COMBINING TRADITIONAL MUTAGENESIS WITH NEW HIGH-THROUGHPUT SEQUENCING AND GENOME EDITING TO REVEAL HIDDEN VARIATION IN POLYPLOID WHEAT

Cristobal Uauy[1], Brande B. H. Wulff[1], Jorge Dubcovsky[2,3]

[1] John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, United Kingdom

[2] Department of Plant Sciences, University of California, Davis, CA 95616, USA

[3] Howard Hughes Medical Institute, Chevy Chase, MD 20815 USA

Email: cristobal.uauy@jic.ac.uk; brande.wulff@jic.ac.uk; jdubcovsky@ucdavis.edu

Running title: Induced mutations in polyploid wheat

Corresponding author:
Cristobal Uauy
John Innes Centre
Norwich Research Park
NR4 7UH, UK
+44 (0)1603-450195
cristobal.uauy@jic.ac.uk

Keywords: Polyploid, hidden variation, mutants, reverse genetics, genome editing, wheat

Abstract:

Induced mutations have been used to generate novel variation for breeding purposes since the early 1900's. However, the combination of this old technology with the new capabilities of high-throughput sequencing has resulted in powerful reverse genetic approaches in polyploid crops. Sequencing genomes or exomes of large mutant populations can generate extensive databases of mutations for most genes. These mutant collections, together with genome editing, are being used in polyploid species to combine mutations in all copies of a gene (homoeologs), and to expose phenotypic variation that was previously hidden by functional redundancy among homoeologs. This redundancy is more extensive in recently formed polyploids such as wheat, which can now benefit from the deployment of useful recessive mutations previously identified in its diploid relatives. Sequenced mutant populations and genome editing have changed the paradigm of what is possible in functional genetic analysis in wheat.

1. POLYPLOIDY, DIPLOIDIZATION, AND THEIR EFFECT ON FORWARD AND REVERSE GENETIC APPROACHES

Our physical landscape is constantly modified by relatively fast orogenic events that result in the rise of mountains followed by a slower erosion that over the years reduces the height and sharp edges of the original mountains. Similarly, the plant genome landscape has been punctuated by polyploidization events that result in the immediate duplication of most genes in the genome followed by long periods of diploidization that dilute and gradually erase the previous duplications (2,40). The diploidization process includes the differential deletion or inactivation of duplicated genes in one of the genomes as well as the divergence of the multiple copies by sub- or neo-functionalization (8,52).

Since diploidization is a continuous process, the older the polyploidization event is, the more advanced the diploidization process is. For example, the grass lineage shares an old polyploidization event that occurred roughly 70 million years ago (62). The diploidization process was so extensive after this long period that many of the resulting species (e.g. rice, barley, sorghum, etc.) were considered true diploids until very recently. Only when genomic tools revealed extensive duplications among chromosome regions it became evident that species such as rice were ancient polyploids (28,62). Younger polyploidization events, such as the one in maize 5 to 15 million years ago (7,91), were discovered earlier, when colchicine treatments showed the formation of occasional cells with five quadrivalents (57), and confirmed later by genomic studies (72). Finally, more recent polyploid species such as tetraploid wheat, which originated less than 700,000 years ago (54) have been recognized as a polyploid species for a long time based on extensive gene duplication (56). Bread wheat, a hexaploid species (genomes AABBDD) that originated less than 10,000 years ago from the hybridization of tetraploid wheat (genomes AABB) with the diploid species *Ae. tauschii* (genome DD)(reviewed in 70), has had almost no time for diploidization, and most genes are expected to have overlapping functions. There is, of course, the divergence among the diploid genomes since their last common ancestor and up until their polyploidization, which in the case of the bread wheat genomes is less than 7 million years (54).

Polyploids are frequently classified as allopolyploids when the diploid species involved in the polyploidization event are different and have diverged enough to result in a clear diploid inheritance. By contrast, autopolyploids originate from whole genome duplications or duplication of hybrids within the same species or between closely related species, which results in a tetraploid inheritance. There is a recognized continuous gradation between these two extreme categories. We argue here, that it is important to classify polyploids also by the age of the polyploidization event, since young and old polyploids exhibit different characteristics that are relevant for strategies to improve polyploid crop species. For

example, loss-of-function mutations in single copy genes in ancient polyploid species frequently result in observable changes, and forward mutation screens are very effective tools to identify mutations affecting particular traits (61). By contrast, the effects of gene mutations in a recent polyploid species such as bread wheat are more frequently masked by redundant gene copies of the other genomes (designated homoeologs) compared to diploid species. Therefore, the importance of reverse genetic tools relative to forward genetic tools increases in young polyploid species.

2. SELECTION OF DIFFERENT TYPES OF VARIATION IN DIPLOID AND POLYPLOID SPECIES

Since the domestication of barley and wheat in the Fertile Crescent ~10,000 years ago, both species expanded westward and eastward with the spread of agriculture (15,70). This brought barley and wheat into new environments that expanded south and north of the original range of the wild species, exposing these crops to different photoperiod and temperature oscillations. These changes imposed a strong selection pressure on genes and pathways critical for reproductive success in the new environments. Since ancient times, wheat growers have been selecting mutations in reproductive regulatory genes that are critical to obtain plants that flower at the optimal time and maximize grain production (Figure 1a). Ancestral photoperiod sensitive winter barley and wheat forms evolved to generate different types of photoperiod sensitivity and different degrees of vernalization requirements. Since wheat and barley expanded together (70) and are still frequently grown in similar locations, the adaptive changes in the reproductive systems of these two related species provides a unique opportunity to compare changes selected in diploid and polyploid species.

Allelic differences in the same gene, *PPD-1*, are responsible for most of the natural variation in the photoperiodic response in barley and wheat. In barley, a loss-of-function recessive mutation in *PPD-H1* (the H indicates the H genome of barley) results in limited acceleration of flowering under long days (LD, 96). By contrast, the deletions selected in the promoter

regions of the wheat *PPD-A1* and *PPD-D1* genes result in the mis-expression of these genes during the night (5,105), which is sufficient to accelerate flowering under short days (SD) (63). Although natural polyploid wheats with simultaneous loss-of-function mutations in all *PPD-1* homoeologs have not been reported so far, the combination of mutations in the three *PPD-1* homoeologs in hexaploid wheat using molecular markers, results in a similar reduction in the acceleration of flowering under LD as the one reported in barley (63,79).

The difference in the type of mutations selected for *PPD-1* between diploid and polyploid species is also evident in *VRN2* (Figure 1). Natural variation in vernalization requirement is frequently associated with deletions or mutations in the regulatory regions of the *VRN1* gene, in both barley and wheat (22,30,43,46,59). However, spring growth habit associated with natural loss-of-function mutations or deletions of the *VRN2* gene have been observed so far only in barley and diploid species of wheat (17,99,108). The combination of non-functional or deleted *VRN2* genes in all three homoeologous *VRN2* loci in hexaploid wheat using molecular markers results in a spring growth habit (42), suggesting that the lack of similar combinations in nature were likely due to the inability to select phenotypically the individual mutations in the presence of other functional homoeologs.

These examples suggest that recessive mutations are more difficult to select in the young polyploid genomes than in diploid wheat and barley species. This is likely because of the extensive functional overlap among wheat homoeologous genes, which had very limited time to diverge since the recent polyploidization events that gave rise to the polyploid wheat species. This has favoured the selection of dominant or semi-dominant mutations in the polyploid wheat species during domestication and more recent breeding efforts. These mutations are usually epistatic over the other homoeologs facilitating the rapid detection of the favourable phenotypes. The selection of the *Q* gene (*APETALA2* on chromosome 5A) responsible for free-threshing grains in wheat (81), is another good example of these type of mutations. The free-threshing *Q* allele originated from a silent mutation in the microRNA miR172 binding site in the 5A homoeolog that reduces the cleavage of this specific

transcript. This results in high levels of AP2 thus conferring a dominant effect on the phenotype that is not dependent on additional mutations in the other *AP2* homoeologs (13).

Genes for resistance to wheat and barley rusts represent an abundant group of well-studied genes that provide an additional example of the different types of mutations selected in polyploid and diploid *Triticeae* species. In tetraploid and hexaploid wheat, almost all characterized resistance genes are dominant or semi-dominant (19). Among the 309 rust resistance genes listed in the 2017 Catalogue of Gene Symbols for Wheat, only 8 genes are characterized as recessive (2.6%)(56). The proportion of reported recessive rust resistance alleles in barley is higher (28%, Supplemental Table S1, $\chi 2$ $P < 0.0001$). In diploid wheat, *T. monococcum*, we recently mapped the recessive stem rust resistance gene *SrTm4* on a region of the long arm of homoeologous group 2 where no recessive rust resistance genes have been reported so far in polyploid wheat (9). The barley rust resistance gene *Mildew Locus O* (*MLO*) is also a good example of a mutation-induced recessive resistance gene selected only in a diploid species (10). Recently, Wang et al. (102) and Acevedo-Garcia et al. (1) demonstrated the feasibility of using either gene editing or induced-mutants to generate triple mutants for the A, B, and D genome homoeologs of *MLO* in wheat by crossing, which conferred strong powdery mildew resistance.

