uOttawa

Targeted Gene Editing Using CRISPR/Cas9 in a Wheat Protoplast System

Xiucheng Cui

Supervised by

Dr. Thérèse Ouellet

A thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements for the degree of

Master of Science in Biology

Department of Biology
Faculty of Science
University of Ottawa

Abstract

The clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 system has become a promising tool for targeted gene editing in a variety of organisms including plants. In this system, a 20 nt sequence on a single guide RNA (sgRNA) is the only gene-specific information required to modify a target gene. Fusarium head blight (FHB) is a devastating disease in wheat caused by the fungus *Fusarium graminearum*. The trichothecene it produces, deoxynivalenol (DON), is a major mycotoxin contaminant causing food production loss both in quality and yield. In this project, we used the CRISPR/Cas9 system to modify three wheat genes identified in previous experiments, including an ABC transporter (*TaABCC6*), and the Nuclear Transcription Factor X box-binding-Like 1 (*TaNFXL1*), both associated with FHB susceptibility, and a non-specific Lipid Transfer Protein (nsLTP) named *TansLTP9.4* which correlates with FHB resistance. Two sgRNAs were designed to target each gene and were shown in an *in vitro* CRISPR/Cas9 assay to guide the sequence-specific cleavage with high efficiency. Another assay for CRISPR/Cas9 was established by the optimization of a wheat protoplast isolation and transformation system. Using a construct expressing a green fluorescent protein (GFP) as a positive control, estimated transformation efficiencies of about 60% were obtained with different batches of protoplasts. High-throughput sequencing of PCR amplicons from protoplasts transformed with editing constructs clearly showed that the three genes have been successfully edited with efficiencies of up to 42.2%. In addition, we also characterized by RT-qPCR the expression pattern of 10 genes in DON-treated protoplasts; seven of the genes were induced by DON in the protoplasts, consistent with their previously identified DON induction in treated wheat heads, while three genes expressed differentially between DON-treated wheat heads and

protoplasts. Preliminary bioinformatics analyses showed that these differentially expressed genes are involved in different plant defense mechanisms.

Abstrait

Le système de répétitions palindromiques courtes groupées et régulièrement espacées (CRISPR) / Cas9 est devenu un outil prometteur pour l'édition ciblée de gènes dans divers organismes, y compris les plantes. Dans ce système, une séquence de 20 nt sur un guide unique d'ARN (sgRNA) est la seule information spécifique du gène nécessaire pour modifier un gène cible. La brûlure de l'épi ou fusariose (FHB) est une maladie dévastatrice du blé causée par le champignon *Fusarium graminearum*. Le trichothécène qu'il produit, le désoxynivalénol (DON), est une mycotoxine majeure qui cause une perte de production alimentaire à la fois en qualité et en rendement. Dans ce projet, nous avons utilisé le système CRISPR / Cas9 pour modifier trois gènes de blé identifiés dans des expériences antérieures, y compris un transporteur ABC (*TaABCC6*) et le facteur de transcription nucléaire X box-binding-Like 1 (*TaNFXL1*), tous deux associés à la susceptibilité à FHB, et une protéine de transfert de lipide non spécifique (nsLTP) nommée *TansLTP9.4* qui corrélèle avec la résistance à FHB. Pour chaque gène, deux sgRNA spécifiques ont été conçus, et un essai in vitro CRISPR / Cas9 a été utilisé pour montrer qu'ils peuvent guider le clivage spécifique de séquence avec un rendement élevé. Un autre essai pour CRISPR / Cas9 a été établi par l'optimisation d'un système d'isolement et de transformation de protoplastes de blé. En utilisant une construction exprimant une protéine fluorescente verte (GFP) comme un contrôle positif, des efficacité de transformation d'environ 60% ont été obtenus avec différents lots de protoplastes. Le séquençage à haut débit des amplicons de PCR provenant de protoplastes transformés avec des constructions pour édition spécifique de gènes a clairement montré que les trois gènes ont été édités avec succès dans le système de protoplastes avec des rendements allant jusqu'à 42,2%. De plus, nous avons caractérisé par RT-qPCR le profil d'expression de 10 gènes dans des protoplastes traités avec DON; sept des gènes ont été induits par le DON dans les protoplastes, en cohérence avec leur induction de DON précédemment identifiée dans les têtes de

blé traitées, tandis que trois gènes ont été exprimés différemment entre les têtes de blé traitées par DON et les protoplastes. Des analyses bioinformatiques préliminaires ont montré que ces gènes différentiellement exprimés sont associés avec différents mécanismes de défense dans les plantes.

Acknowledgement

Two years of master study is coming to an end with the completion of this thesis. Continuing my study in plant molecular biology is a precious experience since it allows me to deepen the insights of the life secrets. Starting a new life abroad alone is always difficult and it wouldn't be possible without the people who supported and journeyed with me.

First, I express my grateful thanks to my supervisor, Dr. Thérèse Ouellet, for her impartial support and patience in guiding the experiments, as well as the enormous efforts in revising thesis drafts. I would like to thank our lab technicians, Margaret Balcerzak and Hélène Rocheleau, for solving my technical problems. The mentoring and encouragement from our lab members was especially valuable for the successful completion of my master study.

Next, I would like to show gratitude to my thesis committee members, Dr. Douglas Johnson and Dr. Gopal Subramaniam, for reading my committee meeting reports and giving comments which were helpful to continue my experiments and complete them on time.

I also thank Jhadeswar Murmu and Natalie Labbé, who helped me with protoplast isolation and transformation experiments.

Lastly, my special and warmest thanks go to my parents, who supported me to pursue my dreams, and for their encouragement during tough times.

