaff_4956646	Cadenza1772	252	G	A	Y	Hom[1]	Hom	gcaggttgacttcccggag	gcaggttgacttcccggaa	tgaggtacgagctaaagaaagc
IWGSC_CSS_4DS_scaff_1715962	Cadenza1772	1225	G	A	Y	Hom	Hom	cagctgtggtatctcaactgg	cagctgtggtatctcaactga	ccctgaaacaccgtttggat
IWGSC_CSS_5AL_scaff_2763407	Cadenza1772	2119	G	A	Y	Hom	Hom	gcgacgaacctcgagatctg	gcgacgaacctcgagatcta	gatggcaatcgtcgtgca
IWGSC_CSS_5AS_scaff_1548786	Cadenza1772	12625	C	T	Y	Het	Het	ataggcacattgctagactgag	ataggcacattgctagactgaa	ggattgggtgttgcacgc
IWGSC_CSS_5BL_scaff_10849226	Cadenza1772	2289	C	T	Y	Hom[1]	Hom	cctgacatcattgttcacgatc	cctgacatcattgttcacgatt	cactccgaggtgtccatgat
IWGSC_CSS_5BS_scaff_2270737	Cadenza1772	2262	G	A	---[2]	Hom	---[2]	attcctgtgttgtggcaaatgag	attcctgtgttgtggcaaatgaa	taagcacaaaccctccagctgg
IWGSC_CSS_1AL_scaff_3022915	Cadenza1661	891	C	T	Y	Hom	Hom	ccacagtgagactcctattgacg	ccacagtgagactcctattgaca	atgtctgattcgtcgtagtcc
IWGSC_CSS_1AS_scaff_3297240	Cadenza1661	1970	C	T	Y	Het	Het	catcccgccgtttcctcc	catcccgccgtttcctct	gctcgccgatgaagagct
IWGSC_CSS_1BL_scaff_3828996	Cadenza1661	1340	G	A	Y	Hom	Hom	agccggatgttagtgttaacc	agccggatgttagtgttaact	agcagcttgtcgcgttaac
IWGSC_CSS_1DS_scaff_1884529	Cadenza1661	10575	G	A	Y	Hom	Hom	acagatacaattgtcatgcaggc	acagatacaattgtcatgcaggt	acctgggttgtccaatacttc
IWGSC_CSS_2AL_scaff_6318370	Cadenza1661	19142	C	T	---[2]	Het	---[2]	cgtggccgaatctcgacg	cgtggccgaatctcgaca	ttcttgtgggagccgggc
IWGSC_CSS_2AS_scaff_5213460	Cadenza1661	1358	G	A	Y	Hom	Hom	gtcacgaacccgctcagg	gtcacgaacccgctcaga	aggaaagagaggaaaagagcg
IWGSC_CSS_2BS_scaff_5179331	Cadenza1661	5604	G	A	Y	Het	Het	actctcgtcaagaactgatacag	actctcgtcaagaactgatacaa	gcagagaatgttcttgcaact
IWGSC_CSS_2DS_scaff_5341235	Cadenza1661	4673	G	A	Y	Het	Het	ggtgaggatctcggagctg	ggtgaggatctcggagcta	gcgcggtcgtacgagttg
IWGSC_CSS_3AL_scaff_4250995	Cadenza1661	7046	G	A	Y	Hom	Hom	ccaagaaacgggtggtccag	ccaagaaacgggtggtccaa	ctgcagctgtcccatcatcgt
IWGSC_CSS_3B_scaff_10404421	Cadenza1661	4303	G	A	Y	Het	Het	ccttcgtcgacaggacctg	ccttcgtcgacaggaccta	gccagtactcacatgctctc
IWGSC_CSS_5DL_scaff_2390496	Cadenza1538	2125	C	T	Y	Hom	**Het**	gcagttttatcctcagtagtcttgg	gcagttttatcctcagtagtcttga	ttctgagaatgtaatgtgcgatg
IWGSC_CSS_6AL_scaff_5753680	Cadenza1538	3920	C	T	Y	Hom	Hom	tgctccaaatttgagcacaataac	tgctccaaatttgagcacaataat	aaatgcaaggggtaagtttttgt
IWGSC_CSS_6AS_scaff_4425792	Cadenza1538	4307	G	A	Y	Hom	**Het**	agatgcttgtcgggccag	agatgcttgtcgggccaa	gctgaagcaacgcgatcaat
IWGSC_CSS_6BS_scaff_3003630	Cadenza1538	6933	C	T	Y	Hom[1]	**Het**	ggcagtaatgtggtgctgagc	ggcagtaatgtggtgctgagt	ttgacttctggtttggtggca
IWGSC_CSS_6DL_scaff_3246988	Cadenza1538	9186	G	A	Y	Het	Het	gctaaagaagagcttgagagaattc	gctaaagaagagcttgagagaattt	aatttctgaagagaggtgttgtatg
IWGSC_CSS_7AL_scaff_4480114	Cadenza1538	3446	C	T	---[2]	Het	---[2]	gatatctcccacacggcgg	gatatctcccacacggcga	tgagccactcttgcagtttt
IWGSC_CSS_7AS_scaff_4193541	Cadenza1538	8359	C	T	Y	Hom	**Het**	agcaattctttggctatcaattagc	agcaattctttggctatcaattagt	tcatctgtcttaactctactgctg

IWGSC contig	Line	Pos.	WT	Mut.	Val.	Pred.	Obs M_4	Primer 1 (Cadenza)	Primer 2 (mutant)	Common Primer
IWGSC_CSS_7BL_scaff_6721572	Cadenza1538	9223	C	T	Y	Het	Het	gctcagggaggaagacaagaag	gctcagggaggaagacaagaaa	tgctatgaagaattccgacctc
IWGSC_CSS_7BS_scaff_3152545	Cadenza1538	3960	G	A	---[2]	Hom	---[2]	tcagcaaaatcacctgccgc	tcagcaaaatcacctgccgt	gctgccccatcatcgtttat
IWGSC_CSS_7DS_scaff_3963838	Cadenza1538	2913	G	A	Y	Het	Het	tcgttgcaagccttttgtgc	tcgttgcaagccttttgtgt	agagttatcaagctactgtcaca
IWGSC_CSS_1AL_scaff_3903380	Cadenza1469	6193	G	A	Y	Hom	Hom	ctcttcagagatgaacgcgg	ctcttcagagatgaacgcga	tcgtgagatggtggtttgtta
IWGSC_CSS_1AS_scaff_3287728	Cadenza1469	3817	C	T	Y	Hom[1]	Hom	ccgaccaattcactaaccgg	ccgaccaattcactaaccga	accctctttcccagacatgat
IWGSC_CSS_1BL_scaff_3815304	Cadenza1469	513	G	A	Y	Hom	Hom	aacatttgccttaccaaaacgc	aacatttgccttaccaaaacgt	acacagcaagttataatgcaagc
IWGSC_CSS_1DL_scaff_2266648	Cadenza1469	5926	C	T	Y	Het	Het	caacatgagacacaacaccttc	caacatgagacacaacaccttt	gtcaacgcgtgaggattgtc
IWGSC_CSS_1DS_scaff_1906671	Cadenza1469	3697	C	T	Y	Hom	Hom	tggtgtagacacttggcgag	tggtgtagacacttggcgaa	catggcgaccaccacctg
IWGSC_CSS_2AL_scaff_6337088	Cadenza1469	7334	G	A	Y	Hom[1]	Hom	acaatgccaagttgacaggttg	acaatgccaagttgacaggtta	gggagtgttggttcagaacat
IWGSC_CSS_2BL_scaff_7972799	Cadenza1469	8995	C	T	Y	Het	**Hom**	gtgctcctcggcatccttc	gtgctcctcggcatccttt	gatccgggcaaactacgtg
IWGSC_CSS_2DL_scaff_9832343	Cadenza1469	3262	G	A	Y	Het	Het	ttgtctaacagcaccgcagg	ttgtctaacagcaccgcaga	agatctcggtcagcctttct
IWGSC_CSS_2DS_scaff_5327939	Cadenza1469	3889	G	A	Y	Het	Het	tttttgccttatgtgactctagtac	tttttgccttatgtgactctagtat	gaggccatcacagatagcg
IWGSC_CSS_3B_scaff_10395219	Cadenza1469	1292	G	A	---[2]	Hom	---[2]	aggtgcttgtgcttgctgg	aggtgcttgtgcttgctga	cctcttctgggggctttatac
IWGSC_CSS_3B_scaff_10592217	Cadenza0580	2994	C	T	---[2]	Het	---[2]	acagcagtatcaagcccctc	acagcagtatcaagcccctt	tgatactgttgtggcggagg
IWGSC_CSS_3DS_scaff_2596771	Cadenza0580	1037	G	A	Y	Het	Het	tggttatgcacaggataatcagg	tggttatgcacaggataatcaga	tggcaaatgtgatgtcattaggt
IWGSC_CSS_4AL_scaff_7093953	Cadenza0580	9881	C	T	Y	Hom	Hom	gacaggaagccggtaacac	gacaggaagccggtaacat	ctccagcaggcatgggat
IWGSC_CSS_4BL_scaff_7037448	Cadenza0580	1837	C	T	Y	Hom	Hom	cgttgaaaagctgcaagaacttaac	cgttgaaaagctgcaagaacttaat	cagttcttccttcagagcagatat
IWGSC_CSS_4BS_scaff_4929479	Cadenza0580	10668	G	A	---[2]	Hom	---[2]	tggattttcccgcactgttc	tggattttcccgcactgttt	gtaaacaaggcatttcaagagtca
IWGSC_CSS_4DL_scaff_14359838	Cadenza0580	1408	G	A	---[2]	Hom	---[2]	gctcattcagggattgtcctatatg	gctcattcagggattgtcctatata	tgacagaacagttggtcatact
IWGSC_CSS_4DS_scaff_2276484	Cadenza0580	8034	G	A	Y	Hom	Hom	gccgtggttgatggagag	gccgtggttgatggagaa	cgtccagattactgatacttgca
IWGSC_CSS_5AL_scaff_2756579	Cadenza0580	5278	G	A	Y	Het	Het	tgaatggatttttcgtcccgttc	tgaatggatttttcgtcccgttt	ggaatcctatgcagaagaaactg
IWGSC_CSS_5BL_scaff_10787208	Cadenza0580	10627	G	A	---[2]	Het	---[2]	gcctctcacatgcggagac	gcctctcacatgcggagat	acgatgtcaggtgggcgt
IWGSC_CSS_5BS_scaff_2282179	Cadenza0580	5267	G	A	---[2]	Het	---[2]	tgatgggctacgacgtgc	tgatgggctacgacgtgt	tcggcgcccttgaaatcc
IWGSC_CSS_5DL_scaff_4498073	Cadenza0423	4937	C	T	Y	Hom	Hom	gcaccctctggttggtcatc	gcaccctctggttggtcatt	tgagcagcaaagcagccg
IWGSC_CSS_5DS_scaff_2738970	Cadenza0423	2319	C	T	---[2]	Het	---[2]	cgtgaggtgggtgatttgc	cgtgaggtgggtgatttgt	tggaactagttacactgcagttc
IWGSC_CSS_6AL_scaff_5757109	Cadenza0423	2788	G	A	Y	Hom	Hom	caggagcctggcaaataaagg	caggagcctggcaaataaaga	ctttcgcagtctcttagtttcg
IWGSC_CSS_6AS_scaff_4387871	Cadenza0423	2543	G	A	Y	Hom	Hom	gcatgctaacaggcgaaaagg	gcatgctaacaggcgaaaaga	ctcatgctcctgatcttaaggtt
IWGSC_CSS_6BL_scaff_4271391	Cadenza0423	4660	C	T	Y	Hom	Hom	tacgtgcatgatgtggtagtcgtac	tacgtgcatgatgtggtagtcgtat	gtttgaagtgcatcagatgtacca
IWGSC_CSS_6DS_scaff_1880206	Cadenza0423	9159	G	A	Y	Het	Het	ctgcgaaggctccacaag	ctgcgaaggctccacaaa	ggatgagaagtttgcattgctc