The fact that mutations in a single wheat homoeolog are frequently masked by redundancy from other homoeologs suggests that recessive variation in polyploid wheat has remained hidden, for the most part, from natural and human selection (8,47). In many cases, single and double mutants in hexaploid wheat are indistinguishable from wild-type plants making selection for allelic variation in any individual homoeolog largely irrelevant. This includes the *MLO* gene mentioned previously and variation in the *Starch Branching Enzyme II* (*SBEII*) gene. Here, single or double *SBEII* mutants have similar starch composition compared to wild type plants, whereas triple mutants for all three homoeologs show a significant increase in amylose (74,83). In other genes, dosage effects are sufficient to result in visible phenotypes that can be selected by breeders (80). This is the case for the *Grain Protein*

Content (*GPC-1*) gene where combining an increasing number of mutant homoeologs leads to a progressive delay in leaf senescence which is maximised in the triple mutant (4,97).

In summary, recessive mutations are difficult to select in young polyploid crops given the subtle phenotypic differences that they confer compared to wildtype plants. However, the extensive catalogue of sequenced mutations now available in wheat (47) provides the opportunity to combine loss-of-function mutations in multiple homoeologs to reveal this hidden variation. The tools are now available, to transfer beneficial recessive mutations identified in diploid species to the commercial polyploid wheats.

3. GENERATING INDUCED MUTATIONS IN POLYPLOID WHEAT

Radiation has been used to generate variability in plant research since the late 1920's. Work by LJ Stadler between 1928 and 1930 showed the use of X-ray treatments to induce visible mutants in diploid and polyploid grasses such as barley, maize, oat and wheat (84-86). Stadler's seminal work established that the rate of visible mutants recovered after mutagenesis in diploids was proportional to the radiation intensity. However, polyploid oat and wheat species yielded little or no visible mutant phenotypes when subjected to otherwise lethal doses in diploid species. An increase in the ploidy level in wheat, from diploid to hexaploid, led to a progressive decrease in the number of visible mutants (85). Based on these observations and inspired by Nilsson-Ehle's earlier work on seed colour in wheat (60), Stadler assumed that the 21 chromosomes of hexaploid wheat represented three groups of 7 pairs each and that these groups shared a proportion of identical genes. Already at this early stage, Stadler (85) concluded that the presence of gene "reduplication" (i.e. homoeologs) would not allow recessive mutations to "appear" as visible effects in polyploids. This buffering effect of polyploidy was supported by later studies in diploid and polyploid *Avena* (oat) and additional wheat species (21,26).

Chemical mutagenesis agents such as ethyl methanesulphonate (EMS) became commonplace in wheat research in the 1960s (77,90). EMS was shown to produce point mutations in wheat (G to A and C to T), contrary to radiation which produced large deletions and chromosome breaks (53), and was adopted both for developing new alleles for breeding (27,45), as well as for functional studies (20,23). Its use, however, has the same limitation of a reduced number of visible mutations as confronted by Stadler in the 1920's.

An important breakthrough came with the development of the TILLING (targeting induced local lesions in genome) approach (55) which provides a relatively simple strategy to identify mutants ('lesions') in a target sequence independently of its phenotypic effect. TILLING approaches require a population of, typically, EMS-induced mutants and a screening method to identify individuals with mutations in the target gene (reviewed in 100). Mutations are then prioritized based on their potential to disrupt protein function. Truncated proteins resulting from a premature termination codon or a change to a canonical splice site provide a high probability of functional disruption. However, non-synonymous mutations resulting in amino acids with different chemical properties (31) and located in evolutionary conserved positions of the protein also have a good probability of altering protein function (48). The higher the number of mutant alleles recovered, the higher the probability of identifying an allele leading to a non-functional protein.

Young polyploids are especially well suited for TILLING due to their tolerance of high densities of induced mutations compared to diploid species (100). We examined the mutation rates of 54 published studies that, in most cases, used the highest possible dose of EMS to develop TILLING populations (Figure 2, Supplementary Table S2). We found a significant difference (Kruskall-Wallis $P < 0.001$) between the mutation rates in diploid (4.3 ± 0.8 mutations per Mb) compared to young polyploid (26 ± 3.8 mutations per Mb) species. Within polyploid species, tetraploids had significantly lower ($P < 0.02$) mutations rates (22.4 ± 6.1) than hexaploids (31.1 ± 2.8 mutations per Mb) (Figure 2). Among "diploids", two populations stand out for their high mutation rates, flax (*Linum usitatissimum*; 24.4 mutations

per MB (11)) and yellow sarson/mustard (*Brassica rapa*; 16.7 mutations per Mb (87)). However, these two species went through whole genome duplication (flax, 103) or triplication (*B. rapa* (101)) events that occurred within the last 9 million years. As in maize, these species should be better classified as polyploids in an intermediate stage of diploidization.

The comparison of mutation densities between diploid and polyploid species supports the hypothesis that loss-of-function mutations in polyploids are masked by genetic redundancy among homoeologs. The direct link between ploidy level and tolerance to increasing mutation densities was established in an elegant study by Tsai et al. (93). Here, mutation densities were compared between a diploid Arabidopsis and a newly synthetized autopolyploid from the same accession. The autopolyploid lines tolerated 5-fold higher dosage of EMS treatments than the original diploid lines and had higher mutation densities.

The high density of mutations tolerated by polyploid species increases the number of mutant alleles recovered and reduces the cost of implementing high-throughput sequencing approaches to identify them. However, the utilization of these strategies in wheat requires an intermediate step to reduce the complexity of the large wheat genomes.

4. COMPLEXITY REDUCTION TO IDENTIFY INDUCED, RANDOM MUTATIONS IN TARGETED GENOMIC REGIONS

Direct whole genome shotgun (WGS) sequencing to identify and catalogue induced mutations is a feasible approach for species with small genomes (39,69,94). However, accessing induced variation in wheat by WGS sequencing is not economically feasible for a large number of individuals because of the large genome size (12 Gb in tetraploid and 16 Gb in hexaploid wheat (6)). Genes represent less than 2% of the wheat genome (38) and the rest is mainly represented by highly repetitive sequences. Mutations in the repetitive sequences cannot be easily identified, uniquely mapped to a physical position, nor assigned a biological function. The sheer size of the hexaploid wheat genome makes WGS

sequencing of mutant lines a costly and computationally intensive proposition. Even using a Hi-Seq X platform, today's cost of sequencing the genome of a single mutant hexaploid line will be approximately $8,000.

A smaller, specific portion of a plant genome can be captured for re-sequencing if the target sequences are known. This is achieved by the use of small biotinylated nucleotide probes, typically RNA, with homology to the target sequence. Hybridisation of the probes to the WGS library followed by probe capture enriches for the target sequence prior to sequencing (36). The technology is well established, with several commercial service providers offering probe synthesis. Multiple labs around the world have designed and captured the protein coding regions of the wheat genome (exome capture) for SNP discovery (3,106), and for cataloguing induced mutations (32,41). Most recently, Krasileva and colleagues (47) described a wheat TILLING resource for pasta and bread wheat containing 2,735 lines with more than 10,000,000 EMS-induced mutations. In this study, DNA from 4 to 8 mutant individuals were pooled, captured, and sequenced to obtain at least 20 million 100 bp paired-end reads per sample. Although this would be equivalent to 0.4 - 0.6 X coverage across the whole genome, the complexity reduction using exome-capture resulted in a median coverage at mutation sites of 21 X, a forty to sixty-fold enrichment.

Exome captures are biased by the gene space annotated in a reference sequence. Automated de-novo gene-prediction programs can often miss real genes and include artificial ones. Prediction programs that incorporate expression data (e.g. RNA-seq) perform better at predicting genes, but can still fail to detect those with low expression or with a very restricted expression profile. In addition to the mutations not detected in genes excluded from the capture assay, further mutations will not be detected if a gene sequence present in the capture assay is absent from the reference genome used to map the reads. A single reference genome only represents a certain proportion of the pan-genome space of a plant species (35,50,58,89). Recent resequencing and annotation of 19 wheat genomes suggests that the reference genome of the Chinese Spring landrace is missing about 12,000 genes

present among the pan-genome of elite, commercial cultivars (58). In the study from Krasileva et al. (47), the exome capture assay was enriched using transcriptomes from multiple RNA-seq data from different varieties. To take advantage of the expanded capture assay, the reference genome was also expanded by reassembling the reads that did not map to the original reference in each of the TILLING populations. The quality of the assembly of the unmapped reads was improved by using unmapped reads from multiple mutant plants. This strategy increased the proportion of targeted captured sequences relative to the non-targeted DNA included in individual captures, which is different in each line. Re-sequencing of multiple wheat varieties (18) will ultimately improve future exome capture designs and generate a more complete reference to map the captured reads.

5. STRATEGIES TO LINK INDUCED VARIATION WITH PHENOTYPES

5.1 SEQUENCED MUTANT POPULATIONS

Reverse genetic approaches, such as TILLING, are well suited to polyploid species, since mutations in individual homoeologs can be identified and combined before analysing the phenotype. However, the screening of TILLING polyploid populations by PCR amplification and sequencing requires the development of efficient genome specific primers, a task that requires sequences from the different homologs and experience (92,98). The advent of next-generation sequencing provides a new platform for detection of mutations in the different homoeologs and eliminates the need to generate genome specific primers. The simultaneous sequencing of most of the mutations present in the different homoeologs also eliminates the need to access the DNAs of the TILLING populations to search for a mutation.

