PNAS

AGRICULTURAL SCIENCES

Sequencing 4.3 million mutations in wheat promoters to understand and modify gene expression

,1 ,b,1 ,d e,
f i j e ,2

Whea is an impo an con ib o o global food ec i , and f he imp o emen s a e eq i ed o feed a g o ing h man pop la ion. F nc ional gene ics and genomics ools can help s o nde s and he f nc ion of di e en genes and o enginee bene cial changes. In his s d , e sed a p omo e cap e assa o seq ence 2-kb egions ps eam of all high-con dence anno a ed genes f om 1,513 m ageni ed plan s f om he e aploid hea a ie K onos. We iden i ed 4.3 million ind ced m a ions i h an acc ac of 99.8%, es l ing in a m a ion densi of 41.9 m a ions pe kb. We also emapped K onos e ome cap e eads o Chinese Sp ing RefSeq 1.1, iden i ed 4.7 million m a ions, and p edic ed hei e ec s on anno a ed genes. Using hese p edic ions, e iden i ed 59% mo e nons non mo s s bs ia ions and 49% mo e nca ion m a ions han in he o iginal s d . To sho he biological al e of he p omo e da a se, e selec ed o m a ions i hin he p omo e of he *VRN-A1* e nali a ion gene. Bo h m a ions, loca ed i hin an sc ip ion fac o binding si es, signi can l al e ed *VRN-A1* e p ession, and one ed ced he n mbe of spikele s pe spike. ese p blicl a ailable seq enced m an da ase s p o ide apid and ine pensi e acces s o ind ced a ia ion in he p omo e s and coding egions of mos hea genes. ese m a ions can be i sed o nde s and and mod la e gene e p ession and pheno pes fo bo h basic and comme cial applica ions, he e limi ed go e nmen al eg la ions can facili a e deplo men . ese m an collec ions, oge he i h gene edi ing, p o ide al able ools o accele a e f nc ional gene ic s dies in his economicall impo an c op.

wheat | promoter capture | exome capture | functional genomics |

Wheat provides one- fth of the calories and proteins consumed worldwide (1), and continuous improvements in wheat productivity and nutritional value are required to feed a growing human population. Crop improvement relies on natural or induced variation that alters the function or expression of genes controlling relevant traits. To induce mutations with detectable e ects, wheat researchers used ionizing radiation in the late 1920s (2, 3) and expanded into chemical mutagenesis in the 1960s utilizing reagents such as ethyl methane sulfonate (EMS). EMS induces point mutations (G to A and the reciprocal C to T) that are less deleterious than the large deletions and chromosome breaks generated by radiation (4).

In early mutagenesis studies, researchers found that the use of induced mutations in polyploid wheat was not as e ective in generating phenotypic changes as in diploid grasses, such as barley or rice. ese researchers hypothesized that gene redundancy, resulting from polyploidy likely masks the cause of the phenotypic e ects of the mutations (2), a hypothesis that was con rmed in more recent studies (5, 6). Duplicated genes (homeologs) result in reduced selection pressure and in opportunities for subfunctionalization or elimination of duplicated genes. Di erences between homeologs are limited in wheat due to the recent origin of the polyploid species. is results in higher levels of functional overlap among duplicated genes and reduced e ects of induced mutations than in older polyploid species (6). Tetraploid wheat (2n = 28, genomes AABB) originated less than 0.8 Mya (7, 8), whereas hexaploid wheat (2n = 42, genomes AABBDD) originated ~10,000 y ago (9, 10), providing little time for functional di erentiation of homeologs.

e extensive functional overlap among homeologs allows the polyploid wheat species to tolerate higher doses of mutagen and more mutations per plant, which reduces the number of mutagenized plants required to saturate the gene space with mutations relative to diploid or old polyploid species (11). is property was used to develop saturated EMS mutagenized populations for both tetraploid and hexaploid wheat (5, 11–13). ese populations were initially screened using DNA pools and the endonuclease *Cel*I, a technology called TILLING (targeting induced local lesions in genomes) (14). e *Cel*I assays

Significance

Induced mutations in regulatory regions can be used to modulate the spatial or temporal expression of genes, whereas mutations in the coding regions can be used to generate knockouts or allelic variants to study gene function. Given the absence of regulatory limitations, these mutations can be also used in commercial applications. In this study, we sequenced 4.3 million induced mutations in the promoters and 4.7 million in the coding regions of most genes from the tetraploid wheat variety Kronos and deposited them in public databases. We provide examples of how this resource can be used to understand gene function, modulate gene expression, and generate changes in valuable wheat agronomic traits.