IWGSC contig	Line	Pos.	WT	Mut.	Val.	Pred.	Obs M_4	Primer 1 (Cadenza)	Primer 2 (mutant)	Common Primer
IWGSC_CSS_7AS_scaff_4227506	Cadenza0423	952	G	A	---[2]	Het	---[2]	ccatgtgtttccaatgttagagc	ccatgtgtttccaatgttagagt	tgccctagctggtatgct
IWGSC_CSS_7BL_scaff_6681782	Cadenza0423	1486	C	T	Y	Hom	Hom	agtaagcgtgacagcaatggg	agtaagcgtgacagcaatgga	atgtctttggtggaagtacatca
IWGSC_CSS_7BS_scaff_3160328	Cadenza0423	7801	C	T	Y	Het	Het	tgttaaatgatacagcctgcagc	tgttaaatgatacagcctgcagt	tggaatggtgcgttgttttt
IWGSC_CSS_7DS_scaff_407428	Cadenza0423	2051	G	A	Y	Het	Het	gtcgcgccatcctgacag	gtcgcgccatcctgacaa	actcatcaggtcagcccaa
IWGSC_CSS_3AL_scaff_442479	Cadenza0364	3198	C	T	Y	Het	Het	gagtcattaagttggtaagattggc	gagtcattaagttggtaagattggt	gcagataacaacaggatcacg
IWGSC_CSS_3AL_scaff_4447942	Cadenza0364	11917	G	A	Y	Het	Het	gtcataaagattgctcctgtgaag	gtcataaagattgctcctgtgaaa	ctcggatgtgggaggaaga
IWGSC_CSS_3AS_scaff_1557483	Cadenza0364	2547	C	T	Y	Het	Het	aaagtcacatcatgcttaccataag	aaagtcacatcatgcttaccataaa	cgaaatccaacgcctcatca
IWGSC_CSS_3AS_scaff_2648747	Cadenza0364	2688	G	A	Y	Het	Het	tggaagcacaaggggccc	tggaagcacaaggggcct	gccgccgatggagactcg
IWGSC_CSS_3AS_scaff_3304956	Cadenza0364	1017	G	A	Y	Het	Het	gtcccttgcacacagctttg	gtcccttgcacacagcttta	cctgctggactacaacttcaat
IWGSC_CSS_3AS_scaff_3321091	Cadenza0364	4585	C	T	Y	Het	Het	caagaatgatgctgatgttggag	caagaatgatgctgatgttggaa	acatgctgaatcgccgaatc
IWGSC_CSS_3AS_scaff_3371333	Cadenza0364	538	G	A	Y	Het	Het	gggaaacgagacgagcgg	gggaaacgagacgagcga	ccgtgccttcctcaccct
IWGSC_CSS_3AS_scaff_3371815	Cadenza0364	1061	C	T	Y	Het	Het	atccccacggcacagagg	atccccacggcacagaga	aattggcccttggtgattcc
IWGSC_CSS_3AS_scaff_3440912	Cadenza0364	4498	G	A	Y	Het	Het	ccgtaaaactttctgtgcttgc	ccgtaaaactttctgtgcttgt	atactgacaaactacatgatgtgc
IWGSC_CSS_3B_scaff_10343586	Cadenza0364	2242	G	A	---[2]	Het	---[2]	ggttctgtcctctcttccactg	ggttctgtcctctcttccacta	tgtgttgaacccgcaagca
IWGSC_CSS_5DL_scaff_242342	Cadenza0281	2433	C	T	Y	Hom	Hom	catggcgacggtgtcctg	catggcgacggtgtccta	aaccctcattttggctacttct
IWGSC_CSS_5DL_scaff_4538822	Cadenza0281	1208	G	A	---[2]	Hom	---[2]	acgtcagaacaaccgtttgac	acgtcagaacaaccgtttgat	ttaaattggttggcgccacc
IWGSC_CSS_6AL_scaff_5813297	Cadenza0281	4532	C	T	---[2]	Hom	---[2]	gggagagggacgtctcgg	gggagagggacgtctcga	ttcttctgccaacgattccg
IWGSC_CSS_6AS_scaff_4378990	Cadenza0281	6748	C	T	Y	Hom	Hom	cccaggttctgcttcttttcc	cccaggttctgcttcttttct	caagtatcaagaaaatgaagggtgt
IWGSC_CSS_6BL_scaff_4360781	Cadenza0281	5426	C	T	Y	Het	Het	actactcaaatggcttggtgtag	actactcaaatggcttggtgtaa	tcagtccaacatgtcaagagatt
IWGSC_CSS_7AL_scaff_4488310	Cadenza0281	3808	G	A	Y	Hom	Hom	gttctcttgtagtagcagccg	gttctcttgtagtagcagcca	ggcgctttcttcggccta
IWGSC_CSS_7BL_scaff_6696509	Cadenza0281	9232	G	A	Y	Het	Het	gctctaggggtggcaaaagg	gctctaggggtggcaaaaga	ggcttgaggtcgcagtgt
IWGSC_CSS_7BS_scaff_3143575	Cadenza0281	1866	C	T	Y	Het	Het	agatgttgagagggcgcttc	agatgttgagagggcgcttt	gcttggatggtggcaagtt
IWGSC_CSS_7DL_scaff_3346250	Cadenza0281	1663	G	A	Y	Het	Het	acgtgcagcaacatcctaac	acgtgcagcaacatcctaat	tttcccaccaggcccaaga
IWGSC_CSS_7DS_scaff_3933917	Cadenza0281	1243	C	T	Y	Het	Het	tgctgagcctttcaccttgc	tgctgagcctttcaccttgt	agaggtttggttccatcgg
IWGSC_CSS_3B_scaff_10626860	Cadenza0148	7847	G	A	Y	Het	Het	gcagctctgggaaggagg	gcagctctgggaaggaga	gttaatgtacctcctagcctcg
IWGSC_CSS_3DL_scaff_6915683	Cadenza0148	6904	C	T	Y	Het	Het	cgtcaacctgtgggcaattg	cgtcaacctgtgggcaatta	tcatgctcataatgtcatagggt
IWGSC_CSS_4AS_scaff_5929057	Cadenza0148	4238	G	A	Y	Hom	Hom	gcgcaacgtagcacctacc	gcgcaacgtagcacctact	ttatctggtgaagtgacaggttca
IWGSC_CSS_4AS_scaff_5950625	Cadenza0148	10590	C	T	Y	Het	Het	agatattcaaatcggtggattggc	agatattcaaatcggtggattggt	cctgctcccctcacgtcc
IWGSC_CSS_4AS_scaff_5967119	Cadenza0148	11626	C	T	Y	Hom	Hom	cgtggacaccccgagctg	cgtggacaccccgagcta	gacgacgcactgcacgac

IWGSC contig	Line	Pos.	WT	Mut.	Val.	Pred.	Obs M_4	Primer 1 (Cadenza)	Primer 2 (mutant)	Common Primer
IWGSC_CSS_4DL_scaff_14455742	Cadenza0148	1946	C	T	Y	Hom	Hom	gcctgagggagatcgcgc	gcctgagggagatcgcgt	aaccggtaactgtgggca
IWGSC_CSS_4DS_scaff_2318993	Cadenza0148	4000	C	T	Y	Hom	Hom	tccagtttgacacagattgaatggg	tccagtttgacacagattgaatgga	tgagattctgtttcctttcacattg
IWGSC_CSS_5AL_scaff_2750707	Cadenza0148	4603	G	A	Y	Het	Het	ccttggtgctagccatttcaagtag	ccttggtgctagccatttcaagtaa	ccaggatgcagtgcaatatttcaag
IWGSC_CSS_5BL_scaff_10794137	Cadenza0148	9235	C	T	Y	Hom	Hom	gaagctgcttctgcgttg	gaagctgcttctgcgtta	agtatcccttccatataagcagtg
IWGSC_CSS_5BS_scaff_1646558	Cadenza0148	2916	C	T	Y	Het	Het	gccgtacactcacctatcctttg	gccgtacactcacctatccttta	gcaatgtccacttatcatccct
IWGSC_CSS_1AL_scaff_3883106	Cadenza0110	27536	C	T	Y	Het	Het	accttccatcactggctgg	accttccatcactggctga	gtgaagaacaacaggttgaagc
IWGSC_CSS_1BL_scaff_3812829	Cadenza0110	10770	G	A	Y	Hom[2]	Hom	cccccactccattccagg	cccccactccattccaga	ggatgttgttctgtgctggaa
IWGSC_CSS_1DL_scaff_2266648	Cadenza0110	6156	G	A	Y	Het	Het	actgcgtggttatgggacc	actgcgtggttatgggact	ccccatcactgaacacaaca
IWGSC_CSS_1DS_scaff_1889435	Cadenza0110	8826	C	T	Y	Hom	Hom	aaccatgaattactcggacagg	aaccatgaattactcggacaga	gccctgaagaattgtatcaaaacag
IWGSC_CSS_2AS_scaff_5268634	Cadenza0110	4636	G	A	Y	Het	Het	gatccatgtgattggcatgtttg	gatccatgtgattggcatgttta	tgctgttggatatgcagttact
IWGSC_CSS_2BL_scaff_7965110	Cadenza0110	15801	C	T	Y	Hom	Hom	cattgaagcatacacaattgcatac	cattgaagcatacacaattgcatat	gccagagtatccagataaggttta
IWGSC_CSS_2DL_scaff_9852812	Cadenza0110	13788	G	A	Y	Hom	Hom	atttttgtatggtctcaatcttcgc	atttttgtatggtctcaatcttcgt	gaacgttcattcttgtacttgct
IWGSC_CSS_2DS_scaff_5371379	Cadenza0110	2166	C	T	Y	Hom	Hom	agacacaaaactagtgatgcgc	agacacaaaactagtgatgcgt	gctgctgagaatgttttgtatttg
IWGSC_CSS_3AL_scaff_4384278	Cadenza0110	1276	C	T	Y	Het	Het	agctgaactgcccctgtag	agctgaactgcccctgtaa	agggacctcggtggatgaa
IWGSC_CSS_3AS_scaff_3340122	Cadenza0110	1467	C	T	Y	Hom	Hom	attcctagtgttgtcggaacatg	attcctagtgttgtcggaacata	gagaagactagaaagttttcagcat
IWGSC_CSS_5DL_scaff_4554222	Cadenza2103	6528	C	T	Y	Hom[1]	Hom	gctgccctacaaagaaacaaaattg	gctgccctacaaagaaacaaaatta	atcccaactatcgattttgtcatac
IWGSC_CSS_6AL_scaff_5833640	Cadenza2103	7346	C	T	Y	Hom	Hom	aagaaaagccacaatggtttctc	aagaaaagccacaatggtttctt	actctgtcagtgtttcccagc
IWGSC_CSS_6AS_scaff_4429974	Cadenza2103	3867	G	A	Y	Hom	Hom	gagatgaatttattgagcatgtggc	gagatgaatttattgagcatgtggt	ggttccggctgcataagt
IWGSC_CSS_6DL_scaff_3307626	Cadenza2103	4970	C	T	Y	Hom	Hom	tgcagatgttgtcctgtgtag	tgcagatgttgtcctgtgtaa	ctaggaaggtgattttgtactgtc
IWGSC_CSS_6DS_scaff_2059604	Cadenza2103	5224	G	A	---[2]	Het	---[2]	gctcaatgcatgctgagtgg	gctcaatgcatgctgagtga	tgtcaagtattattttcctgctctg
IWGSC_CSS_7AL_scaff_4552322	Cadenza2103	1412	C	T	Y	Het	Het	gcaaaggctgatactccaacag	gcaaaggctgatactccaacaa	ggcaagccagtataaaagtaagc
IWGSC_CSS_7BS_scaff_3147455	Cadenza2103	4607	G	A	---[2]	Het	---[2]	gcaccttaggatgtgagttatgc	gcaccttaggatgtgagttatgt	gcatgtagggtttatttgactgtta
IWGSC_CSS_7DL_scaff_3382467	Cadenza2103	3473	C	T	---[2]	Hom	---[2]	ggttctgcagttcataactcatc	ggttctgcagttcataactcatt	attgaatcaactgatacgaagactc
IWGSC_CSS_3B_scaff_10457010	Cadenza0277	10599	G	A	Y	Het	Het	aaccttggccgcagaacac	aaccttggccgcagaacat	actggctgcacgagaggg
IWGSC_CSS_3B_scaff_10593852	Cadenza0277	10124	C	T	Y	Het	Het	tgacaggggacgctatacag	tgacaggggacgctatacaa	gtctaacttacattacccatcagc
IWGSC_CSS_3DS_scaff_2583390	Cadenza0277	663	G	A	Y	Hom	Hom	actgcactcatacaatacttctgc	actgcactcatacaatacttctgt	tccacctggacagcaagtg
IWGSC_CSS_4AL_scaff_7093953	Cadenza0277	10004	C	T	Y	Hom	Hom	ccttgtattcaatggattgttttgg	ccttgtattcaatggattgttttga	ttccccaaataaaaaggaagagc
IWGSC_CSS_4AL_scaff_7176064	Cadenza0277	6220	C	T	Y	Het	Het	gtgccgtattccgcctgg	gtgccgtattccgcctga	atgttcgaggggatgggg
IWGSC_CSS_4DL_scaff_14122349	Cadenza0277	1010	C	T	Y	Hom	Hom	gtcgctgctgcttgtgag	gtcgctgctgcttgtgaa	ggaacaggcccaaggagg