In wheat, the coding regions of homoeologous genes are on average 97.2% identical (SD 1.8%, 75). Therefore, 100 bp paired-end reads are expected to include on average 5.6 SNPs, which is sufficient to map most of the reads from the mutant lines to unique places in the reference (Figure 3a-b). As the size of the paired-end reads continues to increase, the

proportion of reads mapped to unique sequences will also increase, and it will be possible to expand this approach to polyploid species with even more similar genomes. Even for those individual reads that are mapped to more than a single genome location, it is still possible to identify mutations. A dedicated bioinformatics pipeline was developed to keep track of the multiple locations where those reads map and to provide users with those alternative locations (multimap reads, 47).

Krasileva et al. (47) used an 84-Mb exome capture assay to sequence the coding regions of 1,535 tetraploid (cultivar 'Kronos', 98) and 1,200 hexaploid (cultivar 'Cadenza', 66) mutagenized lines. Using the MAPS bioinformatics pipeline (32) more than 10 million uniquely mapped high-confidence (P >0.998) EMS-type mutations were identified (Figure 3c). Across the complete populations, this represents a density of 35 mutations per kb in tetraploid wheat and 40 mutations per kb in hexaploid wheat. Using the published IWGSC gene models (38), effects were predicted for 7.1 million mutations (67%) and of these, 3% corresponded to truncations (premature termination codon or splice variant) and 38% were missense. The mutation density was sufficient to find an average of 23 to 24 predicted missense and truncation alleles per population per gene. Using the "sorting intolerant from tolerant" algorithm (SIFT, 48), we found that over 85% of genes captured had at least one putative reduced- or loss-of-function mutation (SIFT score < 0.05). When truncation mutations were added, >90% of the captured wheat genes were expected to have mutations severely affecting their function (47). To facilitate the utilization of these mutations, homoeolog-specific SNP assays (68,76) were designed for the majority of the mutations and seeds from all sequenced mutant lines were deposited in public repositories. These sequenced mutant populations are a valuable resource for functional studies and similar strategies can be applied to other crop species (sidebar: in silico functional genomic resources).

Although this review focuses on reverse genetic approaches, natural and induced variation at a single locus can lead to sufficient phenotypic differences between wildtype and mutant

individuals to allow traditional (forward) genetic mapping. Map-based cloning to identify genetic variants in wheat has been successfully applied (97,109,110), however, these methods are laborious and time-consuming, limiting the rate at which genetic loci can be identified. Combining high-throughput sequencing and genomic technologies provide an opportunity to greatly accelerate this process (67,71,73,88). However, these methods have to date been restricted to single locus loss-of-function mutations and have not been implemented for more complex quantitative traits in polyploid wheat.

5.2 TARGETED MUTAGENESIS VIA GENOME EDITING

Genome editing technologies provide new routes to induce variation in a more targeted manner than random chemical mutagenesis. The first examples of gene editing in wheat aimed to generate double strand breaks in precise genomic positions to induce small deletions. Early work in wheat used the transcription activator-like effector nucleases (TALENs) to simultaneously target the three homoeologous copies of *MLO* (102). This method was rapidly superseded by Cas9-mediated gene editing which was shown to be effective for inducing small deletions in wheat protoplast (78) and in stably transformed plants when a single homoeolog was targeted (102). Since then, several studies have shown the feasibility of generating transgene-free edited wheat using the Cas9 systems. These include transiently expressed plasmid DNA (TECCDNA, 112), in vitro synthesised mRNA of the Cas9 endonuclease and guide RNAs (TECCRNA, 112), and pre-assembled ribonucleoprotein complexes of purified Cas9 and in vitro transcribed gRNA (51). The latter two methods avoid transgene integration altogether since no DNA is introduced into the plant. Between 1 and 3% of bombarded embryos show an on-target deletion event using these methods, resulting almost exclusively in frameshift alleles and truncated proteins. Across these studies (51,102,112), seven genes have been edited in both tetraploid and hexaploid wheat.

Most recently, base editing was documented in wheat using a mutant nickase version of Cas9 fused to a cytidine deaminase (114). In this system, cytosine residues in a fixed position within the protospacer sequence (13 to 18 bp from the protospacer adjacent motif; PAM) were substituted to thymine without double strand breaks. Although only a single homoeolog was targeted in this study, it expands the applications of genome editing in wheat by shifting from targeted deletions into targeted base editing. Imminently, allele replacement (113), targeted gene insertion (25), and additional applications (65) are likely to be implemented in wheat. This raises the possibility of developing dominant alleles in single homoeologs in a targeted manner.

The sequenced mutant populations and the Cas9-mediated gene editing technologies provide complementary approaches to generate and utilise induced variation in polyploid wheat. Each approach has its own merits and demerits relating to the initial investment by researchers, the access to the technology, the range of mutations that are identified and their use in breeding (47). These are summarised in Table 1.

6. STRATEGIES FOR THE USE OF MUTANTS IN GENE ANALYSIS AND BREEDING

To overcome the functional overlap among homoeologs, loss-of-function mutations in the A and B genome copies in tetraploid wheat, and in the A, B and D genome copies in hexaploid wheat, need to be identified and then combined through crossing to select homozygous double or triple mutants in the F_2 generation. Since these crosses take time, it is critical to select mutations with a strong effect on gene function. As discussed before, truncation mutations have the best chance to achieve this objective, followed by mutations in conserved regions of the protein. However using missense mutations is riskier as it is difficult to predict a priori if the change in amino acid will completely eliminate protein function (see sidebar, Missense mutations, deletions and natural mutants). If no phenotype is observed after multiple crosses, it is difficult to determine if this is due to the ineffectiveness of the

missense allele, a different biological function of the gene in question, or redundant function with a close paralogous genes.

The high frequency of truncations and deleterious mutations detected in the sequenced polyploid wheat mutant populations is advantageous for selecting mutations with a high likelihood of disrupting protein function. It also provides an extended allelic series for detailed functional characterisation (e.g. *WKS1* (23)) or for more subtle modifications of the phenotype. However, once the desired mutant is identified, this advantage becomes a drawback because of the high number of background mutations. For example, in the sequenced wheat TILLING populations an average tetraploid mutant line has 2,705 high confidence EMS-type mutations which on average result in ~50 truncation alleles and over 670 missense mutations per individual mutant line in currently annotated genes. For Cadenza, in which the mutation load is higher, the expectation is ~110 truncation alleles and close to 1,400 missense mutations per individual in annotated genes. Put another way, between 1.5 and 2% of the genes in any given mutant line will have a truncation or missense allele (47).

Several strategies can be used to minimize the confounding effects of high background mutation in functional genetic studies. These strategies include the analysis of independent mutations, the use of sibling mutant lines and the backcrossing to the non-mutagenized line (Figure 4).

Multiple independent lines: The use of multiple independent mutants is a good practice to avoid the potentially confounding effect of other mutations present in the same plant. In polyploid wheat, the mutants in the different homoeologs represent a first level of replication. If the phenotype is observed only in plants combining mutation in all homoeologs, it is very unlikely that the effect will be caused by independent mutations in the same separate gene. For example, the probability of two mutant lines sharing by chance a truncation or missense mutation in the same second gene is $P < 0.0003$ in Kronos and $P < 0.0005$ in Cadenza. This

drops dramatically to a probability of less than 1 in 100,000 or 1 in 5 million when three or four independent lines are examined (Supplementary File S3).

Sibling lines: In tetraploid wheat, homozygous null-mutants and wild-type sibling plants can be selected using molecular markers from segregating F_2 individuals from the cross between A and B genome mutants. These F_2 sibling plants differ at the gene of interest, but share many of the same background mutations, resulting in a more valid comparison. The analysis of the segregating populations can be also used to rule out effects caused by unlinked mutations. In hexaploid wheat, the same strategy is effective although an additional cross is required to generate the F_2 triple mutants.

Backcrossing: Background mutations can not only confound gene-trait associations, but also can have detrimental effects on overall plant performance. Therefore, for quantitative traits or those that need to be assessed in plants grown under field conditions it is advisable to backcross the mutants to the original parent one or two times to reduce the mutation load. The trade-off between reducing mutation load and time required for the backcrossing will ultimately depend on the objective of the experiment and the phenotype being studied.

The time required to reduce the mutation load to acceptable levels can be shortened by selecting plants with the minimum number of background mutations in each generation. Since the positions of the mutations in the genome are known (16,104), background selection can be used to minimize the length of the chromosome segments from the mutant line. Additionally, the use of accelerated growth conditions (33,34) can further reduce generation time to ~10 weeks in Kronos and Cadenza (A. Watson, S. Ghosh, C. Uauy, B. Wulff & L. Hickey, unpublished data). This accelerated growth cycle is particularly useful for breeding applications, where more extensive backcrossing is justified to further reduce the load of background mutations. When mutations across multiple homoeologs need to be combined, it is prudent to backcross the single mutants separately and then inter-cross them to generate the desired double or triple mutants (29,80).

SUMMARY POINTS

1) Functional redundancy among homoeologs is more extensive in recently formed (young) polyploid species compared to older polyploids with more extensive diploidization.

2) A large number of recessive mutations are not detected by phenotypic selection in young polyploid species due to functional redundancy among homoeologs.

3) Domestication and previous breeding efforts have favoured selection of dominant or semi-dominant mutations in polyploid wheat since these are usually epistatic over the other homoeologs, allowing the rapid detection of favourable phenotypes.