Table of Contents

Abstract ii

Abstrait iv

Acknowledgement vi

Table of Contents vii

List of Figures xi

List of Tables xiv

List of Abbreviations xv

1. Introduction 1

1.1 Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) 1

1.1.1 CRISPR/Cas9 genomic editing system in plants 5

1.2 Introduction to wheat (*Triticum aestivum*) 8

1.3 Fusarium Head Blight (FHB) 9

1.4 Genes involved in responses to FHB disease in wheat 13

1.4.1 ABC transporter 15

1.4.2 Nuclear Transcription Factor, X box-binding, like 1 (NFXL1) 19

1.4.3 Non-specific Lipid Transfer Protein (nsLTP) 21

1.5 Objectives 24

2. Materials and Methodology 25

2.1. Plant materials 25

2.2. Design and test of sgRNAs *in vitro* 25

2.2.1. Generation of sgRNA-encoding templates 26

2.2.1.1. Design of two sgRNAs for each gene 26

2.2.1.2. The amplification of sgRNA-encoding templates and *in vitro* transcription 28

2.2.2. Design primers to amplify and sequence DNA cleavage template 30

2.2.3. Testing of sgRNA efficiency *in vitro* 31

2.3. Testing of sgRNAs in a wheat protoplast system 32

2.3.1. Plasmid construction 32

2.3.1.1. Design and assembly of gBlocks 32

2.3.1.2. Cloning of assembled gBlock pairs into pJet1.2/blunt vector 33

2.3.1.3. Transfer of assembled gBlock pairs into an expression vector 35

2.3.2. Protoplast isolation and transformation via polyethylene glycol (PEG) mediated method 36

2.3.2.1. Wheat protoplast isolation 37

2.3.2.2. Protoplast transformation via PEG mediated method 38

2.4. Quantification of gene editing in modified protoplasts by high throughput sequencing 39

2.4.1. PCR amplicons preparation and sequencing 39

2.4.2. Sequencing data analyses 41

2.4.3. Sequence alignment using the latest wheat reference sequence database 42

2.5. Expression analysis of nine DON-induced genes in protoplast system by RT-qPCR 43

2.5.1. Concentration course experiments 43

2.5.2. Time course experiments 44

2.5.3. Gene functional analysis 45

3. Results 46

3.1. *In vitro* test of candidate sgRNAs 46

3.1.1. sgRNA design 46

3.1.2. Targeted gene fragment amplification 48

3.1.3. Sanger sequencing results of PCR amplicons 48

3.1.4. *In vitro* test for six sgRNAs 49

3.2. Testing of sgRNAs in a wheat protoplast system 50

3.2.1. gBlocks designing and expression vector construction 51

3.2.2. Wheat protoplast isolation and transformation 54

3.2.2.1. The optimization of wheat protoplast isolation and transformation 54

3.2.3. Mutation detection and quantification by high throughput sequencing 57

3.2.3.1. Sequencing results for pCam-nsLTP and pCam-NFXL1 target regions 57

3.2.3.2. Sequencing results for ABCC6 and a Cas9 efficiency comparison at NFXL1 target regions 61

3.2.4. Off-target analyses using the latest wheat reference sequence database 68

3.3. Expression analyses of nine DON-induced genes in a transient protoplast system 70

3.3.1. Concentration course experiment 71

3.3.2. Time course experiment 73

3.3.3. Functional annotation of the homologs of DON-induced genes in *Arabidopsis* 76

4. Discussion 79

4.1. Considerations in optimizing CRISPR for future directions 88

4.1.1. The optimization of sgRNA design 88

4.1.2. The diversity of Cas9 nucleases 89

4.1.3. The delivery system of CRISPR/Cas9 complex 90

4.2. Conclusion 92

References 93

Appendices 108

List of Figures

Figure 1. Naturally occurring CRISPR-Cas9 system..... 3

Figure 2. Illustration of the bio-engineered CRISPR/Cas9 system.. 4

Figure 3. Types of ABC transporters 15

Figure 4. The relative expression level for *TaABCC6* in two FHB-susceptible (S) and two FHB-resistant (R) wheat cultivars.. 18

Figure 5. FHB symptoms after virus-induced gene silencing of the genes *TaNFXL1* or *TaABCC6*..... 18

Figure 6. The relative expression level for *TaNFXL1* in two FHB-susceptible (S) and two FHB-resistant (R) wheat cultivars. 21

Figure 7. Schematic representation of the sgRNA *in vitro* test.. 26

Figure 8. Schematic illustration of a selected sgRNA position on three wheat genomes A, B and D. 28

Figure 9. A 69 nt sequence used as a forward primer. 29

Figure 10. Key steps to test sgRNAs in a wheat protoplast system. 32

Figure 11. Illustration of the gBlock components..... 33

Figure 12. Visual representation of wheat protoplasts isolation (A) and transformation (B)..... 37

Figure 13. Examples of output files from InDel, Structural Variant (SV) and breakpoint (BP) analyses. 42

Figure 14. PCR amplification of fragments for three genes. 48

Figure 15. The structure of amplified regions for *TaABCC6* (A), *TansLTP9.4* (B) and *TaNFXL1* (C) genes..... 49

Figure 16. *In vitro* test of six sgRNAs for the *TaABCC6* (A), *TansLTP9.4* (B) and the *TaNFXL1* (C)..... 50

Figure 17. Gel electrophoresis of PCR amplicons from 27 clones for ABC-pJet constructs. 52

Figure 18. Gel electrophoresis of isolated plasmids ABC-pJet 1~11 before and after restriction enzyme digestion. 52

Figure 19. Gel electrophoresis of PCR amplicons from 19 clones for pCambia-ABC constructs. 53

Figure 20. The effect of incubation period on transformation efficiency. 55

Figure 21. The effect of the cell amounts on transformation efficiency. 56

Figure 22. Alignments of the high-throughput sequencing reads with deletion for *TansLTP9*.4 target regions and their respective editing frequency at target regions. 59

Figure 23. Alignments of the high-throughput sequencing reads with deletion for *TaNFXL1* target regions and their respective editing frequency at target regions. 60

Figure 24. The BLAST results for three insertions found at the NFXL1-sgRNA1 site. 61

Figure 25. Alignments of the high-throughput sequencing reads with deletion for *TaABCC6* target regions and their respective editing frequency at target sites. 64

Figure 26. Alignments of the high-throughput sequencing reads with deletion for the pcoNFXL1 target regions and their respective editing frequency at target sites. 65

Figure 27. Alignments of the high-throughput sequencing reads with deletion for the resequenced *TaNFXL1* target regions and their respective editing frequency at target sites. 66

Figure 28. The viability of protoplasts with water or DON treatment 72

Figure 29. The concentration course experiments of four tested genes. 73

Figure 30. Relative expression levels of six highly up-regulated genes under DON treatment in the time course experiments. 75

Figure 31. Relative expression levels of four genes with lower expression change under DON treatment in the time course experiments.. 76

List of Tables

Table 1. A subset of genes induced by DON in wheat heads 14

Table 2. Database searching results to identify predicted coding and genomic sequences for the genes of interest. 47

Table 3. Selected sgRNAs for each gene. 47

Table 4. The estimated transformation efficiency of protoplasts using different number of cells and two incubation periods..... 56

Table 5. A summary of estimated editing efficiency for the 10 sequenced samples in the first round of amplicon sequencing.. 58

Table 6. A summary of estimated editing efficiency for the 15 sequenced samples in the second round of sequencing.. 67

Table 7. The estimated editing efficiency on sgRNA1 or 2 site for each sample from two rounds of sequencing..... 68

Table 8. Comparison of the six sgRNAs and target gene fragments used as the reference in NGS to the IWGSC Reference Sequence v1.0 database.. 69

Table 9. The predicted gene functions in *Arabidopsis*. 78

List of Abbreviations

Abbreviations	Explanation
AAFC	Agriculture and Agri-Food Canada
ABC transporter	ATP-binding cassette transporter
BLAST	Basic Local Alignment Search Tool
BSA	Bovine Serum Albumin
Cas9	CRISPR associated protein 9
crCas9	*Chlamydomonas reinhardtii* codon optimized spCas9
CRD	C-terminal Regulatory Domain
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CV	Coefficient of Variation
DNA	Deoxyribonucleic Acid
DON	Deoxynivalenol
DSB	Double-strand Break
EST	Expressed Sequence Tag
FHB	Fusarium Head Blight
GST	Glutathione S-transferase
HDR	Homology-directed Repair
IDT	Integrated DNA Technologies
IWGSC	International Wheat Genome Sequencing Consortium
MES	2-(N-morpholino) ethanesulfonic acid
MRP	Multidrug Resistance Protein
NBD	Nucleotide-binding Domain

NFXL1	Nuclear Transcription Factor, X box-binding, Like 1
NGS	Next Generation Sequencing
NHEJ	Nonhomologous End Joining
nsLTP	non-specific Lipid Transfer Protein
OD	Optical Density
ORDC	Ottawa Research and Development Center
PAM	Protospacer Adjacent Motif
pcoCas9	Plant codon optimized spCas9
PCR	Polymerase Chain Reaction
PDR	Pleiotropic Drug Resistance
PEG	Polyethylene Glycol
ppm	parts per million
PR	Pathogenesis Related
QTL	Quantitative Trait Locus
RNA	Ribonucleic Acid
RNP	Ribonucleoprotein
RPM	Revolutions Per Minute
RT-qPCR	Quantitative Reverse Transcription PCR
RVD	Repeat Variable Di-residue (amino acids)
SBP	Substrate-binding Protein
SD	Standard Deviation
sgRNA	single guide RNA
spCas9	*Streptococcus pyogenes* Cas9

TALENs	Transcription Activator-like Effector Nucleases
TMD	Transmembrane Domain
URGI	Unité de Recherche Génomique Info
UTR	Untranslated Region
ZFNs	Zinc Finger Nucleases

1. Introduction

Fusarium head blight (FHB) is a major disease in wheat and other crops. So far, many wheat genes have been identified which are closely associated with this disease. However, the roles of these genes are still poorly understood. Clustered regularly interspaced short palindromic repeats (CRISPR)/Cas system is a new emerging way for efficient gene editing. Wheat gene modification was difficult before the appearance of this technology. In this project, this technology is used to initiate the characterization of three genes contributing to FHB susceptibility or resistance.

This thesis includes four parts. The most relevant literature and recent research progress will be discussed in the first chapter, including CRISPR technology, a short introduction to the wheat genome, FHB and genes involved in regulating this disease. Detailed objectives and methodologies will be presented in the second chapter. Chapter 3 includes experimental results and data analysis. Finally, a discussion about this project and future work are addressed in Chapter 4.