IWGSC contig	Line	Pos.	WT	Mut.	Val.	Pred.	Obs M_4	Primer 1 (Cadenza)	Primer 2 (mutant)	Common Primer
IWGSC_CSS_5AL_scaff_2736916	Cadenza0277	4296	G	A	Y	Het	Het	aagaactatgaaagtaacacacgac	aagaactatgaaagtaacacacgat	ttcgcttttaaggcattctcg
IWGSC_CSS_5BL_scaff_10883744	Cadenza0277	2080	C	T	Y	Hom	Hom	gcctctttctgtttagcctcag	gcctctttctgtttagcctcaa	cgacaaggttcgtgattgca
IWGSC_CSS_1AL_scaff_3932013	Cadenza0548	11765	C	T	Y	Hom	Hom	accgccaacccaagacag	accgccaacccaagacaa	cccattagccgtgcaacg
IWGSC_CSS_1BS_scaff_3417505	Cadenza0548	373	C	T	Y	Het	Het	gtggtgaggagggtggag	gtggtgaggagggtggaa	tggtcggccagttgttga
IWGSC_CSS_2AS_scaff_5305619	Cadenza0548	2786	C	T	Y	Hom	Hom	atacagatgccctaagtggttc	atacagatgccctaagtggttt	ggaagacaatgctccaggtac
IWGSC_CSS_2AS_scaff_5306489	Cadenza0548	46953	T	G	N	Het	**wt**	aggttccatgtccatagaaggt	aggttccatgtccatagaaggg	aggctatagactcctgtacagt
IWGSC_CSS_2BL_scaff_7984123	Cadenza0548	11660	G	A	Y	Het	Het	cattgtggcatagtaatcagtacag	cattgtggcatagtaatcagtacaa	aatacattgaggaatcaaagccc
IWGSC_CSS_2DL_scaff_9907477	Cadenza0548	1363	C	T	Y	Hom	Hom	tgcctccctttgccagaac	tgcctccctttgccagaat	ggcaaacctgatgtggcatc
IWGSC_CSS_2DS_scaff_5330886	Cadenza0548	5449	G	A	Y	Hom	Hom	gcatgtccatttatactgaacgtg	gcatgtccatttatactgaacgta	catgctgcttcttctggacc
IWGSC_CSS_3AL_scaff_4449951	Cadenza0548	633	C	T	Y	Het	Het	tccaaacctaacagtctaacactag	tccaaacctaacagtctaacactaa	gtctgcagtgcaatgtgc
IWGSC_CSS_3B_scaff_10479889	Cadenza0097	3339	C	T	---[2]	Hom	---[2]	ttgtttctggagaagatgccg	ttgtttctggagaagatgcca	ggtgctcattcaacggca
IWGSC_CSS_3B_scaff_10562262	Cadenza0097	7819	C	T	Y	Het	Het	agaggggtgctatccatattgg	agaggggtgctatccatattga	agcgatgccaaggcttcc
IWGSC_CSS_4AL_scaff_7040796	Cadenza0097	10772	G	A	Y	Hom	Hom	acacaacattgccaccagag	acacaacattgccaccagaa	caatcgattgcttgcttctcc
IWGSC_CSS_4AL_scaff_7063488	Cadenza0097	6360	C	T	Y	Het	Het	gcctctcaccttaatttgaagctgc	gcctctcaccttaatttgaagctgt	aggcagtggagtatgtgaagttt
IWGSC_CSS_4AL_scaff_7091701	Cadenza0097	5050	G	A	Y	Het	Het	catgagcatctgggaggaaaatg	catgagcatctgggaggaaaata	agcaagggaataatgaacggaaa
IWGSC_CSS_4DS_scaff_1845841	Cadenza0097	7110	G	A	Y	Hom	Hom	aatgtagctccccataccgg	aatgtagctccccataccga	actgaaactgcaatcgtttatgga
IWGSC_CSS_5AL_scaff_2767581	Cadenza0097	3737	G	A	Y	Het	Het	gagaggtcctcactatcggc	gagaggtcctcactatcggt	cgtcatcacaaatattgctggg
IWGSC_CSS_5BL_scaff_10784643	Cadenza0097	1568	C	T	Y	Hom	Hom	agaaatacatggatggatggacg	agaaatacatggatggatggaca	catctcccttccacggaaag
IWGSC_CSS_1AL_scaff_3952258	Cadenza2092	8107	C	T	---[2]	Het	---[2]	tgagtagagaaattgacagtgtgg	tgagtagagaaattgacagtgtga	tgccaccattgacatgagag
IWGSC_CSS_1BL_scaff_3858008	Cadenza2092	10278	G	A	Y	Hom	Hom	tttgagcaggcaggatcgc	tttgagcaggcaggatcgt	actcacggcctatatcactattc
IWGSC_CSS_1DL_scaff_2265172	Cadenza2092	9094	C	T	Y	Hom	Hom	tgcatgtcatttgttcttatcagc	tgcatgtcatttgttcttatcagt	agtgtccaacttccgttcatc
IWGSC_CSS_2AL_scaff_6435867	Cadenza2092	16201	G	A	Y	Hom	Hom	tttctgtaccttaacgtcaattgac	tttctgtaccttaacgtcaattgat	gtgaggatgatgaggtaagacc
IWGSC_CSS_2AL_scaff_6439430	Cadenza2092	25101	C	T	---[2]	Het	---[2]	caagaaagggcagctcagc	caagaaagggcagctcagt	tcgttactctttcactggtgaa
IWGSC_CSS_2DL_scaff_9760848	Cadenza2092	4733	C	T	Y	Het	Het	gcaccatgggtctcaggtac	gcaccatgggtctcaggtat	tcagtcagtttgctctgtctg
IWGSC_CSS_3AL_scaff_4407012	Cadenza2092	2785	C	T	Y	Hom	Hom	acatatagtgttctcatccaccatc	acatatagtgttctcatccaccatt	acctctctcatgttaataggtttgt
IWGSC_CSS_3AS_scaff_3441108	Cadenza2092	541	G	A	Y	Het	Het	gtgatgaccttgagacggag	gtgatgaccttgagacggaa	aggcatgacaacgcgcaa
IWGSC_CSS_3B_scaff_10449827	Cadenza1551	4779	G	A	Y	Hom	Hom	ggcaaggtcaagaaacggtc	ggcaaggtcaagaaacggtt	acagagtgggttagaggcag
IWGSC_CSS_3B_scaff_10550638	Cadenza1551	3250	C	T	Y	Het	Het	ctccttcacttgttgcggc	ctccttcacttgttgcggt	gcaacattttgatactgcaaagg
IWGSC_CSS_3DL_scaff_6945816	Cadenza1551	589	C	T	Y	Hom	Hom	agcatctcacctgcaacaatac	agcatctcacctgcaacaatat	tgtgccctctgaatattttcatg