4) Young polyploid species tolerate higher mutation densities than diploids making them especially well-suited for the development of sequenced mutant populations. This tolerance increases the number of mutant alleles recovered per plant, reducing the cost to identify them by high-throughput sequencing.

5) Useful recessive mutations identified in diploid relatives, are valuable targets to combine mutant alleles in all homoeologs in the polyploid crop. Sequenced mutant populations and genome editing facilitate the identification of these loss-of-function mutant alleles in all the homoeologs of a polyploid crop.

6) The combination of traditional mutagenesis with new high-throughput sequencing and genomics approaches has resulted in new powerful tools for functional genetic analysis in wheat. Alongside genome editing, these tools are changing the paradigm of what is possible in functional genetic analysis in wheat

7) The strategies developed for the generation of sequenced mutant populations in wheat can be transferred to other polyploid crops.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENT

The author's work is supported by the UK Biotechnology and Biological Sciences Research Council (BBSRC) grants BB/J003557/1, BB/J004553/1, BB/J004588/1 and BBL011794/1(CU and BBHW); Two Blades Foundation (BBHW); NRI Competitive Grant 2017-67007-25939 from the USDA National Institute of Food and Agriculture (NIFA) (JD) and Howard Hughes Medical Institute (JD).

LITERATURE CITED

1. Acevedo-Garcia J, Spencer D, Thieron H, Reinstädler A, Hammond-Kosack K, et al. 2017. *mlo*-based powdery mildew resistance in hexaploid bread wheat generated by a non-transgenic TILLING approach. *Plant Biotechnology Journal* 15:367-78
2. Adams KL, Wendel JF. 2005. Polyploidy and genome evolution in plants. *Current Opinion in Plant Biology* 8:135-41
3. Allen AM, Barker GLA, Wilkinson P, Burridge A, Winfield M, et al. 2013. Discovery and development of exome-based, co-dominant single nucleotide polymorphism markers in hexaploid wheat (*Triticum aestivum* L.). *Plant Biotechnology Journal* 11:279-95
4. Avni R, Zhao R, Pearce S, Jun Y, Uauy C, et al. 2014. Functional characterization of *GPC-1* genes in hexaploid wheat. *Planta* 239:313-24
5. Beales J, Turner A, Griffiths S, Snape JW, Laurie DA. 2007. A *Pseudo-Response Regulator* is misexpressed in the photoperiod insensitive *Ppd-D1a* mutant of wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics* 115:721-33
6. Bennett MD, Smith JB. 1976. Nuclear DNA amounts in angiosperms. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 274:227
7. Blanc G, Wolfe KH. 2004. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *The Plant Cell* 16:1667-78
8. Borrill P, Adamski N, Uauy C. 2015. Genomics as the key to unlocking the polyploid potential of wheat. *New Phytologist* 208:1008-22
9. Briggs J, Chen S, Zhang W, Nelson S, Dubcovsky J, Rouse MN. 2015. Mapping of *SrTm4*, a recessive stem rust resistance gene from diploid wheat effective to Ug99. *Phytopathology* 105:1347-54
10. Büschges R, Hollricher K, Panstruga R, Simons G, Wolter M, et al. 1997. The barley *Mlo* gene: A novel control element of plant pathogen resistance. *Cell* 88:695-705

11. Chantreau M, Grec S, Gutierrez L, Dalmais M, Pineau C, et al. 2013. PT-Flax (phenotyping and TILLinG of flax): development of a flax (*Linum usitatissimum* L.) mutant population and TILLinG platform for forward and reverse genetics. *BMC Plant Biology* 13:159
12. Chen A, Li C, Hu W, Lau MY, Lin H, et al. 2014. *PHYTOCHROME C* plays a major role in the acceleration of wheat flowering under long-day photoperiod. *Proceedings of the National Academy of Sciences USA* 111:10037-44
13. Dibernardi J, Faris J, Dubcovsky J. 2017. miR172 plays a critical role in the origin of free-threshing wheat. *submitted*
14. Dobrovolskaya O, Pont C, Sibout R, Martinek P, Badaeva E, et al. 2015. *FRIZZY PANICLE* drives supernumerary spikelets in bread wheat. *Plant Physiology* 167:189-99
15. Dubcovsky J, Dvorak J. 2007. Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science* 316:1862
16. Dubcovsky Lab. 2017. *Wheat TILLING: Dubcovsky Lab* http://dubcovskylab.ucdavis.edu/wheat-tilling
17. Dudnikov AJ. 2003. Allozymes and growth habit of *Aegilops tauschii*: genetic control and linkage patterns. *Euphytica* 129:89-97
18. Earlham Institute. 2017. *Grassroot Genomics*. http://www.earlham.ac.uk/grassroots-genomics
19. Ellis JG, Lagudah ES, Spielmeyer W, Dodds PN. 2014. The past, present and future of breeding rust resistant wheat. *Frontiers in Plant Science* 5:641
20. Feuillet C, Travella S, Stein N, Albar L, Nublat A, Keller B. 2003. Map-based isolation of the leaf rust disease resistance gene *Lr10* from the hexaploid wheat (*Triticum aestivum* L.) genome. *Proceedings of the National Academy of Sciences USA* 100:15253-8
21. Fröler K, Gustafsson Å, Tedin O. 1942. The relation of mitotic disturbances to x-ray dosage and polyploidy. *Hereditas* 28:165-70
22. Fu D, Szűcs P, Yan L, Helguera M, Skinner JS, et al. 2005. Large deletions within the first intron in *VRN-1* are associated with spring growth habit in barley and wheat. *Molecular Genetics and Genomics* 273:54-65
23. Fu D, Uauy C, Distelfeld A, Blechl A, Epstein L, et al. 2009. A *Kinase-START* gene confers temperature-dependent resistance to wheat stripe rust. *Science* 323:1357
24. Gao L, Cox DBT, Yan WX, Manteiga J, Schneider M, et al. 2016. Engineered Cpf1 Enzymes with Altered PAM Specificities. *bioRxiv* https://doi.org/10.1101/091611
25. Gil-Humanes J, Wang Y, Liang Z, Shan Q, Ozuna CV, et al. 2017. High-efficiency gene targeting in hexaploid wheat using DNA replicons and CRISPR/Cas9. *The Plant Journal* 89:1251-62
26. Gustafsson Å. 1947. Mutations in agricultural plants. *Hereditas* 33:1-100
27. Guttieri M, Bowen D, Dorsch JA, Raboy V, Souza E. 2004. Identification and characterization of a low phytic acid wheat. *Crop Science* 44:418-24
28. Guyot R, Keller B. 2004. Ancestral genome duplication in rice. *Genome* 47:610-4
29. Hazard B, Zhang X, Colasuonno P, Uauy C, Beckles DM, Dubcovsky J. 2012. Induced mutations in the *Starch Branching Enzyme II* (*SBEII*) genes increase amylose and resistant starch content in durum wheat. *Crop science* 52:1754-66
30. Hemming MN, Fieg S, James Peacock W, Dennis ES, Trevaskis B. 2009. Regions associated with repression of the barley (*Hordeum vulgare*) *VERNALIZATION1* gene are not required for cold induction. *Molecular Genetics and Genomics* 282:107-17
31. Henikoff S, Henikoff JG. 1992. Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences USA* 89:10915-9
32. Henry IM, Nagalakshmi U, Lieberman MC, Ngo KJ, Krasileva KV, et al. 2014. Efficient genome-wide detection and cataloging of ems-induced mutations using exome capture and next-generation sequencing. *The Plant Cell* 26:1382-97

33. Hickey LT, Germán SE, Pereyra SA, Diaz JE, Ziems LA, et al. 2017. Speed breeding for multiple disease resistance in barley. *Euphytica* 213:64
34. Hickey LT, Wilkinson PM, Knight CR, Godwin ID, Kravchuk OY, et al. 2012. Rapid phenotyping for adult-plant resistance to stripe rust in wheat. *Plant Breeding* 131:54-61
35. Hirsch CN, Foerster JM, Johnson JM, Sekhon RS, Muttoni G, et al. 2014. Insights into the maize pan-genome and pan-transcriptome. *The Plant Cell* 26:121-35
36. Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, et al. 2007. Genome-wide *in situ* exon capture for selective resequencing. *Nature Genetics* 39:1522-7
37. Hopf TA, Ingraham JB, Poelwijk FJ, Scharfe CPI, Springer M, et al. 2017. Mutation effects predicted from sequence co-variation. *Nature Biotechnology* 35:128-35
38. IWGSC. 2014. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* 345
39. Jiao Y, Burke JJ, Chopra R, Burow G, Chen J, et al. 2016. A sorghum mutant resource as an efficient platform for gene discovery in grasses. *The Plant Cell* http://dx.doi.org/10.1105/tpc.16.00373
40. Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473:97-100
41. King R, Bird N, Ramirez-Gonzalez R, Coghill JA, Patil A, et al. 2015. Mutation scanning in wheat by exon capture and next-generation sequencing. *PLOS ONE* 10:e0137549
42. 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. *Theoretical and Applied Genetics* 129:1417-28
43. Kippes N, Debernardi JM, Vasquez-Gross HA, Akpinar BA, Budak H, et al. 2015. Identification of the *VERNALIZATION 4* gene reveals the origin of spring growth habit in ancient wheats from South Asia. *Proceedings of the National Academy of Sciences USA* 112:E5401-E10
44. Kleinstiver BP, Prew MS, Tsai SQ, Topkar VV, Nguyen NT, et al. 2015. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* 523:481-5
45. Knott D. 1980. Mutation of a gene for yellow pigment linked to *Lr19* in wheat. *Canadian Journal of Genetics and Cytology* 22:651-4
46. Konopatskaia I, Vavilova V, Kondratenko EY, Blinov A, Goncharov NP. 2016. *VRN1* genes variability in tetraploid wheat species with a spring growth habit. *BMC Plant Biology* 16:244
47. Krasileva KV, Vasquez-Gross HA, Howell T, Bailey P, Paraiso F, et al. 2017. Uncovering hidden variation in polyploid wheat. *Proceedings of the National Academy of Sciences USA* 114:E913-E21
48. Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols* 4:1073-81
49. Li C, Lin H, Dubcovsky J. 2015. Factorial combinations of protein interactions generate a multiplicity of florigen activation complexes in wheat and barley. *The Plant Journal* 84:70-82
50. Li Y-h, Zhou G, Ma J, Jiang W, Jin L-g, et al. 2014. *De novo* assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nature Biotechnology* 32:1045-52
51. Liang Z, Chen K, Li T, Zhang Y, Wang Y, et al. 2017. Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. *Nature Communications* 8:14261
52. Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290:1151
53. MacKey J. 1968. Mutagenesis in vulgare wheat. *Hereditas* 59:505-17
54. Marcussen T, Sandve SR, Heier L, Spannagl M, Pfeifer M, et al. 2014. Ancient hybridizations among the ancestral genomes of bread wheat. *Science* 345
55. McCallum CM, Comai L, Greene EA, Henikoff S. 2000. Targeted screening for induced mutations. *Nature Biotechnology* 18:455-7

