Application of CRISPR/Cas9 to edit genes affecting seed morphology traits in wheat

by

Qianli Pan

B.S., China Agricultural University, 2012

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Genetics Interdepartmental Program
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2019

Approved by:

Major Professor
Eduard Akhunov

Abstract

The CRISPR/Cas9-based genome editing system holds a great promise to accelerate wheat improvement by helping us to understand the molecular basis of agronomic traits and providing means to modify genes controlling these traits. CRISPR/Cas9 is based on a synthetic guide-RNA (gRNA) that can guide Cas9 nuclease to specific targets in the genome and create double strand breaks (DSB). The DSB are repaired through the error-prone non-homologous end joining process causing insertions or deletions that may result in loss-of-function mutations. Here, we have developed an effective wheat genome editing pipeline. We used next-generation sequencing (NGS) data to estimate genome editing efficiency of many gRNAs using the wheat protoplast assay and choose the most efficient gRNAs for plant transformation. We successfully applied this pipeline to five wheat orthologs of the rice yield component genes that have been identified previously. We obtained edited plants for all these genes and validated the effect of the mutations in *TaGW7* on wheat traits, which showed trends similar to those in rice. These results suggest that transferring discoveries made in rice to improve wheat is a feasible and effective strategy to accelerate breeding.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to express my deep gratitude to my major professor, Dr. Eduard Akhunov, for giving me the opportunity to study and work in his lab, as well as the great guidance and advice he gave me on my project and thesis writing. He helped me build up a good overview of the field and encouraged me to be an excellent student. He also offered me a lot of support in the process of my program study. Appreciation also needs to be sent to Dr. Harold Trick and Dr. Allan Fritz, who served as my committee members. Their valuable and constructive comments in the progress of my project and at the final defense improved my thesis and broadened my perspective on the subject.

I am particularly grateful for the assistance I received from Dr. Wei Wang, who is the person that established the genome editing pipeline in our lab, and many experimental conditions had been well adjusted by him before I came to Manhattan. He welcomed me joining the project and answered my endless questions on a daily basis about how to design and conduct the experiments. I am glad that eventually I was able to independently finish many experiments, but without Wei's guidance, it would not be possible. Wei also helped editing my thesis even though he had already enough on his plate and just got a newborn baby daughter. I feel so lucky to have a knowledgeable colleague and a very nice friend like him.

In my project, there are two crucial steps that were finished outside of our lab, for which I am sincerely grateful. The next-generation sequencing was performed by K-State Integrated Genomics Facility with the help provided by Dr. Alina Akhunova and her team; the biolistic transformation of wheat plants was conducted by Dr. Harold Trick's lab, also with help from many of their lab members. I wish to thank all of them for making this project possible.

Assistance provided by our lab members, Fei He, on bioinformatic analyses and Dwight Davidson, on greenhouse management was greatly appreciated as well. Moreover, I love and thank all our lab members for forming a warm and lovely team together, where I enjoyed my stay and learned much about either science or friendship.

Last but not least, I would like to dedicate my special thanks to my family, especially my husband Huan and my son Ryder. Both of them gave me infinite courage and firm support. Thank all our family members far away in China, who care every progress of us.

1 Introductions

Humans have been breeding crops for hundreds or thousands of years. When it came to 20th century, with the knowledge of genetics, biology and plant pathology growing, enormous contributions were made to increase food production throughout the world. However, there is still a gap between the supply and the need. Even today, food security is a concern. Data in the 2018 worldwide food security annual report from Food and Agriculture Organization (http://www.fao.org/publications/sofi) shows that number of people suffering food and nutrition deficiency grew from 2014 to 2017. This trend can be attributed to increasing population, unstable environments as well as lag in increasing crop yield potential using traditional breeding strategies relying on beneficial alleles from natural populations. If no steps to address these challenges are taken, the situation for the coming decades may become worse [1]. Therefore, breeders and geneticists are looking for ways to improve crop yield using novel approaches.

1.1 The novel approaches in crop breeding

1.1.1 Traditional marker-based methods for crop trait improvement

Conventional crop breeding is based on hybridization and phenotypic selection, which relies on discovery of new favorable allelic combinations resulting from recombination [2]. This process was lately accelerated by the usage of molecular markers, either markers linked with individual loci positively affecting agronomic traits of interest (marker assisted selection, or MAS) [3] or genome-wide markers that together can be used for predicting the performance of breeding lines (genomic selection) [4]. The markers are usually DNA polymorphisms segregating in population detected using either genotyping assays [5] or next-generation sequencing [6]. By combining the genotypic and phenotypic information, the association

between markers and traits can be established to select marker loci linked or closest to the trait-controlling genes. The usage of markers greatly increases the efficiency and precision of trait selection during breeding by alleviating the need to collect phenotypic data for difficult traits or traits having quantitative nature. For example, the collection of root trait data requires root excavation from soil [7] which is not suitable for high-throughput phenotyping. Thus the development of markers for root morphological traits holds great promise to accelerate the efficiency of breeding for improved root system which is crucial for drought tolerance [8]. Likewise, the development of markers predictive of wheat quality traits can help to accelerate breeding by reducing the need for time-consuming phenotyping [9].

The success of breeding depends on influx of new allelic diversity into breeding programs. The efficiency of novel allele transfer is strongly affected by the frequency of recombination [10]. Recombination breaks an allele's linkage with the genetic backgrounds which many harbor other alleles with opposite effects, and thus is a prerequisite of successful selection in breeding. However, suppression of recombination rate in some part of the genome [11] can render the beneficial alleles inaccessible for selection [12]. Moreover, the genetic divergence between wheat and wild ancestors broadly used as a reservoir of useful alleles for breeding, also can result in reduced recombination limiting the utility of this kind of variation in breeding [13].

Thus, even though marker-based selection has been relatively effective at accelerating crop breeding process, its efficiency often was limited by natural variation in recombination frequency around the beneficial alleles.

1.1.2 Random mutagenesis to produce variations positively affecting traits of interest

Selection imposed by breeders during the development of "elite" cultivars resulted in loss of genetic diversity [14]. This genetic erosion eventually became a bottleneck for crop improvement [15]. Various techniques to artificially increase variations in organisms emerged in the 20th century, including radiation and chemical mutagens [16]. Followed by phenotypic screening, plant mutagenesis can be used for either breeding of novel traits of agricultural importance or analysis of gene function, making this technique an important tool in breeding. Since the early last century when X-ray was shown to have potential for mutant breeding [17], different mutagenesis methods have been developed and applied to different crops. Of these methods the EMS (ethyl methanesulfonate) treatment is the most widely used system [16]. Mutant Varieties Database (https://mvd.iaea.org) shows information of thousands of mutant crop cultivars already released officially or commercially worldwide in breeding programs.

In terms of mutation effect, the radiation-based mutagenesis usually causes larger scale deletions and translocations. Such major chromosome changes may reduce plant viability, and also negatively impact allele transfer in breeding program by affecting recombination. In comparison, EMS is an alkylating agent that causes G/C to A/T transitions in the genome. These point mutations are more desired in breeding programs, in addition to the easiness and effectiveness of EMS mutagenesis system [16].

Wheat, as an allopolyploid, represents a unique case for EMS-based mutagenesis. On one hand, wheat can tolerate high mutation load because the extra sets of genomes can compensate each other to maintain basic gene functions to help plants survive, which fits the goal of broadening allelic diversity for breeding selection. For example, an EMS treated rice population

carried ~2 mutations per Mb [18] whereas in wheat the mutation load was close to ~20 mutations per Mb [19]. On the other hand, also because of the polyploidy characteristic, the phenotypic effects of mutations caused by EMS can be masked in wheat [20], making phenotypic selection less effective than in other diploid species. It is therefore usually necessary to cross wheat lines with different mutated homoeologs to get double or triple knock-out mutants for phenotypic evaluation, for which the process is time-consuming [16].

The main problem of applying random mutagenesis methods to breeding is that mutations from these methods are randomly distributed across genome, affecting multiple genes in each mutated individual. Positive phenotypic changes created by some mutations can be masked by other background mutations, which require several rounds of backcrossing for their elimination. Similar to natural beneficial alleles, the causal mutations need to be genetically mapped for marker development and deployment in breeding programs [21]. Therefore, although the random mutagenesis methods have the ability to introduce new variations into crops to improve traits, they require labor-intensive screening and selection procedures.

1.1.3 Gene editing strategy to introduce targeted genome modification and accelerate breeding

Since the 1970s, during which the recombinant DNA technology was developed, scientists have gained the ability to manipulate DNA molecules, study genes and harness them to develop novel biotechnology-based solutions to practical problems. In recent years, we witnessed another revolution in molecular biology field – the possibility of genome editing. A series of programmable nuclease-based genome editing technologies have been developed, enabling targeted and efficient genome modification of a variety of species [22].

Zinc finger nucleases (ZFNs) were developed during 1990s by combining the DNA-binding domains from Zinc finger transcription factors and the cleavage domain of restriction endonuclease Fok I [23]. This was followed by the development of a new gene editing tool referred to as transcription activator-like effector (TALE) nucleases (TALENs) [24]. Similar to ZFNs in principle, TALENs include TALEs as a DNA binding domain and Fok l as a cleavage domain. Multiple studies showed that TALENs are generally easier to design, more mutagenic and specific than ZFNs [24] [25].

The most recent addition to the gene editing toolbox was the development of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) /CRISPR-associated system (Cas) [26], in which Cas9 is the most popular subcategory [27]. DNA binding specificity in the CRISPR system is conveyed by the short pieces of RNAs called guide RNAs or gRNAs, instead of the DNA-binding protein domains in ZFNs and TALENs. Because the RNA-DNA complementary principle is more straightforward and RNA engineering is much easier, CRISPR/Cas9 quickly superseded ZFNs and TALENs and became the technology of choice for most medical and breeding applications [22] [28].

CRISPR/Cas9 can introduce DNA modifications to specific regions of the genome simply based on the sequence. Through selective targeting of coding sequences or regulatory elements, CRISPR/Cas9 can be used to change either amino acid sequences or gene expression to improve crop traits [29]. By targeting the α-gliadin gene family of wheat on a conserved region adjacent to coding sequence for the main immunodominant peptides triggering coeliac disease, CRISPR/Cas9 has been used to successfully generate low-gluten wheat lines, which were previously unable to be obtained using traditional mutagenesis and plant breeding approaches [30]. Another example that shows the utility of CRISPR/Cas9 for breeding is the

modification of regulatory elements of genes regulating major productivity traits in tomato [31]. In this study, researchers created a continuum of phenotypic trait distribution by editing the promoter region of a single gene. Using this strategy, it should be possible to create allelic series of genes with varying effect sizes. A number of genes controlling pathways and processes underlying important agronomic traits should be amenable to this kind of gene engineering. Therefore, CRISPR/Cas9 based genome editing has the potential to create new variation for diverse agricultural traits.

When compared with the traditional breeding approaches, CRISPR/Cas based genome editing has its limitations. The main one is that CRISPR/Cas requires prior knowledge about the targets of editing. This information, unfortunately, is limited for the majority of agronomic traits of interest. On the contrary, MAS or genomic selection and random mutagenesis don't require any information about the genes involved in trait control. In addition, traditional methods do not raise safety concerns from public, whereas regulatory aspects of genome editing technology remain hotly debated among public, academy and government regulatory agencies [32].

Today, advances in genome sequencing and the development of functional genomic resources accelerated the discovery of genes controlling agronomic traits in all major crops [1]. In the light of these developments, genome editing is becoming an attractive technology to accelerate the trait improvement in breeding programs. At the same time, government agencies of several countries, including USA, have decided not to regulate gene edited crops [33], making broad deployment of CRISPR/Cas9 in crop research and breeding programs possible.

1.2 Resources available for selecting candidate genes for editing

Many resources are available for selecting candidate genes for editing. For crop breeding, the most useful information includes the functionally characterized genes in model plants along with comparative genomics analyses, the intervals from QTL (quantitative trait loci) mapping or GWAS (genome wide association study) with annotated genes and the information from gene network studies.

The genomic information available for crops, such as rice, wheat, maize, barley and sorghum, from grass family (Poaceae) has helped to establish syntenic relationships among their genomes. Comparative analyses have shown that they still share high levels of gene content and order conservation [34]. Because orthologs have the same ancestral origins, information about gene function can be transferred among orthologs in different species. The utility of functional information gathered in *Arabidopsis* and rice, which have the best-studied genomes, for understanding gene function in other plant species was demonstrated in multiple studies [35]. Gene editing can be conveniently used in such situations, using information from model organisms to find targets in other species for mutagenesis and functional analysis.

Genomic intervals identified in QTL mapping or GWAS can also be used to select candidate gene for editing. Although the mapping studies in crop species with complex genomes, like wheat, are complicated, a lot of efforts has been devoted to mapping regions of the wheat genome that contribute many important agronomic traits including yield [36] [37]. Gene editing can be a very effective method for follow up functional analyses and can help to identify a causal gene within the region.

Gene-interaction and gene network studies [38] can also provide information for choosing editing candidates. With the development of sequencing technology, sequencing the

entire transcriptomes of multiple biological samples has become possible and this approach, known as RNA-seq [39], has become a popular method for building gene co-expression network [40] [41]. By adding eQTL or ChIP-seq data, these networks can be converted to directional gene regulatory networks that provide a more straightforward evidence for regulatory interactions [42] [43]. A number of studies made an attempt to connect regulatory networks with phenotypic variation [44]–[46] thereby identifying numerous genes involved in controlling molecular processes underlying some of the important agronomic traits. These trait-connected networks can provide promising targets for gene editing beyond those identified by mapping studies.

1.3 CRISPR/Cas9 system

1.3.1 History of CRISPR/Cas9

CRISPR is an array of DNA repeats found in bacterial and archaeal genomes that usually follows a clustered set of CRISPR-associated (Cas) genes. They were first reported as early as the 1980s [47], and referred to as “CRISPR” after 2002 [48]. However, the function of these repeats remained unknown until 2005, when different groups nearly concurrently made the observation that the seemingly random sequences (spacers) separating identical direct repeats actually show homology to invasive DNA sequences, such as viruses and plasmids DNA pieces (protospacers) [49]–[51]. These observations led to the hypothesis that CRISPR are central components of an adaptive immune system, the hypothesis that was later verified [52]. By 2010, the basic function and mechanisms of CRISPR systems were solved. The CRISPR loci are transcribed and processed into a library of small CRISPR RNAs (crRNAs) that guide Cas nucleases to complementary nuclear acid sequence targets, which, in most cases, creates DNA

double strand breaks (DSBs) [22]. This sequence-specific nuclease activity manifests itself not only in bacteria, but also in eukaryotes, which makes the system valuable for biotechnology. Several CRISPR-associated nucleases, e.g., Cas9 and Cpf1, have been quickly adopted and developed into effective genome editing tools, which have literally brought in a revolution in biology.

Doudna and Charpentier were first who found trans-activating CRISPR RNA (tracrRNA) that is complementary to the sequence of the CRISPR repeats in crRNA and plays a critical role in the processing of pre-crRNA [53]. By combining purified Cas9 and a single guide RNA (sgRNA, afterwards also referred to as gRNA, a fusion of crRNA and tracrRNA) they showed the potential of CRISPR/Cas9 system for programmable genome editing [53]. And the first case of CRISPR/Cas genome engineering in eukaryotic cells was reported in 2013 by Feng Zhang's group from MIT. They used codon optimized S. pyogenes Cas9 with the attached nuclear localization signals (NLS) to direct Cas9 to the nuclei of mammalian cells [26]. This study successfully tested the system for targeted gene editing, as well as showed that the mutated variant of Cas9 called Cas9n nickase can be used for homology-directed repair (HDR), and that a single CRISPR array encoding multiple crRNAs enables multiplex editing [26]. These two studies founded the base for the on-going CRISPR Craze.

1.3.2 Application of CRISPR/Cas9 based genome editing in plants

When it comes to the practice of applying CRISPR/Cas system to research, there are many details to learn and consider, including how to choose a correct version of Cas enzymes at the outset [27]. In many cases, it depends on the protospacer adjacent motifs (PAMs) you can find within the sequences you want to target. A PAM is a 2-5 nucleotide motif next to the target

sequences that's required for target recognition in most CRISPR/Cas system. It is 5'-NGG-3' for Cas9 and 5'-TTN-3' for Cpf1, a Cas9 ortholog. Cpf1 recognizes its PAM at the 5' end of target sequences, while Cas9 recognizes that at the 3' end of targets. Next, since these enzymes were borrowed from prokaryotic organisms, we need to optimize the system before we can use it in animals or plants, such as codon optimization and NLS addition. What's more, we should know that even one Cas enzyme can have different forms. dCas9 (dead Cas9) and Cas9n (Cas9 nickase) are the Cas9 variants that lose both or one of the functional nuclease domains, respectively. Such variants highly broaden the range of application of Cas9. For example, tagging a fluorescent dye to the dCas9 enables visualization of specific DNA fragments in the genome [54], and using two nickases targeting opposite strands of DNA each with different guide RNAs could create a double-strand break with high specificity [55].

When DNA DSBs are created by Cas9 in cells, they can be repaired by the endogenous, error-prone non-homologous end joining (NHEJ) pathway, which usually introduces a small insertion or deletion (InDel) at the site. Therefore, we are able to knock out a gene by targeting its coding sequence with Cas9 to introduce InDels that may disrupt translation. If we want to do more than just knocking out a gene, such as accurate insertions or substitutions, then we need to utilize the other DNA repairing pathway, homology-directed repair (HDR). That is, if a donor DNA fragment with homology to the flanking sequence is present, the DSB created by Cas9 can be repaired through HDR using the donor DNA as a template. As a result, the sequence of the donor fragment is integrated into the genome at the DSB site [22].

Cas9 can be used not only to edit genomic sequence, but also to regulate transcription. Early experiments in animals suggested that dCas9 fused with a transcriptional activation domain, e.g., VP64, and guided by gRNAs to the promoter can activate the expression of the

target gene. In plants, gene activation and suppression using dCas9 and gRNA was verified using the agroinfiltration transient expression system in *Nicotiana benthamiana* leaves [29], [56].

Multiplex genome editing is of great value in both basic research and practical applications since higher organisms often use complicated genetic networks to fine tune cellular processes. Cas9 system is easy to use for multiplex editing because different gRNA expression cassettes can be integrated into one plasmid, thus guiding Cas9 to different targets [57]. Nevertheless, limitations of plasmid capacity and delivery method determined that it's still a challenge to deliver numerous gRNA units into cells at the same time for most organisms, especially crops [58]. Therefore, two RNA processing systems that can produce numerous gRNAs from a single polycistronic gene, the Csy4 and the tRNA processing system, have been engineered to increase CRISPR/Cas9 multiplex editing capability and efficiency [58]. Csy4 is an RNA endonuclease from the CRISPR/Cas system in Pseudomonas aeruginosa and specifically binds and cleaves a 28-nt RNA sequence [59]. When this RNA sequence flanks gRNAs in presence of Csy4, single gRNAs will be released after Csy4 cleavage. Successful application of this method in human cells has been reported [60] but not yet in plants. The other strategy is smartly using the endogenous tRNA processing system existing in all living cells, where primary transcripts of tRNA are precisely cut at both ends by endogenous RNases. Because the tRNA structure itself is sufficient for the recognition and cleavage, artificial genes with multiple gRNAs sharing one promoter but separated by tRNA sequences could directly be used for multiplex genome editing. The tRNA method was first reported in rice in 2015 [58] and has been applied to different crops, such as maize [61] and wheat [62].