IWGSC contig	Line	Pos.	WT	Mut.	Val.	Pred.	M_4	Primer 1 (Cadenza)	Primer 2 (mutant)	Common Primer
IWGSC_CSS_3DL_scaff_6954177	Cadenza1551	3508	C	T	Y	Het	Het	tgtagcatcacattaactttcctg	tgtagcatcacattaactttccta	gcttggtataaacccttacgaca
IWGSC_CSS_4AS_scaff_5938272	Cadenza1551	19080	G	A	Y	Hom	Hom	agaccccgatcgccatgg	agaccccgatcgccatga	gggagatacaggtaaaactcttcg
IWGSC_CSS_4AS_scaff_5977594	Cadenza1551	11092	C	T	Y	Hom[1]	**Het**	gccttgattcggaacaacaaac	gccttgattcggaacaacaaat	gcgtctctcagtcctgca
IWGSC_CSS_5AL_scaff_2671035	Cadenza1551	5859	C	T	Y	Het	Het	cggtgatatttttagacttcgacgc	cggtgatatttttagacttcgacgt	ggcagttcagcgacccatt
IWGSC_CSS_5BL_scaff_10889480	Cadenza1551	2530	G	A	Y	Hom	Hom	gagcttaactcgcagatggag	gagcttaactcgcagatggaa	tccatgcaacgccttggt
IWGSC_CSS_3B_scaff_10528396	Cadenza2088	8059	G	A	---[2]	Hom	---[2]	cttttccgtccgtaagcaatag	cttttccgtccgtaagcaataa	gtgcactgttcaggcctga
IWGSC_CSS_3B_scaff_10637573	Cadenza2088	16815	G	A	Y	Het	Het	agcaagcttaccggtctgc	agcaagcttaccggtctgt	cgagcaactacgagcagctt
IWGSC_CSS_4AL_scaff_7086469	Cadenza2088	6697	G	A	Y	Het	Het	gccgtctacttcaacgcg	gccgtctacttcaacgca	ccagaggcttgttgcattttt
IWGSC_CSS_4AL_scaff_7126302	Cadenza2088	3627	G	A	Y	Hom	Hom	gttcaaaaacaagtggctaatttgc	gttcaaaaacaagtggctaatttgt	cacaaggatatgaagctcttctaga
IWGSC_CSS_4BL_scaff_7041808	Cadenza2088	10234	G	A	Y	Hom	Hom	tcaatggatgagggtgcttc	tcaatggatgagggtgcttt	ccatagcagcatcagccaca
IWGSC_CSS_5AL_scaff_2794167	Cadenza2088	13162	G	A	---[2]	Het	---[2]	agtattcaggacaagcatcttcag	agtattcaggacaagcatcttcaa	caatgaaacctctcgaagaagag
IWGSC_CSS_5BL_scaff_10889232	Cadenza2088	3885	G	A	Y	Het	Het	ctcaaccacaatgggcaaatc	ctcaaccacaatgggcaaatt	tccttcatcaatcatcaattgttgg
IWGSC_CSS_5BS_scaff_2267405	Cadenza2088	11113	C	T	Y	Hom	Hom	ctttgatgatcctaggcctcttg	ctttgatgatcctaggcctctta	tgatttggtctggttagagtttga
IWGSC_CSS_3B_scaff_10475354	Cadenza1409	2203	G	A	Y	Hom	Hom	agcgaacaagaggtcaaacg	agcgaacaagaggtcaaaca	ctgaaacacactagacaattaccg
IWGSC_CSS_3B_scaff_10674115	Cadenza1409	4555	C	T	Y	Het	Het	gcttcagtgcatgccttcag	gcttcagtgcatgccttcaa	cttcacacccgagataatgtattg
IWGSC_CSS_4AL_scaff_7153568	Cadenza1409	13073	C	T	Y	Hom	Hom	tccgaccgatcaaccttgg	tccgaccgatcaaccttga	gaccggaactcctcggcc
IWGSC_CSS_4DL_scaff_14314966	Cadenza1409	2010	G	A	Y	Het	**Hom**	gtaggtcccctcctcaggg	gtaggtcccctcctcagga	cggcgtcacaagttgcct
IWGSC_CSS_4DS_scaff_2324074	Cadenza1409	7606	G	A	Y	Het	Het	tgcatgaaaatgtgtgcagag	tgcatgaaaatgtgtgcagaa	gggtaagttcaaaactgaagtgaag
IWGSC_CSS_5AS_scaff_1517889	Cadenza1409	3561	G	A	Y	Het	Het	tctcgacatcttcccgtgtac	tctcgacatcttcccgtgtat	gtgcctggaacattgcttattta
IWGSC_CSS_5AS_scaff_1523866	Cadenza1409	8054	G	A	---[2]	Hom	---[2]	ggtgatctaccgccaggac	ggtgatctaccgccaggat	tcctgcagcctctcctca
IWGSC_CSS_5BL_scaff_10917655	Cadenza1409	19073	G	A	Y	Hom	Hom	caaatgacatgcaaaagaagttgc	caaatgacatgcaaaagaagttgt	cgcttcatcactacaaaatatgtct
IWGSC_CSS_1AL_scaff_3886649	Cadenza1599	5204	C	T	Y	Het	Het	tgatgccaaccacaatgcc	tgatgccaaccacaatgct	ggactgactgctgaccatatttag
IWGSC_CSS_1BL_scaff_3810267	Cadenza1599	6634	C	T	Y	Hom	Hom	cccaggaaatgagcacctc	cccaggaaatgagcacctt	cgcaggcgaagatgtgattg
IWGSC_CSS_1DL_scaff_2291677	Cadenza1599	12856	C	T	Y	Hom	Hom	ggtagacaagtcgccgag	ggtagacaagtcgccgaa	cctcctccttcaacgccg
IWGSC_CSS_2AL_scaff_6354492	Cadenza1599	7566	G	A	Y	Het	Het	ggagaatgcacagtaacttctgg	ggagaatgcacagtaacttctga	ttccgaagaaccacatcctg
IWGSC_CSS_2AS_scaff_5282937	Cadenza1599	9736	G	A	Y	Het	Het	gctgtagattttatagctgctatgc	gctgtagattttatagctgctatgt	caccagaattgttcactgattttc
IWGSC_CSS_2BL_scaff_7952427	Cadenza1599	19249	G	A	Y	Hom	Hom	cgtccctccctagcacgac	cgtccctccctagcacgat	atcactccattagcgcgag
IWGSC_CSS_2DL_scaff_9897981	Cadenza1599	5627	C	T	Y	Het	Het	cttggtgcttgattgcttactc	cttggtgcttgattgcttactt	gtttgctctctctgatctttgtg
IWGSC_CSS_3AL_scaff_4446105	Cadenza1599	1765	G	A	---[2]	Hom	---[2]	aaatgctttcctaccgctagtg	aaatgctttcctaccgctagta	ttctagaggcaatagcttatatgct

[1] Heterozygous corrected to homozygous by bioinformatics filter applied after MAPS (*SI Appendix*, Figs. S2 and S5).

[2] KASP assays that failed to produce valid cluster.

***SI Appendix*, Table S11.** Weighted scores used to identify residual heterogeneity (RH) regions. All mutations in an interval are classified as RH if the interval score is greater than or equal to the threshold score (12.5). Colors indicate the relative weight of each category: green indicates the lowest weight and red the highest.

Metric	Criteria	Points
Number of mutations	1 mutation	0
	≥ 2 mutations	7
Percent canonical EMS mutations	100%	0
	≥50 and <100%	2.5
	<50% and >0%	3.5
	0%	6
Average number of individuals with the same mutation (MI)	1 MI	0
	>1 and ≤4 MI	0.5
	>4 and <6 MI	3.5
	≥6 MI	4.5
Mutation density	≤ 0.2 density	0
	>.2 and <.4 density	1.5
	≥0.4 and <0.7 density	2
	≥0.7 density	3
Combined	%hom ≥ 0.25 and %EMS ≤ 0.75	12.5
	%hom ≥ 0.25, %EMS >0.75 & < 100%, and MI≥ 4	12.5
	Threshold	**12.5**

***SI Appendix*, Table S12**. Uniquely mapped mutations in RH regions of tetraploid wheat Kronos and hexaploid wheat Cadenza at different levels of stringency. Default values at *HetMC5/HomMC3* are indicated by bold type.

Coverage level	No. lines	No. SNPs in RH	%RH[1]	Het/ Hom	EMS SNPs	Avg. EMS SNPs/line	%EMS
Kronos							
HetMC3/HomMC2	1,535	312,725	5.66%	0.94	53,483	35	17.1%
HetMC4/HomMC3	1,534	86,528	1.88%	0.64	21,072	14	24.4%
HetMC5/HomMC3	**1,530**	**69,651**	**1.66%**	**0.33**	**16,412**	**11**	**23.6%**
HetMC6/HomMC4	1,517	62,640	1.66%	0.27	14,747	10	23.5%
Cadenza							
HetMC3/HomMC2	1,200	269,300	3.13%	1.05	50,350	224	18.7%
HetMC4/HomMC3	1,200	62,689	0.87%	0.99	12,744	52	20.3%
HetMC5/HomMC3	**1,199**	**38,626**	**0.60%**	**0.30**	**6,023**	**32**	**15.6%**
HetMC6/HomMC4	1,199	30,980	0.53%	0.21	4,637	26	15.0 %

[1] Number of SNPs in RH divided by the total number of SNPs (*SI Appendix*, Tables S5 and S6).

***SI Appendix*, Table S13.** Approximate proportion of available G/C sites[1] affected by EMS mutagen in tetraploid wheat (unique EMS-type mutations without RH) [2].

No. ind. with same mutation	Poisson prediction 4X using 100% GC sites to calculate mutation prob.	Poisson prediction 4X using **18%** of GC sites to calculate mutation prob.	**Observed tetraploid wheat**
2	8427	237489	248109
3	185	32971	22545
4	3	3433	2916
5	0	286	455
6	0	20	125
7	0	1	51

[1] GC% of captured space = 46.8%.
[2] See *SI Appendix*, Method S6

***SI Appendix*, Table S14.** Approximate proportion of available G/C sites [1] affected by EMS mutagen in hexaploid wheat (unique EMS-type mutations without RH)[2].

No. ind. with same mutation	Poisson prediction 6X using 100% GC sites to calculate mutation prob.	Poisson prediction 6X using **21%** of GC sites to calculate mutation prob.	**Observed hexaploid wheat**
2	16643	354903	372987
3	415	48223	31101
4	8	4914	3642
5	0	401	590
6	0	27	105
7	0	2	25

[1] GC% of captured space = 46.8%.
[2] See *SI Appendix*, Method S6.

***SI Appendix*, Table S15**: Validated multi-map mutations in Kronos and Cadenza M_4 families.

Scaffold	chr	Mutant	Pos.	WT	Mut.	Predicted	Sanger	Left primer	Right primer
IWGSC_CSS_2AL_scaff_6322392	2AL	Cadenza0900	3530	C	T	Hom	Hom	agagacatccagggttcggt	ccgctgtaatgcaggaatgc
IWGSC_CSS_3AL_scaff_4358608	3AL	Cadenza2088	1876	C	T	Het	Het	tggaagccacaattcgtgtc	caccgatcattagttggtacacc
IWGSC_CSS_1DS_scaff_1899554	1DS	Cadenza1538	4559	C	T	Hom	Het	acgtgagtaggttccaatcaaa	acttgctcttttcaggattgtgt
IWGSC_CSS_3DS_scaff_2603373	3DS	Cadenza0110	4858	G	A	Het	Het	tttgcgcgctgtttgtttct	gccaaaaatattgccatggctaatc
IWGSC_CSS_2DS_scaff_5320193	2DS	Cadenza0281	3969	G	A	Het	Het	acatctatctagctatgacatcgc	tccctatagcacaatagccttttat
IWGSC_CSS_4AL_scaff_7098096	4AL	Cadenza1469	2806	G	A	Het	Het	ccagaaagaactaatatctggcgat	gtgaatgccactgactaaaaagac
IWGSC_CSS_2AL_scaff_6407808	2AL	Cadenza0423	5356	C	T	Het	Het	tctgacaacacttccttgagtt	gcagtggagaaaccggaaaaa
IWGSC_CSS_7DS_scaff_3963303	7DS	Cadenza0866	3989	G	A	Hom	Hom	ggtatcttctgcattgtcgttgtt	tgcatgtgaatgtgaagctgg
IWGSC_CSS_6AL_scaff_5769446	6AL	Cadenza0580	989	C	T	Het	Het	cttcttccgggaccttccac	catcgacaagacagtgggct
IWGSC_CSS_2BS_scaff_5186503	2BS	Cadenza0995	13502	G	A	Het	Het	taaaggtttgccgaatgcgc	catccccatgtgtccgaaca
IWGSC_CSS_2AL_scaff_6385463	2AL	Cadenza0423	861	C	T	Het	wildtype	gcactgcttctgatgtcacc	gctgtctttgccatctccttc
IWGSC_CSS_2AS_scaff_5300197	2AS	Cadenza1538	11300	G	A	Het	wildtype	tccaatgagcagaacaatgtcatc	tcatccctgagcttagaagaatttt
IWGSC_CSS_7DL_scaff_3344608	7DL	Cadenza0110	8177	G	A	Het	wildtype	tcagaagacccatagctcctattg	tgacagcttctgggtaatgatct
IWGSC_CSS_4BL_scaff_6981190	4BL	Kronos3085	8282	G	A	Hom	Hom	tccacatagctatttgtgtttgaca	gcgtttaccctctggccac
IWGSC_CSS_2AL_scaff_6327190	2AL	Kronos3288	17196	C	T	Het	Het	agggattaatcgacaagataactgg	acacttcaccggatgaatcatct
IWGSC_CSS_3B_scaff_10740364	3B	Kronos 3825	890	G	A	Hom	Hom	gacacctgatgttaatgctat	caggcaagaaagatggaagg
IWGSC_CSS_2BL_scaff_8054105	2BL	Kronos1096	7445	G	A	Hom	Het	ctccattccaatcatcaccagc	cgatttcaatgttctgaaaaagctt
IWGSC_CSS_3AS_scaff_3416436	3AS	Kronos1194	4471	C	T	Het	Het	tcgaacttgaacttgctagaaaca	gttgtgcttatgaatggactcaata
IWGSC_CSS_2AS_scaff_5294902	2AS	Kronos1194	8985	C	T	Hom	Hom	accaaaccataccaaatcagacg	acaagaacccaaaggcagtag
IWGSC_CSS_2BS_scaff_5209080	2BS	Kronos1344	15239	C	T	Hom	Hom	tgtcaataaaggttaattagctgcc	gcgcctcccaggattacaa
IWGSC_CSS_2BL_scaff_7974262	2BL	Kronos3126	4231	G	A	Hom	Hom	accctctagagtacgaacaaca	tagccgggttctgactacag
IWGSC_CSS_2BL_scaff_7948635	2BL	Kronos3191	1026	C	T	Het	Het	accaagctttgtcaaaattaaccag	gctactaaggcattacttgacca
IWGSC_CSS_4AS_scaff_5942330	4AS	Kronos3191	5456	C	T	Hom	Hom	gtcatttttggtggatgtcagaag	cgactgagatttaggctcctatcat
IWGSC_CSS_5BL_scaff_10882912	5BL	Kronos4346	2007	G	A	Het	Het	tggaatggggttgatgaaacag	ccgtatttaacatgcattccacatc
IWGSC_CSS_2AL_scaff_6319489	2AL	Kronos4346	8848	C	T	Hom	Hom	gttttatgccctcactaaggtttat	ttgttgctgcatcgagtgtg