56. McIntosh R, Dubcovsky J, Rogers W, Morris C, Xia X. 2017. *Catalogue of gene symbols for wheat*. http://shigen.nig.ac.jp/wheat/komugi/genes/symbolClassList.jsp
57. Molina MdC, Naranjo CA. 1987. Cytogenetic studies in the genus *Zea*. *Theoretical and Applied Genetics* 73:542-50
58. Montenegro JD, Golicz AA, Bayer PE, Hurgobin B, Lee H, et al. 2017. The pangenome of hexaploid bread wheat. *The Plant Journal* doi: 10.1111/tpj.13515
59. Muterko A, Kalendar R, Salina E. 2016. Allelic variation at the *VERNALIZATION-A1*, *VRN-B1*, *VRN-B3*, and *PHOTOPERIOD-A1* genes in cultivars of *Triticum durum* Desf. *Planta* 244:1253-63
60. Nilsson-Ehle H. 1909. *Kreuzungsuntersuchungen an hafer und weizen*. Lunds universitets aersskrift N.F
61. Page DR, Grossniklaus U. 2002. The art and design of genetic screens: *Arabidopsis thaliana*. *Nature Review Genetics* 3:124-36
62. Paterson AH, Bowers JE, Chapman BA. 2004. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proceedings of the National Academy of Sciences USA* 101:9903-8
63. Pearce SP, Lin H, Li C, Shaw LM, Dubcovsky J. 2017. Night-break experiments shed light on the *PPD1*-mediated photoperiodic response in wheat. *under review*
64. Poursarebani N, Seidensticker T, Koppolu R, Trautewig C, Gawroński P, et al. 2015. The genetic basis of composite spike form in barley and 'Miracle-wheat'. *Genetics* 201:155
65. Puchta H. 2017. Applying CRISPR/Cas for genome engineering in plants: the best is yet to come. *Current Opinion in Plant Biology* 36:1-8
66. Rakszegi M, Kisgyörgy BN, Tearall K, Shewry PR, Láng L, et al. 2010. Diversity of agronomic and morphological traits in a mutant population of bread wheat studied in the Healthgrain program. *Euphytica* 174:409-21
67. Ramirez-Gonzalez RH, Segovia V, Bird N, Fenwick P, Holdgate S, et al. 2015. RNA-Seq bulked segregant analysis enables the identification of high-resolution genetic markers for breeding in hexaploid wheat. *Plant Biotechnology Journal* 13:613-24
68. Ramirez-Gonzalez RH, Uauy C, Caccamo M. 2015. PolyMarker: A fast polyploid primer design pipeline. *Bioinformatics* 31:2038-9
69. RevGenUK. 2017. *RevGenUK*. http://revgenuk.jic.ac.uk/
70. Salamini F, Ozkan H, Brandolini A, Schafer-Pregl R, Martin W. 2002. Genetics and geography of wild cereal domestication in the near east. *Nature Review Genetics* 3:429-41
71. Sánchez-Martín J, Steuernagel B, Ghosh S, Herren G, Hurni S, et al. 2016. Rapid gene isolation in barley and wheat by mutant chromosome sequencing. *Genome Biology* 17:221
72. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, et al. 2009. The B73 maize genome: Complexity, diversity, and dynamics. *Science* 326:1112
73. Schneeberger K. 2014. Using next-generation sequencing to isolate mutant genes from forward genetic screens. *Nature Review Genetics* 15:662-76
74. 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. *Journal of Plant Registrations* 10:200-5
75. Schreiber AW, Hayden MJ, Forrest KL, Kong SL, Langridge P, Baumann U. 2012. Transcriptome-scale homoeolog-specific transcript assemblies of bread wheat. *BMC Genomics* 13:492
76. Semagn K, Babu R, Hearne S, Olsen M. 2014. Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. *Molecular Breeding* 33:1-14
77. Shama Rao HK, Sears ER. 1964. Chemical mutagenesis in *Triticum aestivum*. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 1:387-99

78. Shan Q, Wang Y, Li J, Gao C. 2014. Genome editing in rice and wheat using the CRISPR/Cas system. *Nature Protocols* 9:2395-410
79. Shaw LM, Turner AS, Herry L, Griffiths S, Laurie DA. 2013. Mutant alleles of *Photoperiod-1* in wheat (*Triticum aestivum* L.) that confer a late flowering phenotype in long days. *PLOS ONE* 8:e79459
80. Simmonds J, Scott P, Brinton J, Mestre TC, Bush M, 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. *Theoretical and Applied Genetics* 129:1099-112
81. Simons KJ, Fellers JP, Trick HN, Zhang Z, Tai Y-S, et al. 2006. Molecular characterization of the major wheat domestication gene *Q*. *Genetics* 172:547
82. Slade AJ, Fuerstenberg SI, Loeffler D, Steine MN, Facciotti D. 2005. A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. *Nature Biotechnology* 23
83. Slade AJ, McGuire C, Loeffler D, Mullenberg J, Skinner W, et al. 2012. Development of high amylose wheat through TILLING. *BMC Plant Biology* 12:69
84. Stadler LJ. 1928. Mutations in barley induced by X-rays and radium. *Science* 68:186-7
85. Stadler LJ. 1929. Chromosome number and the mutation rate in *Avena* and *Triticum*. *Proceedings of the National Academy of Sciences USA* 15:876-81
86. Stadler LJ. 1930. Some genetic effects of X-rays in plants. *Journal of Heredity* 21:3-19
87. Stephenson P, Baker D, Girin T, Perez A, Amoah S, et al. 2010. A rich TILLING resource for studying gene function in *Brassica rapa*. *BMC Plant Biology* 10:62
88. Steuernagel B, Periyannan SK, Hernandez-Pinzon I, Witek K, Rouse MN, et al. 2016. Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture. *Nature Biotechnology* 34:652-5
89. Sun C, Hu Z, Zheng T, Lu K, Zhao Y, et al. 2017. RPAN: rice pan-genome browser for ~3000 rice genomes. *Nucleic Acids Research* 45:597-605
90. Swaminathan M, Chopra V, Bhaskaran S. 1962. Chromosome aberrations and the frequency and spectrum of mutations induced by ethylmethane sulphonate in barley and wheat. *Indian Journal of Genetics and Plant Breeding* 22:192-207
91. Swigoňová Z, Lai J, Ma J, Ramakrishna W, Llaca V, et al. 2004. Close split of sorghum and maize genome progenitors. *Genome Research* 14:1916-23
92. Tsai H, Howell T, Nitcher R, Missirian V, Watson B, et al. 2011. Discovery of rare mutations in populations: TILLING by sequencing. *Plant Physiology* 156:1257-68
93. Tsai H, Missirian V, Ngo KJ, Tran RK, Chan SR, et al. 2013. Production of a high-efficiency TILLING population through polyploidization. *Plant Physiology* 161:1604-14
94. Tsuda M, Kaga A, Anai T, Shimizu T, Sayama T, et al. 2015. Construction of a high-density mutant library in soybean and development of a mutant retrieval method using amplicon sequencing. *BMC Genomics* 16:1014
95. Turck F, Fornara F, Coupland G. 2008. Regulation and identity of florigen: *FLOWERING LOCUS T* moves center stage. *Annual Review of Plant Biology* 59:573-94
96. Turner A, Beales J, Faure S, Dunford RP, Laurie DA. 2005. The *Pseudo-Response Regulator Ppd-H1* provides adaptation to photoperiod in barley. *Science* 310:1031
97. Uauy C, Distelfeld A, Fahima T, Blechl A, Dubcovsky J. 2006. A NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat. *Science* 314:1298
98. Uauy C, Paraiso F, Colasuonno P, Tran RK, Tsai H, et al. 2009. A modified TILLING approach to detect induced mutations in tetraploid and hexaploid wheat. *BMC Plant Biology* 9
99. von Zitzewitz J, Szűcs P, Dubcovsky J, Yan L, Francia E, et al. 2005. Molecular and structural characterization of barley vernalization genes. *Plant Molecular Biology* 59:449-67
100. Wang TL, Uauy C, Robson F, Till B. 2012. TILLING *in extremis*. *Plant Biotechnology Journal* 10:761-72
101. Wang X, Wang H, Wang J, Sun R, Wu J, et al. 2011. The genome of the mesopolyploid crop species *Brassica rapa*. *Nature Genetics* 43:1035-9