1.4 Wheat genome editing

1.4.1 Wheat genome structure and its effect on gene editing

Bread wheat (Triticum aestivum L., AABBDD) is an allohexaploid (2n=6x=42), with three sets of homoeologous genomes (A, B and D). The polyploid genome structure may be one of the reasons for the successful adaption of wheat in diverse climate conditions all over the world. Evolutionary studies showed that this structure is the result of two recent, natural polyploidization events. The first hybridization event occurred several hundred thousand years ago between *T. urartu* (AA genome) and a species of the *Sitopsis* group (BB genome) and created wild tetraploid wheat *Triticum turgidum* (AABB genome). The other hybridization event was between *T. turgidum* and a wild goatgrass *Ae. tauschii* (DD genome) less than ten thousand years ago, forming the allohexaploid bread wheat *T. aestivum*, which currently accounts for more than ninety percent of the wheat grown worldwide [63].

The complexity of wheat genome requires special gRNA designs for gene editing. For example, gRNAs can be designed to target regions conserved among the three wheat genomes allowing for editing all copies simultaneously and producing triple knock-out mutants. gRNAs designed to a genome-specific region will result in only the selected gene copy being edited, thereby facilitating functional analyses of each gene copy and allowing for studying the interactions among wheat subgenomes. Thus, different design strategies give us opportunities to apply genome editing to address various aspects of polyploid genome biology.

In addition to the polyploid nature, wheat genome harbors a number of repetitive sequences constituting more than 80% of the 17 Gb-size genome [64]. This complexity was one of the main factors for slow progress with wheat genome assembly, which was only recently released [65]. The availability of genome sequence substantially facilitates genome editing

efforts by providing a resource for highly specific target editing without significant off-target effects, which should be avoided as much as possible in a genome engineering studies.

1.4.2 Cas9 editing in wheat – case studies

Application of CRISPR/Cas9 to wheat was first reported on wheat protoplasts [66] and cell suspension cultures [67] in 2013. In the latter research, gRNAs targeting multiple genes were expressed from one construct and caused InDel mutations at all tested sites. This study also showed that deletions of a gene fragment can be achieved by simultaneous cleavage at two targeted sites within one gene, which is a useful method for generating long deletions at the specific regions of genome and developing knockout mutants.

In 2014, a stepwise protocol for genome editing in rice and wheat was published [68]. The same author also collaboratively published another paper that presented a successful simultaneous editing of three homoeoalleles of *TaMLO* in hexaploid bread wheat that confers heritable resistance to powdery mildew [69]. They proved that both TALENs and CRISPR/Cas9 system can be used to generate beneficial genetic traits in wheat, establishing genome editing as a promising molecular breeding strategy.

An optimized multiplex gene editing strategy for wheat was established by our lab in 2016 [62]. We used the transient expression assay in wheat protoplasts followed by next-generation sequencing (NGS) to assess the targeting efficiency of gRNAs. We also assembled wheat codon-optimized Cas9 as well as the Golden Gate-compatible tRNA processing-based gRNA cassette to expand the CRISPR/Cas9 toolbox capability for editing the hexaploid wheat genome.

1.5 Yield component traits with focus on seed morphology

1.5.1 Yield and yield components

Yield is a complex trait defined by a number of components including the number of spikes per unit area, number of spikelets per spike, biological yield, seed size and so on [70]. Seed size is the main trait selected during wheat domestication and breeding [71]. A trait closely related to seed size is Thousand Grain Weight (TGW), which is a key grain yield component and often used as a grain receival and marketing standard [72]. In addition, seed size can be divided into several factors including seed length, seed width, and seed area. The length/width ratio determines seed shape, which has recently become a breeding target in wheat because it influences the milling performance [71].

Each of the yield component traits is controlled by molecular pathways with many participating genes. Genetic studies in many crops start unraveling the gene components of these pathways building foundation for targeted selection of yield component genes in breeding programs. It remains critical to sustain efforts aimed at the identification and functional validation of yield component genes in all crops, including wheat, and at understanding how these genes interact with each other. These efforts will help to build foundation for modern precision breeding strategies relying on understanding the molecular pathways that underlie yield component traits.

1.5.2 Identified yield component genes and adapted editing strategies

Recent advances in molecular genetics have enabled the mapping and the molecular characterization of many QTLs that control yield/yield components in model crops, such as rice. As a major crop for human consumption, rice has 40-fold smaller genome size than wheat. Since

early 2000s, following the completion of rice genome sequencing [73], more than 40 genes controlling seed size have been discovered in rice [74]. On the contrary, only a few wheat genes related to yield components have been identified or validated to date [74]. Considering the high level of rice-wheat synteny, wheat orthologs of rice yield component genes would be good candidates for wheat genome engineering to accelerate wheat breeding and genome study.

These rice genes can be generally categorized into three pathways: proteasomal degradation, G-protein signaling, and phytohormone signaling [74]. They regulate seed size in either a positive way or a negative way. Therefore, the effect and biological nature of a beneficial allele should be considered while developing gene editing strategy for wheat. Only those genes that negatively regulate yield component traits can be easily modified using Cas9-based targeted knock-out mutations. Otherwise, Cas9 system need to be adapted to produce duplicated genes by inserting an additional copy, to activate gene expression, or to create nucleotide substitutions within coding regions to create gene variants that can affect yield component traits in a positive way.

GW2 [75], *GW5* [76], *GW7* [77], *An-1* [78], *CKX2* [79], *GS3* [80], and *GSE5* [81] are some of the genes identified within the major yield component QTLs in rice that can affect grain size. *GW2* gene, first characterized in 2007, was shown to encode a RING-type E3 ubiquitin ligase and negatively regulate grain width and weight in rice [75]. Our lab has applied our wheat genome editing pipeline [62] on *TaGW2*, and demonstrated the effect of *TaGW2* mutations on wheat grain size [82]. In that study, single gRNA targeting the first exon of all three homoeologs of *TaGW2* was used to successfully create single, double, and triple mutants. Phenotypic evaluation of these wheat mutants not only validated *TaGW2* as a negative regulator of grain size, but also displayed the additive effects of the homoeologous alleles, which is valuable

information for breeding. This pipeline was proven to be an effective approach for gene editing in wheat, and therefore, the other rice genes affecting yield component traits, which have not been validated or well-studied in wheat, were selected as the candidates for editing in our project.

2 Methods

2.1 Candidates selection

The genes which are proved to be negative regulators of yield component traits in rice were selected. Their CDS (coding sequence) were used to perform BLASTN searches against the wheat genome (TGAC v1) at the Ensembl Plants website (http://plants.ensembl.org/Triticum_aestivum/Tools/Blast?db=core) and the recent released wheat genome reference IWGSC RefSeq v1.0 at URGI (https://urgi.versailles.inra.fr/blast_iwgsc/?dbgroup=wheat_iwgsc_refseq_v1_chromosomes&program=blastn). The wheat orthologs of these genes were determined by the highest similarity and the synteny between rice and wheat [83]. Only those genes that have >80% of the sequence covered by alignment and are located in syntenic chromosomal regions are considered as orthologs. The annotated sequences of the wheat orthologs were downloaded and used for pre-editing analyses.

2.2 gRNA design

The CDS of the genes of interest were submitted to sgRNA Scorer website (https://crispr.med.harvard.edu/) to design gRNAs. For all the returned gRNA target candidates, the top scored ones were BLASTN-compared against the IWGSC RefSeq v1.0. Then we selected those gRNAs that can specifically target our genes of interest. If more than 13 out 20 bp from 3' end of the gRNA guide sequence completely matched any other site in the wheat genome, we discarded these gRNAs. In rare cases, when gRNAs with few off-targeting sites was not discarded, the off-targeting information was saved for further analysis.

2.3 Plasmids construction

We used plasmid constructs to deliver CRISPR/Cas9 system into wheat. Our lab developed a plasmid, named as pA9Cas9sg, containing both a wheat codon optimized Cas9 expressing cassette and an incomplete gRNA expressing cassette (lacking the guide sequence of gRNA). Complete CRISPR/Cas9 construct was created by inserting a synthesized guide sequence into the designed site, as shown in Figure 1.

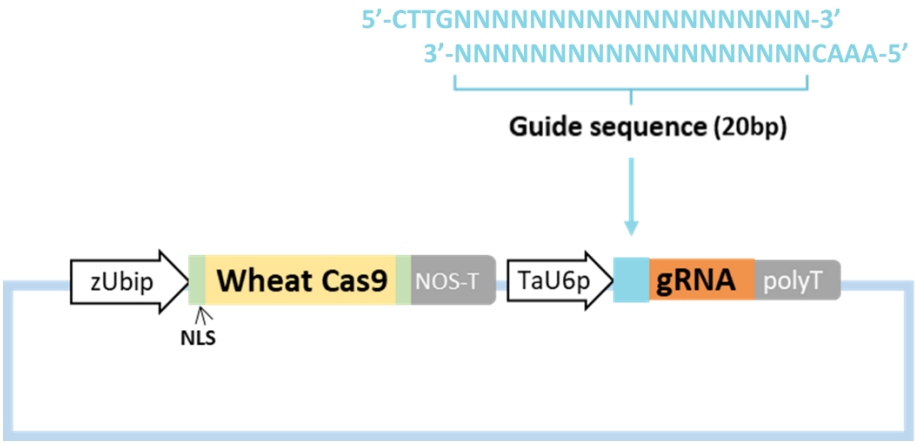

Figure 1. Strategy for building CRISPR/Cas9 constructs.

Promoters are shown as white arrows. Terminators are shown as gray rectangles. zUbip, maize ubiquitin gene promoter; TaU6p, wheat U6 promoter. Green rectangles correspond to nuclear location signal signals (NLS). The light blue rectangle corresponds to the 20 bp guide sequence insertion site in the synthetic single guide RNA (gRNA). The 5'-tails on both forward and reverse strands of the synthesized guide sequence fragment were used for Golden Gate ligation.

The details of pA9Cas9sg design were described previously [62]. It contains a wheat codon optimized Cas9 driven by maize ubiquitin promoter and a gRNA scaffold driven by wheat U6 promoter. The construct pBUN421 is ordered from Addgene (https://www.addgene.org/) and contains a maize codon optimized Cas9 driven by maize ubiquitin promoter and a gRNA scaffold driven by wheat U3 promoter.

We ordered 20 bp guide sequence of gRNAs as synthesized oligonucleotides from Integrated DNA Technology (IDT, https://www.idtdna.com/). Here, the forward strand specifically refers to the strand with PAM following the 20 bp target site in the genome. Synthesized guide sequences had CTTG added to their 5' ends on the forward strand and AAAC added to 5' end on the reverse strand (Figure 1). These two tails serve as over-hangs in the Golden Gate cloning reaction. Because the U6 promoter initializes transcription from nucleotide G, if the guide sequence (forward strand) itself starts with G, we would add only CTT to its 5' end, and delete the last base of 3' end of the reverse strand to keep the 4-nucleotide sticky end for ligation in Golden Gate reaction.

The oligos were annealed in a reaction mix containing 2μL of 10× annealing buffer (200mM Tris-HCl, 100 mM $(NH_4)_2SO_4$, 100 mM KCl, 1% Triton X-100, 20 Mm $MgCl_2$, pH 9.0), 2μL of 100 μM forward oligos, 2 μL of 100 μM reverse oligos and 14 μL ddH_2O. The reaction mix was incubated at 94℃ for 5 min, followed by the gradual 1℃ decrease per min until it reached 30℃.

In the Golden Gate reaction [84], the overhangs at both ends of the oligos were designed to match the sticky ends of BsaI-cut pA9Cas9sg vector. The Golden Gate reaction mix included 2 μL of annealed target, 0.5 μL of vector (0.6 μg/μL), 2 μL of 10× T4 DNA ligase buffer, 2 μL of BSA (1 mg/mL), 1 μL of T4 ligase (400 U/μL), 0.5μL of BsaI-HF (20 U/μL), and 12 μL ddH_2O. The mix was incubated for 50 cycles of 5min in 37℃ and 10min in 16℃, followed by final incubation at 50℃ for 30 min. The Golden Gate reaction products were transformed into competent cells of *E. coli* strain DH5α by heating for 1 min at 42℃. The bacteria was plated on 2YT media with ampicillin and grown at 37℃ overnight. Then colony PCR was conducted to test for the presence of the inserted targeting sequence, using primers NosF and targetR. For the

positive colonies, plasmid DNA was extracted using QIAprep Spin Miniprep Kit (Qiagen) from increased bacteria culture and the gRNA sequences were validated by Sanger sequencing.

For the double-gRNA constructs built before plant transformation, the construction procedure was similar. The only difference was that instead of one single gRNA guide sequence, a unit "gRNA1-tRNA-gRNA2 was inserted into pBUN421. This unit was produced by PCR using gRNA1 and gRNA2 sequence tails in the forward and reverse primers, respectively, to amplify the sequence of gRNA scaffold fused with a tRNA from a polycistronic tRNA-gRNA (PTG) construct serving as a template. See more details in the reference [62].

2.4 Protoplast transformation

Protoplast transformation was performed using a modified protocol developed for *Arabidopsis* [85]. Seedlings of *Triticum aestivum* cv. Bobwhite were grown in dark at 25°C for more than 2 weeks. Shoot tissues from about 50-100 seedlings were cut into 1-mm pieces with a razor blade and transferred into a flask with W5 solution (0.1 % glucose, 0.08 % KCl, 0.9 % NaCl, 1.84 % $CaCl_2$•$2H_2O$, 2 mM MES-KOH, pH 5.7) that can immerse all the tissues. To infiltrate solution into the tissues, vacuum (-600MPa) was applied to the flask for 30 min in dark. Then the W5 was replaced by 30 mL digestion solution, which includes 0.15 M Sorbitol, 0.25 M sucrose, 35 mM $CaCl_2$, 20 mM KCl, 0.5% Cellulase R10 (From Trichoderma Viride, 7.5 U/mg), 0.25 % Macerozyme (R10 Macerating enzyme from Rhizopus sp. RPI) and 10mM MES-KOH (pH 5.7). The chemicals were ordered from Sigma Aldrich. After incubation at room temperature for 2 hours under dark with gentle shaking at 20-30 rpm, the digested tissues were filtered through a 40 μm nylon mesh into a centrifuge tube, the tissues were rinsed with 30 mL W5 and filtered again. The tube was centrifuged for 7 min at 100 g at room temperature. Then the

precipitated protoplasts were washed 2 times with 10 mL W5 and 5 min centrifugation at 50 g. Finally, protoplasts were resuspended in 5 mL of W5 and incubated on ice for 30 min.

Before transformation, W5 was removed and protoplasts were resuspended in MMG solution (0.4 M mannitol, 15 mM $MgCl_2$, 4 mM MES-KOH, pH 5.7). The protoplast count was adjusted to 106 protoplasts/mL. 10µg of plasmid DNA in 10-20 µL volume was gently mixed with 100µL of protoplasts followed by adding 130 µL of PEG-calcium transformation solution (40 % PEG4000, 0.2 M mannitol, 100 mM $CaCl_2$). The resulting solution was mixed by gently shaking the tube and incubated at room temperature for 30 min. Then the mix was diluted with 500 µL W5 to terminate the transformation, after which protoplasts were collected by centrifuging at 100 g for 2 min and resuspended in 1 mL W5.

After 18 hours of incubation at room temperature, GFP signal in the protoplasts transformed with pA9-GFP [62] was detected using the Zeiss LSM 780 confocal microscope with the excitation wave length of 488 nm. The efficiency of transformation was calculated as the fraction of fluorescence-positive protoplasts. When the efficiency was at least 10%, we continued the experiment. Usually it was 20-40 %; this transformation efficiency was later used to normalize the gRNA editing efficiency. The CRISPR/Cas9 transformed protoplasts were kept in dark at room temperature for 48 hours after transformation.

W5 solution was removed by carefully pipetting it out and leaving the protoplasts at the bottom of the tubes for DNA isolation. We isolated protoplast DNA with the PureLink Genomic DNA Mini Kit (Thermo Fisher Scientific, Cat. No. K182002) following the manufacture's protocol.

2.5 Estimating editing efficiency by NGS

To detect mutations induced by CRISPR/Cas9, primers were designed to amplify genomic regions harboring the gRNA targets. To simultaneously analyze the large number of samples, NGS (next generation sequencing) was applied. Two rounds of PCR were used to add barcode to each amplicon. During the first round, we used target-specific forward and reverse PCR primers modified by adding to the 5' ends a 30-nucleotide long sequence complimentary to a part of the Illumina's TruSeq adaptors [62]. The second round of PCR was performed using primers that can complete the TruSeq adaptor sequences and add amplicon-specific barcodes. PCR was performed in 15 µL reaction volume containing 7.5 µL NEB Next High Fidelity 2× PCR Master Mix, 1-6µL template DNA, 0.5µL primers (5µM for each primer) and ddH_2O. All amplicons were pooled in equimolar proportions sequenced at the K-State Integrated Genomics Facility on a MiSeq instrument using the 600 cycles MiSeq reagent v3 kit.

Quality-trimmed and de-multiplexed reads were aligned to each of the three homoeologous copies of the genes using BWA followed by SNP and indel calling with SAMtools [62]. The mutation rates in the targeted regions were estimated using a custom Perl script. The proportion of reads with indels in the 20 bp range of the targeted regions relative to the total number of aligned reads was calculated.

2.6 Regeneration and genotyping of transgenic plants

Wheat immature embryo transformation and plant regeneration were performed as previously described [86]. To isolate DNA of the regenerated plants, leaf tissues were sampled and homogenized in 400 µL of TPS buffer (100 mM Tris-HCl, 10 mM EDTA, 1 M KCl, pH8.0), then incubated 20 min in 75°C water bath. After centrifugation 5 min under 4500 rpm, 140 µL of

the supernatant was mixed with 140 μL of isopropanol and incubated 20 min at room temperature. DNA was precipitated, washed with 70% ethanol, and re-suspended in 150 μL of deionized water.

The presence of CRISPR-Cas9 constructs in the transgenic plants was validated by PCR using primers amplifying the Cas9 and gRNA expression cassettes [87]. The CRISPR-Cas9-induced mutations were examined using the NGS-based procedure. To increase the multiplexing capacity of barcoding, during first-round PCR we used primers with internal barcodes. For this purpose, four barcoding bases were added between the target-specific primer and Illumina TruSeq adaptors to index multiple plants under each Illumina TruSeq barcode [87].