Author contributions: J.D. designed research; J.Z., H.X., Q.L., J.M.D., A.A., and E.A. performed research; K.G.-C., S.F.K., G.B.-G., C.P., and J.D.F. contributed new reagents/analytic tools; J.Z., G.F.B., H.V.-G., and J.D. analyzed data; and J.Z. and J.D. wrote the paper.

Reviewers: G.M., John Innes Centre; T.S., Leibniz-Institut fur Pflanzengenetik und Kulturpflanzenforschung; and B.T., Commonwealth Scientific and Industrial Research Organisation.

Competing interest statement: Dr. Pozniak and Dr. Trevaskis are co-authors on a 2020 review including 25 co-authors, but did not directly collaborate which each other.

[1]J.Z. and H.X. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: jdubcovsky@ucdavis.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2306494120/-/DCSupplemental.

Published September 13, 2023.

0.7% even in the lowest quality libraries, so we called mutations in each library at the lowest stringency level that resulted in ≥98%-EMS (henceforth EMS98%, *SI Appendix, Method S3*).

A total of 940 among the 1,535 sequenced lines exceeded the EMS98% threshold at the lowest stringency level (HetMC3/HomMC2), yielding 3,166,422 EMS mutations (74% of the total EMS mutations). is result indicated that most of the libraries were of good quality. We selected 834,662 EMS mutations at HetMC4/HomMC3, 205,755 EMS mutations at HetMC5/HomMC3 and 80,522 EMS mutations at HetMC6/HomMC4. Finally, we eliminated 22 libraries that did not reach the EMS98% threshold even at the highest stringency level (HetMC6/HomMC4), resulting in 1,513 lines used in the PC dataset.

By eliminating more SNPs from the low-quality libraries and selecting more from the good quality libraries, we were able to call 561,605 more EMS-type mutations than by using the HetMC5/HomMC3 across all libraries, while maintaining a very low estimated error (0.21%, *SI Appendix*, Table S1). e use of the EMS98% method adjusted by library also reduced the correlation between the promoter and EC studies for percent EMS (from R = 0.154 to R = 0.034) and non-EMS transitions (from R = 0.183 to R = 0.119), suggesting a reduced e ect of the di erences in sequencing library quality on the mutation identi cation process (*SI Appendix*, Table S2).

In summary, using the EMS98% method, we identi ed 4,287,361 EMS-type promoter mutations from 1,513 libraries (2,834 EMS mutations per line) with an estimated error rate of 0.21% (Table 1). Based on an estimated mapping space of 102,378,005 bp (quality ≥ 20, coverage ≥ 3), we estimated a mutation density of 41.9 EMS mutations/kb across the complete population, or 23.8 EMS mutations per Mb per individual line (Table 1).

Trait	Promoter	Exome
Number of lines used (overlap 1,465)	1,513	1,521
Total uniquely mapped SNPs	4,345,625	4,748,394
Uniquely mapped EMS-type mutations*	4,287,361	4,690,454
Valid mapping space [quality ≥20, min. coverage 3, no RH (Residual heterogeneity)]	102,378,005	131,190,164
Avg. EMS-type mutations per kb (population)	41.9	35.8
Avg. EMS-type mutations/line*	2,833.7	3,083.8
% EMS-type*	98.7%	98.8%
% Heterozygous for EMS SNPs (expected 66.66% at M_2)*	64.67%	65.07%
% Heterozygous for non-EMS SNPs	99.99%	99.99%
Non-EMS-type transitions (estimated error)*†	0.21%	0.34%
RH SNPs	40,823	73,239
% RH SNPs	0.93%	1.52%
% Heterozygous in RH	14.7%	26.8%
% EMS-type in RH	16.9%	23.0%
Non-EMS-type transitions in RH	25.4%	29.9%

*Excluding RH and deletions.
†Estimated from the percent of reciprocal A>G and T>C transitions among total mutations. The comparison between parameters for the PC obtained by methods EMS98% adjusted by library and HetMC5/HomMC3 global adjustment is provided in *SI Appendix*, Table S1.

e previous EC data were mapped to a fragmented wheat genome with limited annotation (5). Since there are no conversion tables from previous assemblies to RefSeqv1.0, we decided to remap the reads to the CS RefSeq v1.1 using the EMS98% error threshold, and to reannotate the mutation e ects using the gene models in RefSeq v1.1. is involved the remapping of the ECs of 1,535 Kronos lines and running the MAPS pipeline. We called SNPs using the same four stringency levels as in the PC described above, and then used the EMS98% threshold method to select high-con dence mutations.

e use of the EMS98% threshold resulted in the elimination of 14 libraries that showed less than 98%-EMS even at the highest stringency level (HetMC6/HomMC4), but still yielded 537,629 more EMS mutations relative to the previous HetMC5/HomMC3 threshold across all libraries (*SI Appendix*, Table S1). Using the EMS98% threshold, we identi ed 4,690,454 unique EMS-type mutations from 1,521 libraries with an estimated error rate of 0.34% (Table 1). is represents an average of 3,084 EMS mutations per line, and a mutation density of 36.7 EMS mutations/kb across the complete population (Table 1, EC).