1.1 Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)

Known as an adaptive immune system, type II prokaryotic clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) system was firstly identified in 2007 (Barrangou et al., 2007). It widely exists in both Eubacteria and Archaea. It uses short RNA to detect and to degrade invading viral or plasmid DNAs during further attacks by similar viruses (Bhaya et al., 2011; Wiedenheft et al., 2012). Initial attempts in gene editing used zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) as the two major approaches in higher eukaryotic organisms. FokI nuclease domains are essential in both of those

techniques for DNA cleavage. In the ZFN system, 3-6 zinc finger modules are often customized for target recognition while in the TALEN system, the target sequence is often recognized by assembled repeat variable di-residues (RVDs) (Carlson et al., 2012). Both approaches will induce double strand break at target sites and trigger DNA endogenous repairing. However, design, synthesis and screening of customized zinc finger modules or RVDs are time consuming and expensive (Mahfouz et al., 2011). Following ZFNs and TALENs, the CRISPR technology has become one of the most widespread ways to specifically modify genomic DNA due to its higher efficiency and lower cost. In 2013, the first application of CRISPR as a gene editing tool was reported in human and mouse cells, demonstrating that Cas9 nucleases could induce precise cleavage at genome loci with the presence of short RNA guiding sequences and that such activities could convey simultaneous editing events at different sites within the genome, indicating the wide applicability and simplicity of the CRISPR/Cas9 system (Cong et al., 2013). Recent experiments also revealed that the CRISPR technology could be used in various life species, from bacteria to plants and animals (Feng et al., 2013; Jiang et al., 2013a; Shalem et al., 2014).

In naturally occurring circumstances, a variable RNA sequence transcribed from invading DNA, known as protospacer, combines with a CRISPR repeat to form a complex called crRNA. Each crRNA combines another RNA piece, called tracrRNA, forming crRNA-tracrRNA hybrids. These two RNAs and a Cas9 nuclease together form a ribonucleoprotein complex, named Cas9:crRNA-tracrRNA complex. If there is a special 5'-NGG structure named PAM (Protospacer Adjacent Motif) adjacent to the 3' end of the target site on the genomic sequence, then the crRNA will guide Cas9 nuclease to cleave the double-stranded genomic DNA template (Figure 1).

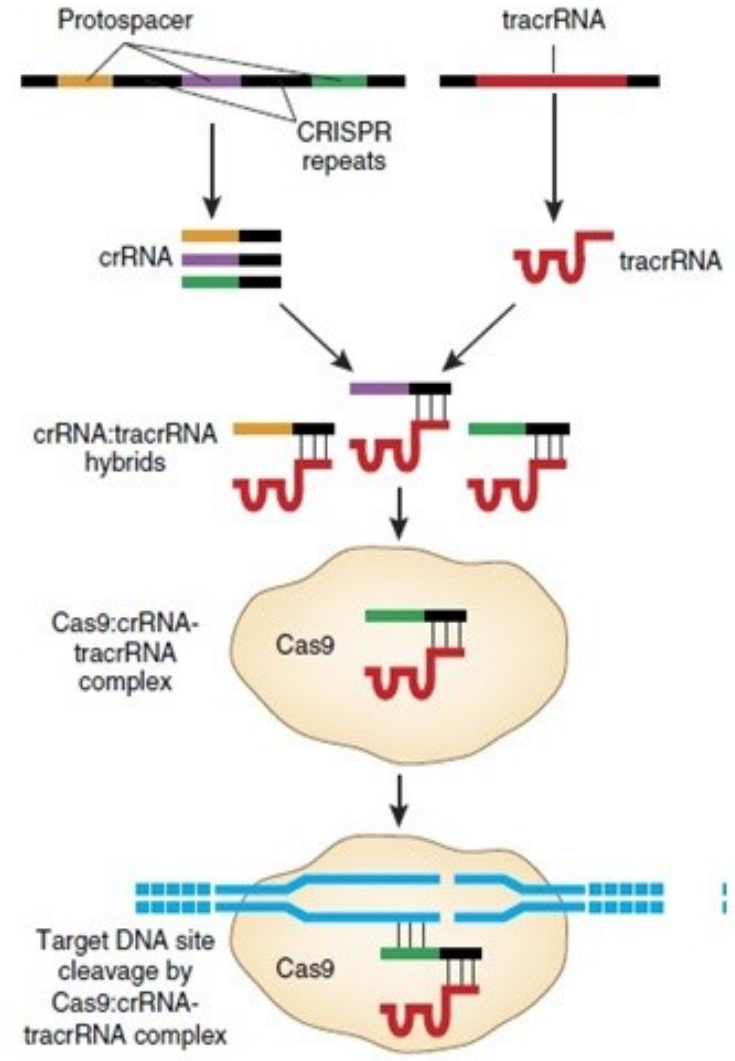


Figure 1. Naturally occurring CRISPR-Cas9 system. crRNA can combine to tracrRNA to form crRNA:tracrRNA complex. This RNA complex can associate with Cas9. The RNA complex is responsible for recognition of foreign DNA and Cas9 nuclease will cleave the DNA sequence at the target site. Reproduced from Sander & Joung, (2014).

Generally, the strategy in genome editing with this technology includes four major steps: selection of targeted sites on genomic DNA; generation of constructs with complementarity to the targets; transfection and evaluation of editing efficiency. The most widely applied CRISPR/Cas9 system uses a fusion of crRNA and a part of tracrRNA (Figure 2) instead of two

individual RNAs, which makes this system simpler and easier to apply. This complex is known as a single guide RNA (sgRNA) and, after design for target specificity, the sgRNAs can be ordered directly from commercial companies. By recruiting a Cas9 nuclease, the function of this complex is similar to the Cas9:CrRNA-tracrRNA system. A sgRNA includes two parts: the guide sequence which is 20 nt in length, and a scaffold structure. The 20 nt spacer sequence determines the specificity of sgRNA, and the scaffold structure is required for Cas9 binding. The guide sequence needs to be complementary to the target sequence in the genomic DNA, and a PAM is required to be located immediately downstream of target sites for successful sgRNA binding and Cas9 cleavage.

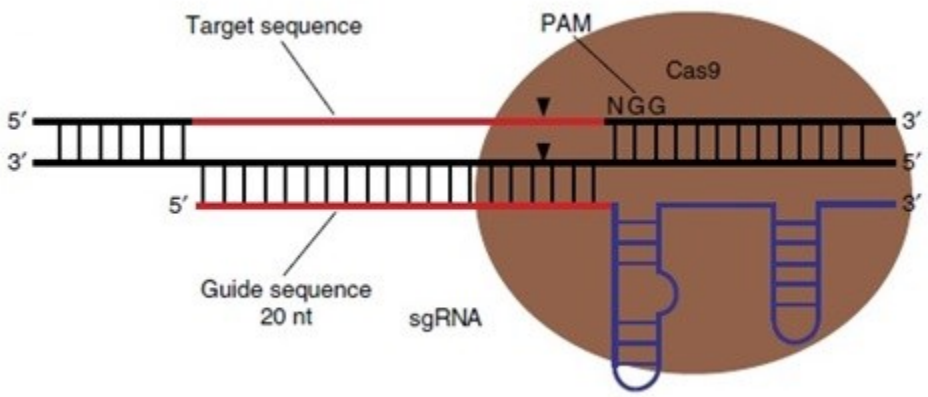


Figure 2. Illustration of the bio-engineered CRISPR/Cas9 system. The 20 nt sgRNA sequence (red) is adjacent in 5' to the PAM (NGG) structure. It will recognize the target sequence on genomic DNA and recruit a Cas9 nuclease (in brown) to cleave the target sequence on each strand (arrow heads). The blue part on the sgRNA represents the scaffold which is used to maintain the hairpin structure required by Cas9. Reproduced from Shan et al. (2014).

Eukaryotes have two ways for repairing double-strand breaks (DSBs), nonhomologous end joining (NHEJ) and homology-directed repair (HDR). These two pathways can both work in CRISPR/Cas9 system (Ran et al., 2013b). The NHEJ pathway is imprecise, which could cause

variable lengths of deletions or insertions at the DSB site to disrupt a specific gene, leading to silencing of the target genes by reading frame shifting. In contrast, the HDR pathway could precisely cause point mutations (Daley et al., 2005; Gaj et al., 2013). By this repairing method, an artificially created DNA fragment showing homology with both sequences at the break sites can be introduced into the targeted site. These exogenous fragments can be reporter genes or antibiotic resistance genes which would help for mutation detection and analysis (Nemudryi et al., 2014).