2.7 Plant growth and phenotypic analysis

Plants were grown in the greenhouse under 12 h light per day for 1 month, and then grown until seed harvesting under 16 h of light per day at a temperature 21-24°C. For each experiment, a randomized complete block design with three biological replications was applied with plants randomly assigned to benches (blocks) in greenhouse. The MARVIN seed analyzer (GTA Sensorik GmbH) was used to measure the TGW, grain width, length, and area traits in the wild-type plants and the mutants. The mean phenotypic values were estimated for each plant and used for further analyses. A two-tailed Student's t-test was applied to assess the significance of the phenotypic differences between the wild-type and gene edited plants. Box and whisker plots were generated with the “boxplot” package implemented in R.

2.8 The *TaGW7* gene expression analysis

To estimate the relative gene expression levels of *TaGW7*-A1, -B1 and -D1 homoeologs in cv. Bobwhite, the tissues from spikes during anthesis were sampled and frozen immediately in liquid nitrogen. In total, nine different tissues for five biological replicates were collected. Seven tissues were sampled from the main tiller: 1-week-old endosperm, spike without endosperm, flag leaf, flag leaf sheath, lower leaf and leaf sheath (two leaves lower than flag leaf), and stem. One non-pollinated spike was sampled from another tiller of the same plants. The roots of the same plants were sampled as well. RNA was extracted using Trizol (Thermo Fisher Scientific, Catalog#: 15596018) following the manufacture's protocol. The first-strand cDNA was synthesized using kit "SuperScript™ III First-Strand Synthesis SuperMix for qRT-PCR" (Thermo Fisher Scientific, Catalog #: 11752-250) following the manufacture's protocol. The *TaGW7*-A, -B and -D homoeolog-specific primers were designed and validated by PCR using the DNA of Chinese Spring nulli-tetrasomic lines (Appendices Figure 9). Real-time PCR was performed using iQ™ SYBR Green Supermix (BIO-RAD, Catalog #: 170-8882) following the manufacture's protocol, using the TaActin gene as reference.

3 Editing and functional analysis of the candidate genes affecting wheat yield component traits

3.1 Results

3.1.1 Selection of candidate genes controlling yield component traits

Previous studies showed that the wheat ortholog of rice *GW2* gene contributes to the thousand grain weight (TGW) and grain size [82], suggesting that genes controlling yield component traits can be conserved between rice and wheat. To find more genes which contribute to TGW and grain size in wheat, gene editing by CRISPR/Cas9 was applied to further study the wheat orthologs of six genes that were shown to affect yield component traits in rice (Table1). All these genes, except for *GW7*, are negative regulators of TGW or grain number in rice. *GW7* regulates rice grain shape and higher expression of this gene results in longer grains with better quality [77] [88]. Here, we planned to use CRISPR/Cas9 to knock out all these genes to validate their functions in wheat.

Table 1. The list of candidate genes for editing using the Cas9-based system

Gene locus	Mapped in crop	Trait[a]	Chromosomes of wheat orthologs	Validated in wheat	Nature of beneficial allele[b]	Cas9 strategy [c]	Ref.	Gene function
GW7	rice	GW, GL, L/W	2AS, 2BS, 2DS	No	IE	KO	[77] [88]	TONNEAU1-recruiting motif proteins
An-1	rice	GL, GN	2AL, 2BL, 2DL	No	LF	KO	[78]	Basic Helix-Loop-Helix protein
CKX2	rice	GN / TGW	3AS*2, 3BS*2, 3DS*4	*TaCKX6*	RE	KO	[79]	Cytokinin oxidase／dehydrogenase
GS3	rice	GL, GW, GS, TGW	7AS, 4AL, 7DS	*TaGS-D1*	LF	KO	[80]	G-protein γsubunit
GSE5	rice	GW, TGW	1AS, 1BS, 1DS, (3AS, 3BS, 3DS)	No	RE	KO	[81]	Plasma membrane-associated protein /calmodulin-binding
GW5	rice	GW, TGW	-	No	LF	KO	[76]	Unknown nuclear protein

[a]GW: grain width; GL: grain length; L/W: grain length-to-width ratio; GS: grain size; GN: grain number; TGW: thousand grain weight. [b]IE: increased expression; LF: loss of function; RE: reduced expression. [c] KO: knock-out.

The BLASTN search detected wheat homologs of all these genes, except for *GW5*, which may have been deleted in the wheat genome. For *GW7*, *An-1* and *GS3*, only one copy of a gene was detected in each wheat subgenome (Table 1). The locations of these genes also had no conflicts with synteny between rice and wheat [83], suggesting that these genes are likely orthologs of rice genes.

More than one group of homoeologs was found in wheat for *CKX2* (*TaCKX6*) and *GSE5* (Table 1). To investigate the relationship among the duplicated gene copies, The NJ (neighbor-joining) tree based on the aligned coding sequences was built using the rice sequence as an outgroup (Figures 2 and 3).

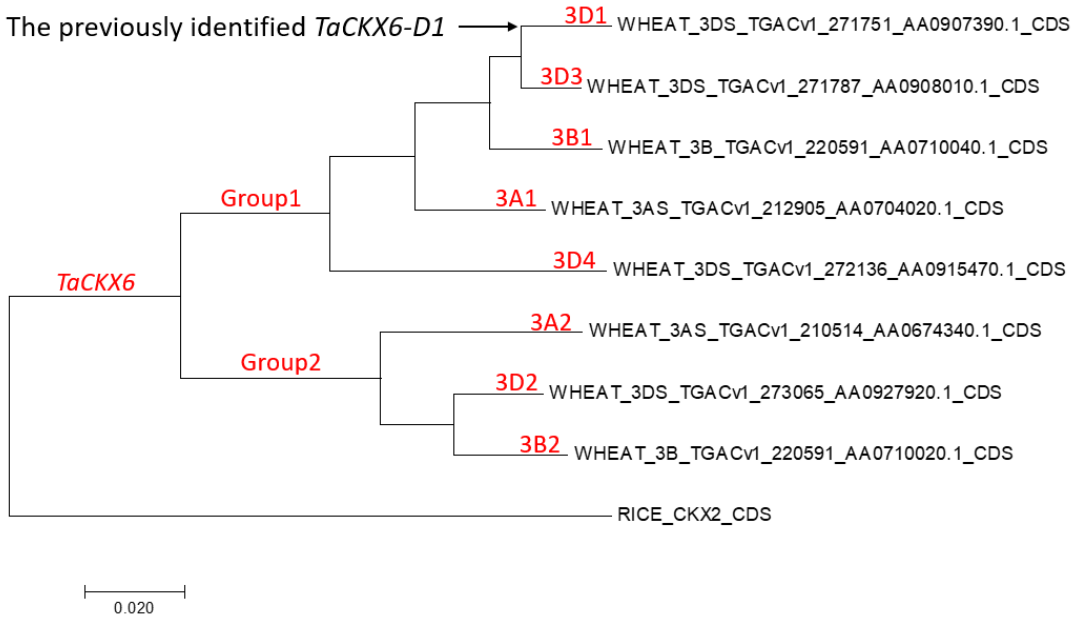

Figure 2. Neighbor-joining tree for *CKX2* homologs

The CDS of wheat homologs of rice *CKX2* gene were aligned by Cluster W (http://www.clustal.org/clustal2/) as implemented in Mega version 7.0 (Kumar et al., 2016). The tree was also generated using Mega version 7.0 with the default parameters.

By comparing rice *CKX2* CDS against the whole genome sequence, we identified eight gene copies located on chromosome 3 group. Two copies were located on 3AS, two on 3BS and four on 3DS. On the constructed NJ tree, gene copies formed two large groups (Figure 2). One of

the 3D gene copies in group 1 was previously shown to increase TGW in wheat and was named *TaCKX6*-D1 by association analyses [79]. Given the difficulties with identifying an exact orthologous copy of the rice *CKX2* gene, we opted to mutate all gene copies. A unique name based on the chromosome location was assigned to each gene for the convenience of further analyses: *TaCKX6*-A1, *TaCKX6*-B1, *TaCKX6*-D1, *TaCKX6*-D3, and *TaCKX6*-D4 fall into group 1; *TaCKX6*-A2, *TaCKX6*-B2, and *TaCKX6*-D2 fall into group 2 (Figure 2).

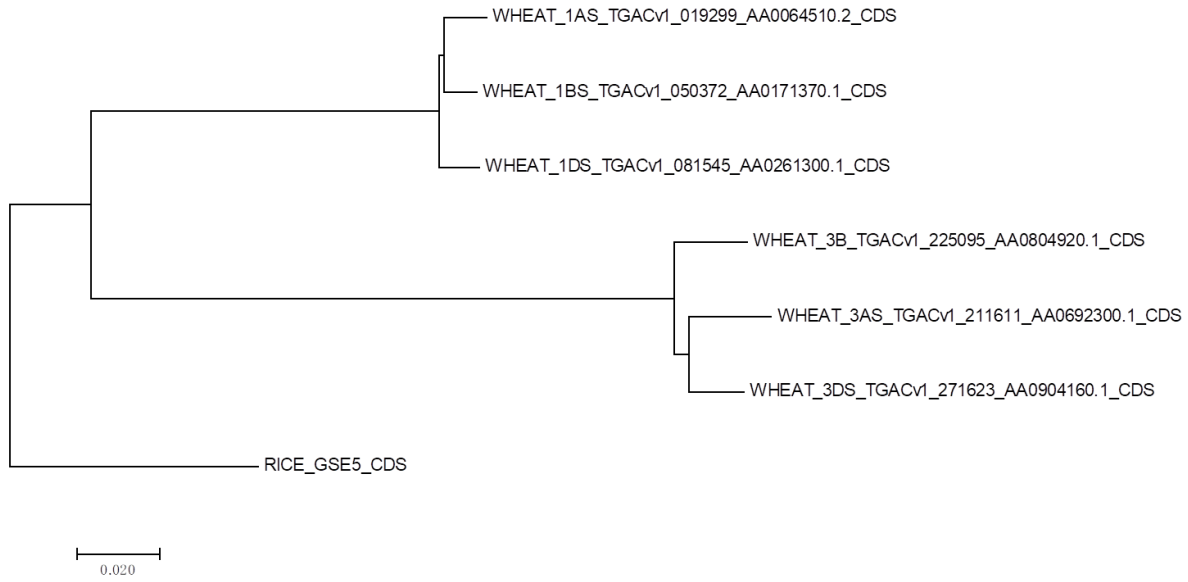

Figure 3. NJ tree for *GSE5* homologs

The CDS of *GSE5* homologs in wheat and rice were aligned by Cluster W (http://www.clustal.org/clustal2/) as implemented in Mega version 7.0 (Kumar et al., 2016). The tree was also generated using Mega version 7.0 with the default parameters.

There are two groups of rice *GSE5* gene homologs located on wheat chromosomes 1 and chromosome 3. According to syntenic relationships between rice and wheat genomes [83], *GSE5* located on rice chromosome 5 should correspond to the copy located on wheat chromosome 1. The NJ tree (Figure 3) shows that rice gene *GSE5* is closer to genes on chromosome 1. Hence, the homoeologous copies on chromosome 1 were orthologs of rice *GSE5* gene and were selected for CRISPR/Cas9-based mutagenesis.

3.1.2 gRNAs Design

The coding sequences (CDS) of selected wheat genes were submitted to sgRNA Scorer 1.0/2.0 website to search for the best target sites. From the top ranked target sites for each gene, 5 to 10 target sites were chosen for further validation in protoplasts. To find the most reliable CRISPR/Cas9 targets, we applied the following rules. First, top ranked target sequences were searched using BLASTN against the whole wheat genome, and only those that are specific to the homoeologs of the target gene were kept for further analyses. Second, target sites were ranked according to their position relative to the 5' end of CDS, because earlier stop codon or frameshift mutations are likely to have a more severe effect on gene function. Third, priority was given to the target sites located in the conserved regions of the homoeologous gene copies. Mutating all gene homoeologs simultaneously can increase our chances to detect phenotypic effects that can be masked by redundant non-mutated gene copies. The selected target sites were summarized in Table 2. Their positions on the genes were shown in Figure 4 and their genomic locations were shown in Appendices Table 9.

Table 2. List of the designed gRNAs

Gene	Target name	Target sequence (5' to 3')	Score[a]	Targeted gene copies[b]
TaGW7	GW7_T1	CGGCCATGCAGCCCATCTGG	66	2A,2B,2D
	GW7_T2	GCGCGGTGAGCATCCGCTGG	88	2A,2B,2D
	GW7_T3	CAGCAGAAGAGACTGCCCGC	95	2A,2B,2D
	GW7_T4	ACGAAGAACTCTTTCCGCAG	83	2A,2B
	GW7_T5	GAAACCGACGTTCGGCTGTG	98	2A,2B,2D
	GW7_T6	TCCATCAACCGGGACTCGGG	95	2A,2B,2D
TaGW7	GW7_T7	GTTGGAACGCCTGTGGCCCG	86	2A,2D
	GW7_T8	AGAGATTGCCAGGATGCGTG	74	2A,2B,2D
	GW7_T9	GCTCCCCTTGCAGGGCTGAG	99	2A,2B,2D
TaAn-1	An1_T1	CGAGACCAGAGAGCTGAGCG	76	2B,2D
	An1_T2	GAAGCCTAAGCCTCCGCCGG	88	2A,2B,2D
	An1_T3	GGCTTAGGCTTCACGCACGG	97	2A
	An1_T4	CCTATCAGCTCCCCTCCACG	79	2A,2B,2D
	An1_T5	GGGGCGCCGCAGGCGTTCCG	87	2A,2B,2D
	An1_T6	GCGCTTCCGCGCGGAACCGG	92	2A,2B,2D
	An1_T7	GGCCGCCTTGCGCTTCCGCG	67	2A,2B,2D

	An1_T8	TTCCGCGCGGAAGCGCAAGG	89	2A,2B,2D
	An1_T9	CGAGCCCGCCGAAGCGGCCG	95	2A,2B
	An1_T10	CGGCGACGAGGAAAACGCCG	98	2A,2D
TaCKX6 - Group1	CKX6_T1	ACACACGCAGGAGATGATCA	60	3B2,3D2,3D3
	CKX6_T2	GTCAGCCCGTGCGCCCTGGA	77	3D2
	CKX6_T3	GCAGCGCGGCAATGTCTGCG	99	3B2,3D2
	CKX6_T4	CCGGGCGGACACCGGGAACG	94	3A2,3B2,3D2
	CKX6_T5	GGACGCGGCCACTCCGTACG	83	3B2,3D2,3D3
	CKX6_T6	CCGCCGAGGCGTCCGAGCGA	73	3A2,3B2,3D2,3D3,3D4
	CKX6_T7	ACGCAGTGGAGCCTCCGCCG	95	3B2,3D2
	CKX6_T8	CACGGAGAGGCGGGACGCAG	80	3A2, 3B2,3D2,3D3
	CKX6_T9	GCGGCGTTAGGCCGTGCGCG	96	3D2,3D3
TaCKX6 - Group2	CKX6_T10	GCTTTACCCGTCGTGCCCCG	98	3A1,3B1,3D1
	CKX6_T11	GGCAGACACGGGGAACGGGG	98	3A1,3B1,3D1
	CKX6_T12	GGCTGCGGCCACTCGACCCG	96	3A1,3B1,3D1,3DX
	CKX6_T13	GGCCAGGCGTCCGCACCCCG	92	3A1,3B1,3D1
	CKX6_T14	CGCAGCCCCGGGCAGACACG	92	3A1,3B1,3D1
TaGS3	GS3_T6	GGCCGGCAATGGCGGCGCCC	84	7A,4A,7D
	GS3_T7	AAGTCCCCGCTCGACCCCTG	83	7A,4A,7D
	GS3_T8	GCGGCCGCAGGGGTCGAGCG	92	7A,4A,7D
	GS3_T9	GCAGCCGGTGGCGGCCGCAG	93	7A,4A,7D
TaGSE5	GSE5_T1	GCGGCTACGTCGGCGCGCGG	1.68	1A,1B,1D
	GSE5_T2	CCGCCGCGGCCGCCGGTCCG	1.37	1A,1B,1D
	GSE5_T3	GCGGAGCGAGCACGGGGTGG	1.40	1A,1B,1D
	GSE5_T4	CTGCCGTGCCACATGCCAGG	1.30	1A,1B,1D
	GSE5_T5	CGCACCAGTCGTACTCCGGG	1.31	1A,1B,1D
	GSE5_T6	CGACGGCGCAGTGCACGCCG	1.41	1A,1B,1D

[a]The scores were obtained using sgRNA Scorer versions 1.0 and 2.0. The version 1.0 returned scores ranging from 0 to 100 and version 2.0 returned scores ranging from -2.0 to 2.0.
[b]In the last column, letter A, B, and D and the numbers before indicate the targeted wheat chromosome and genome; the additional numbers in the target names for *TaCKX6* correspond to gene copy numbers from Figure1. Black color indicates target sites whose sequence exactly matches the sequence of gRNA sequence; grey color indicates minor difference between the target and gRNA sequences at the 5' end of the gRNA. These differences have minor effect on Cas9 function but may still reduce targeting efficiency. Red color indicates the presence of off-target sites.

Finally, 9, 10, 9, 5, 4, and 6 target sites were chosen for *TaGW7*, *TaAn-1*, *TaCKX6*-group 1, *TaCKX6*-group 2, *TaGS3*, and *TaGSE5* genes, respectively (Table 2). They are all located close to the 5' end of CDS (Figure 4) have relatively high scores based on sgRNA Scorer analyses. Most of them target 3 homoeologous gene copies simultaneously, while some of them target only 1 or 2 genome (Table 2). The ones not targeting conserved regions either have very high scores or are close to start codons. Out of these gRNAs, only CKX6_T12 has a possible off-

targeting site located on chromosome 3D at position different from that of the *TaCKX6* locus. After the gRNAs targeting sites were chosen, DNA oligos for all guide sequences were synthesized and subcloned into construct pA9Cas9sg for experimental validation of their gene editing efficiencies.