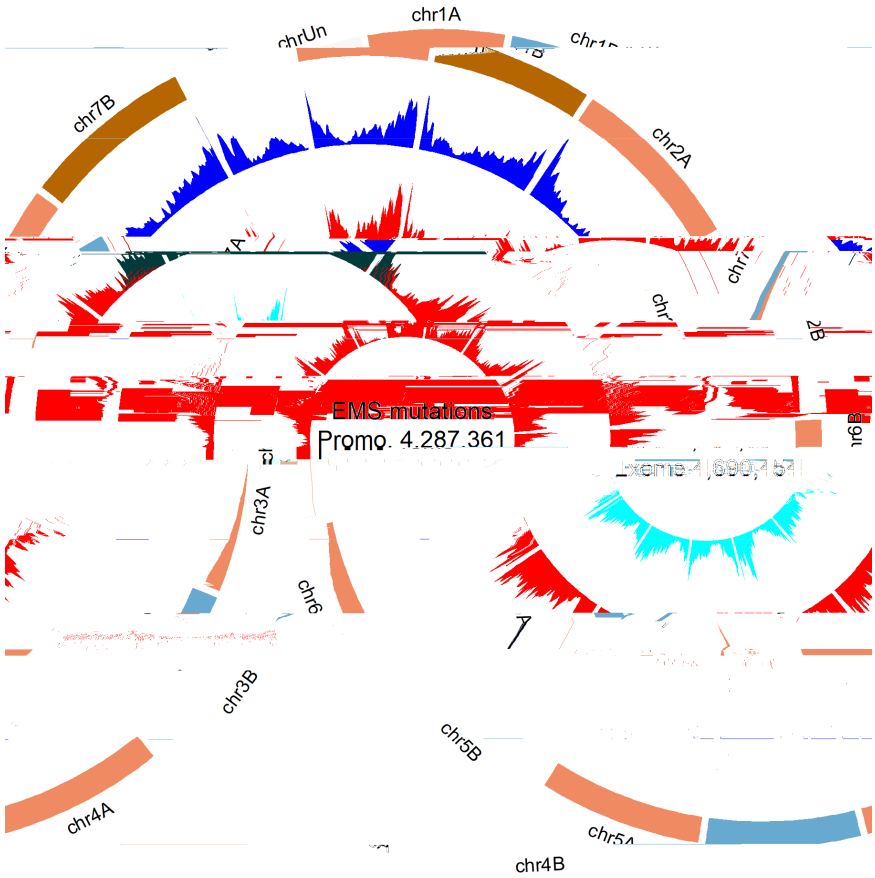
We determined mutation e ects using the Variant E ect Predictor (VEP) program (43) (*SI Appendix, Method S4*). VEP was also used to predict Sorting Intolerant from Tolerant scores (44), which predict the potential impact of amino acid substitutions on protein function. e improved e ect prediction tools, together with the improved gene models of CS RefSeq v1.1, increased the number of annotated missense variants to 1,637,961 (59% increase) and of truncation mutations to 113,880 (49% increase) relative to the previously published study (5) (*SI Appendix*, Table S3). As a result of the additional annotated e ects, the number of gene models with at least one missense mutation increased 22.9%, and the number of genes with predicted truncations (premature stop codons plus splice site mutations) increased by 34.9%, relative to previous results (5) (*SI Appendix*, Table S3).