***SI Appendix*, Table S16.** Small indels identified by MAPS.

Coverage	Mutation set	# lines	Indel SNPs	# indels
Kronos				
HetMC3/HomMC2 [1]	Non-RH [2]	1,535	2,952	2,947
	RH-only	1,535	12,495	7,702
HetMC4/HomMC3	Non-RH	1,535	931	928
	RH-only	1,534	2,197	1,179
HetMC5/HomMC3 [3]	**Non-RH**	**1,535**	**618**	**616**
	RH-only	1,532	1,853	1,046
HetMC6/HomMC4	Non-RH	1,535	501	500
	RH-only	1,520	1,355	813
Cadenza				
HetMC3/HomMC2	Non-RH	1,209	12,027	11,945
	RH-only	1,209	33,992	11,856
HetMC4/HomMC3	Non-RH	1,209	3,270	3,249
	RH-only	1,209	8,309	2,245
HetMC5/HomMC3	**Non-RH**	**1,209**	**1,274**	**1,268**
	RH-only	1,208	4,408	1,041
HetMC6/HomMC4	Non-RH	1,209	960	955
	RH-only	1,207	3,082	690

[1] *MC*= minimum coverage.

[2] RH= residual heterogeneity.

[3] Default stringency level *HetMC5/HomMC3* used in the main text is indicated in bold.

***SI Appendix*, Table S17.** Primers and validation of small deletion in Kronos and Cadenza M_4 mutants (called by MAPS).

Mutant	Target region	Expected Sequence (deletion in CAPS)	Val.	Zyg.	Left Primer	Right Primer
Kronos2273	IWGSC_CSS_1AL_scaff_3977540:4559-4759:G100c	aaaaa**G**cgaga	Yes	Het	cgggtgtgaaactgtgatatctg	ggaatggggaaagggcaatttc
Kronos2398	IWGSC_CSS_3AS_scaff_3294685:2381-2581:C100g	aaaga**C**actac	Yes	Hom	ccgttggaccaaacagagga	aaacaggtgatgccgtagct
Kronos2398	IWGSC_CSS_7BS_scaff_3074057:6265-6465:A100t	taaac**A**tagcc	Yes	Hom	tgcttggccacttgctctt	cttcaaggactggatttgactgt
Kronos2421	IWGSC_CSS_4BS_scaff_4908743:13536-13736:G100c	ttgac**G**aaagc	Yes	Het	agcgccatatactgcagcat	gccatgcactctggtgtagt
Kronos2644	IWGSC_CSS_5BL_scaff_10868431:4277-4477:G100c	aaaaa**G**ctaac	Yes	Het	ttagacagccatggcaagca	aaagcctcccaactcaccag
Kronos2644	IWGSC_CSS_5BS_scaff_2240909:6748-6948:C100g	cctga**C**gatgt	Yes	Hom	caggtatgatatctatgcgccttaa	gccttcattgttgtctcagtgc
Kronos3085	IWGSC_CSS_5AL_scaff_2706605:6811-7011:G100c	aaaaa**G**atggg	Yes	Het	cccagtagatagaggtgtgctatc	gcctcagtaaaggatggatttcttt
Kronos3288	IWGSC_CSS_2AS_scaff_5201411:5980-6182:T100a	gaatt**TTC**actcc	Yes	Het	cgacaatccgccgaattagaa	gcgggtgcattcttcttcatc
Kronos598	IWGSC_CSS_3AS_scaff_3295530:920-1120:C100g	catct**C**ttttt	Yes	Hom	tgtgttctccaagggtgagc	aggaaccaaggcctttcgtt
Kronos640	IWGSC_CSS_7BL_scaff_6668812:13070-13270:C100g	cttttt**C**agttt	Yes	Het	tccttctcacggtccttctca	tgctgacatctcatccaaacataat
Cadenza0110	IWGSC_CSS_2BL_scaff_8076667:2335-2535:C100g	cgatt**C**cccgg	No	-	agctgcatgcatgttcattgg	acctacaagaaactgagggaaata
Cadenza0148	IWGSC_CSS_5DS_scaff_2289990:9895-10095:C100g	accgt**C**tttta	Yes	Het	gggaaactccgaaattcagaagttt	tacaggtctggttctcgcaa
Cadenza0580	IWGSC_CSS_4AL_scaff_7167665:8070-8270:C100g	atttg**G…C**tcttc[1]	Yes	Het	tttacaatgcaaacggaatcctt	gtcactgtatttacttgggttaact
Cadenza0580	IWGSC_CSS_5DL_scaff_4537294:7891-8091:A100t	agacc**AACAT**acatg	Yes	Hom	ggaataggcagtaaggtttgtgg	ccaaacaagtagtcaccacatgtc
Cadenza1661	IWGSC_CSS_2DL_scaff_9850151:1008-1208:g100c	aaaaa**G**ttagg	Yes	Hom	tgtcctttggtttgccacaaa	tctttttggatgtgcacagga

[1] **GAGCATAAACTTCCGTATC**

***SI Appendix*, Table S18:** Summary of large homozygous deletions in Kronos and Cadenza populations.

	Kronos		Cadenza	
Mutant lines analysed ($sd(\text{XNORM}_j) \leq 0.3$)	1,494		1,011	
Total deletions (scaffolds x mutant lines)	870		7,971	
Lines with no deletions	1,379	(92.3%)	718	(71.0%)
Lines with at least 1 deletion	115	(7.7%)	293	(29.0%)
Between 1 and 10 scaffolds deleted	87	(75.7%)	165	(56.3%)
Greater than 10 scaffolds deleted	28	(24.3%)	128	(43.7%)
Scaffolds with 5 or more exons	15,629		19,191	
Scaffolds with at least 1 deletion	785	(5.0%)	5,433	(28.3%)
Scaffolds deleted in a single mutant line	722	(92.0%)	3,872	(71.3%)
Scaffolds deleted in two mutant lines	42	(5.4%)	1,248	(23.0%)
Scaffolds deleted in 3 or more mutants	21	(2.7%)	313	(5.8%)
Genes deleted in large deletions	832		6,657	

***SI Appendix*, Table S19**: Characteristics of large deletions validated using KASP assays.

Mutant Line		Scaffold	Independent assays	Valid Exons	*3Sigma Del*	Scaffold Score	Relative coverage exons in scaffold	Del. per scaffold	Del. per mut.
Cadenza0423	1DL	IWGSC_CSS_1DL_scaff_2208937	2	12	11	0.92	0.05	1	18
	5BL	IWGSC_CSS_5BL_scaff_10847976	2	14	13	0.93	0.02	1	18
	5BL	IWGSC_CSS_5BL_scaff_10865441	1	10	9	0.90	0.03	1	18
Cadenza0580	2AL	IWGSC_CSS_2AL_scaff_6354142	2	11	9	0.82	0.04	1	9
	2AL	IWGSC_CSS_2AL_scaff_6358197	1	18	17	0.94	0.01	1	9
	4BS	IWGSC_CSS_4BS_scaff_4879496	2	10	8	0.80	0.05	1	9
Kronos1017	1AL	IWGSC_CSS_1AL_scaff_3976129	3	6	6	1.00	0.00	1	1
Kronos376	3B	IWGSC_CSS_3B_scaff_10571389	1	7	7	1.00	0.00	1	64
	3B	IWGSC_CSS_3B_scaff_10412973	1	11	10	0.91	0.00	1	64
	3B	IWGSC_CSS_3B_scaff_10520549	1	9	7	0.78	0.02	1	64
Kronos682	2BL	IWGSC_CSS_2BL_scaff_7986332	1	10	10	1.00	0.02	1	7

***SI Appendix*, Table S20. Variant Effect Prediction (VEP).** Variant effect predictions for Kronos and Cadenza uniquely mapped EMS-type mutations (excluding RH regions).

Type	4x Kronos	6x Cadenza
splice_donor_variant	15,074	26,783
splice_acceptor_variant	14,624	20,889
stop_gained	46,580	85,985
initiator_codon_variant	943	1,953
missense_variant	1,030,287	1,668,693
splice_region_variant[1]	84,890	146,929
synonymous_variant	550,556	871,675
coding_sequence_variant[2]	33	30
5_prime_UTR_variant	85,448	133,732
3_prime_UTR_variant	146,474	251,644
intron_variant	676,648	1,228,055
stop_lost	0	53

[1] Splice_region_variant includes mutations within 1-3 bases of the exon or 3-8 bases of the intron border.

[2] Coding_sequence_variant includes non-coding variants or variants affecting non-coding genes, where predictions are difficult or there is no evidence of impact. Many of these regions have mutations in codons that have at least one ambiguous base in the reference.

***SI Appendix*, Table S21**. Summary of variant effects per available gene models with at least one mutation.

	Kronos	**(%)**	**Cadenza**	**(%)**
Valid gene models with at least 1 mutation (GM_1) [1]	48,172		73,895	
Genes models with at least one truncation	28,604	(59%)	45,311	(61%)
Premature stop codon	22,536	(47%)	37,452	(51%)
Mutant acceptor or donor splice site	16,000	(33%)	24,909	(34%)
Avg. number of truncations per GM_1	**1.58**		**1.81**	
Additional GM_1 with at least one splice_region_variant	5820	(12%)	7742	(10%)
Average number of splice_region_variant	**1.76**		**1.99**	
GM_1 with at least one missense mutations	46,198	(96%)	69,543	(94%)
Avg. No. missense mutations per GM_1	**21.41**		**22.61**	
GM_1 with initiator_codon_variant	907	(2%)	1,862	(3%)
GM_1 with at least one deleterious missense allele [2]	40,913	(85%)	66,734	(90%)
GM_1 with at least one truncation and/or deleterious missense allele	43,787	(91%)	67,830	(92%)

[1] If we add genes that only have mutations downstream or upstream of the gene the total count is 50,258.

[2] Predicted deleterious missense mutations by SIFT (<0.05).

***SI Appendix*, Table S22**. Mutations in genes from the starch biosynthesis pathway. Chromosome locations are shown in *SI Appendix*, Fig. S12.