102. Wang Y, Cheng X, Shan Q, Zhang Y, Liu J, et al. 2014. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nature Biotechnology* 32:947-51
103. Wang Z, Hobson N, Galindo L, Zhu S, Shi D, et al. 2012. The genome of flax (*Linum usitatissimum*) assembled *de novo* from short shotgun sequence reads. *The Plant Journal* 72:461-73
104. Wheat TILLING. 2017. *in silico wheat TILLING*. www.wheat-tilling.com
105. Wilhelm EP, Turner AS, Laurie DA. 2009. Photoperiod insensitive *Ppd-A1a* mutations in tetraploid wheat (*Triticum durum* Desf.). *Theoretical and Applied Genetics* 118:285-94
106. Winfield MO, Wilkinson PA, Allen AM, Barker GLA, Coghill JA, et al. 2012. Targeted re-sequencing of the allohexaploid wheat exome. *Plant Biotechnology Journal* 10:733-42
107. Xu R, Qin R, Li H, Li D, Li L, et al. 2017. Generation of targeted mutant rice using a CRISPR-Cpf1 system. *Plant Biotechnology Journal* 10.1111/pbi.12669:n/a-n/a
108. Yan L, Helguera M, Kato K, Fukuyama S, Sherman J, Dubcovsky J. 2004. Allelic variation at the *VRN-1* promoter region in polyploid wheat. *Theoretical and Applied Genetics* 109:1677-86
109. Yan L, Loukoianov A, Blechl A, Tranquilli G, Ramakrishna W, et al. 2004. The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. *Science* 303:1640
110. Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J. 2003. Positional cloning of the wheat vernalization gene *VRN1*. *Proceedings of the National Academy of Sciences USA* 100:6263-8
111. Zetsche B, Gootenberg Jonathan S, Abudayyeh Omar O, Slaymaker Ian M, Makarova Kira S, et al. 2015. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* 163:759-71
112. Zhang Y, Liang Z, Zong Y, Wang Y, Liu J, et al. 2016. Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. *Nature Communications* 7:12617
113. Zhao Y, Zhang C, Liu W, Gao W, Liu C, et al. 2016. An alternative strategy for targeted gene replacement in plants using a dual-sgRNA/Cas9 design. *Scientific Reports* 6:23890
114. Zong Y, Wang Y, Li C, Zhang R, Chen K, et al. 2017. Precise base editing in rice, wheat and maize with a Cas9- cytidine deaminase fusion. *Nature Biotechnology* doi:10.1038/nbt.3811

SIDEBARS

in silico functional genomic resources

It seems likely that in the coming years multiple mutant populations will be sequenced to establish *in silico* functional genomic resources in crops. This approach requires an up-front investment but generates a lasting and easy to use community resource. The specific sequencing strategy will depend on the genome size, the cost-effectiveness of sequencing platforms and complexity reduction methods, the similarity of the genomes, and the objective of the resource (e.g. complete genome vs. protein coding exome). Recent examples include the complete genome sequence for 256 mutant lines of sorghum (39), which has a relatively small diploid genome (750 Mb) compared to polyploid wheat (16,000 Mb). Despite the small population size in the sorghum study, 57% of the genes harboured at least one disruptive mutation (truncation or missense SIFT<0.05). In soybean (975 Mb), twelve mutant lines were sequenced to verify mutation density in a new mutant population that was then screened with amplicon resequencing (94). More recently, a population of over 1,000 *Brassica rapa* (diploid, 485 Mb) mutant lines was exome-captured and sequenced and made available online (69,87). Together with the sequenced mutant resources in tetraploid and hexaploid wheat (47), these examples document a shift towards whole genome/exome-sequenced mutant populations in the major crop species.

Missense mutations, deletions and natural mutants

Given the time required in wheat to combine mutants in all homoeologs, initial efforts to determine the effect of individual missense mutations on the activity of the protein are justified. For well-characterised protein domains one can select mutations which affect known active sites (12,23). However, for most proteins, computational methods are used to predict the effect of mutations on protein function (e.g. 48). The majority of these rely on the principle that proteins of related species and which perform a common function have

evolutionary conservation in key amino acid residues. Mutations affecting these positions are predicted to be deleterious. However, recent methods, which account for residue dependencies between all positions in a protein, have been shown to outperform traditional single-position models (37). In some cases, natural variation may be available to complement induced mutants (4,23,82,97) whereas in others, one of the homoeologs may not be expressed simplifying the development of null-mutants (14,64). Gamma-irradiated deletion populations can also be used to complement natural and induced variation. These large deletions are particularly useful when closely related genes are duplicated in tandem. Ultimately, the best strategy will depend on the available induced mutants and deletions, natural variation, and the specific gene being targeted.

FIGURE LEGENDS

Figure 1 Natural and induced variation for flowering in wheat and barley

(Top) The ancestral wheat and barley accessions have a winter growth habit, determined by the presence of functional *VRN2* repressors and wild type alleles of *VRN1* that are induced only after long exposure to cold temperatures (vernalization). The vegetative shoot apical meristem represents the lack of flowering during the fall. *VRN2* acts as a repressor of *FLOWERING LOCUS 1* (*FT1*) (109), which encodes a mobile protein that carries the photoperiod signal from leaves to the shoot apical meristem (95). In the apex, FT1 forms a floral activation complex that induces the meristem identity gene *VERNALIZATION 1* (*VRN1*) (49,110), initiating the transition of the apex to the reproductive stage. The upregulation of *VRN1* during the winter prevents the induction of *VRN2* in the spring, which results in the upregulation of *FT1* and the initiation of flowering. (**Middle**) Mutations or deletions in the *VRN1* promoter or first intron (red diagonal lines) eliminate the requirement for vernalization. The early expression of a single *VRN1* homoeolog is sufficient to repress all *VRN2* homoeologs, induce *FT1*, and initiate flowering without vernalization (dominant spring growth habit). This shoot apical meristem induced to develop a spike represents the initiation of

flowering. (**Bottom**) Natural deletions (dashed box) of the single *VRN2* locus in diploid barley, or the targeted combination using molecular markers (MAS) of the non-functional *VRN-A2* allele with a *Vrn-B2* deletion from *T. turgidum* subsp. *dicoccon* and a *VRN-D2* loss-of-function mutation from *Ae tauschii*, result in the upregulation of *FT1* and the induction of flowering without vernalization (recessive spring growth habit). No hexaploid wheat combining all three recessive mutations has been found so far (likely due to functional redundancy of the individual *VRN2* homoeologs). The dashed line in *VRN-A2* indicates that this locus is not functional in most polyploid wheats. The new sequenced TILING populations and genome editing tools will accelerate the combination of recessive mutations in polyploid wheat.

Figure 2 Mutation rates in mutant populations according to their ploidy level

Mutant populations from diploid (2x, red) species have an average density of 4.3 mutations per Mb. Flax and *B. rapa* (purple), two species that underwent whole genome duplications within the last 9 million years, should be considered polyploids in an intermediate state of diploidization. Tetraploid (4x, blue) species have an average density of 22.4 mutations per Mb. The high spread of the tetraploid values seems to be originated in part by different dosages of the mutagen used in these species (see threshold line). Hexaploid (6x, green) species have average densities of 31.1 mutations per Mb and include seven bread wheat and an oat mutant population. Links to the publications used in this figure are provided in Supplemental Table S2.

Figure 3 Types of mutations found in a mutagenized polyploid plant

(a) The presence of homoeologous SNPs (light blue) complicates the identification of EMS mutations in polyploid wheat. Different wheat varieties (e.g. Chinese Spring and Cadenza) share the majority of these homoeologous SNPs, but also have varietal SNPs (green). (b)

When next-generation sequencing (NGS) reads from one variety are mapped to the reference genome sequence, reads from a different homoeolog (bottom read in italics; D genome) may map to the same reference sequence in conserved regions or in regions with a missing homoeolog in the reference. This generates homoeologous SNPs that can confound the analysis (41). (c) The MAPS pipeline (32) solves this problem by eliminating variants that appear in more than one mutant included in each analysed batch (normally 24-30 mutants, three mutants in this example). This removes both varietal (green) and homoeologous SNPs (blue). The residual variants present in a single mutant can be attributed to sequencing errors (orange) and real EMS-type mutations (red). The two can be distinguished by their frequency: homozygous and heterozygous EMS mutations are expected to be present in all or half of the reads, whereas errors are randomly distributed in one read (exceptionally in more). In addition, EMS only generates G to A and C to T mutations in wheat. The MAPS pipeline allows users to define the minimum coverage (MC) required to call a homozygous (HomMC) or heterozygous (HetMC) mutation. Using HetMC=5 and HomMC=3 resulted in >99.8% accuracy in the detection of mutations (47).