Some of the mapped reads in the EC extend upstream of the start codon, so we expected some overlap between the mutations detected in the PC and EC studies. We used this overlap to con rm the correct identity of the lines in both captures. We identi ed 174,677 mutations in 1,465 lines shared between the PC and EC (*SI Appendix*, Table S4), con rming the correct tracking of the sequencing libraries and le identi cation numbers in the two experiments, and the quality of the called mutations.

e number of EMS mutations per chromosome was highly correlated with the number of annotated genes for both the EC (R = 0.942) and PC (R = 0.926, *SI Appendix*, Table S5).

ese results indicate that the probes captured mainly the gene regions as intended, and that the di erences in the number of SNPs across chromosomes were mainly driven by their di erences in gene content.

e EMS mutation density is relatively uniform, so the average number of detected mutations per kilobase is directly correlated with the length of DNA captured for a particular region. Since the PC and ECs are focused on genes, and wheat genes are more abundant in the distal regions of the chromosomes, we expected more mutations in those regions. To visualize the distribution of EMS mutations within chromosomes, we generated a circle graph including both the promoter and remapped EC data (Fig. 2). is graph showed that the EMS mutations in both studies are, as expected, more abundant toward the distal regions of the chromosomes, re ecting the higher gene density in the distal regions of the wheat chromosomes (45).



Genome-

e previous Kronos EC revealed that 1.64% of the detected SNPs were not induced by EMS mutagenesis but were instead originated from RH present in the mutagenized Kronos seed (5). We identi ed these regions by their higher proportion of linked non-EMS mutations, higher SNP density, higher proportion of homozygous SNPs, and by the presence of the same SNPs in multiple individuals. We combined those four criteria in one index and used a bioinformatics pipeline developed in a previous study (5) to identify and eliminate mutations in the predicted RH regions. is lter resulted in the elimination of 40,823 SNPs in the PC (0.93%) and 73,239 SNPs in the EC (1.52%, Table 1). ese percentages of RH are consistent with pooled seeds from di erent plants after six generations of self-pollination, a common practice in wheat breeding.

We collected the DNAs used in this study from M_2 plants, so we expected 1/4 of the plants to be homozygous for the wild-type allele, 1/4 homozygous for the mutant allele, and 1/2 heterozygous. Since we can only identify the mutant alleles, we expect 2/3 of the identi ed mutations to be heterozygous (66.7%). However, we initially observed 68.1% heterozygous EMS mutations in the PC and 68.2% in the reanalyzed EC. is excess heterozygosity is likely caused by the low threshold used by the MAPS program to classify mutations as heterozygous (5). Even when a single wild-type read is mapped to a mutation site, MAPS calls that mutation heterozygous even if it is at a very low frequency relative to the reads showing the mutation.

is can happen, for example, when a read from a closely related homeolog is erroneously mapped to the region carrying the mutation and is not eliminated by the MAPS pipeline.

To correct these errors, we used a heterozygous lter pipeline developed in the initial EC project (5). is lter reclassi es heterozygous mutations as homozygous when the frequency of the wild-type allele is less than 15% of the reads (5). After we applied this lter, the average percentage of heterozygous EMS mutations (%-het) was reduced to 64.7% for the PC mutations and to 65.1% for the EC mutations.

Unexpectedly, 100 lines showed an average proportion of heterozygous mutation (>90%) in both the PC and ECs that was much higher than expected for M_2 plants (66%, Fig. 3 *A* and *B*).

ese 100 lines showed similar average % EMS (98.4 to 98.7%) as the other 1,365 shared lines (98.6 to 98.7%), but they had on average 17% more mutations than the rest of the lines both in the PC and ECs ($P < 0.0001$, Fig. 3*C*). We hypothesize that the M_2 plants with >90% heterozygous mutations originated by hybridization between two mutant M_1 gametes. e maximum proportion of shared mutations between each of the 100 lines with >90% heterozygous and any of the other 1,365 lines was 0.51% indicating that none of the 100 lines originated from crosses with any of the other sequenced Kronos mutant lines (which will be expected to share ~50% of the mutations). erefore, we hypothesize that hybridization occurred between gametes from chimeric mutant tillers from the same M_1 plant or between mutant plants not included in our subset of sequenced lines. e percent cross-pollination in wheat is usually small (46), but the increased sterility of the mutant lines likely increased this probability.