1.1.1 CRISPR/Cas9 genomic editing system in plants

The CRISPR/Cas9 system has been widely used in editing both monocot and dicot plant genomes. And it has been shown to have a high potential for multiple gene editing (Xie et al., 2015). To date, model plants such as *A. thaliana* and tobacco (Jiang et al., 2013b) have been successfully edited using this technology. In crops, successful modifications of rice (*Oryza sativa*) (Miao et al., 2013), wheat (Zhang et al., 2016), maize (*Zea mays*) (Feng et al., 2016) and sorghum (Jiang et al., 2013b) genomes have also been reported.

Many efforts have been done to optimize this system in plants, including enhancing editing efficiency, minimizing the off-target effect and the development of rapid validation system, since obtaining transgenic plants can be very time-consuming.

Optimizing Cas9 is an approachable way to enhance editing efficiency in plants. The Cas9 protein from *Streptococcus pyogenes* (spCas9) is the first endonuclease shown to induce locus-specific editing in mammalian cells (Cong et al., 2013). In plant systems, a human-codon optimized Cas9 was firstly used in editing *A. thaliana* genome, and three genes, *BRI1*, *JAZ1* and *GAI* were targeted, with efficiencies ranging from 30% to 84% (Feng et al., 2013). Subsequently,

different kinds of Cas9 proteins including a variety of plant codon-optimized (Jiang et al., 2013b), rice codon-optimized (Zhou et al., 2014) and maize codon-optimized (Xing et al., 2014) have been used, which appear to be more active and with a higher efficiency. Ran et al. (2013a) demonstrated that by using a Cas9 mutant named D10A nickase (Cas9n) and a pair of sgRNAs targeting opposite strands of genomic DNA at a site of interest, the editing specificity can be over 100-fold greater than with the normal Cas9/sgRNA complex.

Designing highly specific sgRNAs is an alternative choice to minimize the off-target effect. The 20 bp single nucleotide sequence adjacent to the PAM structure is a key element in the CRISPR system. sgRNAs are generally driven by U3 or U6 promoters in plants to enhance and stabilize their expression (Shan et al., 2013). Studies on optimizing sgRNA activities have found that the sequence features could affect off-target effects. Anderson et al. (2015) found that 2 mismatches on the crRNA-target complex was enough to stop the cleavage reactions therefore reduce the off-target effect. In contrast, Lin et al. (2014) found that, on target sequences, Cas9 cleavage could tolerate up to a 5-bases difference in insertions (DNA bulge because target sequence is longer than sgRNA) or in deletions (RNA bulge because target sequence is shorter). By screening 1,841 sgRNAs in mammalian cells, base G at the 20^{th} position, and base C at the 16^{th} position were found to be favored by functional sgRNAs (Doench et al., 2014). In addition, software has been developed to select appropriate sgRNAs for genome editing, such as WU-CRISPR tool, E-CRISP, sgRNA Scorer 1.0 and CRISPR MultiTargeter (Wang et al., 2016).

Since transgenic plants can take several months to regenerate, transient expression assays in a protoplast system have been developed as an efficient way to detect possible genomic editing events. The mutations in protoplasts can be detected by either a PCR/restriction enzyme (PCR/RE) or a T7 endonuclease I (T7EI) assay. T7EI can cleave mismatches on double strand

DNA sequences. In order to form heteroduplexes, PCR products containing both wild-type and mutant amplicons from modified genomes are denatured and then renatured. Therefore, mutations caused by CRISPR/Cas9 system will be validated if two smaller fragments can be observed on the gel separation after T7E1 cleavage. Such assays have been used in gene editing experiments in rice (Shan et al., 2013) and *A. thaliana* (Feng et al., 2013) protoplasts and showed editing efficiencies from 26.5% to 28%, and 18.8%, respectively. More recently, the next generation sequencing (NGS) technology has been employed to detect mutation events and editing efficiency (Wang et al., 2016). This technology was firstly developed in the mid-1990s to early-2000s, and presents advantages especially for large-scale sequencing with lower cost. The NGS usually aims at sequencing large DNA pieces such as chromosomes, or a large number of shorter DNA fragments like PCR products. For gene editing efficiency validation, targeted regions are firstly amplified from modified genomic DNA by PCR, after which thousands or even millions of PCR amplicons from targeted regions can be analyzed by NGS to determine editing efficiency.

It is more challenging to edit the hexaploid wheat genome than it is to edit dicot species such as *Arabidopsis*, rice and maize. Reasons include the genome size of wheat, its polyploidy (6X), incomplete genomic sequencing information and high-level sequence similarity between its three genomes (see Subsection 1.2 for more about wheat genome). Before the CRISPR system was developed, only one study reported on a wheat gene, mildew-resistance locus (MLO), that was successfully edited by TALENs (Wang et al., 2014b). More reports concerning wheat gene editing emerged after the development of the CRISPR system. Even though, such occurrences are still limited in number. Only a few research teams have reported successful gene editing in

wheat so far (Shan et al., 2013; Upadhyay et al., 2013; Wang et al., 2016). Gene editing in wheat will likely benefit from optimization of the CRISPR/Cas9 system.

1.2 Introduction to wheat (*Triticum aestivum*)

Wheat (*Triticum aestivum*) plays a critical role in food and feed crop consumption worldwide. It is now cultivated on more land than any other crops around the world (McMullen et al., 2012). The estimated Canadian wheat production was 30.5 million tonnes in 2016, up 10.5% compared with the previous year ("The Daily — Production of principal field crops, July 2016,"). Currently, wheat is second to rice as a major crop for human consumption and accounts for 20% of all calories consumed around the world. It is expected that food demand would double by the year of 2050 (Green et al., 2005). With the increase of population, available areas for cultivation have decreased significantly during the past decades. Food security and shortage is becoming a major problem worldwide. Other factors like water pollution, the abuse of pesticides and soil contamination make the problem even worse. Many avenues are being considered to alleviate this problem, including the characterization of interactions between host and pathogens, to develop disease-resistant cultivars.

Bread wheat is a hexaploid plant, with three homoeologous genomes (A, B and D), each containing seven chromosomes. It originated from hybridization between emmer wheat (AABB, *T. dicoccon*) and the goat grass (DD, *Aegilops tauschii*) (Petersen et al., 2006). Those homoeologous genomes share over 95% sequence identity across coding regions in the wheat genome and in bread wheat, most genes have copies on each of the three genomes, A, B and D (Borrill et al., 2015). The size of the wheat genome is extremely large, up to 17 Gb, and over 80% of it is composed of highly repetitive sequences and transposable elements. More than 120,000

gene loci have been annotated (IWGSC, 2014). In contrast, the genome sizes for rice and *Arabidopsis thaliana* are around 430 Mb and 135 Mb, respectively. All of those factors make genome analysis more difficult in wheat than in other crops like rice, barley (H*ordeum vulgare L.)* or maize.

1.3 Fusarium Head Blight (FHB)

Fusarium head blight (FHB) is a devastating disease occurring worldwide, especially in wheat and other small grain cereals (Stack, 2000). This disease has caused great loss in food production, both in quality and yield. The accumulation of mycotoxins, especially the trichothecene deoxynivalenol (DON), a potential inhibitor of protein biosynthesis, has a negative impact on both human and animal health (Sobrova et al., 2010).

FHB mainly caused by the filamentous fungus *Fusarium graminearum* Schwabe (teleomorph *Gibberella zea*). This fungus infects cereal crops, including wheat, barley and maize (Kazan et al., 2017). The early symptoms of FHB usually appear shortly after flowering. The infected spikelets show premature bleaching as the fungus spread within the heads. Diseased plants may fail to produce grains, or produce grains that are poorly filled. In addition, the grain quality is negatively affected by DON, which is a major mycotoxin produced by *F. graminearum*. This chemical compound is quite stable and can withstand high temperatures up to 350 °C. If DON-contaminated crops are consumed by humans, it may cause nausea and vomiting, as well as abdominal pain, fever and dizziness (Pestka, 2007; Sobrova et al., 2010). Further, it may also induce other severe symptoms such as chromosome aberrations and failure in immune systems. Therefore, this disease has deep economic impacts on agriculture and economy, causing food safety problems and great loss in food production. Previous studies have shown that *F.*

graminearum can survive on the residue of grains from preceding years and re-infect wheat at flowering time, which is the most susceptible stage in its lifecycle (McMullen et al., 2012). The life style of this fungus could be defined as hemibiotrophic, having the ability to first infect the host without killing the cells and then to transition to a necrotrophic mode where there is cell death. According to recent studies in wheat, the wild-type *F. graminearum* strain PH-1 could stay in the intercellular space between rachis cells during the early phases of infection before invading host cells, which may subsequently cause cell death and necrosis (Brown et al., 2010). Crown infection caused by *F. graminearum* also seems to follow a similar process (Stephens et al., 2008).