Figure 4 Strategies for the use of mutants in gene analysis and breeding

Mutant lines carry large numbers of mutations (black triangles) across all chromosomes (seven chromosome groups x three genomes depicted) that can confound the phenotypic analyses. Three strategies are described to minimize that risk. **(Top)** The use of multiple lines carrying independent mutations in the same gene (red triangle), since it is unlikely that all will share the same confounding mutations elsewhere ($P < 0.0005$ for two independent mutants). **(Middle)** The mutant line can be crossed to the wild type parent to generate a segregating F_2 population. Using markers, F_2 siblings that differ in the target gene but share background mutations, can be compared. **(Bottom)** To further remove background mutations, the mutant lines can be backcrossed (BC) for 'n' generations to the wild type parent and progeny carrying the mutation can be selected using marker assisted selection

(MAS). On average, each backcross generation reduces unlinked background mutations by half. When the desired reduction of background mutations is achieved, BC_nF_2 siblings differing for the target mutation are selected.

TABLE LEGEND

Table 1 Characteristics of identifying and using induced variation from sequenced mutant populations and gene editing

Characteristic	Sequenced Mutant Population	Gene editing
Initial investment by researcher	Mutations are searchable online (16,104). Immediate access to mutant seed.	Requires additional time for construct design and optimisation. Delivery into wheat is dependent on access to technology; relatively high cost.
Specificity	Specific to 'C to T' and 'G to A' transitions. Has local sequence dependent bias which affects probability of C/G positions to be mutated (32,47).	Dependent on presence of PAM motif (5'-NGG-3'). New Cas9 specificities have been published (44). New nucleases (Cpf1 (111)) have different range of PAM motifs (24,107).
Off-target effects	Thousands of mutations outside gene of interest with many potential deleterious mutations.	Very specific with limited off-target effects.
Development of triple mutant	Mutants in individual homoeologs can be combined through traditional crossing and marker assisted selection.	Triple mutants in first generation not likely (~0.5%; (112)); will require crossing of single homoeologs to generate triple mutant.
Range of varieties	Original mutants restricted to sequenced populations. Can be transferred to locally-relevant germplasm by crossing.	Dependent on transformation efficiency of variety. Requires crossing to locally-relevant germplasm to deploy in agriculture.
Use in breeding	Currently deployed and not subject to regulation.	Non-transgenic classification is still uncertain in many countries. This may be more problematic for globally traded crops.

Figures

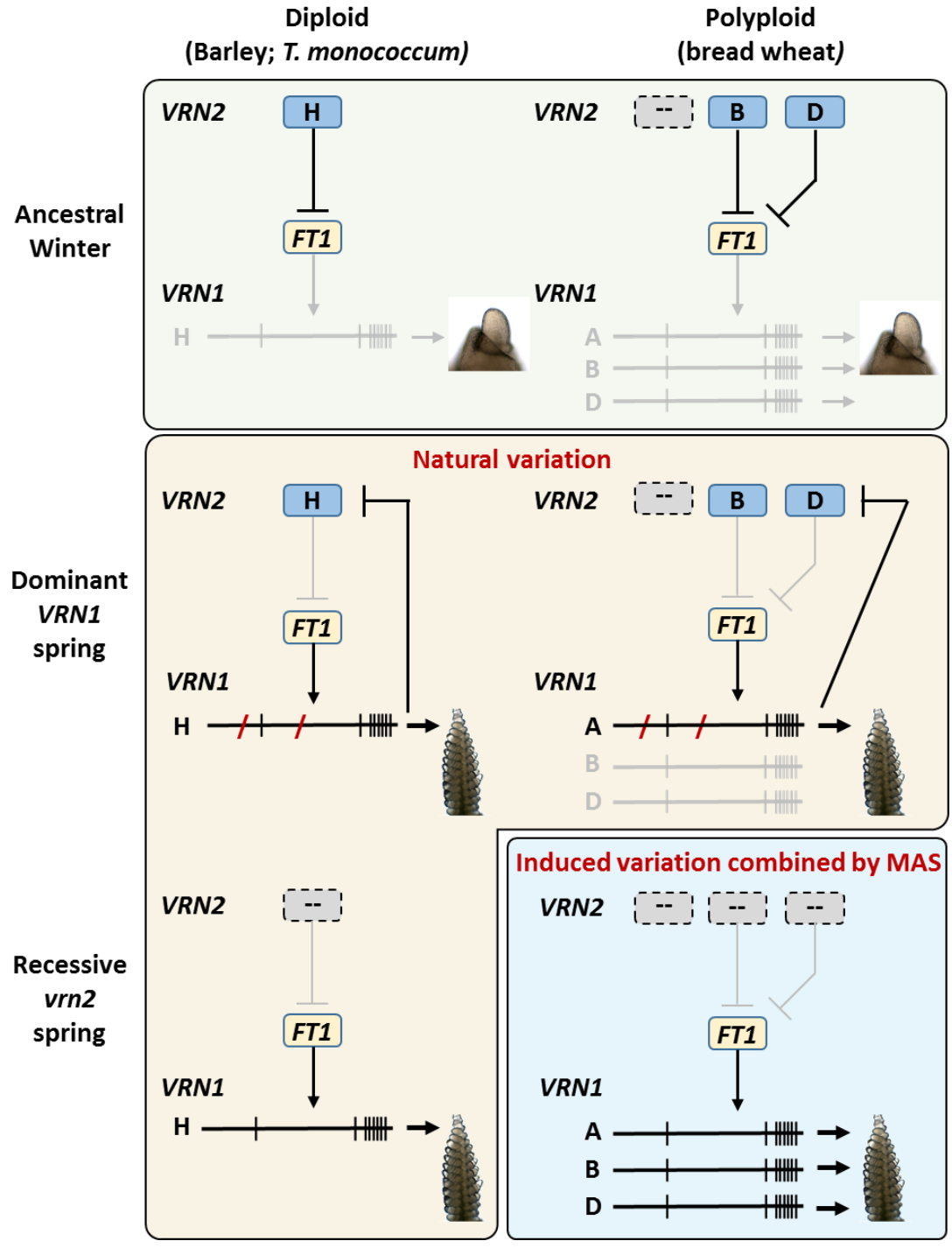

Figure 1

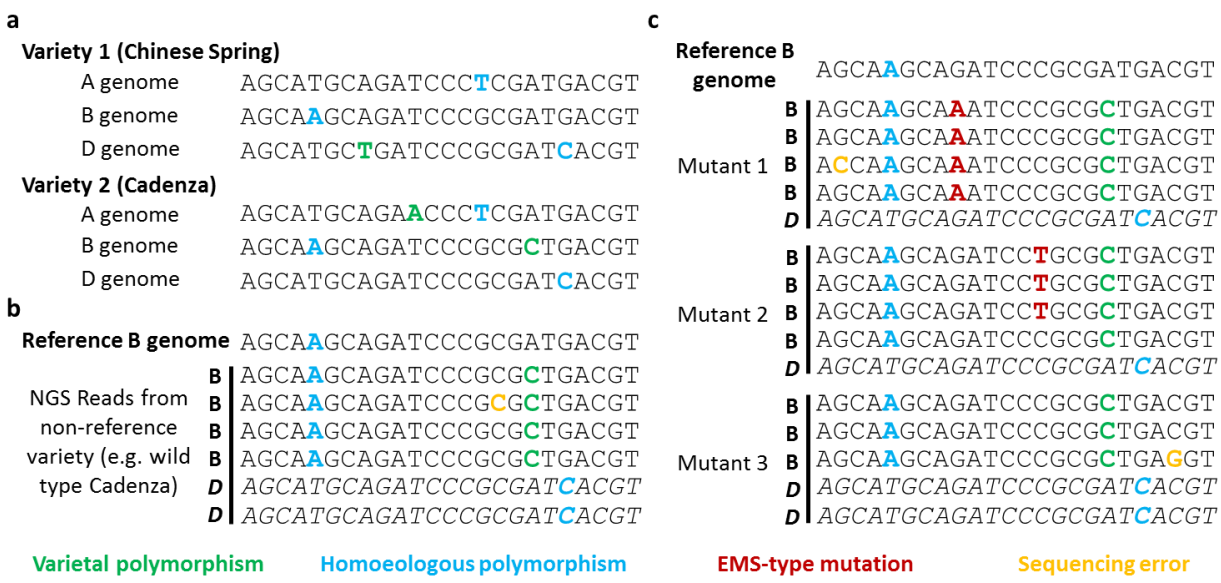
a
Variety 1 (Chinese Spring)
A genome AGCATGCAGATCCCTCGATGACGT
B genome AGCAAGCAGATCCCGCGATGACGT
D genome AGCATGCTGATCCCGCGATCACGT
Variety 2 (Cadenza)
A genome AGCATGCAGAACCCTCGATGACGT
B genome AGCAAGCAGATCCCGCGCTGACGT
D genome AGCATGCAGATCCCGCGATCACGT
b
Reference B genome AGCAAGCAGATCCCGCGATGACGT
NGS Reads from non-reference variety (e.g. wild type Cadenza)
B AGCAAGCAGATCCCGCGCTGACGT
B AGCAAGCAGATCCCGCGCTGACGT
B AGCAAGCAGATCCCGCGCTGACGT
B AGCAAGCAGATCCCGCGCTGACGT
D AGCATGCAGATCCCGCGATCACGT
D AGCATGCAGATCCCGCGATCACGT
Varietal polymorphism
Homoeologous polymorphism
c
Reference B genome AGCAAGCAGATCCCGCGATGACGT
Mutant 1
B AGCAAGCAAATCCCGCGCTGACGT
B AGCAAGCAAATCCCGCGCTGACGT
B ACCAAGCAAATCCCGCGCTGACGT
B AGCAAGCAAATCCCGCGCTGACGT
D AGCATGCAGATCCCGCGATCACGT
Mutant 2
B AGCAAGCAGATCCTGCGCTGACGT
B AGCAAGCAGATCCTGCGCTGACGT
B AGCAAGCAGATCCTGCGCTGACGT
B AGCAAGCAGATCCCGCGCTGACGT
D AGCATGCAGATCCCGCGATCACGT
Mutant 3
B AGCAAGCAGATCCCGCGCTGACGT
B AGCAAGCAGATCCCGCGCTGACGT
B AGCAAGCAGATCCCGCGCTGAGGT
D AGCATGCAGATCCCGCGATCACGT
D AGCATGCAGATCCCGCGATCACGT
EMS-type mutation
Sequencing error