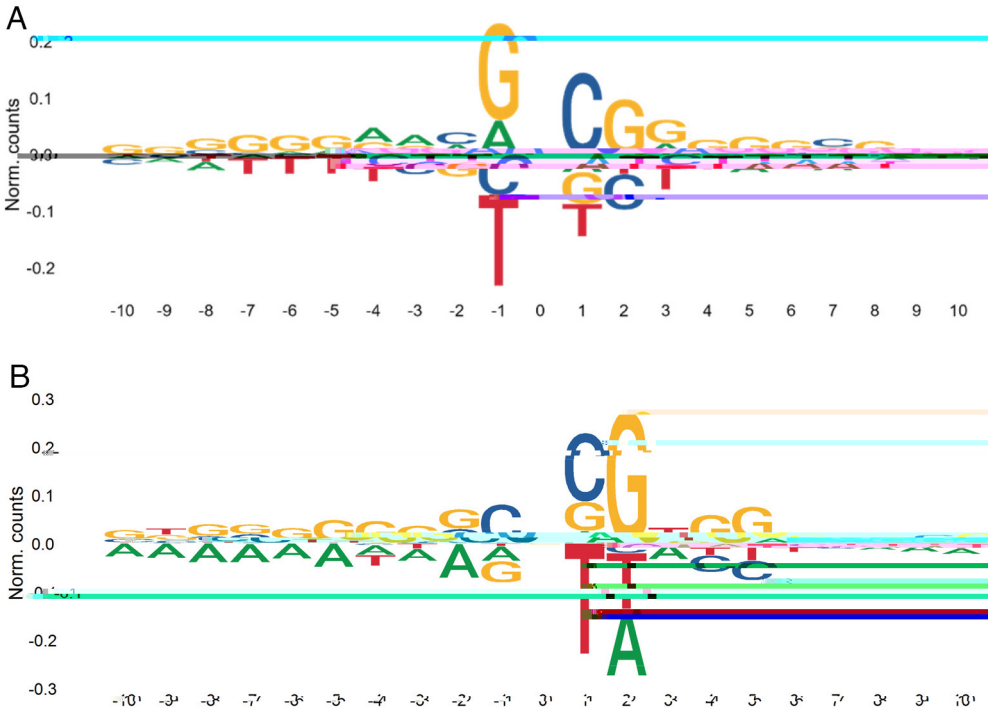
We analyzed the 10 bp anking the mutated G sites in the PC to determine if there was a preferred sequence context for the

mutations. We found a 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; whereas a negative bias for T was observed at position −1 (Fig. 4*A*). ese preferences are similar to those reported in the previous EC study (5), and indicate that not all the G bases in the genome have a similar probability of being mutated by EMS.

Although this variation is likely continuous, we arti cially divided them into two hypothetical populations of EMS-accessible and EMS-inaccessible GC bases to simulate a Poisson distribution and estimate the average proportion of accessible GC positions. We wanted to use this value to calculate the percent of the EMS-accessible GC positions for which we already have mutations in our population, a parameter we will refer hereafter as the percent saturation. is parameter is useful to decide if it is worth sequencing additional wheat mutant lines treated with EMS or if it is better to switch to a di erent mutagen that can access di erent sequences in the genome. As the percent saturation increases, additional sequencing results in diminishing returns because the proportion of mutations found in more than one line (henceforth, duplicated mutations) increases.

Among the large number of mutations identi ed in this study, we found 291,577 EMS mutations in the PC and 281,686 in the EC that were present in two lines. ose numbers decayed rapidly for mutation shared by three (PC = 31,539 and EC = 25,214) or four (PC = 4,503 and EC = 3,228) lines (*SI Appendix*, Table S6), following an approximate Poisson distribution. To estimate the proportion of EMS accessible GCs, we rst tested di erent means (λ = average mutations per site across the population) to identify the Poisson distribution that better t the observed data. We found that $\lambda = 0.182$ and $\lambda = 0.153$ minimized the di erences between observed and expected values for the PC and EC, respectively (*SI Appendix*, Table S6 and *Method S5*).

Using $\lambda = 0.182$, we estimated the existence of 23,556,929 "accessible" GC sites among the 47.9 M predicted GC in the total mapping space of the PC (102.4 Mb, GC content of 46.8%). Using a similar procedure, we estimated that, on average, 50.7% of the GC sites in the EC space were accessible to EMS (*SI Appendix*, Table S6).



Sequence context of G>A EMS mutations and C>A errors. (*A*) Sequence preference in regions flanking EMS-type G>A calculated based on all EMS mutations ( ) C>A sequence context was calculated based on the 60 sequencing libraries in both captures with the highest %C>A/G>T. The *x* axis indicates the number of nucleotides upstream (negative) and downstream (positive) from the mutated site.

e lower estimate for this parameter presented in our previous EC paper (5) was due to a calculation error, and a correction to that paper has been published (47), showing similar values to those presented here. In summary, these estimations suggest that approximately half of the GC sites in the sequenced space were accessible to the EMS treatment performed in this study.