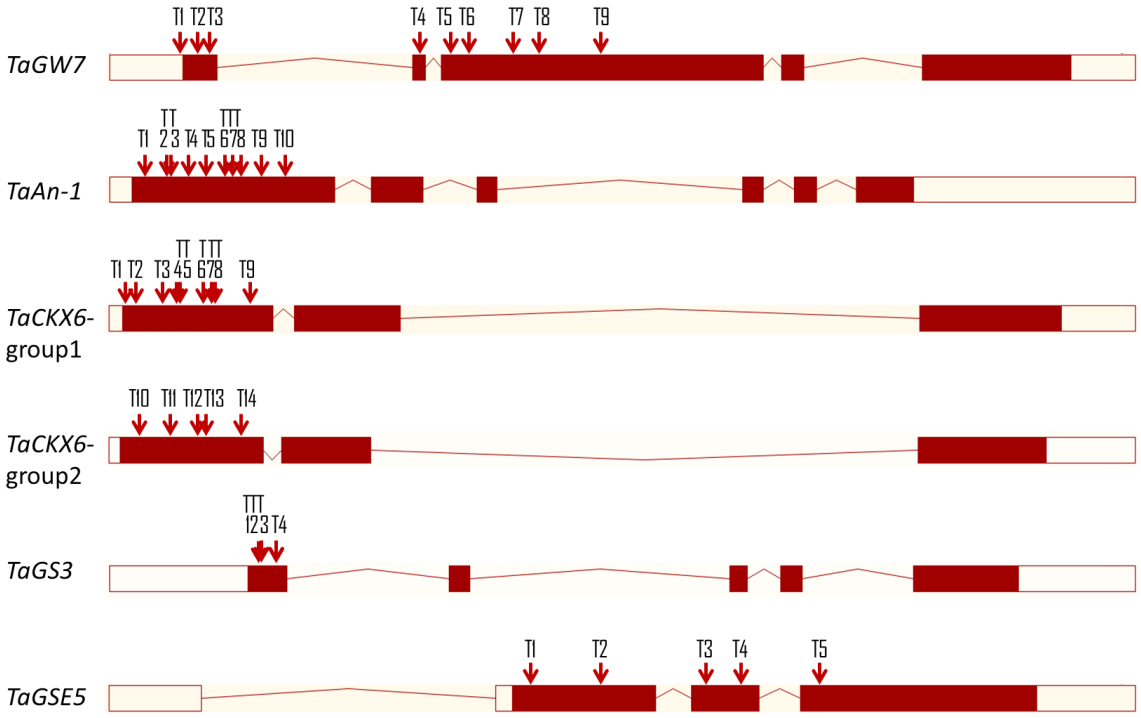

Figure 4. Positions of the gRNA target sites in the genes selected for editing

The arrows denote the positions of the gRNA target sites on each gene. The gene models were downloaded from Ensembl Plants. Red block, coding sequence; empty block, UTR; connecting line, intron.

3.1.3 Assessments of gRNA targeting efficiency in wheat protoplasts

To check the editing efficiency of the designed gRNAs, the pA9Cas9sg constructs with different guide sequences were transiently expressed in the wheat protoplasts. And then the genome editing efficiencies were calculated based on ratio of mutated reads and total reads generated by NGS (next generation sequencing) from PCR products flanking the target sites.

Table 3. Editing efficiency for each gRNA in wheat protoplasts

Protoplast transformation- 01312017 (transformation efficiency 30%)											
	replicate 1					replicate 2					
Gene	Target	Genome	total reads #	Efficiency	Average efficiency	Target	Genome	total reads #	Efficiency	Average efficiency	Average between replicates
GW7	T1	2A	23238	1.46%	1.07%	T1	2A	15202	0.72%	0.46%	0.77%
		2B	36197	0.83%			2B	19575	0.27%		
		2D	4234	0.94%			2D	1593	0.31%		
	T2	2A	28463	1.62%	1.74%	T2	2A	6086	0.36%	0.48%	1.11%
		2B	37615	1.86%			2B	7795	0.49%		
		2D	5272	1.50%			2D	716	1.40%		
	T3	2A	69009	0.16%	0.33%	T3	2A	49364	0.15%	0.51%	0.42%
		2B	76962	0.60%			2B	49559	0.85%		
		2D	75406	0.22%			2D	49869	0.53%		
	T4	2A	4187	1.41%	0.10%	T4	2A	886	0.00%	0.06%	0.08%
		2B	54360	0.05%			2B	14386	0.06%		
		2D	56867	0.05%			2D	15010	0.05%		
	T5	2A	3345	0.69%	0.92%	T5	2A	2124	0.47%	0.35%	0.63%
		2B	45162	0.54%			2B	32220	0.27%		
		2D	48838	1.28%			2D	33213	0.42%		
	T6	2A	3500	8.60%	10.94%	T6	2A	1091	4.22%	5.73%	8.33%
		2B	56707	10.95%			2B	14588	5.62%		
		2D	58901	11.07%			2D	14992	5.94%		
	T8	2A	5467	0.29%	0.24%	T8	2A	2466	0.00%	0.09%	0.16%
		2B	64555	0.14%			2B	27739	0.16%		
		2D	66941	0.32%			2D	31034	0.04%		
	T9	2A	48599	1.29%	1.52%	T9	2A	34911	1.23%	0.93%	1.23%
		2B	46394	1.56%			2B	34707	0.73%		
		2D	46096	1.72%			2D	36070	0.83%		
	replicate 1					replicate 2					
Gene	Target	Genome	total reads #	Efficiency	Average efficiency	Target	Genome	total reads #	Efficiency	Average efficiency	Average between replicates
An-1	T2	2A	34303	19.14%	15.23%	T2	2A	35027	3.84%	2.72%	8.98%
		2B	38511	13.41%			2B	37590	2.41%		
		2D	37078	13.52%			2D	35340	1.95%		
	T3	2A	30911	0.01%	0.02%	T3	2A	23636	0.05%	0.04%	0.03%
		2B	34211	0.04%			2B	23546	0.04%		
		2D	33683	0.02%			2D	22385	0.04%		
	T4	2A	45618	0.26%	0.27%	T4	2A	33521	0.16%	0.12%	0.20%
		2B	46653	0.40%			2B	33239	0.11%		
		2D	49287	0.15%			2D	33095	0.10%		
	T6	2A	27413	7.32%	7.51%	T6	2A	7872	2.57%	2.83%	5.17%
		2B	66029	7.32%			2B	29963	2.88%		
		2D	60751	7.80%			2D	27263	2.85%		
	T7	2A	15954	0.72%	0.85%	T7	2A	15529	0.40%	0.53%	0.69%
		2B	49348	0.91%			2B	51435	0.54%		

		2D	47877	0.84%			2D	50398	0.56%		
	T8	2A	30241	4.35%		T8	2A	9984	0.96%		
		2B	65912	4.31%	4.43%		2B	43454	1.31%	1.35%	2.89%
		2D	60895	4.59%			2D	41720	1.48%		
	T9	2A	66620	0.45%		T9	2A	35806	0.43%		
		2B	63370	0.91%	0.52%		2B	34503	0.70%	0.40%	0.46%
		2D	54456	0.13%			2D	29733	0.03%		
	T10	2A	45422	3.05%		T10	2A	15150	0.88%		
		2B	48195	0.04%	3.64%		2B	16054	0.04%	1.04%	2.34%
		2D	41322	4.30%			2D	13025	1.22%		
	replicate 1					**replicate 2**					
Gene	**Target**	**Genome**	**total reads #**	**Efficiency**	**Average efficiency**	**Target**	**Genome**	**total reads #**	**Efficiency**	**Average efficiency**	**Average between replicates**
CKX6 group 1	T1	3A2	15334	0.00%		T1	3A2	4	0.00%		
		3B2	15671	4.15%			3B2	8	0.00%		
		3D2	15016	2.68%	2.66%		3D2	9	22.22%	14.29%	8.47%
		3D3	14434	3.85%			3D3	7	28.57%		
		3D4	xxxx				3D4	xxxx			
	T2	3A2	7	14.29%		T2	3A2	14463	1.59%		
		3B2	5	0.00%			3B2	13629	0.41%		
		3D2	7	0.00%	4.00%		3D2	11966	1.35%	0.95%	2.48%
		3D3	6	0.00%			3D3	11594	0.38%		
		3D4	xxxx				3D4	xxxx			
	T3	3A2	41580	0.12%		T3	3A2	23644	0.12%		
		3B2	43371	0.06%			3B2	28017	0.05%		
		3D2	41763	0.07%	0.09%		3D2	25638	0.06%	0.09%	0.09%
		3D3	37546	0.12%			3D3	22977	0.13%		
		3D4	4	0.00%			3D4	9	0.00%		
	T4	3A2	30839	3.45%		T4	3A2	11275	0.90%		
		3B2	45951	5.38%			3B2	34406	1.69%		
		3D2	42324	4.50%	3.15%		3D2	33900	1.86%	1.15%	2.15%
		3D3	41420	0.09%			3D3	33239	0.04%		
		3D4	13620	0.14%			3D4	2867	0.21%		
	T5	3A2	55056	0.11%		T5	3A2	17127	0.06%		
		3B2	87388	4.75%			3B2	40400	3.51%		
		3D2	82675	4.13%	3.44%		3D2	40440	2.21%	2.22%	2.83%
		3D3	75592	4.12%			3D3	38950	2.16%		
		3D4	11851	0.03%			3D4	6275	0.05%		
	T6	3A2	31698	1.19%		T6	3A2	6176	0.24%		
		3B2	56961	0.50%			3B2	33495	0.14%		
		3D2	57359	0.64%	0.69%		3D2	36765	0.11%	0.12%	0.40%
		3D3	55183	0.83%			3D3	32618	0.09%		
		3D4	15622	0.03%			3D4	1574	0.00%		
	T7	3A2	22694	0.47%		T7	3A2	14833	0.40%		
		3B2	64507	0.57%			3B2	48734	0.24%		
		3D2	64264	0.52%	0.50%		3D2	46148	0.26%	0.28%	0.39%
		3D3	63194	0.47%			3D3	43556	0.32%		
		3D4	6294	0.10%			3D4	4538	0.07%		
	T8	3A2	36360	1.96%	0.96%	T8	3A2	12496	0.42%	0.22%	0.59%

		3B2	61686	1.21%			3B2	41955	0.27%		
		3D2	59335	1.10%			3D2	44439	0.30%		
		3D3	53646	0.08%			3D3	38127	0.03%		
		3D4	16443	0.13%			3D4	4376	0.05%		
	T9	3A2	6104	0.00%		T9	3A2	5840	0.00%		
		3B2	27984	0.01%			3B2	12378	0.01%		
		3D2	27911	1.42%	0.85%		3D2	12318	0.02%	0.01%	0.43%
		3D3	27457	1.31%			3D3	10191	0.01%		
		3D4	xxxx				3D4	80	0.00%		
	replicate 1										
Gene	**Target**	**Geno me**	**total reads #**	**Efficiency**	**Average efficiency**						
CKX6 group 2	T10	3A1	5988	0.67%							
		3B1	11453	0.12%	0.45%						
		3D1	5367	0.56%							
	T11	3A1	12543	2.90%							
		3B1	24973	3.92%	3.15%						
		3D1	9998	2.63%							
	T12	3A1	4138	3.77%							
		3B1	19278	3.76%	3.62%						
		3D1	2370	3.33%							
	T13	3A1	4883	4.65%							
		3B1	7802	3.73%	4.23%						
		3D1	3116	4.30%							
	T14	3A1	10113	1.21%							
		3B1	17851	1.30%	1.31%						
		3D1	7527	1.42%							
Protoplast transformation- 02232017 (transformation efficiency 10%)											
	replicate 1					**replicate 2**					
Gene	**Target**	**Geno me**	**total reads #**	**Efficiency**	**Average efficiency**	**Target**	**Geno me**	**total reads #**	**Efficiency**	**Average efficiency**	**Average between replicates**
GS3	T6	4A	26041	0.38%		T6	4A	22467	0.05%		
		7A	22066	0.24%	0.35%		7A	17753	0.37%	0.40%	0.37%
		7D	26783	0.42%			7D	23386	0.78%		
	T7	4A	28041	0.45%		T7	4A	20078	0.29%		
		7A	23658	0.00%	0.35%		7A	17138	0.00%	0.16%	0.26%
		7D	28367	0.59%			7D	20951	0.20%		
	T8	4A	9539	0.73%		T8	4A	22228	1.16%		
		7A	7350	1.51%	1.07%		7A	19455	0.70%	1.06%	1.06%
		7D	9961	0.97%			7D	22540	1.30%		
	T9	4A	24160	0.02%		T9	4A	24391	0.21%		
		7A	19425	0.27%	0.45%		7A	20604	0.15%	0.86%	0.65%
		7D	24773	1.06%			7D	25117	2.21%		
An-1	T5	2A	20963	0.41%		T5	2A	37377	0.51%		
		2B	20091	0.34%	0.34%		2B	36593	0.15%	0.37%	0.35%
		2D	21441	0.26%			2D	35556	0.45%		
Protoplast transformation- 05192017 (transformation efficiency 40%)											

	replicate 1					replicate 2					
Gene	Target	Genome	total reads #	Efficiency	Average efficiency	Target	Genome	total reads #	Efficiency	Average efficiency	Average between replicates
GSE5	T1	1A	8866	6.10%	2.05%	T1	1A	15629	1.88%	0.66%	1.36%
		1B	800	0.00%			1B	1976	0.10%		
		1D	38833	0.06%			1D	38674	0.01%		
	T2	1A	4960	0.71%	1.35%	T2	1A	16489	2.24%	2.20%	1.78%
		1B	1290	0.93%			1B	5039	1.07%		
		1D	16973	2.43%			1D	40351	3.28%		
	T4	1A	8	0.00%	0.00%	T4	1A	67	1.49%	11.61%	5.80%
		1B	xxxx	xxxx			1B	3	33.3%		
		1D	4	0.00%			1D	61	0.00%		
	T5	1A	1467	0.27%	0.11%	T5	1A	10172	0.28%	0.20%	0.16%
		1B	8416	0.00%			1B	7337	0.07%		
		1D	1436	0.07%			1D	4600	0.26%		
	T6	1A	11041	0.01%	0.00%	T6	1A	19420	0.01%	0.00%	0.00%
		1B	20834	0.00%			1B	16617	0.00%		
		1D	2517	0.00%			1D	661	0.00%		

Results from the same protoplast transformation were shown together, with the transformation efficiency of that time shown in green words at the top. In the "genome" column, red color indicates the on-target copies; blue color means that copy has no PAM; xxxx means no corresponding reads aligned to reference. The yellow highlighted are the chosen gRNA targeting sites.

Two biological replicates were applied for most of the gRNAs in protoplast transient expression assay. Though there is variation between the replicates, the results clearly show the gene editing efficiency of each CRIPSR/Cas9 target sites are different. The editing efficiency reported in Table 3 is based on the percentage of mutated NGS reads, and it can be normalized with the transformation efficiency. For example, after taking into account the protoplast transformation efficiency (10%), the editing efficiency of the GS3T6 target, which was less than 0.5%, became close to 5%. By normalizing the averaged genome editing efficiency with protoplast transformation efficiency, the highest gene editing efficiency that can be reached was about 30% (Table 4), which is promising for getting mutants for the gene by transforming wheat with the CRISPR/Cas9 construct. To choose the reliable CIRSPR/Cas9 target sites, the gRNAs showing extreme low editing efficiency (<0.5%) were discarded. Meanwhile, if no or small number of NGS reads were generated for a target site, these gRNAs were also excluded from the

further study. By counting in the gene editing efficiency and the proximity to the translation start codon, we finally selected two gRNAs target sites for each gene, which are highlighted in Table 3 with yellow color. In most cases, these two RNAs together can target all the homoeologs of the gene, except for *TaCKX6*-group1, where 4 out of 5 gene copies can be targeted by the chosen two gRNAs and only the copy *TaCKX6*-3D4 is missing. All the selected gRNAs do not have off-target chance.

3.1.4 Plant transformation with gene editing constructs

For each gene, the two chosen gRNAs were fused into a single polycistronic gene, in which gRNAs were separated by a tRNA, and subcloned into a construct pBUN421 (Table 4). The transcripts of the polycistronic gene on construct pBUN421could be processed by the endogenous tRNA processing system, which releases mature gRNAs [58]. We have previously demonstrated that this gRNA multiplexing strategy can be effectively applied in wheat [62]. This construct design guarantees that both gRNAs are delivered simultaneously into the plant cells during transformation, and thereby increase our chances to get mutations in the targeted gene. The constructs were used for biolistic transformation of immature wheat embryos followed by regeneration of transgenic plants.

Table 4. The chosen gRNAs for plant transformation

Construct name for bombardment	Gene locus	gRNA targets	Normalized editing efficiency in protoplast[a]	Targeted gene copies
YLD2	*TaGW7*	GW7_T2	3.69%	2A,2B,2D
		GW7_T6	27.78%	2A,2B,2D
YLD3	*TaAn-1*	An-1_T2	29.93%	2A,2B,2D
		An-1_T6	17.23%	2A,2B,2D
YLD4	*TaCKX6-*	CKX6_T1	28.24%	3B1, 3D1, 3D3

	group1	CKX6_T4	7.18%	3A1, 3B1, 3D1
YLD5	*TaCKX6*-group2	CKX6_T11	10.50%	3A2,3B2,3D2
		CKX6_T13	14.09%	3A2,3B2,3D2
YLD12	*TaGS3*	GS3_T6	3.73%	4A,7A,7D
		GS3_T8	10.65%	4A,7A,7D
YLD15	*TaGSE5*	GSE5_T1	3.40%	1A,1B
		GSE5_T2	4.44%	1A,1B,1D

[a]The editing efficiencies were normalized by dividing averages of the detected mutation rates from NGS by the protoplast transformation efficiencies, both of which were shown in Table 3.

3.1.5 Edited plants

The regenerated plants were screened using PCR for the presence of the pA9Cas9sg construct. In total, we screened 661 transgenic plants, and 180 of them were Cas9-positive plants (Table 5). More than 20 Cas9-positive transgenic plants were received for each gene; the frequency of Cas9-positives among the regenerated plants ranged from 20% to 40.2%.

Table 5. Numbers of the transgenic plants

Gene locus	gRNA targets	Construct name for bombardment	Regenerated plant number	Cas9-positive plant number	Edited (T0)/ NGS checked plant number
TaGW7	T2+T6	**YLD2**	**79**	**29**	**9/23**
TaAn-1	T2+T6	**YLD3**	**115**	**23**	**7/20**
TaCKX6-group1	T1+T4	**YLD4**	**123**	**36**	**3/36**
TaCKX6-group2	T11+T13	**YLD5**	**89**	**22**	**1/17**
TaGS3	T6+T8	**YLD12**	**168**	**35**	**2/33**
TaGSE5	T1+T2	**YLD15**	**87**	**35**	**8/27**

The Cas9-positive plants were further genotyped using NGS to identify gene editing events in the targeted regions (Table 6). In total, we received 30 edited plants out of 156 Cas9-postive plants. Of these, 9, 7 and 8 edited T0 plants for *TaGW7*, *TaAn-1* and *TaGSE5* were discovered from 29, 23 and 35 Cas9-positive plants, respectively (Table 5). This indicates that

about 23%-32% of the Cas9-positive plants showed the evidence of gene editing, which is consistent with our previous estimates [87]. For the *TaCKX6* and *TaGS3* genes, however, the frequencies of edited plants among Cas9-positive plants were lower, 3/36, 1/17, and 2/33, all less than 10%.