ID	Gene	IWGSC Scaffold	Gene Model	Kronos (AB)			Cadenza (ABD)		
				Syn.	Miss.	Trun.	Syn.	Miss.	Trun.
1	*Sucrose synthase*	IWGSC_CSS_6AL_scaff_5742333	Traes_6AL_02ECEFFB7	32	71	10	10	19	3
		IWGSC_CSS_6BL_scaff_4220015	Traes_6BL_FC54B7E8F	28	44	6	Deleted in Cadenza		
		IWGSC_CSS_6DL_scaff_3210843	Traes_6DL_3C0C05516	-	-	-	63	96	10
2	*Sucrose synthase*	IWGSC_CSS_4AS_scaff_5984059	Traes_4AS_EF48BBCCF	42	56	5	39	64	4
		IWGSC_CSS_4BL_scaff_7007100 IWGSC_CSS_4BL_scaff_6986922	Traes_4BL_B49C2A51C Traes_4BL_D1AA966DE	29	30	3	29	43	5
		IWGSC_CSS_4DL_scaff_14460089 IWGSC_CSS_4DL_scaff_14349308	Traes_4DL_9F7C8C343 Traes_4DL_F08462AD2	-	-	-	36	50	7
3	*Sucrose synthase*	IWGSC_CSS_7AS_scaff_4255448 IWGSC_CSS_7AS_scaff_4256374	Traes_7AS_5D84FA56B Traes_7AS_A80BE362A	Deleted in Kronos			32	45	3
		IWGSC_CSS_4AL_scaff_2713402	Traes_4AL_2BC235062	5	8	1	46	87	8
		IWGSC_CSS_7DS_scaff_3920687	Traes_7DS_7094F3B4D	-	-	-	63	94	3
4	*Sucrose synthase*	IWGSC_CSS_2AL_scaff_6333913	Traes_2AL_8E8343BA5	36	72	3	57	71	6
		IWGSC_CSS_2BL_scaff_8026696	Traes_2BL_C963272C8	38	62	9	50	57	11
		IWGSC_CSS_2DL_scaff_9853726	Traes_2DL_22482812B	-	-	-	45	82	9
5	*Sucrose synthase*	IWGSC_CSS_7AL_scaff_4536617	Traes_7AL_3F2C16688	41	57	5	39	55	5
		IWGSC_CSS_7BL_scaff_6751305	Traes_7BL_FBEAF8C41	42	57	7	37	64	6
		IWGSC_CSS_7DL_scaff_3364136	Traes_7DL_471B4134B	-	-	-	40	56	4
6	*Sucrose synthase*	IWGSC_CSS_7AS_scaff_4255196	Traes_7AS_2742DDF6C	42	41	5	42	67	3
		IWGSC_CSS_7BS_scaff_3131167	Traes_7BS_182F2A1F1	44	44	5	39	59	5
		IWGSC_CSS_7DS_scaff_3893149	Traes_7DS_529BAB150	-	-	-	50	65	9
7	*Sucrose synthase*	IWGSC_CSS_2AS_scaff_5216970	Traes_2AS_F2967D6F7	55	66	10	53	71	11
		IWGSC_CSS_2BS_scaff_5178880	Traes_2BS_96ECE84C2	51	56	6	43	58	7
		IWGSC_CSS_2DS_scaff_5360578	Traes_2DS_ECBFB4D8C	-	-	-	56	76	4
8	*Fructokinase*	IWGSC_CSS_3AL_scaff_2780612	Traes_3AL_7FE1083A4	5	19	2	13	14	2
		3B	TRAES3BF078000040CFD_g[1]	1	10	3	12	26	5
		IWGSC_CSS_3DL_scaff_6826059	Traes_3DL_CBD8FCDB3	-	-	-	7	17	3
9	*Fructokinase*	IWGSC_CSS_5AL_scaff_2806837	Traes_5AL_35901C90B	7	12	1	9	12	3
		IWGSC_CSS_5BL_scaff_10818071	Traes_5BL_62C79EBEC	19	20	3	19	27	2
		IWGSC_CSS_5DL_scaff_4606422	Traes_5DL_3B71B1E8C	-	-	-	10	17	2
10	*Fructokinase*	IWGSC_CSS_5AL_scaff_2740220	Traes_5AL_ADBCA73F2	26	27	0‡	63	68	1
		IWGSC_CSS_5BL_scaff_10835859	Traes_5BL_9789C37DD	4	10	0‡	20	27	0‡
		IWGSC_CSS_5DL_scaff_4530232	Traes_5DL_C538965B2	-	-	-	5	17	0
11	*Fructokinase*	IWGSC_CSS_7AL_scaff_4480746	Traes_7AL_3B2536995	6	10	1	10	12	1
		IWGSC_CSS_5BL_scaff_10796109	Traes_5BL_91FBE67D3	10	6	0‡	6	11	0‡
		IWGSC_CSS_7DL_scaff_3394360	Traes_7DL_1A847FDCC	-	-	-	25	33	0‡
12	*PGI*	IWGSC_CSS_1AS_scaff_990437	Traes_1AS_682A666AE	14	23	7	24	47	9
		IWGSC_CSS_1BS_scaff_3469276	Traes_1BS_757804D58	9	11	4	15	36	7
		IWGSC_CSS_1DS_scaff_1910343	Traes_1DS_0435E9A1F	-	-	-	17	49	7

13	*PGI*	IWGSC_CSS_5AL_scaff_2774618	Traes_5AL_D3B6FE48E	12	33	5	8	31	1
		IWGSC_CSS_5BL_scaff_10809883	Traes_5BL_A2B216782	16	46	5	21	55	6
		IWGSC_CSS_5DL_scaff_4537294	Traes_5DL_BD1C8E19E	-	-	-	18	29	4
14	*UGPase*	IWGSC_CSS_5AL_scaff_2769163	Traes_5AL_E97939490	19	32	4	16	32	6
		IWGSC_CSS_5BL_scaff_10822911	Traes_5BL_AEEB6621B	10	34	2	7	9	2
		IWGSC_CSS_5DL_scaff_4567798	Traes_5DL_CFFABFAA6	-	-	-	26	32	5
15	*UGPase*	IWGSC_CSS_6AS_scaff_4340916	Traes_6AS_3A8E07254	14	24	1	40	55	6
		IWGSC_CSS_6BS_scaff_2504270	Traes_6BS_0AFE47E4B	20	27	3	16	34	4
16	*AGPase (cyt.)*[2]	IWGSC_CSS_1AL_scaff_3947987	Traes_1AL_A1B2A8EB0	28	46	3	38	66	2
		IWGSC_CSS_1BL_scaff_3798340	Traes_1BL_190920E1E	17	38	3	27	41	1
		IWGSC_CSS_1DL_scaff_2268051	Traes_1DL_844FE40E6	-	-	-	36	48	7
17	*AGPase (cyt.)*[2]	IWGSC_CSS_7AS_scaff_4256551	Traes_7AS_1B2A8C929	13	46	3	6	26	5
		IWGSC_CSS_7BS_scaff_3108101	Traes_7BS_4FBE4B00A	14	44	3	9	31	3
		IWGSC_CSS_7DS_scaff_3968428	Traes_7DS_02539EB3B	-	-	-	13	36	3
18	*PGM*	IWGSC_CSS_3AL_scaff_4429950	Traes_3AL_7C9E9BA54	13	34	1	13	32	5
		3B	TRAES3BF090300070CFD_g[3]	14	25	4	33	71	5
		IWGSC_CSS_3DL_scaff_6953716	Traes_3DL_3A0E67ECD	-	-	-	12	31	3
19	*PGM*	IWGSC_CSS_3AL_scaff_4350138	Traes_3AL_A3B9A7009	4	15	1	11	14	1
		IWGSC_CSS_3DL_scaff_6951977	Traes_3DL_CA803F9CA	-	-	-	1	9	0‡
20	*PGM*	IWGSC_CSS_1AL_scaff_3915240	Traes_1AL_928DC3A3C	5	7	0‡	5	14	3
		IWGSC_CSS_1BL_scaff_3865511	Traes_1BL_1D467B1C2	2	1	0‡	1	2	0
		IWGSC_CSS_1DL_scaff_2205719	Traes_1DL_1369670F9	-	-	-	10	22	3
21	*PGM*	IWGSC_CSS_7BS_scaff_3101165	Traes_7BS_FE38B3165	2	16	0‡	7	15	0‡
		IWGSC_CSS_7DS_scaff_3874546	Traes_7DS_41AC2C383	-	-	-	Deleted in Cadenza		
22	*ADPG transporter*[4]	IWGSC_CSS_6AS_scaff_2894486	Traes_6AS_F67BFB2A5	41	73	2	34	56	3
		IWGSC_CSS_6DS_scaff_213852	Traes_6DS_CA5464FC5	-	-	-	12	22	0‡
23	*AGPase (plast.)*[5]	IWGSC_CSS_5AL_scaff_260268	Traes_5AL_AE2EF1F20	21	36	6	15	37	3
		IWGSC_CSS_5BL_scaff_10921868	Traes_5BL_3212A5875	12	28	0‡	9	28	2
		IWGSC_CSS_5DL_scaff_4489059	Traes_5DL_CB059AB1D	-	-	-	12	49	9
24	*AGPase (plast.)*[6]	IWGSC_CSS_5AL_scaff_2692522	Traes_5AL_92651A012	16	50	5	19	37	6
		IWGSC_CSS_5BL_scaff_3669803	Traes_5BL_28478C1D3	14	40	2	16	50	4
		IWGSC_CSS_5DL_scaff_4530451	Traes_5DL_FC5E1C178	-	-	-	12	48	4
25	*G6P-Pi translocator*	IWGSC_CSS_2AS_scaff_5307747	Traes_2AS_7B612C346	14	18	3	12	20	5
		IWGSC_CSS_2BS_scaff_5247592	Traes_2BS_94C42CB70	19	25	4	26	32	4
		IWGSC_CSS_2DS_scaff_5388479	Traes_2DS_FCCFA0BE1	-	-	-	23	32	1
26	*G6P-Pi translocator*	IWGSC_CSS_7AS_scaff_4204384	Traes_7AS_0B0DC9CE5	12	28	2	23	36	1
		IWGSC_CSS_7BS_scaff_3089122	Traes_7BS_E7BBA5276	20	34	0‡	17	29	3
		IWGSC_CSS_7DS_scaff_3898661	Traes_7DS_485E7E61C	-	-	-	15	33	2
27	*Starch PPase*	IWGSC_CSS_3AL_scaff_4446000 IWGSC_CSS_3AL_scaff_4435316	Traes_3AL_C40EC1F8D Traes_3AL_184DD77F2	16	40	6	22	52	9
		IWGSC_CSS_3DL_scaff_6956147	Traes_3DL_FFCCD5827	-	-	-	13	25	6
28	*Starch PPase*	IWGSC_CSS_3AL_scaff_4401664 IWGSC_CSS_3AL_scaff_4418482	Traes_3AL_9A3B8E4D9 Traes_3AL_3ACAED752	11	31	5	17	49	11
		IWGSC_CSS_3DL_scaff_6930332	Traes_3DL_468A9D15B	-	-	-	10	21	4