Figure 2

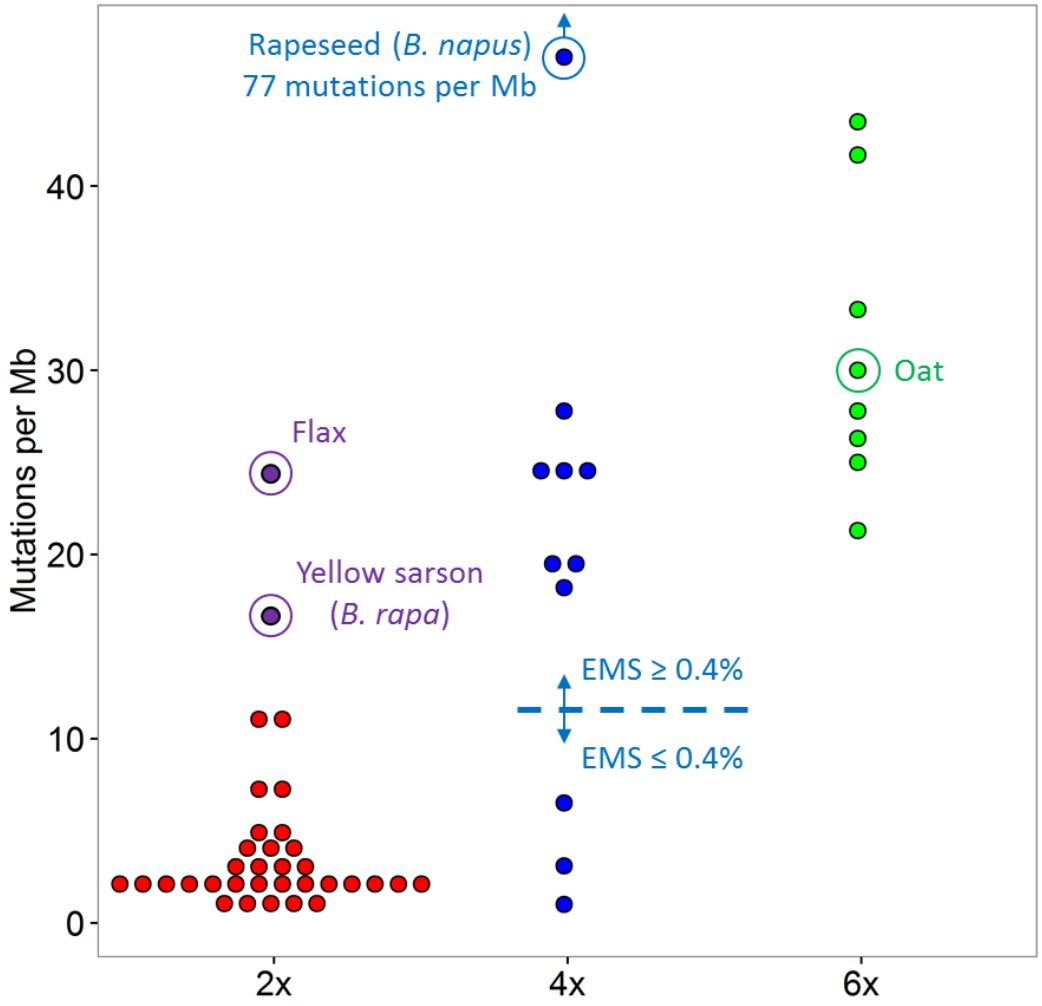
Rapeseed (B. napus)
77 mutations per Mb
Oat
Flax
Yellow sarson
(B. rapa)
EMS ≥ 0.4%
EMS ≤ 0.4%
Mutations per Mb
0
10
20
30
40
2x
4x
6x

Figure 3

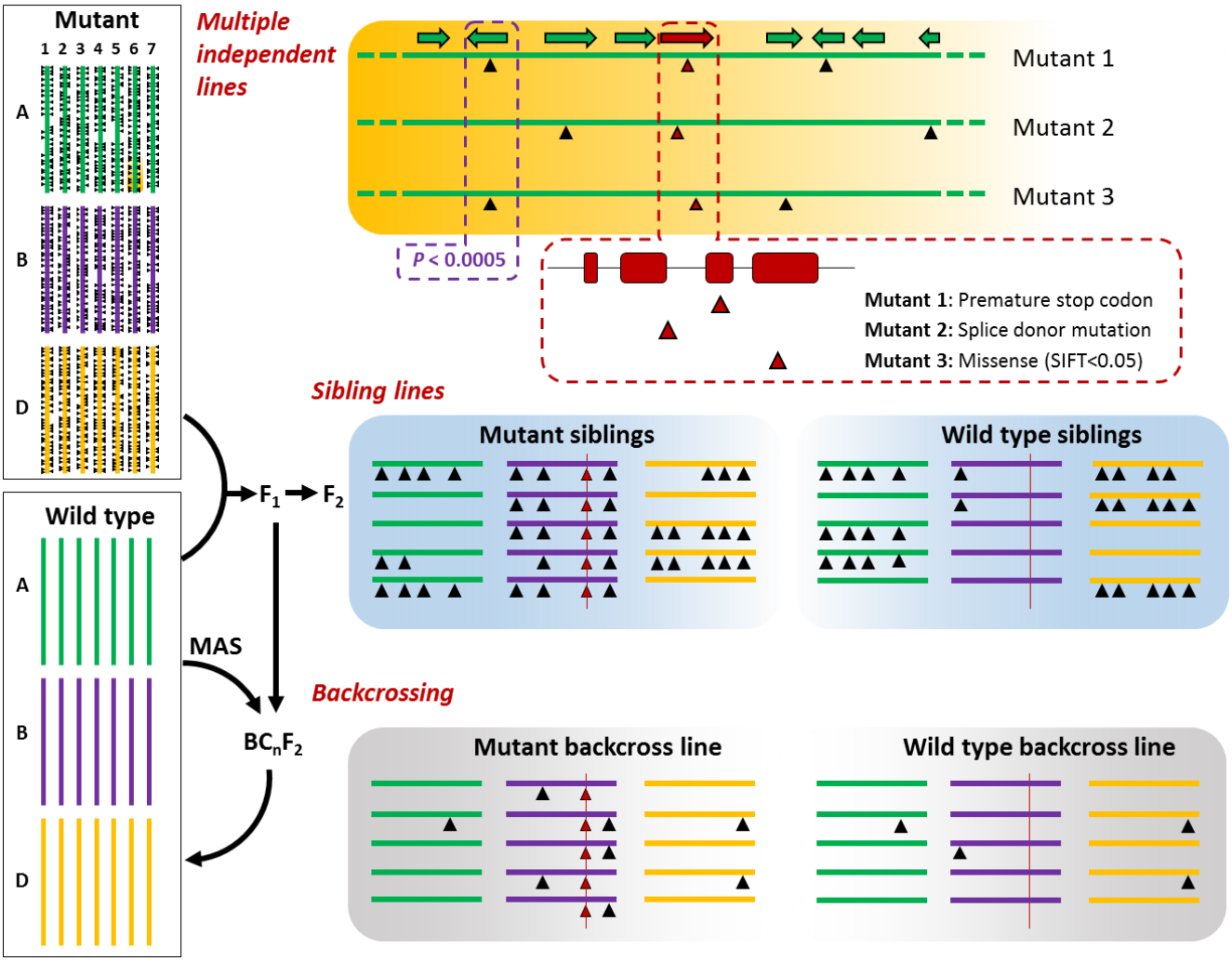
Mutant
1 2 3 4 5 6 7
A
B
D
Wild type
A
B
D
Multiple independent lines
Mutant 1
Mutant 2
Mutant 3
P < 0.0005
Mutant 1: Premature stop codon
Mutant 2: Splice donor mutation
Mutant 3: Missense (SIFT<0.05)
Sibling lines
F1
F2
Mutant siblings
Wild type siblings
MAS
Backcrossing
BCnF2
Mutant backcross line
Wild type backcross line

Figure 4

SUPPLEMENTAL FILES

Table S1: Dominant and recessive mutations in rust resistance genes in diploid barley and hexaploid wheat (.xlsx file).

Table S2: Mutation rate of TILLING populations in diploid, tetraploid and hexaploid plant species (.xlsx file).

File S3: Calculations for probability of shared mutations in mutant lines (.xlsx. file).