To assign the genotypes for the newly mutated alleles (Table 6), the following rules were applied. First, if the total number of NGS reads was smaller than 10, no mutation rate will be counted because insufficient numbers result in high level of errors. Second, a genotype will be assigned if mutation rate of the allele is close to that of either a homozygous wild type allele (mutation rate < 10%), a heterozygous allele (mutation rate between 30% and 70%) or a homozygous mutation allele (mutation rate > 95%). Third, for the remaining plants that carried mutations but could not be assigned to either homozygous or heterozygous category, we showed the mutated read ratios. For example, in transgenic plant Y5-5 shown in Table 6 the ratio of mutated reads was 0.86 in the A genome, between the ratios for heterozygous and homozygous mutants.

Among 30 edited plants, the gene editing levels varied. The heterozygous mutated alleles were found in 20 plants, such as Y3-3 line with mutation at An-1T6 in the D genome, or Y4-1 line with mutations in CKX6T4 in the A genome. There were also several homozygous mutations detected in edited lines Y2-4, Y3-5, Y3-6, Y4-2, Y4-3, Y12-1, Y12-2, Y15-2, and Y15-3. Around half of these plants had mutations in more than one gene copy in the A/B/D genomes. Some plants carried only chimeric mutations with multiple types of mutated reads. For example, in YLD2-1 line we had the low percentage of mutated reads in GW7T6: 4% in the B genome and 2% in the D genome.

Table 6. Genotypes of T0 transgenic plants

T0 Plant ID	Tiller	Assigned genotype by NGS result					
		GW7T2			GW7T6		
Y2-1	2816-1	AA	BB	0/1	AA	0.04	0.02
Y2-2	C239-1	AA	BB	NA	0/4	BB	0.13
Y2-3	C199-2	AA	BB	NA	0/7	BB	0.10
Y2-4	C201-1	5/9	9/9	NA	0/3	BB	DD
Y2-5	C177-3	0/9	BB	NA	AA	0.13	DD
Y2-6	4671-2	AA	BB	NA	0/4	0.13	DD

Y2-7	5097-1	AA	BB	NA		AA	0.22	DD		
Y2-8	C606-2	AA	BB	NA		0.25	BB	0.14		
Y2-9	C585-1	AA	BB	NA		4/8	0.19	DD		
			An-1T2				An-1T6			
Y3-1	C25-1	AA	BB	0.13		AA	BB	DD		
	C25-3	0.14	BB	0.23		AA	BB	DD		
Y3-2	C450-2	0.10	BB	DD		AA	BB	Dd		
Y3-3	5028-1	0.11	BB	DD		AA	BB	DD		
Y3-4	C549-1	AA	BB	dd		AA	BB	dd		
Y3-5	5127-1	AA	0.11	dd		AA	bb	dd		
Y3-6	5263-1	Aa	BB	Dd		0.26	BB	Dd		
Y3-7	5263-2	0.22	BB	Dd		AA	BB	DD		
			CKX6T1				CKX6T4			
		A2	B2	D2	D3	A2	B2	D2	D3	D4
Y4-1	C509-1	AA	BB	0/6	DD	Aa	BB	0/6	DD	0/8
Y4-2	C727-1	AA	BB	DD	DD	8/8	bb	Dd	DD	DD
Y4-3	C727-2	AA	BB	0/5	0/4	4/4	bb	Dd	0/8	DD
			T11&T13							
Y5-5	C413-1	0.86	NA	1/1						
			T6&T8							
Y12-1	4906-1	Aa	bb	Dd						
Y12-2	4906-2	0.70	bb	Dd						
Y12-3	C538-1	AA	BB	Dd						
			T1				T2			
Y15-1	C228-1	Aa	BB	NA		0/2	NA	DD		
Y15-2	4802-1	aa	BB	NA		AA	NA	DD		
	C365-1	aa	BB	0/4		0/6	NA	0/6		
Y15-3	C365-2	aa	BB	NA		AA	NA	DD		
	C365-3	0.75	BB	4/4		AA	NA	DD		
Y15-4	C462-2	Aa	BB	NA		0/6	NA	0/8		
Y15-5	C624-1	0.28	BB	0/2		AA	NA	DD		
	C624-2	0.29	BB	0/6		0/6	NA	DD		
Y15-6	C726-2	0.14	BB	0/8		AA	NA	DD		
	5121-2	Aa	BB	0/4		Aa	NA	DD		
Y15-7	5121-3	Aa	BB	0/6		Aa	NA	DD		
	5121-4	Aa	BB	DD		Aa	0/2	DD		
Y15-8	5213-2	Aa	NA	NA		AA	NA	DD		

We marked a plant as edited (red color) when the ratio of the mutation reads within the targeted regions was more than 10%, except for YLD2-1. If total number of NGS reads for an allele was smaller than 10, the original numbers of mutated reads / total reads were shown. The homozygous wild type allele was assigned when mutated read ratio was < 10%; the heterozygous allele was assigned when the mutated read proportion was 30-70% with only one type of mutated reads; the homozygous mutated allele was assigned when mutated read proportion was >95% with only one or two types of mutated reads. For the remaining alleles we showed their mutation ratios.

Among these gene-edited plants, we select those having heterozygous or homozygous mutated alleles and frameshift mutations for further phenotypic evaluation. Most of these mutants were generated recently, and their phenotypic evaluation is under way. In this study, we are reporting seed morphology and weight traits for the *TaGW7* gene mutants.

3.1.6 Functional analysis of *TaGW7*

3.1.6.1 Recovering mutated alleles in the *TaGW7* gene utilizing the transgenerational CRISPR/Cas9 activity

Our previous study showed that transgenerational CRISPR/Cas9 activity can facilitate recovery of mutated genes from unedited Cas9/gRNA-expressing plants in the next generation [87]. The T0 screening showed low-percentage of mutated reads in YLD2-1 line, which had several different types of mutations (Figure 5). This result provided evidence that Cas9 construct carried by this plant is capable of producing desired mutated alleles. Therefore, we screened 70 T1 lines to identify new mutated alleles in the next generation. By analyzing the NGS data, seven mutant lines that carried mutations in heterozygous state (single type of mutations with approximately 1:1 ratio to WT reads) in different genomes were detected (Figure 5). Six of these plants carried heterozygous mutations in either A or B genomes, and one plant carries heterozygous mutations in both B and D genomes. Most of these mutations were located 3-4 bp before the PAM at the expected target editing site. One- or two-bp deletions were the most common type of mutations, while longer deletions were also observed (Figure 5). Besides these

deletion events, 1-bp long insertion into the B genome in line YLD2-1-26 was also detected (not shown in the figure).

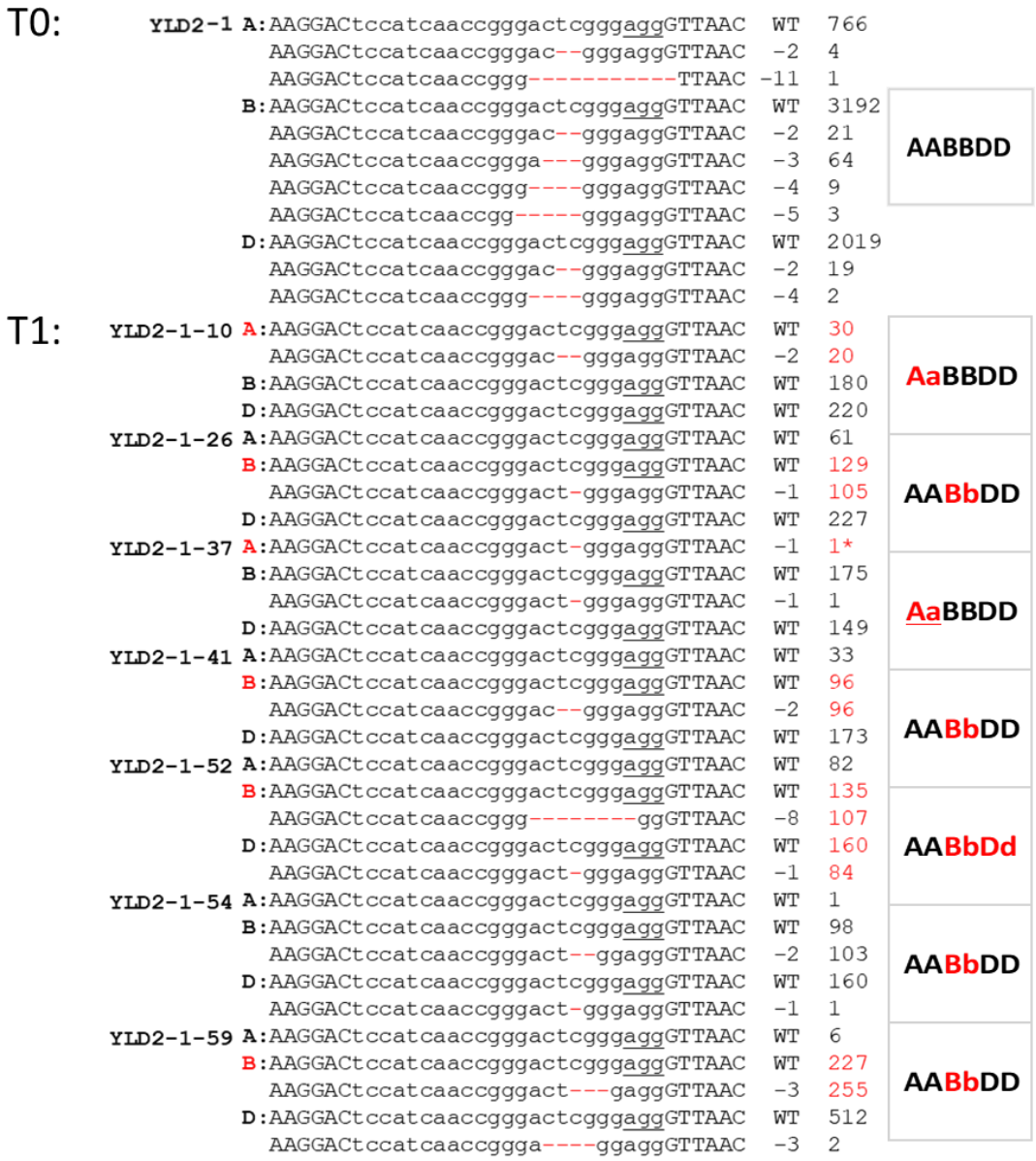

Figure 5. NGS results for the *TaGW7*-edited plants.

NGS reads of PCR amplicons of the DNA from the mutant plants are aligned to wild type reference sequence. Sequences flanking the gRNA target sites and numbers of each type of reads are shown in the figure. The lower-case nucleotides in the alignments are the gRNA targeting sites; the underlined are PAMs; each red dash line means one base deletion. The wild type reads are marked with "WT", the deletions are shown as "-" followed by the number of deleted nucleotides. The numbers shown after genotypes shows the number of reads from NGS that can be aligned to references. The heritably mutated genomes are marked with red color. *, this mutation was verified to be heterozygous by the segregation observed in T2 generation.

3.1.6.2 Effect of Cas9-induced mutations in the *TaGW7* gene on phenotype

In the progenies of the T1 mutants of *TaGW7*, we selected 57 plants with representative genotypes for the phenotypic analysis, about 10 plants for each genotype configuration. We collected data for TGW, seed area (SA), seed length (SL), and seed width (SW) traits for each T2 plant (Appendices Table 8).

The phenotypic data displayed that only D genome mutants have significantly shorter seeds than D-genome wild type plants (AABbdd vs AABbDD) (Figure 6). We found no significant differences in TGW, SA, and SW between the wild type plants and the mutants. And mutations in the A and B genome copies of the *TaGW7* gene showed no significant differences for all measured traits.

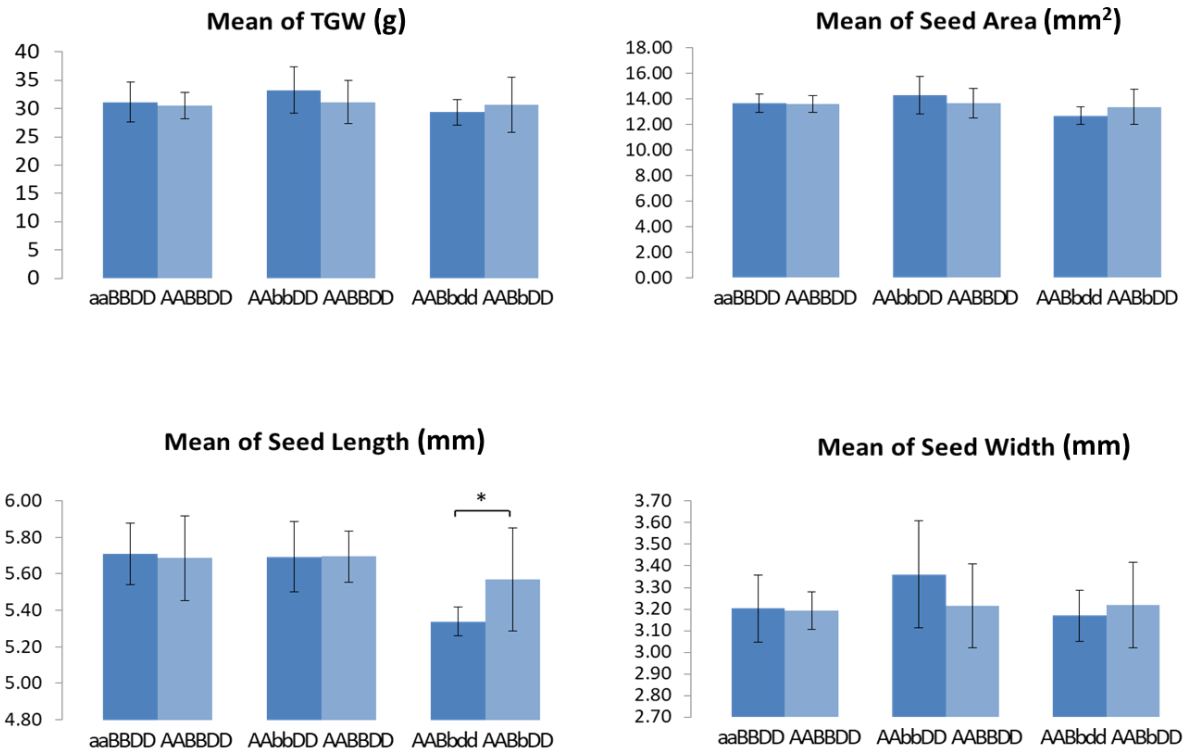

Figure 6. Effect of the *TaGW7* mutations on A/B/D genome estimated in T2 plants

Means and SEMs of the trait values for different genotype groups were shown in each chart, with the comparisons of plants with single-genome mutations and wild type. Genotypes of each plant groups are shown at the bottom of the columns. *, significant at $P < 0.05$, by the two-tailed Student's t test.

We further evaluated T3 generation plants for the effect of single-genome mutations on these traits (Figure 7, Appendices Table 9). Thinking about the possible genetic background effects, the analyses in T3 generation data were done based on genetic families. A genetic family is defined as all the plants derived from a single T1 plant. In the T3 results, there were several significant differences observed between single-genome mutants and wild type plants for different traits. However, the effects of A or B genome mutations were not stable across different families. For example, even though mutation in the A genome in the YLD2-1-37 family affected TGW, SA, and SL traits, mutation in the A genome in the YLD2-1-10 family showed no effect on any of these traits. The difference in GL between the D genome mutants and WT plants was

still significant in the tested family, confirming results obtained from the T2 generation plants. In addition, for the seed length/width (L/W) ratio, we found that mutations in the A and B genomes had no effect, whereas the effect of D genome mutation was highly significant ($P < 0.005$). Loss of wild type allele in the D genome was associated with a lower L/W ratio, namely a wider and/or shorter seeds. These results corroborated the conclusion that the D genome mutation of *TaGW7* causes significant change in wheat grain length.

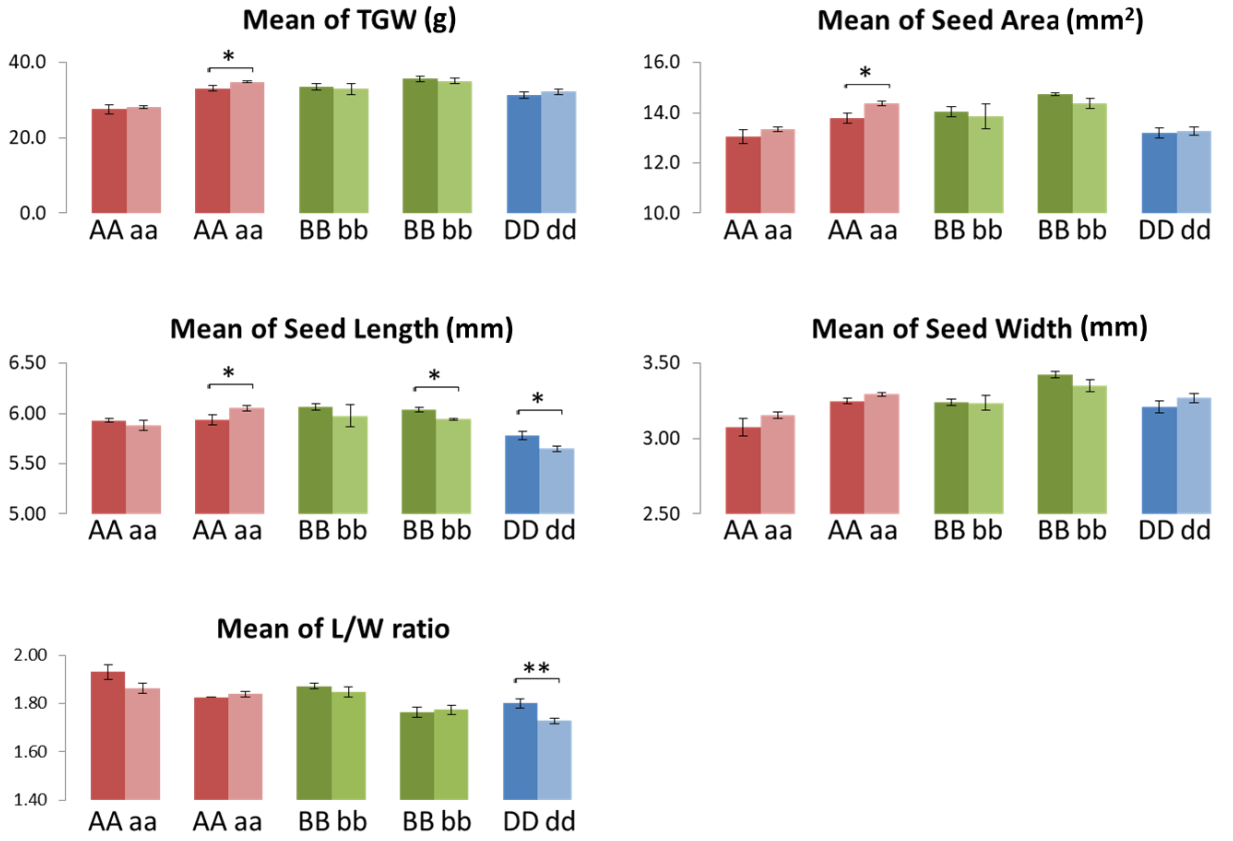

Figure 7. Effect of the *TaGW7* mutations on A/B/D genome estimated in T3 plants

Means and SEMs of the trait values for different genotype groups were shown in each chart, with the comparisons of plants with single-genome mutations and wild type. Five pairs of columns correspond to five families developed from five T1 mutants, which are YLD2-1-10, YLD2-1-37, YLD2-1-26, YLD2-1-54, and YLD2-5-1-52, in the same order as the column clusters. Each red pair of columns shows a phenotypic comparison between plants with AABBDD and aaBBDD genotypes, while only AA and aa are labeled due to the limited space. Likewise, green columns show comparisons between AABBDD and AAbbDD plants and blue columns show comparisons between AAbbDD and AAbbdd plants. Significances of the differences were tested by the two-tailed Student's t test. *, significant at $0.005<P < 0.05$; **, significant at $P<0.005$.