We then used the estimated EMS-accessible GC to calculate the %-saturation values (mutated sites/accessible GC sites). Only 14.2% of the EMS-accessible GC in the EC and 16.6% in the PC are covered by EMS mutations in at least one of the sequenced lines in the mutant population. e larger value observed in the PC re ects the higher mutation density in the PC (41.9 mutations/kb) than in the EC (37.7 mutations per/kb). e percent-saturation values estimate the probability that a new mutation will be the same as a previously identi ed one, a probability that can be also estimated by dividing the number of duplicated mutations by the total number of EMS mutations. ese two independent estimates are almost identical validating the Poisson simulation (*SI Appendix*, Table S6).

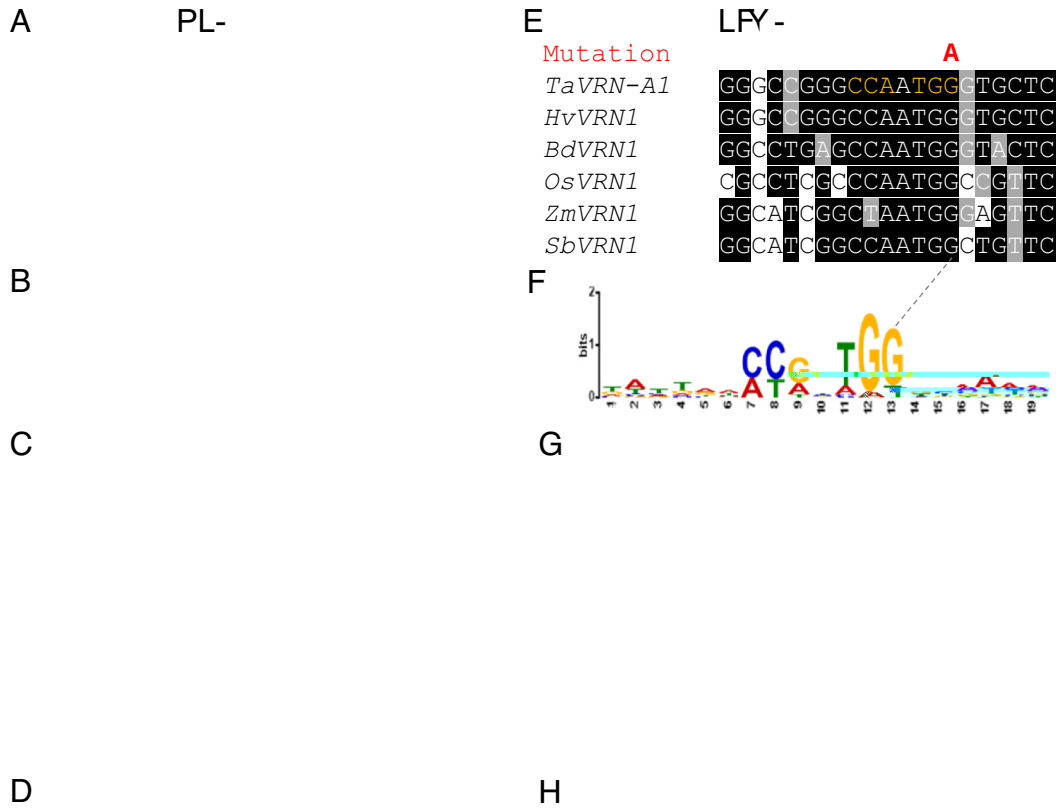
To show the value of the sequenced promoter mutations, we explored the 2 kb upstream of the vernalization gene *VRN1*, which is a central regulator of heading time (48–50) and spike development (24) in wheat. We rst delimited conserved regions in the promoter by aligning *VRN1* orthologs from wheat, barley, rice, maize, sorghum, and *Brachypodium* using T-co ee (https://tco ee.crg.eu/). We focused on two conserved regions (Fig. 5 *A* and *E*) including predicted binding sites for transcription factors (TFs) SQUAMOSA PROMOTER BINDING PROTEIN LIKE (SPL) and LEAFY (LFY) (Fig. 5 *B* and *F*, respectively).

We rst selected a C to T mutation in Kronos mutant K4679 located within the predicted SPL core binding site (sequence GTAC) (51). is mutation, located -368 bp upstream from the *VRN-A1* start codon, changes the core binding motif GTAC to GTAT in a position that is well conserved across grass species (Fig. 5*A*). Since SPL proteins are known to a ect panicle development in rice (52–54), we explored the e ect of the selected mutation on the expression of *VRN-A1* in young developing spikes (5 mm long) dissected from tetraploid wheat Kronos plants. e expression of *VRN-A1* in the mutant line across two experiments was on average 33.6% higher than in the wild-type sister line ($P = 0.002$, Fig. 5*C* and *SI Appendix*, Table S7), con rming that this mutation a ects *VRN-A1* expression. However, we did not detect signi cant di erences in spikelet number per spike (SNS, Fig. 5*D*) or heading time (*SI Appendix*, Table S8) in this particular mutant line.