Fungal invasion is a complex process, with cooperation between different genes and with interactions between the host and the pathogen. A considerable amount of work has been done in this area, including transcriptomic, proteomic and metabolomic analyses. For example, using a transcriptomic approach, researchers have determined the expression level of *F. graminearum* genes during mycotoxin biosynthesis (Brown et al., 2001), the invasion stage (Güldener et al., 2006; Zhang et al., 2012), in infecting different hosts (Harris et al., 2016) and in response to different environmental conditions (Schmidt-Heydt et al., 2011).

It is wildly accepted that the plant resistance to FHB can be divided into two types: the resistance to initial infection (type I), and the resistance to spread within the head (type II) (Schroeder & Christensen, 1963). Both morphological and physical traits may enhance type I resistance, like the thickening of cell walls and the waxy coating of spikelets. Mechanisms of type II resistance have not been elucidated yet.

Quantitative trait locus (QTL) stands for a region of genomic DNA that correlates with phenotypic variations in different individuals. A QTL is typically related to genes which control that designated phenotype. So far, over 100 QTLs associated with FHB resistance have been identified in wheat. *Fhb1*, located on chromosome 3B, and *Qhfs.ifa-5A*, on chromosome 5A, are thought to play a major role in FHB resistance (Kugler et al., 2013; Schweiger et al., 2013). It has been shown that these two QTLs contribute to the type II and type I resistance, respectively. A recent research has successfully cloned *Fhb1* from Sumai 3. Using gene silencing and transgenic overexpression, a pore-forming toxin-like (PFT) gene encoding a chimeric lectin and located at this locus was shown to confer FHB resistance (Rawat et al., 2016). Yet, elucidating the mechanisms by which *Fhb1* and *Qhfs.ifa-5A* operate remains challenging.

RNA expression and profiling analyses have shown that many wheat genes are induced after infection by *F. graminearum*. The expression level of these genes shows a significant difference at different time points. For example, three pathogenesis-related (PR) genes, *PR-2*, *-4* and *-5* were found to have a higher expression level in the resistant wheat cultivar Sumai 3 after *F. graminearum* infection when compared with two susceptible near isogenic lines (NILs). These two NILs are derived from a cross between Sumai 3 and a susceptible cultivar, Chuan 980 (Golkari et al., 2009). In a comparison between FHB-resistant cultivar Wangshuibai and susceptible Meh0106 mutant derived from Wangshuibai, using a combined transcriptomic and proteomic approach, the expression level of 77 proteins showed at least 1.5-fold change in one or both genotypes 12 hours after *F. graminearum* infection and 163 up-regulated genes were found to be defense-related (Ding et al., 2011). Some studies also found that the overexpression of defense responsive genes could enhance FHB resistance in wheat. Loss-of-function analysis of an *Arabidopsis NPR1* (*AtNPR1*) gene in transgenic plants showed a reduced tolerance to

different kinds of pathogens (Dong, 2004). Transgenic wheat overexpressing this gene exhibited higher resistance to FHB, including a faster response of defense-related genes, and a higher wheat seed yield in infected plants (Makandar et al., 2006). Alpha-1-purothionin, thaumatin-like protein 1 (tlp-1), and β-1, 3-glucanase are three wheat genes that can be induced by FHB. The overexpression of these genes in transgenic wheat cultivar Bobwhite led to an enhanced tolerance to FHB both in greenhouse and field conditions, with fewer disease symptoms and lower DON concentration, indicating that these genes play a role in controlling FHB disease positively (Mackintosh et al., 2007).

Recent research showed that some plant hormones are also important for FHB resistance. It is generally considered that salicylic acid (SA) contributes to biotrophic defense, while jasmonic acid (JA) and ethylene (ET) are involved in defense against necrotrophic pathogens (Koornneef & Pieterse, 2008). In-plate experiments showed that *F. graminearum* mycelia growth and spore germination were strongly inhibited by addition of SA. Further, co-inoculation of SA and *F. graminearum* on wheat heads also showed reduced disease symptoms in both resistant and susceptible wheat cultivars (Qi et al., 2012). NAUH117, a susceptible mutant derived from Wangshuibai, has a lower endogenous JA level compared with Wangshuibai. Pretreatment with exogenous methyl JA prior to the inoculation of *F. graminearum* on wheat heads also helped reduce disease symptoms. In contrast, applying exogenous ethephon did not have such effects (Sun et al., 2016). Additionally, still found by Sun et al. (2016), an *Arabidopsis* JA-associated *LTP* gene (*AT5G59310*), which is the homolog of wheat *LTP* gene Ta.23917.3.S1_x_at,.was mutated by T-DNA insertion in *Arabidopsis*. Transgenic *Arabidopsis* mutant plants showed more severe disease symptoms compared with the wild type plants after *F. graminearum* inoculation.

Taken together, this suggested that this wheat LTP gene might contribute to FHB resistance through JA pathways.

1.4 Genes involved in responses to FHB disease in wheat

Previous studies using microarray analysis with confirmation by RT-qPCR have shown that in the susceptible wheat cultivar Roblin, the expression level of some genes can be greatly up-regulated in heads infected with a wild type *F. graminearum* strain (DON+), but remained unchanged when the heads were inoculated with a *F. graminearum* mutant which lost its ability to produce DON (DON-) (Balcerzak et al., 2009). Further studies also showed that many of those genes can be induced by direct treatment of wheat heads with DON, indicating that those genes were responsive to DON directly, not to the pathogen itself (Ouellet et al., 2012). A list of those genes is presented in

Table 1. In this project, I focused on three genes, two of the genes in

Table 1, the ABC transporter (Traes_2AL_2F9F6DB1A.1) and a transcription factor called NFXL1 (Nuclear Transcription Factor, X box-binding, Like 1, Traes_7DL_ACACA513D.1), as well as a non-specific Lipid Transfer Protein (nsLTP) gene associated with FHB resistance. In addition, the expression patterns of the other genes listed in

Table 1 were further characterized in a wheat protoplast system. The three genes are introduced in more details in the following sections.

Table 1. A subset of genes induced by DON in wheat heads

Affymetrix probeset name	Traes name[#]	UniGene name	Annotation from GeneChip
Ta.9831.1.S1_at	Traes_2AL_2F9F6DB1A.1	Ta.9831	MRP-like ABC transporter [*Oryza sativa* (japonica cultivar-group)]
Ta.11647.2.S1_at	Traes_7DL_ACACA513D.1	Ta.11647	TF-like protein [*Oryza sativa* (japonica cultivar-group)]
TaAffx.1095.1.A1_at	Traes_7DL_5968FA56C.1	CK214015.1*	WRKY2 protein [*Hordeum vulgare subsp. vulgare*]
TaAffx.12215.1.A1_at	Traes_1DL_2200EA8BC.1	Ta.8604	AAA-type ATPase [*Oryza sativa* (japonica cultivar-group)]
TaAffx.12876.1.S1_at	Traes_5DL_9D713AD9C.1	BJ220837.1*	Cys2/His2 zinc-finger protein [*Oryza sativa* (japonica cultivar-group)]
Ta.1221.1.A1_at	Traes_5DL_9E028D04A.1	Ta.1221	Aspartyl-tRNA synthetase [*Oryza sativa* (japonica cultivar-group)]
TaAffx.64766.1.S1_at	Traes_7DL_4C3A77D7A.1	Ta.50890	Glutathione transferase
Ta.5297.1.A1_at	Traes_7BL_E0E018397.1	Ta.5297	unknown
Ta.22180.1.S1_x_at	Traes_5BL_F5D379AFC.1	Ta.22180	unknown

[#] From the survey sequence gene models, MIPS v2.2, made public in July 2014 (https://wheat-urgi.versailles.inra.fr/Seq-Repository/Genes-annotations)

* Indicates GeneBank ID.