3.1.6.3 Expression of *TaGW7* gene in WT plants

To better understand the contribution of the *TaGW7* gene homoeologs to phenotypic variation, we investigated homoeolog-specific expression of this gene in different wheat tissues collected at different developmental stages (Figure 8). Our results show that *TaGW7* is most highly expressed in young spikes, which is consistent with a previous report [89].

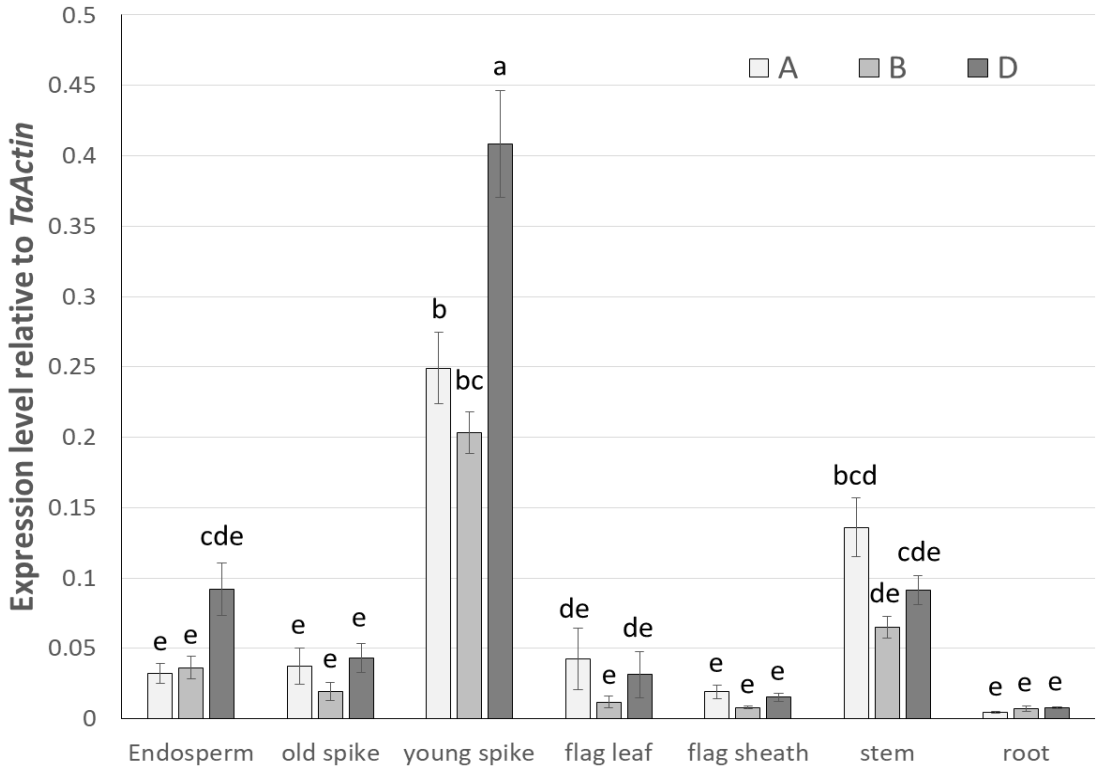

Figure 8. Expression of *TaGW7* gene in wheat cultivar Bobwhite

The quantitive PCR of the *TaGW7* gene homoeologs was conducted using genome-specific primers (Appendices Figure 9), the *TaActin* gene was used as control. Data from five biological and two technical replicates were analyzed and shown in this figure. The relative expression levels are shown as means ± SEM. Different letters above the bars indicate significant differences ($P < 0.05$) using Tukey's multiple-comparison test following one-way ANOVA. All the tissues were sampled from the plants of which the main spike already pollinated and the grain filling started, but other tillers have the spikes pollinating or not pollinated. The old spike is the main spike from which the grains are removed, the young spike is the spike which was not pollinated yet.

Among three homoeologs, the D genome copy was expressed at the highest level in young spikes. Gene expression differences among genomes in the other tissues were not significant. Considering the non-specific A genome primers (Appendices Figure 9), the actual A

genome expressions could be even lower than that shown in Figure 8. These results support the speculation that the higher expression level of the D genome copy of *TaGW7* can be associated with the strong phenotypic effect of the D genome mutations observed in T2 and T3 mutants.

3.2 Discussion

3.2.1 Cas9-based wheat improvement by transferring information from rice

Large numbers of edited plants were obtained in this study using our wheat genome editing pipeline. Our study demonstrated that the usage of genes controlling yield component traits in rice for improving wheat can be an effective strategy. Similar to the results obtained for the *TaGW2* gene, the editing of *TaGW7* produced phenotypic effects in similar direction in both rice and wheat [77].

In rice, *GW7* was identified as a gene involved in the spatial control of cell division [77]. Downregulation of *GW7* expression in rice was correlated with the production of shorter and wider grains, as a result of increased cell division in the transverse direction and decreased cell division in the longitudinal direction of the grain. TGW was not significantly different between the NIL-$gw7^{HJX74}$ (low *GW7*-expression) and NIL-$GW7^{TFA}$ (high *GW7*-expression). We validated the role of *TaGW7* as a grain shape regulator in wheat. Mutation in *TaGW7* resulted in shorter seeds but did not significantly change TGW, which is similar to the findings in rice [77]. These results suggest that the genetic architecture of yield component traits in wheat and rice are similar and discoveries made in rice are likely transferrable to wheat. This strategy tremendously decreases the time and labor needed to identify a gene and allows substantially accelerating analyses of yield component pathways and complementing QTL mapping studies in wheat.

Another finding about rice *GW7* gene is that when in *gs3* (a widely used allele in rice breeding for longer and larger grains) background, NIL-*gs3*-*GW7*^{TFA} could produce significantly higher TGW than NIL-*gs3*-*gw7*^{HJX74} plants [77], indicating that *GW7* is possibly a positive regulator of grain yield. In addition, higher *GW7* expression is associated with better grain quality and hybridization of high *GW7*-expression line with other high-yield breeding lines resulted in significant improvement of grain quality without compromise of grain yield [88]. These results demonstrated the potential of combining beneficial alleles to improve crops. It would be a good strategy to combine *TaGW7* allele with beneficial alleles from other yield component genes for wheat improvement. For this purpose, the genes included into our study and showing positive effects on yield component traits would be a valuable source of novel beneficial alleles for improving wheat yield potential.

Genetic dissection of yield component traits in wheat has been progressed slowly. To date, there are only few genes affecting yield component traits in wheat that have been identified, such as *TaGW2* [82] and *Q* gene [90], [91]. To our best knowledge, *TaGW7* does not overlap with the currently mapped QTL for yield component traits. It is possible that the effect of natural *TaGW7* alleles is not strong enough to be detected in the analyzed QTL mapping populations. Alternatively, beneficial alleles of *TaGW7* can be extremely rare in wheat precluding their identification using genome-wide association mapping approaches. This possibility is consistent with rice studies demonstrating that some alleles of yield component genes with strong effect are rare in natural populations [92], [93]. Therefore, functional validation of the wheat orthologs of rice yield component genes is an effective strategy for expanding the set of genes that can be utilized in wheat breeding.

3.2.2 Wheat genome editing efficiency

Our results indicate that transient expression in protoplasts provide effective screen for the quick assessment of gRNA editing efficiency. By using the pre-screened gRNAs, we successfully obtained mutant plants for all targeted genes. However, editing efficiencies observed in protoplasts and transgenic plants were not always consistent. For example, even though the editing efficiencies of *TaCKX6* targets in protoplasts were similar to those of the *TaGW7* and *TaAn-1* gene targets, all close to 30%, getting *TaCKX6*-edited plants was much more difficult than getting edited plants for the other two genes. Plant editing efficiency for *TaGW7* and *TaAn*-1 were 39% and 35%, respectively, while for *TaCKX6* editing efficiency was only 8% and 6% for groups 1 and 2, respectively. It is possible that multiple factors contribute to this editing efficiency variation. First, there can be variation in the number of copies of Cas9 construct inserted into genome, or in the completeness of the inserted constructs, which will directly influence Cas9 expression and activity. Second, Cas9 constructs can be inserted into different genomic regions. Wheat genome harbors a large number of repetitive elements located in the transcriptionally inactive heterochromatic regions [94]. Insertion of Cas9 into these regions will probably result in either low expression or full silencing of *Cas9*. This has be amply demonstrated in other studies where transgene silencing was associated with the location of a transgene in a genome [95]. Thus, each transformation event may have a specific level of *Cas9* expression, causing variation in gene editing efficiency. Third, chromatin state and cell physiology in the protoplasts and immature embryos can be different. Chromatin accessibility has been proven to have strong effect on editing efficiency by influencing gRNA binding [96]. Therefore, the same gRNA may show different performance in different cell types.

3.2.3 Differential expression of wheat homoeologs

In our study, we found that the contribution of different gene homoeologs to trait variation can be associated with the levels of their relative expression. Mutation in *TaGW7* on the D genome that was expressed at the highest level in young spikes showed strongest effect on grain shape, whereas A and B genome mutants, which had lower expression levels, had relatively minor phenotypic effects. In rice, it was shown that the *GW7* gene is expressed in spikelet hulls during grain growth and regulates the division of the outer epidermal cells to control grain shape [77]. We observed the high expression of *TaGW7* in wheat young spikes, which agrees with the finding in rice. Therefore, the expression levels of the *TaGW7* homoeologs in young spikes can be important indicator of their functional relevance for grain development in wheat.

The relationship between the *TaGW7* gene expression and the phenotypic effects of the mutations is similar to our previous findings for the *TaGW2* gene [82]. These results suggest that polyploidy plays an important role in shaping the genetic architecture of trait variation in wheat, where intergenomic differences among duplicated genes can define their relative contribution to the gene function in biological pathways controlling major agronomic traits. Understanding interplay between gene expression and function in each homoeolog can help researchers and breeders develop better strategies to improve wheat using gene editing approaches.

3.2.4 Utilization of transgenerational activity of CRISPR/Cas9 for breeding

In our study, all *Tagw7* mutants in T1 and T2 generation were derived from an almost unedited T0 plant, with only 2-4% NGS reads from leaf tissue carrying mutations. This result is consistent with our previous study [87], and shows that the low editing efficiencies in wheat

transgenic plants can be offset by screening for editing events in the next generation transgenic plants, where on-going transgenerational gene editing allows for recovering desired mutations.

The transgenerational CRISPR/Cas9 activity can not only help to recover desired mutations in the progenies of transgenic plants, but also facilitate trait transfer in breeding programs. A transgenic plant expressing Cas9/gRNA construct targeting a gene of interest can be used as "vehicle" for delivering gene editing reagents to different wheat cultivars. When such transgenic line is crossed with a breeding line, Cas9/gRNA complex can edit the gene copy coming from a breeding line in F1 individuals. Following this strategy, homozygous edited alleles in the breeding line background as well as transgene-free plants can be easily recovered in few generations. Another advantage of this approach is that it does not rely on recombination rate variation across genome, which strongly influences the efficiency of beneficial allele selection in breeding programs.

3.3 Conclusion

Here, we describe an effective pipeline for wheat genome editing using the CRISPR/Cas9 technology. This pipeline can be used to create mutations in any wheat gene for either their functional validation or for creating new beneficial variants of these genes. We demonstrated that the wheat orthologs of rice genes affecting seed traits tend to have similar effects on the same traits in wheat. This result suggests that the genetic architectures of yield component traits in wheat and rice can be similar. Hence, it's a promising strategy to utilize knowledge accumulated for rice genome to improve wheat traits using genome editing. By studying the *TaGW7* gene, we further corroborate our earlier conclusion that wheat gene homoeologs can be functionally diversified and contribute differently to phenotypic variation depending on the levels of their

expression in relevant tissues. Gene editing provides a powerful tool to dissect these gene-phenotype relationships in the context of allopolyploid genome, and to broaden genetic diversity available for precision trait breeding in wheat.

References

[1] A. Scheben, F. Wolter, J. Batley, H. Puchta, and D. Edwards, “Towards CRISPR/Cas crops - bringing together genomics and genome editing,” *New Phytol.*, pp. 682–698, 2017.

[2] J. M. Poehlman, D. A. Sleper, and J. Rudd, *Breeding field crops*, vol. 378. Springer, 1995.

[3] B. C. Y. Collard and D. J. Mackill, “Marker-assisted selection: An approach for precision plant breeding in the twenty-first century,” *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 363, no. 1491, pp. 557–572, 2008.

[4] E. L. Heffner, M. E. Sorrells, and J.-L. Jannink, “Genomic selection for crop improvement,” *Crop Sci.*, vol. 49, no. 1, pp. 1–12, 2009.

[5] S. Wang *et al.*, “Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array,” *Plant Biotechnol. J.*, vol. 12, no. 6, pp. 787–796, 2014.

[6] S. Kiani, A. Akhunova, and E. Akhunov, “Application of next-generation sequencing technologies for genetic diversity analysis in cereals,” in *Cereal Genomics II*, Springer, 2013, pp. 77–99.

[7] H. Cai *et al.*, “Mapping QTLs for root system architecture of maize (Zea mays L.) in the field at different developmental stages,” *Theor. Appl. Genet.*, vol. 125, no. 6, pp. 1313–1324, 2012.

[8] L. Shen, B. Courtois, K. L. McNally, S. Robin, and Z. Li, “Evaluation of near-isogenic lines of rice introgressed with QTLs for root depth through marker-aided selection,” *Theor. Appl. Genet.*, vol. 103, no. 1, pp. 75–83, 2001.

[9] M. Ahmad, “Molecular marker-assisted selection of HMW glutenin alleles related to wheat bread quality by PCR-generated DNA markers,” *Theor. Appl. Genet.*, vol. 101, no.

5, pp. 892–896, 2000.

[10] E. Wijnker and H. de Jong, “Managing meiotic recombination in plant breeding,” *Trends Plant Sci.*, vol. 13, no. 12, pp. 640–646, 2008.

[11] K. W. Jordan *et al.*, “The genetic architecture of genome-wide recombination rate variation in allopolyploid wheat revealed by nested association mapping,” *Plant J.*, pp. 1039–1054, 2018.

[12] J. King *et al.*, “Comparative analyses between lolium/festuca introgression lines and rice reveal the major fraction of functionally annotated gene models is located in recombination-poor/very recombination-poor regions of the genome,” *Genetics*, vol. 177, no. 1, pp. 597–606, 2007.

[13] N. Arrigo, R. Guadagnuolo, S. Lappe, S. Pasche, C. Parisod, and F. Felber, “Gene flow between wheat and wild relatives: Empirical evidence from Aegilops geniculata, Ae. neglecta and Ae. triuncialis,” *Evol. Appl.*, vol. 4, no. 5, pp. 685–695, 2011.

[14] C. R. Cavanagh *et al.*, “Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars,” *Proc. Natl. Acad. Sci.*, vol. 110, no. 20, p. 8057 LP-8062, May 2013.

[15] M. Van De Wouw, C. Kik, T. Van Hintum, R. Van Treuren, and B. Visser, “Genetic erosion in crops: Concept, research results and challenges,” *Plant Genet. Resour. Characterisation Util.*, vol. 8, no. 1, pp. 1–15, 2010.

[16] P. Sikora, A. Chawade, M. Larsson, J. Olsson, and O. Olsson, “Mutagenesis as a tool in plant genetics, functional genomics, and breeding,” *Int. J. Plant Genomics*, vol. 2011, 2011.

[17] L. J. Stadler, “Some genetic effects of X-rays in plants.,” *J. Hered.*, vol. 21, p. 3–19 pp.,

1930.

[18] B. J. Till *et al.*, “Discovery of chemically induced mutations in rice by TILLING,” *BMC Plant Biol.*, vol. 7, pp. 1–12, 2007.

[19] I. M. Henry *et al.*, “Efficient Genome-Wide Detection and Cataloging of EMS-Induced Mutations Using Exome Capture and Next-Generation Sequencing,” *Plant Cell*, vol. 26, no. 4, pp. 1382–1397, 2014.

[20] K. V. Krasileva *et al.*, “Uncovering hidden variation in polyploid wheat,” *Proc. Natl. Acad. Sci.*, vol. 114, no. 6, pp. E913–E921, 2017.

[21] Y. Mo, T. Howell, H. Vasquez-Gross, L. A. de Haro, J. Dubcovsky, and S. Pearce, “Mapping causal mutations by exome sequencing in a wheat TILLING population: a tall mutant case study,” *Mol. Genet. Genomics*, vol. 293, no. 2, pp. 463–477, 2018.

[22] P. D. Hsu, E. S. Lander, and F. Zhang, “Development and applications of CRISPR-Cas9 for genome engineering,” *Cell*, vol. 157, no. 6, pp. 1262–1278, 2014.

[23] Y.-G. Kim, J. Cha, and S. Chandrasegaran, “Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain,” *Proc. Natl. Acad. Sci.*, vol. 93, no. 3, pp. 1156–1160, 1996.

[24] R. Jankele and P. Svoboda, “TAL effectors: tools for DNA Targeting,” *Brief. Funct. Genomics*, vol. 13, no. 5, pp. 409–419, Sep. 2014.

[25] T. Gaj, C. A. Gersbach, and C. F. Barbas III, “ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering,” *Trends Biotechnol.*, vol. 31, no. 7, pp. 397–405, 2013.

[26] F. Cong, L.*, Ran, F.A.*, Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., & Zhang, “Multiplex Genome Engineering Using CRISPR/Cas Systems,” *Science (80-.).*, vol. 339, no. 6121, pp. 819–823, 2013.

[27] A. V. Wright, J. K. Nuñez, and J. A. Doudna, “Biology and Applications of CRISPR Systems: Harnessing Nature’s Toolbox for Genome Engineering,” *Cell*, vol. 164, no. 1–2, pp. 29–44, 2016.