We also found a mutation in Kronos mutant K2944 within a predicted LFY-binding site (55) located between positions −285 and −281 upstream from the *VRN-A1* start codon. is mutation resulted in a change from the conserved CC(A/G)n(T/A)GG binding site (55) to CC(A/G)n(T/A)GA, in a position that was conserved across the analyzed grass species (Fig. 5 *E* and *F*). We then studied *VRN-A1* transcript levels at the early stages of spike development (lemma primordium present, Waddington scale W3.25, ~2 mm long) by qRT-PCR in two independent experiments (Fig. 5*G* and *SI Appendix*, Table S9). e di erences in expression were signi cant only in the second experiment with a larger number of replications (Fig. 5*G*), but the trend was the same in both experiments resulting in a signi cant di erence in the combined ANOVA using experiments as blocks ($P = 0.010$, *SI Appendix*, Table S9). On average, we observed a 27% higher *VRN-A1* expression in the line with the mutation in the LFY-binding site than in the wild type.

is change was associated with an average reduction of 6.6% in the number of spikelets per spike relative to the sister lines with the wild-type allele ($P = 0.0001$, Fig. 5*H* and *SI Appendix*, Table S10). No signi cant di erences were observed for heading time and leaf number (*SI Appendix*, Table S10).

In summary, these examples suggest that the catalogue of sequenced promoter mutations developed in this study will be useful for the identi cation of regulatory variants to manipulate the levels of gene expression and to ne tune phenotypic changes in wheat.



Discussion

e large and repetitive nature of the polyploid wheat genomes delayed the initial genome sequencing e orts relative to other diploid crop species with smaller genomes. However, since the release of the rst hexaploid wheat reference genome (41), multiple complete wheat genomes have been sequenced (56), exome sequencing data have been completed for ~1,000 wheat accessions (57), and expression data have been generated from multiple tissues and germplasm (58, 59). ese public genomics resources, together with more e cient wheat transformation technologies (60, 61), have signi cantly expanded the opportunities for wheat research and improvement to more groups around the world. In this study, we report the development of a public database of 4.3 million sequenced mutations in the promoters of most genes in durum wheat, and the expansion and improved annotation of the sequenced mutations in the coding gene region from the same species.

e rst EC sequenced mutant populations were developed for tetraploid cv. Kronos (1,535 lines) and hexaploid cv. Cadenza (1,200 lines) in 2017 (5), before the release of the rst wheat genome reference CS RefSeq v1.0 (41). In those studies, reads were aligned to fragmented sca olds assembled using Illumina short reads generated from ow-sorted chromosome arms [Chromosome Survey Sequencing (CSS)] (62). In spite of these initial limitations, the two mutant wheat collections have been extensively used by the international wheat research and breeding community, with >20,000 mutant seed samples distributed until March 2023 by the University of California (USA) and the John Innes Centre (UK) [plus an unknown number from ve other locations that received complete Kronos mutant populations, including laboratories in Australia, Canada and China (5)].

To increase the value of these resources, we remapped the reads from the two ECs to the CS RefSeq v1.0 released in 2018 (41), and predicted the e ects of the mutations using the improved gene annotation from CS RefSeq v1.1. Cadenza reads from the 1,200 lines were called using the DRAGEN system (63), whereas Kronos reads from the 1,535 lines were remapped using the MAPS pipeline (15) as part of this study. Both datasets are available in the GrainGenes Genome browser (https://wheat.pw.usda.gov/GG3/genome_browser) for CS RefSeq v1.0. e Cadenza mutants

are also available in ENSEMBL, whereas the expanded Kronos mutations described in this study will be incorporated in the next ENSEMBL release 111 (VCF les have been submitted).

e remapped Kronos EC detected 561,605 additional mutations compared with the original study (5), while maintaining a very low estimated error (0.34%). is was achieved by using a stringency level adjusted by library rather than the uniform HetMC5/HomMC3 used in the previous study. is change was motivated by the nding of a signi cant correlation in %-EMS between the EC and PC in Kronos, which were generated using the same sequencing libraries. is signi cant correlation suggested that the %-EMS was signi cantly a ected by the quality of the libraries. By adjusting the stringency level based on the quality of the library, we were able to eliminate more mutations from the poor-quality libraries (lower %-EMS) and to extract more mutations from the good quality ones, which greatly exceeded the problematic ones. is method reduced the correlation between the PC and EC studies for %-EMS (*SI Appendix*, Table S2), suggesting a reduced e ect of the di erences in sequencing library quality on the selected mutations.