1.4.1 ABC transporter

ATP-binding cassette (ABC) transporters are members of a protein family, which is one of the largest and the oldest families and can be found in bacteria, fungi, animals and plants. They are involved in multiple cellular processes (Jones & George, 2004). These transporters are transmembrane proteins which can use energy from ATP hydrolysis to carry small metabolites between cell compartments, including in and out of the cell (Davidson et al., 2008). All ABC transporters have four subunits: two transmembrane domains (TMDs) and two nucleotide-binding domains (NBDs). The NBDs are conserved in all ABC transporters, while the TMDs are less conserved (Rice et al., 2014). Most ABC importers also contain a subunit called substrate-binding protein (SBP). In some transporters, additional domains with a regulatory function may exist, named C-terminal regulatory domain (CRD) (ter Beek et al., 2014). It is wildly acknowledged that there are four types of ABC transporters (Figure 3).

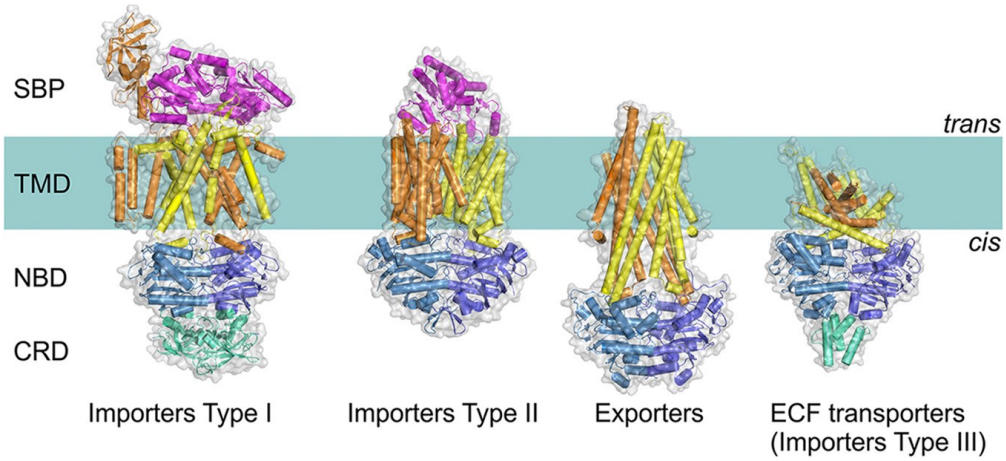


Figure 3. Four different types of ABC transporters. They share similar structures of transmembrane domains (TMDs, two subunits in orange and yellow) and nucleotide-binding domains (NBDs, two subunits in dark and light blue). In some cases, the transporters may have additional substrate-binding proteins (SBPs) or C-terminal regulatory domains (CRDs, in green). Reproduced from ter Beek et al. (2014).

In *Arabidopsis*, over 100 ABC transporters have been found but only 22 have been functionally characterized (Kang et al., 2011). They were first thought to be involved in the detoxification process (Martinoia et al., 1993). Subsequently, plant ABC transporters were shown to have multiple functions during the plant development, including plant growth, interactions between plants and the environment, and in response to abiotic stress (Kang et al., 2011).

In rice, the expression of the ABC transporter *OsABCC1* was shown to reduce arsenic (As) accumulation, suggesting a detoxification role (Song et al., 2014). This gene was expressed in many tissues, including roots, rachis and leaves; The expression pattern of this gene was not affected by low As concentration levels while its expression was up-regulated at higher concentrations of As. By different treatments mimicking different abiotic stresses, it was found that a PDR (pleiotropic drug resistance)-type ABC transporter, *Ospdr9*, can be induced by heavy metal, osmotic and low oxygen stresses in rice, especially in rice roots, suggesting that this ABC transporter may have multiple roles against abiotic stresses (Moons, 2003).

In wheat, the expression of putative ABC transporters has been associated with multiple disease resistance. For example, a PDR-like ABC transporter, *TaPDR1*, was found to be up-regulated by both DON treatment and *F. graminearum* infection. Its expression pattern is different between wildtype Wangshuibai and a Wangshuibai mutant without the QTL FHB1. *TaPDR1* might be closely correlated with FHB resistance in wheat (Shang et al., 2009). In addition, another putative ABC transporter, *Lr34*, contributes to the resistance of multiple fungal pathogens such as *Puccinia triticina*, *P. striiformis* and *Blumeria graminis* (Krattinger et al., 2009).

Using microarray data analysis, an ABC transporter was found to be induced by *F. graminearum* and by DON (Balcerzak et al., 2016). BLAST results showed that this wheat ABC transporter

shared 97 to 99% identity with homoeologous wheat gene models Traes_2DL_E79ECCBC9, Traes_2BL_6FFD5A5E0 and Traes_2AL_2F9F6DB1A (IWGSC, 2014), as well as with Expressed Sequence Tags including CJ859948, CJ668043 and CJ856118 (available at NCBI). This group of ESTs represents partial sequences for the protein *TaABCC6* (Bhati et al., 2015). From now on, *TaABCC6* will be used to refer to the DON-induced ABC transporter discussed in this work.

Preliminary RT-qPCR results with primers targeting all the three copies of *TaABCC6* on genomes A, B and D confirmed that this gene had a higher expression level in susceptible wheat cultivars than in resistant cultivars after inoculation with *F. graminearum* (Figure 4). By comparing the number of infected spikelets and the proportion of bleached heads, reduced expression of this gene by transient virus-induced gene silencing (VIGS) produced a delay in the appearance of FHB disease symptoms in FHB-susceptible Roblin, suggesting that this gene is associated with FHB susceptibility (Figure 5B; Balcerzak et al., 2016).

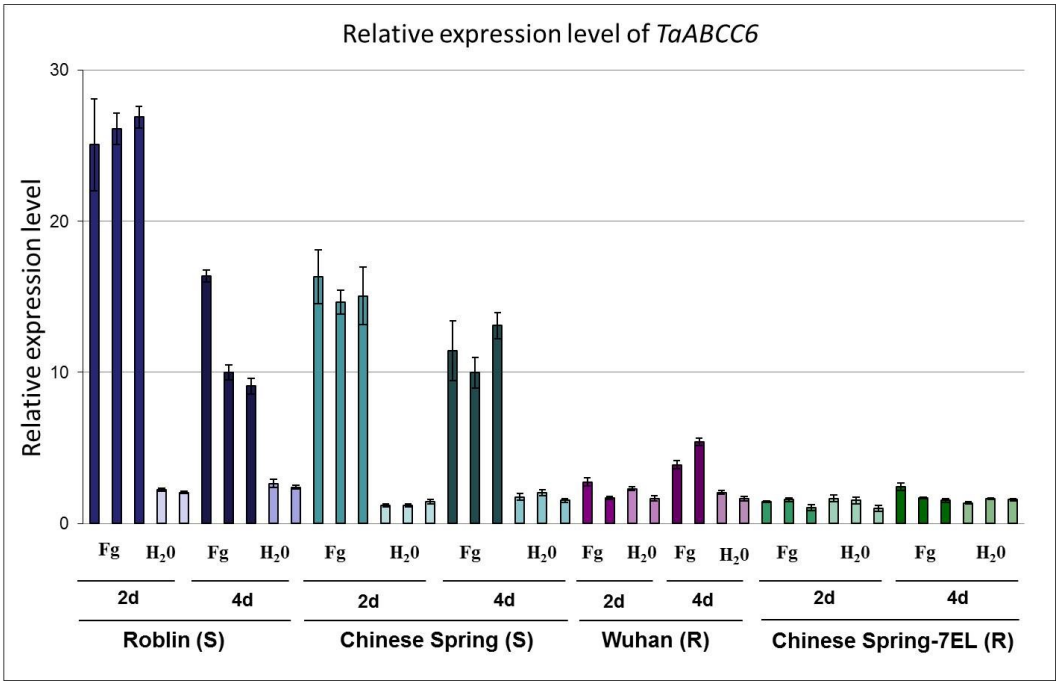


Figure 4. The relative expression level for *TaABCC6* in two FHB-susceptible (S) and two FHB-resistant (R) wheat cultivars. Individual plants from four wheat cultivars were treated either with *F. graminearum* or water and sampled after two or four days of treatment. For each treatment, columns for the same treatment stand for two or three biological replicates; error bars represent standard error between two technical replicates. Modified from Balcerzak et al. (2016).