[28] N. Savić and G. Schwank, “Advances in therapeutic CRISPR/Cas9 genome editing,” *Transl. Res.*, vol. 168, pp. 15–21, Feb. 2016.

[29] Y. Ding, H. Li, L. Chen, K. Xie, and L. Chen, “Recent Advances in Genome Editing Using CRISPR / Cas9,” vol. 7, no. May, pp. 1–12, 2016.

[30] S. Sánchez-León *et al.*, “Low-gluten, nontransgenic wheat engineered with CRISPR/Cas9,” *Plant Biotechnol. J.*, vol. 16, no. 4, pp. 902–910, 2018.

[31] D. Rodríguez-Leal, Z. H. Lemmon, J. Man, M. E. Bartlett, and Z. B. Lippman, “Engineering Quantitative Trait Variation for Crop Improvement by Genome Editing,” *Cell*, vol. 171, no. 2, p. 470–480.e8, 2017.

[32] D. F. Voytas and C. Gao, “Precision Genome Engineering and Agriculture: Opportunities and Regulatory Challenges,” *PLoS Biol.*, vol. 12, no. 6, pp. 1–6, 2014.

[33] N. Duensing *et al.*, “Novel features and considerations for ERA and regulation of crops produced by genome editing,” *Front. Bioeng. Biotechnol.*, vol. 6, p. 79, 2018.

[34] A. H. Paterson *et al.*, “Comparative genomics of plant chromosomes,” *Plant Cell*, vol. 12, no. 9, pp. 1523–1539, 2000.

[35] R. Schmidt, “Synteny: Recent advances and future prospects,” *Curr. Opin. Plant Biol.*, vol. 3, no. 2, pp. 97–102, 2000.

[36] Q.-H. Wu *et al.*, “High-density genetic linkage map construction and QTL mapping of grain shape and size in the wheat population Yanda1817× Beinong6,” *PLoS One*, vol. 10, no. 2, p. e0118144, 2015.

[37] M. Jia *et al.*, “Wheat functional genomics in the era of next generation sequencing: An update,” *Crop J.*, vol. 6, no. 1, pp. 7–14, 2018.

[38] T. Lee, H. Kim, and I. Lee, “Network-assisted crop systems genetics: network inference and integrative analysis,” *Curr. Opin. Plant Biol.*, vol. 24, pp. 61–70, 2015.

[39] Z. Wang, M. Gerstein, and M. Snyder, “RNA-Seq: a revolutionary tool for transcriptomics.,” *Nat. Rev. Genet.*, vol. 10, no. 1, pp. 57–63, Jan. 2009.

[40] F. M. Giorgi, C. Del Fabbro, and F. Licausi, “Comparative study of RNA-seq-and microarray-derived coexpression networks in Arabidopsis thaliana,” *Bioinformatics*, vol. 29, no. 6, pp. 717–724, 2013.

[41] W. B. Rutter *et al.*, “Divergent and convergent modes of interaction between wheat and Puccinia graminis f. sp. tritici isolates revealed by the comparative gene co-expression network and genome analyses,” *BMC Genomics*, vol. 18, no. 1, p. 291, 2017.

[42] F. A. Feltus, “Systems genetics: A paradigm to improve discovery of candidate genes and mechanisms underlying complex traits,” *Plant Sci.*, vol. 223, pp. 45–48, 2014.

[43] J. Huang, S. V. Alegre, L. Shi, and K. McGinnis, “Construction and Optimization of Large Gene Co-expression Network in Maize Using RNA-Seq Data,” *Plant Physiol.*, p. pp-00825, 2017.

[44] I. Lee, B. Ambaru, P. Thakkar, E. M. Marcotte, and S. Y. Rhee, “Rational association of genes with traits using a genome-scale gene network for Arabidopsis thaliana,” *Nat. Biotechnol.*, vol. 28, no. 2, p. 149, 2010.

[45] C. Ma, M. Xin, K. A. Feldmann, and X. Wang, “Machine Learning–Based Differential Network Analysis: A Study of Stress-Responsive Transcriptomes in Arabidopsis,” *Plant Cell*, vol. 26, no. 2, p. 520 LP-537, Feb. 2014.

[46] I. Lee *et al.*, “Genetic dissection of the biotic stress response using a genome-scale gene network for rice,” *Proc. Natl. Acad. Sci.*, Oct. 2011.

[47] Y. Ishino, H. Shinagawa, K. Makino, M. Amemura, and A. Nakata, “Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product.,” *J. Bacteriol.*, vol. 169, no. 12, pp. 5429–5433, 1987.

[48] R. Jansen, J. D. A. van Embden, W. Gaastra, and L. M. Schouls, “Identification of genes that are associated with DNA repeats in prokaryotes,” *Mol. Microbiol.*, vol. 43, no. 6, pp. 1565–1575, 2002.

[49] F. J. M. Mojica, J. García-Martínez, and E. Soria, “Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements,” *J. Mol. Evol.*, vol. 60, no. 2, pp. 174–182, 2005.

[50] C. Pourcel, G. Salvignol, and G. Vergnaud, “CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies,” *Microbiology*, vol. 151, no. 3, pp. 653–663, 2005.

[51] A. Bolotin, B. Quinquis, A. Sorokin, and S. D. Ehrlich, “Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin,” *Microbiology*, vol. 151, no. 8, pp. 2551–2561, 2005.

[52] E. S. Lander, “The Heroes of CRISPR,” *Cell*, vol. 164, no. 1–2, pp. 18–28, 2016.

[53] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier, “A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptice Bacterial Immunity,” *Science*, vol. 337, no. August, pp. 816–822, 2012.

[54] L. Tu, A. Naseri, M. Huisman, S. Zhang, and T. Pederson, “Multiplexed labeling of

genomic loci with dCas9 and engineered sgRNAs," vol. 34, no. 5, pp. 528–530, 2016.

[55] F. A. Ran *et al.*, "Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity F.," *Cell*, vol. 154, no. 6, pp. 1380–1389, 2014.

[56] R. Lino, L. Barbosa, J. Prado, F. Oliveira, L. Reis, and R. Barbosa, *Development of high-strength microalloyed coiled rebar*, vol. 15, no. 8. 2018.

[57] H.-L. Xing *et al.*, "A CRISPR/Cas9 toolkit for multiplex genome editing in plants," *BMC Plant Biol.*, vol. 14, no. 1, p. 327, 2014.

[58] K. Xie, B. Minkenberg, and Y. Yang, "Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system," *Proc. Natl. Acad. Sci.*, vol. 112, no. 11, pp. 3570–3575, 2015.

[59] R. E. Haurwitz, M. Jinek, B. Wiedenheft, K. Zhou, and J. A. Doudna, "Sequence-and structure-specific RNA processing by a CRISPR endonuclease," *Science (80-.).*, vol. 329, no. 5997, pp. 1355–1358, 2010.

[60] L. Nissim, S. D. Perli, A. Fridkin, P. Perez-Pinera, and T. K. Lu, "Multiplexed and Programmable Regulation of Gene Networks with an Integrated RNA and CRISPR/Cas Toolkit in Human Cells," *Mol. Cell*, vol. 54, no. 4, pp. 698–710, 2014.

[61] W. Qi, T. Zhu, Z. Tian, C. Li, W. Zhang, and R. Song, "High-efficiency CRISPR/Cas9 multiplex gene editing using the glycine tRNA-processing system-based strategy in maize," *BMC Biotechnol.*, vol. 16, no. 1, p. 58, 2016.

[62] W. Wang, A. Akhunova, S. Chao, and E. Akhunov, "Optimizing multiplex CRISPR/CAS9Cas9 system for wheat genome editing," *Cold Spring Harb. Lab.*, pp. 15–35, 2016.

[63] T. Marcussen *et al.*, "A chromosome-based draft sequence of the hexaploid bread wheat (

Triticum aestivum) genome Ancient hybridizations among the ancestral genomes of bread wheat Genome interplay in the grain transcriptome of hexaploid bread wheat Structural and functional pa," *Science*, vol. 345, no. 6194, p. 1250092, 2014.

[64] T. Wicker *et al.*, "Frequent Gene Movement and Pseudogene Evolution Is Common to the Large and Complex Genomes of Wheat, Barley, and Their Relatives," *Plant Cell*, vol. 23, no. 5, pp. 1706–1718, 2011.

[65] A. V Zimin, D. Puiu, R. Hall, S. Kingan, B. J. Clavijo, and S. L. Salzberg, "The first near-complete assembly of the hexaploid bread wheat genome, Triticum aestivum," *Gigascience*, vol. 6, no. 11, pp. gix097-gix097, Nov. 2017.

[66] Q. Shan *et al.*, "Targeted genome modification of crop plants using a CRISPR-Cas system," *Nat. Biotechnol.*, vol. 31, no. 8, pp. 686–688, 2013.

[67] S. K. Upadhyay, J. Kumar, A. Alok, and R. Tuli, "RNA-Guided Genome Editing for Target Gene Mutations in Wheat," *G3: Genes|Genomes|Genetics*, vol. 3, no. 12, pp. 2233–2238, 2013.

[68] Q. Shan, Y. Wang, J. Li, and C. Gao, "Genome editing in rice and wheat using the CRISPR/Cas system," *Nat. Protoc.*, vol. 9, no. 10, pp. 2395–2410, 2014.

[69] Y. Wang *et al.*, "Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew," *Nat. Biotechnol.*, vol. 32, no. 9, pp. 947–951, 2014.

[70] E. Harasim, M. Wesołowski, C. Kwiatkowski, P. Harasim, M. Staniak, and B. Feledyn-Szewczyk, "The contribution of yield components in determining the productivity of winter wheat (Triticum aestivum L.)," *Acta Mycol.*, vol. 69, no. 3, 2016.

[71] V. C. Gegas *et al.*, "A Genetic Framework for Grain Size and Shape Variation in Wheat,"

Plant Cell, vol. 22, no. 4, pp. 1046–1056, 2010.

[72] H. Kuchel, K. J. Williams, P. Langridge, H. A. Eagles, and S. P. Jefferies, "Genetic dissection of grain yield in bread wheat. I. QTL analysis," *Theor. Appl. Genet.*, vol. 115, no. 8, pp. 1029–1041, 2007.

[73] S. A. Goff *et al.*, "A draft sequence of the rice genome (Oryza sativa L. ssp. japonica)," *Science (80-.).*, vol. 296, no. 5565, pp. 92–100, 2002.

[74] W. Li and B. Yang, "Translational genomics of grain size regulation in wheat," *Theor. Appl. Genet.*, vol. 130, no. 9, pp. 1765–1771, 2017.

[75] X.-J. Song, W. Huang, M. Shi, M.-Z. Zhu, and H.-X. Lin, "A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase," *Nat. Genet.*, vol. 39, p. 623, Apr. 2007.

[76] J. Weng *et al.*, "Isolation and initial characterization of GW5, a major QTL associated with rice grain width and weight," *Cell Res.*, vol. 18, p. 1199, Nov. 2008.

[77] S. Wang *et al.*, "The OsSPL16-GW7 regulatory module determines grain shape and simultaneously improves rice yield and grain quality," *Nat. Genet.*, vol. 47, no. 8, pp. 949–954, 2015.

[78] J. Luo *et al.*, "An-1 Encodes a Basic Helix-Loop-Helix Protein That Regulates Awn Development, Grain Size, and Grain Number in Rice," *Plant Cell*, vol. 25, no. 9, pp. 3360–3376, 2013.

[79] L. Zhang *et al.*, "TaCKX6-D1, the ortholog of rice OsCKX2, is associated with grain weight in hexaploid wheat," *New Phytol.*, vol. 195, no. 3, pp. 574–584, 2012.

[80] Y. Zhang, J. Liu, X. Xia, and Z. He, "TaGS-D1, an ortholog of rice OsGS3, is associated with grain weight and grain length in common wheat," *Mol. Breed.*, vol. 34, no. 3, pp.

1097–1107, 2014.

[81] P. Duan *et al.*, “Natural Variation in the Promoter of GSE5 Contributes to Grain Size Diversity in Rice,” *Mol. Plant*, vol. 10, no. 5, pp. 685–694, 2017.

[82] W. Wang *et al.*, “Gene editing and mutagenesis reveal inter-cultivar differences and additivity in the contribution of TaGW2 homoeologues to grain size and weight in wheat,” *Theor. Appl. Genet.*, vol. 131, no. 11, pp. 2463–2475, 2018.

[83] M. Sorrells, “Comparative DNA sequence analysis of wheat and rice genomes,” *Genome …*, pp. 1818–1827, 2003.

[84] C. Engler, R. Kandzia, and S. Marillonnet, “A one pot, one step, precision cloning method with high throughput capability,” *PLoS One*, vol. 3, no. 11, 2008.

[85] Z. Zhai, T. Sooksa-nguan, and O. K. Vatamaniuk, “Establishing RNA interference as a reverse-genetic approach for gene functional analysis in protoplasts,” *Plant Physiol.*, vol. 149, no. 2, pp. 642–652, 2009.

[86] C. Saintenac *et al.*, “Identification of wheat gene Sr35 that confers resistance to Ug99 stem rust race group,” *Science (80-.).*, vol. 341, no. 6147, pp. 783–786, 2013.

[87] W. Wang *et al.*, “Transgenerational CRISPR-Cas9 Activity Facilitates Multiplex Gene Editing in Allopolyploid Wheat,” *Cris. J.*, vol. 1, no. 1, pp. 65–74, 2018.

[88] Y. Wang *et al.*, “Copy number variation at the GL7 locus contributes to grain size diversity in rice,” *Nat. Genet.*, vol. 47, no. 8, pp. 944–948, 2015.

[89] J. Ma *et al.*, “Structure and expression of the TaGW7 in bread wheat (Triticum aestivum L.),” *Plant Growth Regul.*, vol. 82, no. 2, pp. 281–291, 2017.

[90] Q. Xie *et al.*, “Pleiotropic effects of the wheat domestication gene Q on yield and grain morphology,” *Planta*, pp. 1–10, 2018.

[91] J. M. Debernardi, H. Lin, G. Chuck, J. D. Faris, and J. Dubcovsky, “microRNA172 plays a crucial role in wheat spike morphogenesis and grain threshability,” *Development,* vol. 144, no. 11, pp. 1966–1975, 2017.

[92] J. Hu *et al.*, “A rare allele of GS2 enhances grain size and grain yield in rice,” *Mol. Plant*, vol. 8, no. 10, pp. 1455–1465, 2015.

[93] X. Zhang *et al.*, “Rare allele of OsPPKL1 associated with grain length causes extra-large grain and a significant yield increase in rice,” *Proc. Natl. Acad. Sci.*, vol. 109, no. 52, pp. 21534–21539, 2012.

[94] K. K. Davison and L. L. Birch, “Epigenetic modifications in plants: an evolutionary perspective,” vol. 64, no. 12, pp. 2391–2404, 2008.

[95] C. D. Day, E. Lee, J. Kobayashi, L. D. Holappa, H. Albert, and D. W. Ow, “Transgene integration into the same chromosome location can produce alleles that express at a predictable level, or alleles that are differentially silenced,” *Genes Dev.*, vol. 14, no. 22, pp. 2869–2880, 2000.

[96] S. A. Verkuijl and M. G. Rots, “The influence of eukaryotic chromatin state on CRISPR–Cas9 editing efficiencies,” *Curr. Opin. Biotechnol.*, vol. 55, pp. 68–73, 2019.

Appendices

Supplemental figure

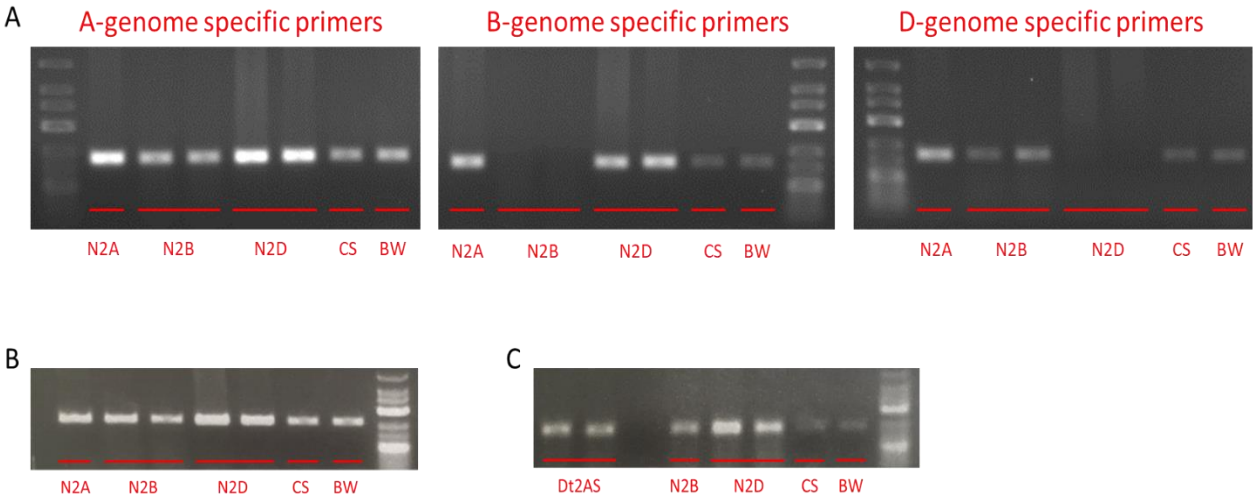

Figure 9. Genome-specific primers for *TaGW7* expression analyses

(A) PCR results for the testing of the genome-specific primers using Chinese Spring nulli-tetrasomic lines. (B) Second repeat of the same testing using a second pair of A-genome-specific primers. (C) Third repeat of the same testing using a third pair of A-genome-specific primers and a different DNA template. The labels under the pictures indicate the template used in each PCR. N2A, nulli-2A; N2B, nulli-2B; N2D, nulli-2D; CS, Chinese Spring; BW, Bobwhite; Dt2AS, 2AS-deleted.