e characterization of the non-EMS mutations (*SI Appendix*, Table S11) revealed that they were all highly heterozygous (>99.9%). is result suggested that these were errors originated in the M_2

VRN-A1 In this study, we explored the e ect of induced mutations in conserved regions of the *VRN1* promoter on heading time and SNS as an example of the value of this sequenced mutant collection. *VRN1* encodes a MADS-box protein that plays a critical role in the regulation of the transition of the shoot apical meristem (SAM) to the reproductive phase (50, 75). e ancestral *VRN1* allele for winter growth habit requires long exposures to low temperature (vernalization) to be expressed, making *VRN1* a central gene in the owering pathway in the temperate grasses (50, 75–78). However, mutations in the *VRN1* promoter or deletions in the rst intron eliminate the vernalization requirement resulting in a spring growth habit (48, 49, 79, 80). e Kronos *VRN-A1* allele targeted in this study has a large intron deletion that eliminates the vernalization requirement.

Loss-of-function mutations in *VRN1* not only delay Kronos heading time but also increase SNS, indicating a role in the regulations of the transition of the in orescence meristem into a terminal spikelet (24). Plants with combined loss-of-function mutants in *VRN1* and its closest paralog *FUL2* cannot form spikelets and have an indeterminate spike, indicating an essential role of these two genes in the formation of both terminal and lateral spikelets (24). Overexpression of rice *MADS15* (the homolog of *FUL2*) reduces the number of primary branches in the rice panicle (81), suggesting a conserved role of these MADS-box genes in in orescence development in grasses.

We focused on the e ect of two EMS mutations located within conserved regions of the *VRN-A1* promoter, which potentially a ect SPL- and LFY-predicted binding sites. In Arabidopsis, these two TFs are known to bind directly to the promoter of *AP1* (82, 83), which is a homolog of wheat *VRN1* (84). SPLs are plant-speci c TFs that bind to a GTAC core sequence that is conserved from Arabidopsis (83) to wheat (85), and are known to a ect owering time (83) and in orescence architecture (52, 86). e mutation in the conserved SPL-binding site in the *VRN1* promoter resulted in a small but signi cant e ect on *VRN1* transcript levels in early spike development (Fig. 5*C*), which was not associated with signi cant changes in heading time or SNS (Fig. 5*D*). We are currently crossing the SPL-binding site mutant with a *VRN-B1* mutant, to test if the expression of the homeologous gene is masking the phenotypic e ect of the mutations.

LFY plays an important role in the speci cation of the oral meristem in Arabidopsis, and a separate role in the transcriptional activation of *AP1* (82). LFY loss-of-function mutations in rice show increased transcript levels of *MADS14* and *MADS15* (rice homologs of *AP1* and *VRN1*) and reduced number of primary branches and spikelet numbers, suggesting that LFY acts as a transcriptional repressor of these MADS-box genes during early panicle development (87). Consistent with the results of studies in rice, the Kronos mutant for the LFY-binding site in the *VRN1* promoter was associated with a slight but signi cant upregulation of *VRN1* during the early stages of spikelet development (before terminal spikelet formation) (Fig. 5*G*) and a signi cant reduction in the number of spikelets per spike (Fig. 5

Author affiliations: [a]Department of Plant Sciences, University of California, Davis, CA 95616; [b]Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China; [c]HHMI, Chevy Chase, MD 20815; [d]Nevada Bioinformatics Center, University of Nevada, Reno, NV 89557; [e]Department of Plant Pathology, Kansas State University, Manhattan, KS 66506; [f]United States Department of Agriculture - Agricultural Research Service, Wheat Health, Genetics and Quality Research Unit, Pullman, WA 99164; [g]United States Department of Agriculture - Agricultural Research Service, Cereal Disease Laboratory, Saint Paul, MN 55108-6086; [h]United States Department of Agriculture - Agricultural Research Service, Plant Science Research Unit, Raleigh, NC 27695; [i]Crop Development Centre, University of Saskatchewan, Saskatoon S7N 5A8, Canada; and [j]United States Department of Agriculture - Agricultural Research Service, Cereal Crops Research Unit, Northern Crop Science Laboratory, Fargo, ND 58102

[illegible] *Triticum aestivum* [illegible] *Science* **345** [illegible]
[illegible] *et al.* [illegible] *Open J. Genet.* **7** [illegible]
[illegible] *et al.* [illegible] *Nucleic Acids Res.* **41** [illegible]