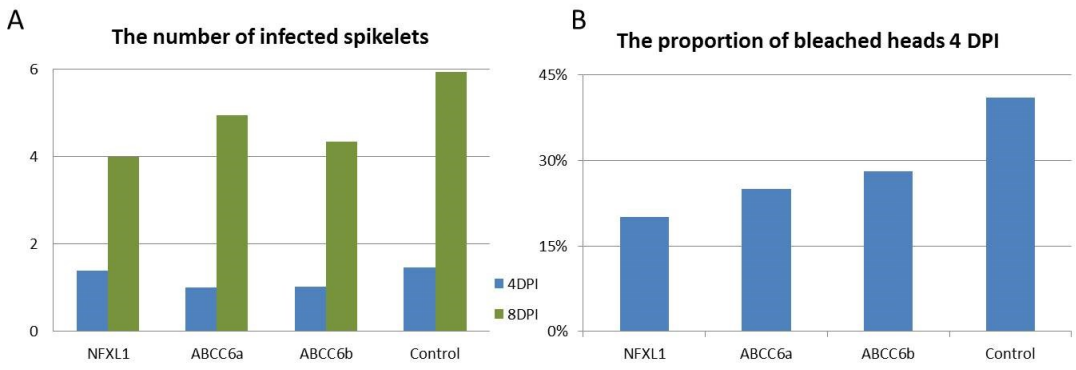


Figure 5. FHB symptoms after virus-induced gene silencing of the genes *TaNFXL1* or *TaABCC6*. (A) The number of infected spikelets at 4 and 8 days post inoculation (DPI) with *F. graminearum*, and (B) the proportion of bleached heads at 4 DPI. About 7 days before inoculation with *F. graminearum*, the plants were treated with barley stripe mosaic virus strains containing either no insertion (control) or a fragment of *TaNFXL1* (NFXL1) or one of two fragments of *TaABCC6* (ABCC6a and ABCC6b). Modified from Balcerzak et al. (2009).

1.4.2 Nuclear Transcription Factor, X box-binding, like 1 (NFXL1)

Transcription factors are proteins that usually bind to a specific DNA sequence, thereby controlling the transcription from DNA to mRNA. Based on their functions, these factors can be divided into two classes: those who can activate the transcription and those who can inhibit this process (Karin, 1990; Latchman, 1997). NFXL1 proteins are transcription repressors.

NFXL1 proteins have been identified in a wide range of species, from yeast to plants to mammals based on their homology to the human *NFX1* gene. Genes in this family are often characterized by *NFX1* type zinc finger motifs. In human, this gene is a repressor of class II MHC genes that can bind to the conserved X-box motif, and then inhibit transcription (Song et al., 1994). *FAP1* is thought to be the yeast homolog of the human *NFX1* gene; the immunosuppressive drug rapamycin can bind to FKBP12, inhibiting cell growth. Researchers have found that *FAP1* can compete with the protein FKBP12 for the binding of rapamycin, thus reducing rapamycin toxicity (Kunz et al., 2000).

In plants, NFXL1 is newly identified; however, its biological role is still largely unknown. In *A. thaliana*, there are two *NFXL* genes, *AtNFXL1* and *AtNFXL2*. They play antagonistic roles in stress responses, including responses to salt and drought (Lisso et al., 2006). *AtNFXL2* can suppress the accumulation of abscisic acid (ABA) while enhancing stomatal aperture (Lisso et al., 2011). In loss-of-function *nfxl2-1 Arabidopsis* mutants, a loss of stomatal density and chlorophyll concentration, as well as higher ABA and hydrogen peroxide accumulation were observed, indicating that this gene might play a role in epidermis development and abiotic stress resistance (Lisso et al., 2012). *AtNFXL1* was firstly found to regulate plant growth positively under salt and osmotic stress (Lisso et al., 2006). More studies on this gene further revealed that

it was up-regulated in *A. thaliana* when the plant was treated with the T-2 toxin, a trichothecene mycotoxin related to DON which is produced by *Fusarium sporotrichioides* (Masuda et al., 2007). Transgenic plants *atnfxl-1*, containing a silenced version of *AtNFXL1* using T-DNA insertion, were hypersensitive to the T-2 toxin. Of great interest is the observation that, when these mutant plants were treated with trichothecenes, defense-related genes were found to be upregulated, suggesting that *AtNFXL1* might be a negative regulator of defense response to trichothecenes (Asano et al., 2008).

In wheat, the information on this gene is limited. Like the *TaABCC6* mentioned above, the expression of this gene correlates with FHB susceptibility according to preliminary RT-qPCR results (Figure 6). Treatment of Roblin plants with DON rapidly induced the expression of this gene. Reduced gene expression of *TaNFXL1* by VIGS in Roblin also led to a delay in apparition of FHB disease symptoms (Figure 5A), suggesting that this gene contributes to FHB susceptibility. Preliminary results have indicated that a reduction in *TaNFXL1* expression level is associated with a reduction in expression of a group of other genes, including a WRKY2-like, an AAA-type ATPase and a Zinc finger protein gene which are also induced by DON treatment (Balcerzak et al., 2016; and T. Ouellet, personal communication). In transgenic *Arabidopsis* transformed with a wheat WRKY2 (*TaWRKY2*) gene, an enhanced tolerance to both drought and high salinity stress was observed (Niu et al., 2012). Zinc finger proteins are often associated with regulation of gene expression. Sakamoto found that in *Arabidopsis*, such proteins could repress the expression of other transcription factors. Further studies using RNA blot also showed that two of these proteins, AZF2 and STZ were highly induced under abiotic stress such as drought and high salinity (Sakamoto et al., 2004). AAA-type ATPases are associated with energy-

dependent remodeling of chromatin and translocation of macromolecules. Taken together, these suggest that *TaNFXL1* could be part of a network modulating plant defense and stress response.

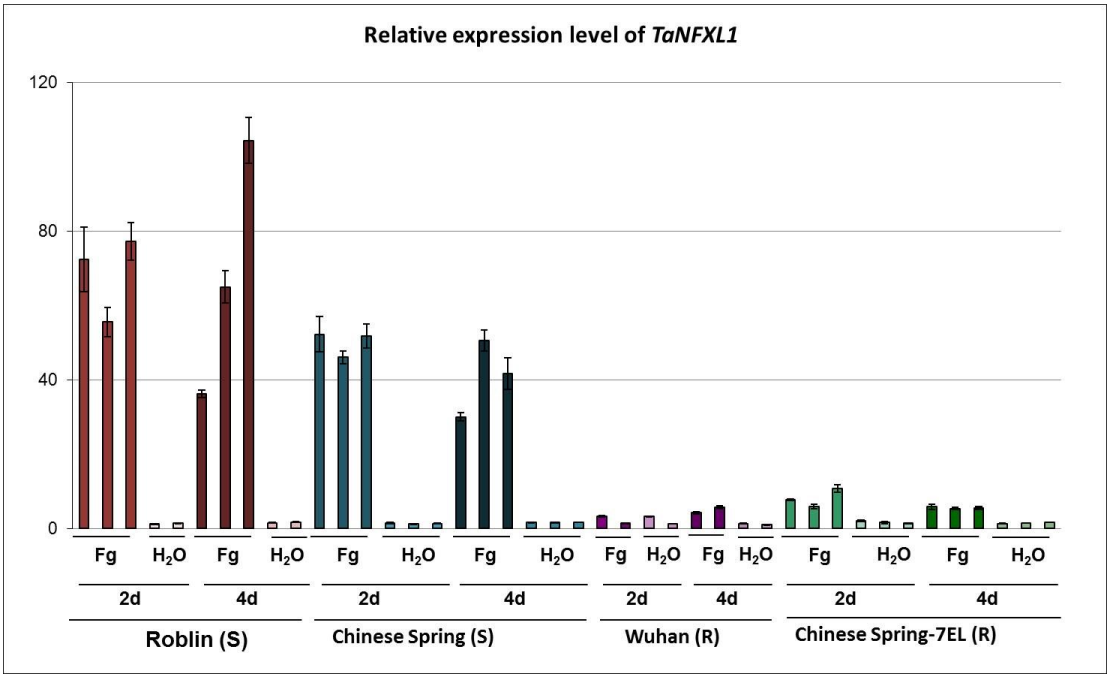


Figure 6. The relative expression level for *TaNFXL1* in two FHB-susceptible (S) and two FHB-resistant (R) wheat cultivars. Individual plants from four wheat cultivars were treated either with *F. graminearum* or water, and sampled after two or four days of treatment. For each treatment, columns for the same treatment stand for two or three biological replicates; error bars represent standard error between two technical replicates. Modified from Balcerzak et al. (2016).