29	*Starch PPase*	IWGSC_CSS_5AL_scaff_2749928	Traes_5AL_ED21C722B	37	59	4	29	72	8
		IWGSC_CSS_5BL_scaff_10925003	Traes_5BL_CA33BB947	29	57	9	33	85	10
		IWGSC_CSS_5DL_scaff_297369	Traes_5DL_64F27292C	-	-	-	24	72	3
30	*GBSSI*[7]	IWGSC_CSS_4AL_scaff_7104101	Traes_4AL_4B9D56131	10	10	0‡	7	9	2
		IWGSC_CSS_7DS_scaff_1340123 IWGSC_CSS_7DS_scaff_3930191	Traes_7DS_60EF68D7A Traes_7DS_B89DCC51A	-	-	-	2	6	1
31	*GBSSII*[7]	IWGSC_CSS_2AL_scaff_6389371	Traes_2AL_06AD35739	17	27	4	43	96	7
		IWGSC_CSS_2BL_scaff_8026821	Traes_2BL_25DAD57C9	13	67	6	25	63	6
		IWGSC_CSS_2DL_scaff_9867583	Traes_2DL_E4B86D0C3	-	-	-	16	63	6
32	*SSI*[8]	IWGSC_CSS_7AS_scaff_4250052 IWGSC_CSS_7AS_scaff_4250052	Traes_7AS_0680C7277 Traes_7AS_70ED86B15	20	62	6	24	66	7
		IWGSC_CSS_7BS_scaff_3069694 IWGSC_CSS_7BS_scaff_3108170	Traes_7BS_5DB99DA42 Traes_7BS_6135B1D85	11	31	4	23	49	5
		IWGSC_CSS_7DS_scaff_3940620 IWGSC_CSS_7DS_scaff_3945994	Traes_7DS_5159E3934 Traes_7DS_854A1DA29	-	-	-	27	41	5
33	*SSIIa*[9]	IWGSC_CSS_7AS_scaff_4251385	Traes_7AS_53CAFB43A	33	24	2	15	40	3
		IWGSC_CSS_7BS_scaff_3162407	Traes_7BS_7BEAF5EC0	3	12	2	12	13	1
		IWGSC_CSS_7DS_scaff_3877787	Traes_7DS_E6C8AF743	-	-	-	16	53	4
34	*SSIIb*[10]	IWGSC_CSS_6AL_scaff_5749784	Traes_6AL_AE01DC0EA	26	56	6	28	56	3
		IWGSC_CSS_6BL_scaff_4224185	Traes_6BL_61D83E262	8	20	2	7	22	2
		IWGSC_CSS_6DL_scaff_3304909 IWGSC_CSS_6DL_scaff_3265653	Traes_6DL_060A32386 Traes_6DL_19F1042C7	-	-	-	24	40	1
35	*SSIIc*[10]	IWGSC_CSS_1AL_scaff_3930028	Traes_1AL_729BF3204	8	28	0‡	8	16	3
		IWGSC_CSS_1BL_scaff_3799401	Traes_1BL_447468BDE	17	35	1	15	30	0‡
		IWGSC_CSS_1DL_scaff_2205619	Traes_1DL_F667ED844	-	-	-	16	34	6
36	*SSIIIa*[11]	IWGSC_CSS_1AS_scaff_3291252 IWGSC_CSS_1AS_scaff_3285944	Traes_1AS_4499E3652 Traes_1AS_83A34BFC8	46	146	17	55	165	8
		IWGSC_CSS_1BS_scaff_2553401 IWGSC_CSS_1BS_scaff_2177025 IWGSC_CSS_1BS_scaff_3468420	Traes_1BS_AF95964AE Traes_1BS_0AECE1698 Traes_1BS_7EC2BDE95	25	44	4	31	56	5
		IWGSC_CSS_1DS_scaff_1898084	Traes_1DS_1B53199CF	-	-	-	63	141	13
37	*SSIIIb*[12]	IWGSC_CSS_2AL_scaff_6321976 IWGSC_CSS_2AL_scaff_3369196 IWGSC_CSS_2AL_scaff_1340591 IWGSC_CSS_2AL_scaff_6325346 IWGSC_CSS_2AL_scaff_6411556	Traes_2AL_66A17401A Traes_2AL_627154C03 Traes_2AL_C787D1A97 Traes_2AL_A0E1D05F1 Traes_2AL_98A9B764F	20	49	6	15	53	6
		IWGSC_CSS_2BL_scaff_7955846	Traes_2BL_93191B70D	9	28	2	13	20	4
		IWGSC_CSS_2DL_scaff_9849597	Traes_2DL_6E75354B9	-	-	-	17	60	6
38	*SSIVb*[13]	IWGSC_CSS_1AL_scaff_3877961	Traes_1AL_5E9E6239D	27	46	4	22	48	6
		IWGSC_CSS_1BL_scaff_3896368	Traes_1BL_709B74768	26	44	3	26	60	10
		IWGSC_CSS_1DL_scaff_2258883	Traes_1DL_B0409DEFD	-	-	-	21	60	3
39	*SBEI*[14]	IWGSC_CSS_7AL_scaff_4552612	Traes_7AL_AA28D70BF	37	64	3	32	81	7
		IWGSC_CSS_7BL_scaff_6702230	Traes_7BL_11C5C7BC4	21	37	2	16	51	5
		IWGSC_CSS_7DL_scaff_3394511	Traes_7DL_50609FDAB	-	-	-	34	73	10
40	*SBEI*[14]	IWGSC_CSS_7AL_scaff_4554621 IWGSC_CSS_7AL_scaff_4554621	Traes_7AL_B17CA28E3 Traes_7AL_2096BB4DF	43	52	3	26	71	5
		IWGSC_CSS_7BL_scaff_6750609 IWGSC_CSS_7BL_scaff_1748736	Traes_7BL_5AFBD701D Traes_7BL_92676B214	9	31	4	7	25	3

		IWGSC_CSS_7DL_scaff_3384282	Traes_7DL_D39A0193F	-	-	-	13	35	8
41	*SBEIIa*[15]	IWGSC_CSS_2AL_scaff_6434929	Traes_2AL_CC968FC52	18	73	6	25	79	12
		IWGSC_CSS_2BL_scaff_8026237 IWGSC_CSS_2BL_scaff_8035556	Traes_2BL_A3FAE4AE7 Traes_2BL_0E1E397A8	13	34	2	13	51	4
		IWGSC_CSS_2DL_scaff_9906577	Traes_2DL_647B61E84	-	-	-	16	62	11
42	*SBEIIb*[16]	IWGSC_CSS_2AL_scaff_6316048 IWGSC_CSS_2AL_scaff_6427577	Traes_2AL_FE11B55BA Traes_2AL_77043B634	6	14	5	16	17	3
		IWGSC_CSS_2BL_scaff_7939720 IWGSC_CSS_2BL_scaff_7983892	Traes_2BL_2C62185EE Traes_2BL_5B0460EA4	31	54	8	22	59	11
		IWGSC_CSS_2DL_scaff_9908528 IWGSC_CSS_2DL_scaff_9908051 IWGSC_CSS_2DL_scaff_8191185	Traes_2DL_6CA449145 Traes_2DL_FBE8DBDCF Traes_2DL_5A1CE8406	-	-	-	28	89	4
43	*SBEIII*[17]	IWGSC_CSS_7AL_scaff_4436291 IWGSC_CSS_7AL_scaff_4555306	Traes_7AL_5D8D7CB60 Traes_7AL_1E9FBDD47	20	41	8	12	45	3
		IWGSC_CSS_7BL_scaff_6642359 IWGSC_CSS_7BL_scaff_6642050	Traes_7BL_06EE82124 Traes_7BL_934073AA8	8	16	4	3	15	4
		IWGSC_CSS_7DL_scaff_3351544 IWGSC_CSS_7DL_scaff_3393537	Traes_7DL_E638906EF Traes_7DL_8368AB176	-	-	-	10	15	3
44	*ISAI*[18]	IWGSC_CSS_7AS_scaff_4100902	Traes_7AS_7C6D1C2D3	25	43	6	29	50	6
		IWGSC_CSS_7BS_scaff_3091538 IWGSC_CSS_7BS_scaff_3091539	Traes_7BS_E27A1E5DD Traes_7BS_DAEC45679	17	56	8	23	48	10
		IWGSC_CSS_7DS_scaff_3873671 IWGSC_CSS_7DS_scaff_3890593	Traes_7DS_22BB72EAC Traes_7DS_1CC708CC8	-	-	-	20	56	8
45	*ISAII*[18]	IWGSC_CSS_1AL_scaff_3908194	Traes_1AL_EC42C010D	6	18	1	6	16	2
		IWGSC_CSS_1BL_scaff_3895723	Traes_1BL_B7652D772	7	17	0‡	7	22	2
		IWGSC_CSS_1DL_scaff_2258215	Traes_1DL_482D35BEE	-	-	-	15	26	0‡
46	*ISAIII*[18]	IWGSC_CSS_5AL_scaff_2808573	Traes_5AL_5D542B354	25	66	10	24	78	8
		IWGSC_CSS_5BL_scaff_10830121	Traes_5BL_1E6720B2B	26	65	11	21	69	12
		IWGSC_CSS_5DL_scaff_4602103	Traes_5DL_49795C4EE	-	-	-	19	47	11

Syn.= synonymous mutation, Miss= missense mutations, Trun. = Truncation (premature stop codon plus splice donor or acceptor site mutants). Wheat genes were identified using rice orthologues from (22) except where noted.

‡ Deleterious missense mutation with SIFT <0.05 available.

Kronos mutations were called using original IWGSC gene Traes_3B_1C1D702D1.

[2] *Cytosolic AGPAse* reported by (30).

[3] Kronos mutations were called using original IWGSC gene Traes_3B_B98A3F969.

[4] *ADPG transporter* identified by orthology to maize *brittle-1* (31) and barley *lys5* (32).

[5] *Plastidial AGPase* identified by orthology to barley plastidial large subunit of *AGPase* (33).

[6] *Plastidial AGPAse* identified by orthology to barley plastidial small submit of *AGPase* (34).

[7] *GBSSI* (AF163319) and *GBSSII* (AF109395) identified using wheat cDNA from (35). *GBSS1-A* on incomplete scaffold was excluded from analysis.

[8] *SSI* (AFO91803) (36).

[9] *SSIIa* (AF155217) (37).

[10] *SSIIb* and *SSIIc* identified by orthology to maize orthologues in *EnsemblPlants*.

[11] *SSIIIa* (AF258608) (38).

[12] *SSIIIb* (ACJ68100) annotated in Uniprot.

[13] *SSIVb* (AY044844) (39).

[14] *SBEI* (Y12320) (40).

[15] *SBEIIa* (AF338432) (41, 42).

[16] *SBEIIb* (AY740401) (42, 43).

[17] *SBEIII* (JQ346193) (44).

[18] *ISAI*, *ISAII* and *ISAIII* identified by orthologue to Arabidopsis genes annotated in TAIR.

***SI Appendix*, Table S23**. Mutations in genes from the wheat flowering pathway. Chromosome locations are shown in *SI Appendix*, Fig. S12.

ID	*Gene*	IWGSC Scaffold	Gene Model	Kronos (AB)			Cadenza (ABD)		
				Syn.	Miss.	Trun.	Syn.	Miss.	Trun.
1	*PHYB-A*	IWGSC_CSS_4AS_scaff_5989820	Traes_4AS_1F3163292	33	100	6	31	46	4
	PHYB-B	IWGSC_CSS_4BL_scaff_7039327 IWGSC_CSS_4BL_scaff_7028025	Traes_4BL_94E5E213F Traes_4BL_4E9C903D0[7]	12	38	1	15	38	3
	PHYB-D	IWGSC_CSS_4DL_scaff_14408483 IWGSC_CSS_4DL_scaff_14465550	Traes_4DL_53442ACE0 Traes_4DL_2B1C9A09C[7]	-	-	-	35	68	3
2	*PHYC-A*	IWGSC_CSS_5AL_scaff_2763108	Traes_5AL_B5FA7CEF0[5]	28	38	2	27	45	3
	PHYC-B	IWGSC_CSS_5BL_scaff_10897570	Traes_5BL_02F8B4BFF	24	52	2	48	94	4
	PHYC-D	IWGSC_CSS_5DL_scaff_934837	Traes_5DL_0B769D390[6]	-	-	-	23	38	6
3	*PPD-A1*	UCW_Kronos_U_jcf7180000439988 IWGSC_CSS_2AS_scaff_5262553	Traes_2AS_2FCD59730	31	76	8	20	59	2
	PPD-B1[1]	IWGSC_CSS_2BS_scaff_2504037 IWGSC_CSS_2BS_scaff_2646347	Traes_2BS_2CA9DB5DE Traes_2BS_8BED816B1	55	128	6	29	60	6
	PPD-D1	IWGSC_CSS_2DS_scaff_2440553 IWGSC_CSS_2DS_scaff_171278 IWGSC_CSS_2DS_scaff_867705	Traes_2DS_2A961F39D Traes_2DS_3A2D5B67D Traes_2DS_00DB0399E[9]	-	-	-	16	47	2
4	*ELF-A3*	IWGSC_CSS_1AL_scaff_3795905	Traes_1AL_52C5531A4	30	56	2	29	60	4
	ELF-B3	IWGSC_CSS_1BL_scaff_3896400 IWGSC_CSS_1DL_scaff_2227465	Traes_1BL_B95F8C666	28	55	5	34	73	4
	ELF-D3[2]	IWGSC_CSS_1DL_scaff_2224292	Traes_1DL_96D83DE2D	-	-	-	Deleted in Cadenza		
5	*FT-A1*	IWGSC_CSS_7AS_scaff_4253310	Traes_7AS_EBD5F1F54	12	24	2	10	26	0
	FT-B1	IWGSC_CSS_7BS_scaff_3140477	Traes_7BS_581AA844D	7	15	1	9	15	1
	FT-D1	IWGSC_CSS_7DS_scaff_3824351	Traes_7DS_12C14942B	-	-	-	7	14	1
6	*ZCCT-D1*	IWGSC_CSS_4DL_scaff_14464484	KM503042	-	-	-	5	17	0[‡]
7	*ZCCT-B2*[3]	IWGSC_CSS_4BL_scaff_6972905 UCW_Kronos_U_jcf7180000435533	FJ173823 & FJ173824	22	36	0	9	9	2
	ZCCT-D2	IWGSC_CSS_4DL_scaff_14450864	KM503043	-	-	-	3	16	1
8	*FDL-A2*	IWGSC_CSS_1AL_scaff_3943384	Traes_1AL_1FFBFB058	18	42	0	15	33	2
	FDL-B2	IWGSC_CSS_1BL_scaff_3907952	Traes_1BL_DE2CF9613	14	30	0	15	24	3
	FDL-D2	IWGSC_CSS_1DL_scaff_2253590	Traes_1DL_D9BA83221	-	-	-	7	27	2
9	*VRN-A1*[4]	IWGSC_CSS_5AL_scaff_1660355 IWGSC_CSS_5AL_scaff_2805369	Traes_5AL_13E2DEC48	16	12	6	Duplicate in reference		
	VRN-B1	IWGSC_CSS_5BL_scaff_10800239 IWGSC_CSS_5BL_scaff_10898038	Traes_5BL_5D2D22E67	19	15	3	13	24	5
	VRN-D1	IWGSC_CSS_5DL_scaff_4550952	Traes_5DL_9CC4EC839	-	-	-	15	18	2
10	*FUL-A2*	IWGSC_CSS_2AL_scaff_6419608	Traes_2AL_20C2D79E1	16	16	4	17	19	1
	FUL-B2	IWGSC_CSS_2BL_scaff_8085378	Traes_2BL_26F24E716	21	20	3	15	24	2