Supplemental table

Table 7. Genomic locations of the targeting sites of all the designed gRNAs

Target	Geome location (IWGSC RefSeq v1.0)	mis-match	Target	Geome location (IWGSC RefSeq v1.0)	mis-match
GW7_T1	128395021-128395002 on chr2D	0	CKX6_T4	550050261-550050280 on chr3A	0
GW7_T1	182016287-182016306 on chr2B	0	CKX6_T5	105891712-105891731 on chr3D	0
GW7_T1	135192238-135192219 on chr2A	0	CKX6_T5	106736928-106736947 on chr3D	0
GW7_T2	128395063-128395044 on chr2D	0	CKX6_T5	157660606-157660625 on chr3B	0
GW7_T2	182016245-182016264 on chr2B	0	CKX6_T6	105891794-105891775 on chr3D	0
GW7_T2	135192280-135192261 on chr2A	0	CKX6_T6	106737010-106736991 on chr3D	0
GW7_T3	128395099-128395118 on chr2D	0	CKX6_T6	157660688-157660669 on chr3B	0
GW7_T3	182016215-182016196 on chr2B	0	CKX6_T6	550050178-550050197 on chr3A	0
GW7_T3	135192328-135192347 on chr2A	0	CKX6_T7	106737024-106737005 on chr3D	0
GW7_T4	135193685-135193704 on chr2A	0	CKX6_T7	157660702-157660683 on chr3B	0
GW7_T5	182014767-182014786 on chr2B	0	CKX6_T8	106737038-106737019 on chr3D	0
GW7_T5	128396550-128396531 on chr2D	0	CKX6_T8	105891822-105891803 on chr3D	1
GW7_T5	135193779-135193760 on chr2A	0	CKX6_T8	157660716-157660697 on chr3B	0
GW7_T6	182014748-182014729 on chr2B	0	CKX6_T8	550050144-550050163 on chr3A	0
GW7_T6	128396569-128396588 on chr2D	0	CKX6_T9	105891910-105891891 on chr3D	0
GW7_T6	135193798-135193817 on chr2A	0	CKX6_T9	106737126-106737107 on chr3D	0
GW7_T7	128396994-128396975 on chr2D	0	CKX6_T10	549906993-549906974 on chr3A	0
GW7_T7	135194223-135194204 on chr2A	0	CKX6_T11	107042290-107042271 on chr3D	0
GW7_T8	128397136-128397155 on chr2D	0	CKX6_T11	157764119-157764100 on chr3B	0
GW7_T8	135194365-135194384 on chr2A	0	CKX6_T11	549906912-549906931 on chr3A	0
GW7_T8	182014181-182014162 on chr2B	0	CKX6_T12	106315397-106315416 on chr3D	0
GW7_T9	128397571-128397590 on chr2D	0	CKX6_T12	107042294-107042313 on chr3D	0
GW7_T9	135194800-135194819 on chr2A	0	CKX6_T12	157764123-157764142 on chr3B	0
GW7_T9	182013752-182013733 on chr2B	0	CKX6_T12	549906908-549906889 on chr3A	0
An1_T1	343555502-343555483 on chr2D	0	CKX6_T13	549906887-549906868 on chr3A	0
An1_T1	408724260-408724241 on chr2B	0	CKX6_T13	107042315-107042334 on chr3D	0
An1_T1	454690516-454690497 on chr2A	1	CKX6_T14	107042300-107042281 on chr3D	0
An1_T2	343555541-343555522 on chr2D	0	CKX6_T14	157764129-157764110 on chr3B	0
An1_T2	408724299-408724280 on chr2B	0	CKX6_T14	549906902-549906921 on chr3A	0
An1_T2	454690555-454690536 on chr2A	0	GS3_T6	733695314-733695295 on chr4A	0
An1_T3	454690544-454690563 on chr2A	0	GS3_T6	6483386-6483405 on chr7D	0
An1_T4	343555641-343555660 on chr2D	0	GS3_T6	7598362-7598381 on chr7A	0
An1_T4	408724399-408724418 on chr2B	0	GS3_T7	6483412-6483431 on chr7D	0
An1_T4	454690655-454690674 on chr2A	0	GS3_T7	7598388-7598407 on chr7A	0
An1_T5	343555662-343555681 on chr2D	0	GS3_T7	733695288-733695269 on chr4A	0
An1_T5	408724420-408724439 on chr2B	0	GS3_T8	733695262-733695281 on chr4A	0
An1_T5	454690676-454690695 on chr2A	0	GS3_T8	6483438-6483419 on chr7D	0
An1_T6	343555895-343555876 on chr2D	0	GS3_T8	7598414-7598395 on chr7A	0
An1_T6	408724653-408724634 on chr2B	0	GS3_T9	6483448-6483429 on chr7D	0
An1_T6	454690903-454690884 on chr2A	0	GS3_T9	7598424-7598405 on chr7A	1
An1_T7	343555904-343555885 on chr2D	0	GS3_T9	733695252-733695271 on chr4A	1
An1_T7	408724662-408724643 on chr2B	0	GSE5_T1	190630571-190630590 on chr1B	0
An1_T7	454690912-454690893 on chr2A	0	GSE5_T1	140961176-140961195 on chr1A	0
An1_T8	343555880-343555899 on chr2D	0	GSE5_T1	126473115-126473134 on chr1D	1
An1_T8	408724638-408724657 on chr2B	0	GSE5_T2	126473361-126473342 on chr1D	0
An1_T8	454690888-454690907 on chr2A	0	GSE5_T2	190630817-190630798 on chr1B	0
An1_T9	408724852-408724871 on chr2B	0	GSE5_T2	140961422-140961403 on chr1A	0
An1_T9	454691102-454691121 on chr2A	0	GSE5_T3	190631308-190631327 on chr1B	0
An1_T10	343556128-343556147 on chr2D	0	GSE5_T3	140961883-140961902 on chr1A	0
An1_T10	454691136-454691155 on chr2A	0	GSE5_T4	126474062-126474081 on chr1D	0
CKX6_T1	105891399-105891418 on chr3D	0	GSE5_T4	190631519-190631538 on chr1B	0
CKX6_T1	106736615-106736634 on chr3D	0	GSE5_T4	140962094-140962113 on chr1A	0
CKX6_T1	157660293-157660312 on chr3B	0	GSE5_T5	190631600-190631581 on chr1B	0
CKX6_T2	106736693-106736674 on chr3D	0	GSE5_T5	140962175-140962156 on chr1A	0
CKX6_T3	106736880-106736861 on chr3D	0	GSE5_T6	126474165-126474184 on chr1D	0
CKX6_T4	106736927-106736908 on chr3D	0	GSE5_T6	190631622-190631641 on chr1B	0
CKX6_T4	157660605-157660586 on chr3B	0	GSE5_T6	140962197-140962216 on chr1A	0

The red color indicates an off-target site. The shown mismatches for some of these targets location at the 5' end of the target and do not destroy gRNA binding.

Table 8. Genotypes and phenotypic data of T2 plants used for phenotypic evaluation

Plant_ID	Cas9	Genotype	# of Seeds	TGW	Seed Area	Seed Width	Seed Length	L/W
YLD2-1-10-13	-	aaBBDD	82	25.6098	12.8989	2.9340	5.9115	2.0148
YLD2-1-10-15	+	aaBBDD	96	26.2500	13.1343	3.0152	5.9358	1.9686
YLD2-1-10-24	+	aaBBDD	97	36.1856	14.9864	3.3684	5.9034	1.7526
YLD2-1-37-1	+	aaBBDD	80	32.0000	13.6450	3.2755	5.5854	1.7052
YLD2-1-37-11	+	aaBBDD	96	31.1458	13.4485	3.2294	5.5688	1.7244
YLD2-1-37-15	+	aaBBDD	75	32.2667	14.1481	3.2680	5.7528	1.7603
YLD2-1-37-20	+	aaBBDD	100	28.6000	12.7046	3.0338	5.5429	1.8271
YLD2-1-37-3	+	aaBBDD	58	35.5172	14.5129	3.3872	5.7225	1.6895
YLD2-1-37-4	+	aaBBDD	98	33.0612	13.9313	3.2852	5.6980	1.7344
YLD2-1-37-6	+	aaBBDD	97	30.2062	13.2964	3.2340	5.4689	1.6911
YLD2-1-10-14	+	AABBDD	94	27.4468	13.3473	3.0744	5.7593	1.8733
YLD2-1-10-2	+	AABBDD	89	28.9888	13.7037	3.1710	5.7679	1.8189
YLD2-1-10-25	+	AABBDD	96	30.2083	13.7895	3.1623	5.8229	1.8414
YLD2-1-10-5	+	AABBDD	73	33.0137	14.3140	3.2894	5.8804	1.7877
YLD2-1-10-7	+	AABBDD	78	32.4359	14.2854	3.1930	6.0453	1.8933
YLD2-1-37-10	+	AABBDD	67	34.0299	14.0470	3.3558	5.5880	1.6652
YLD2-1-37-16	+	AABBDD	88	31.5909	13.5764	3.2390	5.5458	1.7122
YLD2-1-37-22	+	AABBDD	95	29.2632	12.8479	3.1347	5.4670	1.7440
YLD2-1-37-7	+	AABBDD	108	27.9630	12.3596	3.1252	5.2964	1.6947
YLD2-1-26-10	+	AAbbDD	103	33.3981	14.0970	3.3162	5.7531	1.7348
YLD2-1-26-11	+	AAbbDD	157	27.6433	12.6645	3.0711	5.5734	1.8148
YLD2-1-26-14	+	AAbbDD	91	32.9670	14.0585	3.3047	5.6806	1.7190
YLD2-1-26-7	+	AAbbDD	98	27.7551	12.6179	3.0079	5.7407	1.9086
YLD2-1-41-7	+	AAbbDD	72	31.6667	13.6063	3.2813	5.4601	1.6640
YLD2-1-52-26	+	AAbbDD	154	31.6234	13.2212	3.3063	5.3780	1.6266
YLD2-1-54-13	+	AAbbDD	30	38.6667	16.2853	3.6884	5.9168	1.6042
YLD2-1-54-18	+	AAbbDD	30	37.0000	15.5155	3.5639	5.7865	1.6236
YLD2-1-54-19	+	AAbbDD	16	38.1250	16.5380	3.7057	5.9444	1.6041
YLD2-1-26-15	+	AABBDD	84	32.2619	13.7800	3.2757	5.6504	1.7249
YLD2-1-26-16	+	AABBDD	84	32.8571	14.1741	3.2443	5.8490	1.8028
YLD2-1-26-17	+	AABBDD	87	33.2184	14.3518	3.2965	5.8430	1.7725
YLD2-1-26-19	+	AABBDD	95	27.3684	12.6843	3.0318	5.6400	1.8603
YLD2-1-26-8	+	AABBDD	93	25.5914	11.9691	2.9638	5.4421	1.8362
YLD2-1-26-9	+	AABBDD	110	29.6364	13.2662	3.1402	5.6782	1.8082
YLD2-1-54-17	+	AABBDD	46	36.7391	15.4205	3.5550	5.7536	1.6185
YLD2-1-52-18	+	AABbdd	50	31.0000	12.9289	3.2469	5.3117	1.6359
YLD2-1-52-22	+	AABbdd	123	28.8618	12.6253	3.1546	5.3976	1.7110
YLD2-1-52-31	+	AABbdd	96	29.1667	12.6064	3.1566	5.2978	1.6783
YLD2-1-52-46	+	AABbdd	84	31.7857	13.5449	3.2984	5.4387	1.6489
YLD2-1-52-32	+	AABbdd	69	25.9420	11.6667	2.9912	5.2423	1.7526
YLD2-1-52-57	+	AABbdd	89	19.8876	10.6051	2.7372	5.2465	1.9167
YLD2-1-26-1	+	AABbDD	101	31.2871	13.6474	3.1266	5.8921	1.8845
YLD2-1-26-12	+	AABbDD	67	34.6269	14.4677	3.3357	5.6953	1.7074
YLD2-1-26-13	+	AABbDD	76	29.0789	13.0626	3.1812	5.5103	1.7321
YLD2-1-26-4	+	AABbDD	83	25.5422	12.4052	2.9694	5.5636	1.8737
YLD2-1-26-5	+	AABbDD	90	31.8889	13.9338	3.2261	5.8600	1.8164
YLD2-1-41-4	+	AABbDD	156	29.4872	12.9031	3.1746	5.4319	1.7111
YLD2-1-52-1	+	AABbDD	86	29.3023	12.8254	3.0950	5.5316	1.7873
YLD2-1-52-24	+	AABbDD	140	28.1429	12.5930	3.1248	5.5243	1.7679
YLD2-1-52-27	+	AABbDD	47	31.7021	13.3499	3.2531	5.4437	1.6734
YLD2-1-54-1	+	AABbDD	58	21.7241	10.2586	2.8331	4.8957	1.7280
YLD2-1-54-10	+	AABbDD	63	34.2857	14.4261	3.3213	5.7834	1.7413
YLD2-1-54-11	+	AABbDD	11	23.6364	12.0187	3.2997	5.0624	1.5342
YLD2-1-54-14	+	AABbDD	27	39.6296	15.8361	3.5980	5.8674	1.6307
YLD2-1-54-15	+	AABbDD	51	37.8431	15.4931	3.5514	5.8866	1.6576
YLD2-1-54-20	+	AABbDD	80	34.3750	14.0407	3.3607	5.5880	1.6627
YLD2-1-54-5	+	AABbDD	86	28.3721	12.6175	3.0412	5.5787	1.8344

"Cas9": "+", Cas9 gene was detected by PCR using the plant DNA; "-", Cas9 gene was not detected by PCR using the plant DNA. "Genotype" was determined by NGS analyses and these mutations are all fixed heritable mutations. Units of the numbers: TGW, g; Seed area, mm^2; Seed width, mm; Seed length, mm. L/W: length to width ratio.

Table 9. Genotypes and phenotypic data of T3 plants used for phenotypic evaluation

Plant ID	Genotype	# of Seeds	TGW	Seed Area	Seed Width	Seed Length	L/W
BW-1	AABBDD	115	34.5217	14.3906	3.3249	6.0486	1.8192
BW-2	AABBDD	124	34.1129	14.1686	3.2537	6.0468	1.8584
BW-3	AABBDD	106	34.7170	14.2602	3.2732	6.1044	1.8649
BW-4	AABBDD	108	37.3148	15.0563	3.4220	6.1897	1.8088
BW-5	AABBDD	130	36.6923	14.8939	3.4185	6.0091	1.7578
gw7 T3 plants:							
YLD2-1-10-2-1	AABBDD	248	25.4032	12.7066	3.0034	5.9483	1.9805
YLD2-1-10-2-2	AABBDD	147	30.8844	13.8773	3.2461	5.9467	1.8320
YLD2-1-10-2-3	AABBDD	206	27.9126	12.9896	3.0505	5.9660	1.9558
YLD2-1-10-2-4	AABBDD	229	25.9389	12.6667	3.0001	5.8566	1.9521
YLD2-1-10-13-1	aaBBDD	158	28.4810	13.3832	3.1494	6.0359	1.9165
YLD2-1-10-13-2	aaBBDD	214	28.9720	13.6914	3.1884	5.8556	1.8365
YLD2-1-10-13-3	aaBBDD	254	27.5591	13.1097	3.0960	5.9184	1.9116
YLD2-1-10-15-1	aaBBDD	256	28.0078	13.3455	3.2098	5.8147	1.8115
YLD2-1-10-15-2	aaBBDD	219	27.6712	13.1470	3.1352	5.7792	1.8433
YLD2-1-37-7-1	AABBDD	163	32.0245	13.3597	3.2012	5.8047	1.8133
YLD2-1-37-7-2	AABBDD	147	34.0136	14.0392	3.2901	6.0206	1.8299
YLD2-1-37-7-3	AABBDD	154	31.7532	13.5478	3.2150	5.8918	1.8326
YLD2-1-37-7-4	AABBDD	145	34.5517	14.2274	3.2896	6.0115	1.8274
YLD2-1-37-1-1	aaBBDD	160	34.8125	14.3386	3.2833	6.0506	1.8428
YLD2-1-37-1-2	aaBBDD	143	34.1259	14.1694	3.2954	5.9606	1.8087
YLD2-1-37-3-1	aaBBDD	187	35.0802	14.4709	3.2773	6.1157	1.8661
YLD2-1-37-3-2	aaBBDD	124	35.4839	14.5076	3.3152	6.0848	1.8354
YLD2-1-26-8-1	AABBDD	169	33.0769	13.9914	3.2428	6.0325	1.8603
YLD2-1-26-8-2	AABBDD	138	33.4783	13.9361	3.2195	6.0464	1.8780
YLD2-1-26-8-3	AABBDD	189	35.7672	14.5940	3.2981	6.1445	1.8631
YLD2-1-26-8-4	AABBDD	246	31.7480	13.6649	3.1973	6.0386	1.8886
YLD2-1-26-10-1	AAbbDD	152	32.8947	14.0018	3.2149	6.0479	1.8812
YLD2-1-26-10-2	AAbbDD	157	30.3822	12.9576	3.1595	5.7612	1.8234
YLD2-1-26-10-3	AAbbDD	101	35.5446	14.6197	3.3339	6.1221	1.8363
YLD2-1-54-17-1	AABBDD	108	35.9259	14.6663	3.4704	5.9920	1.7266
YLD2-1-54-17-2	AABBDD	101	36.8317	14.8726	3.4084	6.0477	1.7743
YLD2-1-54-17-4	AABBDD	111	34.2342	14.7010	3.3891	6.0727	1.7918
YLD2-1-54-13-1	AAbbDD	123	35.5285	14.6556	3.4286	5.9147	1.7251
YLD2-1-54-13-2	AAbbDD	169	32.9586	13.7346	3.2245	5.9424	1.8429
YLD2-1-54-13-3	AAbbDD	133	36.4662	14.5206	3.3599	5.9406	1.7681
YLD2-1-54-13-4	AAbbDD	122	35.4098	14.5307	3.3957	5.9792	1.7608
YLD2-1-52-26-1	AAbbDD	138	31.9565	13.2925	3.2219	5.8387	1.8122
YLD2-1-52-26-2	AAbbDD	163	29.1411	12.6532	3.1163	5.7289	1.8383
YLD2-1-52-26-3	AAbbDD	118	33.0508	13.6450	3.2927	5.8549	1.7782
YLD2-1-52-26-4	AAbbDD	188	31.4894	13.2185	3.2164	5.7033	1.7732
YLD2-1-52-3-1	AAbbdd	182	33.2967	13.3419	3.2868	5.6353	1.7145
YLD2-1-52-3-2	AAbbdd	146	29.1781	12.7729	3.1604	5.6562	1.7897
YLD2-1-52-3-3	AAbbdd	178	32.3034	13.1651	3.2550	5.6301	1.7297
YLD2-1-52-3-4	AAbbdd	175	33.1429	13.6093	3.3267	5.6764	1.7063
YLD2-1-52-54-1	AAbbdd	155	29.8710	12.7234	3.1931	5.6214	1.7605
YLD2-1-52-54-2	AAbbdd	183	33.4426	13.2638	3.2861	5.5619	1.6926
YLD2-1-52-54-3	AAbbdd	100	35.0000	14.1567	3.3763	5.8517	1.7332
YLD2-1-52-54-4	AAbbdd	148	32.0946	13.1309	3.2740	5.5536	1.6963

"Cas9": "+", Cas9 gene was detected by PCR using the plant DNA; "-", Cas9 gene was not detected by PCR using the plant DNA. "Genotype" was determined by NGS analyses and these mutations are all fixed heritable mutations. Units of the numbers: TGW, g; Seed area, mm^2; Seed width, mm; Seed length, mm. L/W: length to width ratio.