	FUL-D2	IWGSC_CSS_2DL_scaff_9852812	Traes_2DL_903A29CBA[8]	-	-	-	12	19	2
11	*FUL-A3*	IWGSC_CSS_2AS_scaff_5296065 IWGSC_CSS_2AS_scaff_4877451	Traes_2AS_E2C631DBE	17	13	3	11	28	9
	FUL-B3	IWGSC_CSS_2BS_scaff_5242343	Traes_2BS_4818EA1FF	20	23	6	13	24	2
	FUL-D3	IWGSC_CSS_2DS_scaff_5335776 IWGSC_CSS_2DS_scaff_5389210	Traes_2DS_4F6BA4A13	-	-	-	25	29	1

Syn.= synonymous mutation, Miss= missense mutations, Trun. = Syn.= synonymous mutation, Miss= missense mutations, Trun. = Truncation (premature stop codon plus splice donor or acceptor site mutants).

‡ Deleterious missense mutation with SIFT <0.05 available.

[1] Two identical copies exist in Kronos.

[2] *ELF-D3* is deleted in Cadenza (45).

[3] *ZCCT-A1*, *ZCCT-A2* and *ZCCT-B1* encode non-functional proteins. *ZCCT-B2* is duplicated and the available Traes_4BL_1FE4A71E6 model is incorrect (46).

[4] The wheat reference sequence includes *VRN-D4*, a recent duplication of *VRN-A1* (47).

[5-9] Annotations corrected using GenBank accessions: [5] KF859916, [6]AH014104, [7]AY888046, [8] (48), and [9] DQ885766

References *SI Appendix*

1. Krasileva KV, et al. (2013) Separating homeologs by phasing in the tetraploid wheat transcriptome. *Genome Biol* 14:1-19.
2. Schreiber AW, et al. (2012) Transcriptome-scale homoeolog-specific transcript assemblies of bread wheat. *BMC Genomics* 13:1-14.
3. Brenchley R, et al. (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* 491:705-710.
4. Fu L, Niu B, Zhu Z, Wu S, Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28:3150-3152.
5. Mochida K, Yoshida T, Sakurai T, Ogihara Y, Shinozaki K (2009) TriFLDB: A database of clustered full-length coding sequences from Triticeae with applications to comparative grass genomics. *Plant Physiol* 150:1135-1146.
6. Cantu D, et al. (2011) Effect of the down-regulation of the high *Grain Protein Content* (*GPC*) genes on the wheat transcriptome during monocarpic senescence. *BMC Genomics* 12:1-17.
7. Falgueras J, et al. (2010) SeqTrim: a high-throughput pipeline for pre-processing any type of sequence read. *BMC Bioinformatics* 11:1-12.
8. Pertea G, et al. (2003) TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics* 19:651-652.
9. Wicker T, Matthews DE, Keller B (2002) TREP: a database for Triticeae repetitive elements. *Trends in Plant Science* 7:561-562.

10. Mayer KFX, et al. (2012) A physical, genetic and functional sequence assembly of the barley genome. *Nature* 491:711-716.

11. Mayer KFX, et al. (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* 345.

12. Kent WJ (2002) BLAT—The BLAST-like alignment tool. *Genome Res* 12:656-664.

13. Dvorak J, McGuire PE, Cassidy B (1988) Apparent sources of the A genomes of wheats inferred from the polymorphism in abundance and restriction fragment length of repeated nucleotide sequences. *Genome* 30:680-689.

14. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754-1760.

15. Li H, et al. (2009) The Sequence Alignment/Map format and SAM tools. *Bioinformatics* 25:2078-2079.

16. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841-842.

17. Zimin AV, et al. (2013) The MaSuRCA genome assembler. *Bioinformatics* 29:2669-2677.

18. Henry IM, et al. (2014) Efficient genome-wide detection and cataloging of EMS-induced mutations using exome capture and next-generation sequencing. *Plant Cell* 26 1382-1397.

19. Chapman JA, et al. (2015) A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. *Genome Biol* 16:26.

20. Ng PC, Henikoff S (2003) SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res* 31:3812-3814.

21. McLaren W, et al. (2016) The Ensembl Variant Effect Predictor. *Genome Biol* 17:122.

22. Pont C, Murat F, Confolent C, Balzergue S, Salse J (2011) RNA-seq in grain unveils fate of neo- and paleopolyploidization events in bread wheat (*Triticum aestivum* L.). *Genome Biol* 12:R119.

23. Hucl P (1996) Outcrossing rates for 10 Candian spring wheat cultivars. *Can J Plant Sci* 76:423-427.

24. Riley R, Kimber G (1961) Aneuploids and the cytogenetic structure of wheat varietal populations. *Heredity* 16, 275–290.

25. Lukaszewski AJ (1993) Reconstruction in wheat of complete chromosome-1B and chromosome-1R from the 1RS.1BL translocation of Kavkaz origin. *Genome* 36:821-824.

26. Priyam A, et al. (2015) SequenceServer: a modern graphical user interface for custom BLAST databases. *bioRxiv* doi: 10.1101/033142.

27. Goto N, et al. (2010) BioRuby: bioinformatics software for the Ruby programming language. *Bioinformatics* 26:2617-2619.

28. Ramirez-Gonzalez RH, Bonnal R, Caccamo M, Maclean D (2012) Bio-samtools: Ruby bindings for SAMtools, a library for accessing BAM files containing high-throughput sequence alignments. *Source Code Biol Med* 7:6.

29. Skinner ME, Uzilov AV, Stein LD, Mungall CJ, Holmes IH (2009) JBrowse: a next-generation genome browser. *Genome Res* 19:1630-1638.

30. Burton RA, et al. (2002) Characterization of the genes encoding the cytosolic and plastidial forms of ADP-glucose pyrophosphorylase in wheat endosperm. *Plant Physiol* 130:1464-1475.

31. Sullivan TD, Strelow LI, Illingworth CA, Phillips RL, Nelson OE, Jr. (1991) Analysis of maize *brittle-1* alleles and a defective *Suppressor-mutator*-induced mutable allele. *Plant Cell* 3:1337-1348.

32. Patron NJ, et al. (2004) The *lys5* mutations of barley reveal the nature and importance of plastidial ADP-Glc transporters for starch synthesis in cereal endosperm. *Plant Physiol* 135:2088-2097.

33. Eimert K, et al. (1997) Molecular cloning and expression of the large subunit of ADP-glucose pyrophosphorylase from barley (*Hordeum vulgare*) leaves. *Gene* 189:79-82.

34. Johnson PE, et al. (2003) A low-starch barley mutant, riso 16, lacking the cytosolic small subunit of ADP-glucose pyrophosphorylase, reveals the importance of the cytosolic isoform and the identity of the plastidial small subunit. *Plant Physiol* 131:684-696.

35. Vrinten PL, Nakamura T (2000) Wheat granule-bound starch synthase I and II are encoded by separate genes that are expressed in different tissues. *Plant Physiol* 122:255-264.

36. Li Z, et al. (1999) Cloning and characterization of a gene encoding wheat starch synthase I. *Theor Appl Genet* 98:1208-1216.

37. Li ZY, et al. (1999) The localization and expression of the class II starch synthases of wheat. *Plant Physiol* 120:1147-1155.

38. Li ZY, et al. (2000) The structure and expression of the wheat starch synthase III gene. Motifs in the expressed gene define the lineage of the starch synthase III gene family. *Plant Physiol* 123:613-624.

39. Leterrier M, Holappa LD, Broglie KE, Beckles DM (2008) Cloning, characterisation and comparative analysis of a *starch synthase IV* gene in wheat: functional and evolutionary implications. *BMC Plant Biol* 8:98.

40. Repellin A, Nair RB, Baga M, Chibbar RN (1997) Isolation of a *Starch Branching Enzyme I* cDNA from a wheat endosperm library (Accession No. Y 12320) (PGR97-094). *J Plant Physiol* 114:1145-1145

41. Rahman S, et al. (2001) Comparison of starch-branching enzyme genes reveals evolutionary relationships among isoforms. Characterization of a gene for starch-branching enzyme IIa from the wheat D genome donor *Aegilops tauschii*. *Plant Physiol* 125:1314-1324.

42. Hazard B, et al. (2012) Induced mutations in the *Starch Branching Enzyme II* (*SBEII*) genes increase amylose and resistant starch content in pasta wheat *Crop Sci* 52:1754-1766.

43. Regina A, et al. (2005) *Starch branching enzyme IIb* in wheat is expressed at low levels in the endosperm compared to other cereals and encoded at a non-syntenic locus. *Planta* 222:899-909.

44. Kang GZ, Xu W, Liu GQ, Peng XQ, Guo TC (2013) Comprehensive analysis of the transcription of starch synthesis genes and the transcription factor RSR1 in wheat (*Triticum aestivum*) endosperm. *Genome* 56:115-122.

45. Zikhali M, Wingen LU, Griffiths S (2016) Delimitation of the *Earliness per se D1* (*Eps-D1*) flowering gene to a subtelomeric chromosomal deletion in bread wheat (*Triticum aestivum*). *J Exp Bot* 67:287-299.

46. Distelfeld A, Li C, Dubcovsky J (2009) Regulation of flowering in temperate cereals. *Curr Opin Plant Biol* 12:178-184.

47. Kippes N, et al. (2015) Identification of the *VERNALIZATION 4* gene reveals the origin of spring growth habit in ancient wheats from South Asia. *Proc Natl Acad Sci USA* 112:E5401-E5410.

48. Li C, Lin H, Dubcovsky J (2015) Factorial combinations of protein interactions generate a multiplicity of florigen activation complexes in wheat and barley. *Plant J* 84:70-82.