1.4.3 Non-specific Lipid Transfer Protein (nsLTP)

Lipid transfer proteins (LTPs) help transfer lipids between different membranes. They exist in a vast diversity of organisms, from bacteria to human, and from plants to animals (Rueckert & Schmidt, 1990). Some LTPs are substrate-specific while some are not; non-specific LTPs may help transfer different lipids between membranes (Liu et al., 2015). Non-specific lipid transfer proteins constitute a small multi-gene family in plants (Wang et al., 2012). There are two types of nsLTP, *nsLTP-1* and *nsLTP-2*, based on their molecular masses (Kader, 1996).

In plants, previous studies showed that some nsLTPs could response to both abiotic and biotic stresses (Sun et al., 2008); however, the biological role of nsLTPs in plants is still largely unknown. Most researches have shown that these proteins exhibit some intercellular characteristics therefore making intracellular activities very unlikely. In *A. thaliana*, an *nsLTP* was located to the cell wall by immunoelectron microscopy, suggesting that this protein plays a role in intercellular lipid transfer (Thoma et al., 1993). Trichothecin (Tcin) is a type B trichothecene in the same class as DON. The overexpression of an *Arabidopsis* nsLTP gene, *AtLTP4.4*, showed an enhanced resistance to Tcin and a reduction in oxidative stress (Mclaughlin et al., 2015). In transgenic *Populus tomentosa Carr.* (Chinese white poplar), overexpression of a nsLTP-like gene (*LJAMP2*) from *Leonurus japonicas* (motherwort) could enhance resistance to pathogens *Alternaria alternata* and *Colletotrichum gloeosporioides* (Jia et al., 2010). These two pathogens cause poplar leaf blight in northeast China, which lead to environmental and economic losses.

A recent research showed that some nsLTPs exhibit antifungal effects in wheat, possibly by inducing fungal membrane permeabilization, inhibiting fungal growth or spore germination (Sun et al., 2008). By overexpressing a wheat LTP gene, *TaLTP5*, in a susceptible wheat cultivar, Yangmai 158, an enhanced resistance to *Cochliobolus sativus* and *F. graminearum* was observed in transgenic wheat plants (Zhu et al., 2012). Gene silencing by VIGS of a type 1 nsLTP named *TansLTP9.4* (corresponding to the gene model Traes_5BL_CE73B6BFF.1) in the FHB-resistant cultivar Wuhan1 showed that the FHB disease symptoms were increased by 20~30%, when the gene expression level was significantly reduced, suggesting that this gene contributed to FHB resistance in this cultivar (Ouellet et al., 2013). Microarray and RT-qPCR results for *TansLTP9.4* have shown that its expression profile does not considerably change after inoculation with *F.*

graminearum, in both susceptible and resistant cultivars when compared with water treatment; however, there is a large difference in expression level between susceptible and resistant cultivars that carry the FHB-resistance QTL 5A, with the 5A QTL carrying lines expressing *TansLTP9.4* at a higher level (Foroud et al., 2012; T. Ouellet, personal communication). This suggests that *TansLTP9.4* may contribute to resistance to FHB provided through the action of QTL 5A.

1.5 Objectives

My project was part of a larger research program with the long-term goal of understanding molecular mechanisms contributing to resistance and susceptibility to FHB in wheat. The goal of this project was to inactivate three wheat target genes by gene editing using the CRISPR/Cas9 system, as a step towards a better understanding of their respective contribution to the wheat response to *F. graminearum* attack. The three wheat genes to be edited included *TaABCC6*, *TaNFXL1* and *TansLTP9.4*. The major difficulty to achieve this goal was to design efficient sgRNAs targeting all the three genomes, and to evaluate potential off-target effect. The main objectives for this project were as follows.

1) Testing of sgRNA *in vitro*
    - Design sgRNAs with predicted high efficiency for editing each gene
    - Test sgRNAs in an *in vitro* test
2) Testing of sgRNA in a transient wheat protoplast system
    - Build expression vectors for protoplast transformation
    - Isolate and transform wheat protoplasts
    - Detect and quantify modifications at target regions
3) Characterization of nine DON induced genes in a transient wheat protoplast system
    - Analyze gene expression level in DON-treated protoplasts
    - Annotate the functions of these genes by comparison to *Arabidopsis*

2. Materials and Methodology

A brief introduction of this project has been provided in Chapter 1. In this chapter, detailed materials and methodology is presented.

2.1. Plant materials

Seeds from wheat (*Triticum aestivum* cv Fielder, kindly provided by Dr. Jas Singh, ORDC, Agriculture and Agri-Food Canada (AAFC)) were kept in a -20 ℃ freezer under dry conditions in our lab. They were firstly sterilized with 75% ethanol (v/v) for 1 min and then by 50% bleach (v/v) for 10 min and washed with sterile water for 5 times. The plants were grown in magenta boxes with MS media (4.2 g/L Murashige and Skoog Salts, 10 g/L sucrose, 3 g/L phytagel, pH 5.8), in a controlled environment at 21 ℃ under 16 hr-light/ 8 hr-dark light cycle for 12 days. Twelve-day old leaves were used for protoplasts isolation and transformation.

2.2. Design and test of sgRNAs *in vitro*

A schematic representation of the sgRNA *in vitro* test is shown in Figure 7. The key steps include sgRNA *in vitro* transcription, PCR amplification of targeted genomic regions and digestion with a Cas9 nuclease. Detail information for each step is presented below. Briefly, the sgRNA-encoding templates were firstly designed in house then synthesized by Sigma (Canada). Those templates, each containing a sgRNA, a sgRNA scaffold and a T7 promoter, were transcribed *in vitro*. Meanwhile, templates containing the sgRNA target sites were amplified from genomic DNA with gene specific primers. Guided by sgRNAs, the Cas9 nuclease cleaved the DNA templates into two smaller fragments. Finally, the sgRNA efficiency was evaluated by separating cleavage products on an agarose gel.

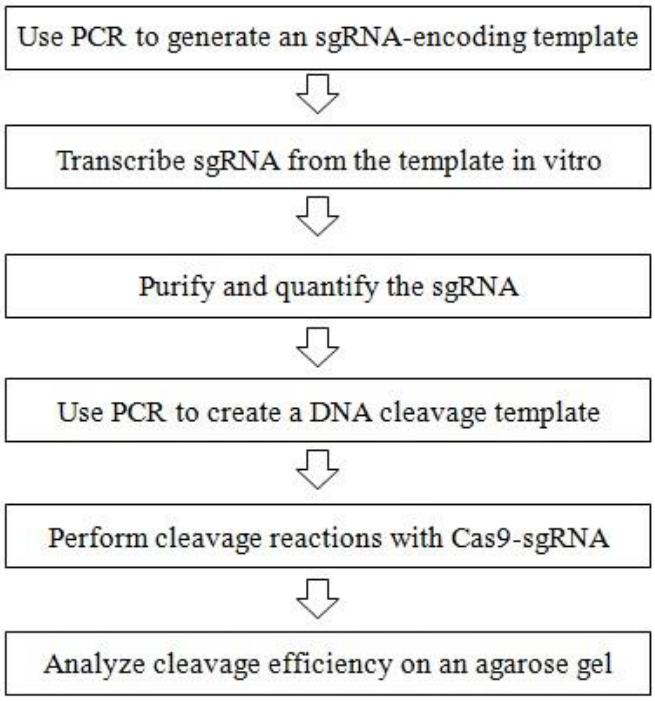


Figure 7. Schematic representation of the sgRNA *in vitro* test. The key steps include *in vitro* transcription of sgRNA and sgRNA scaffold, amplification of targeted genomic regions and *in vitro* cleavage reactions with a Cas9 nuclease.

2.2.1. Generation of sgRNA-encoding templates

2.2.1.1. Design of two sgRNAs for each gene

i. The EST sequences used to design Affymetrix probeset sequences of each gene were firstly downloaded from GrainGenes (http://wheat.pw.usda.gov/cgi-bin/GG3/browse.cgi; sequences were retrieved in January, 2015). To find the longest expressed contig sequence (assembly of overlapping ESTs) for each gene, these EST sequences were blasted against an ORDC local database of assembled public wheat ESTs (Hattori et al., 2005) using the software Bigblast v2.93_2008_0929 (Tinker et al., ECORC, AAFC) built on BLASTN 2.2.17 (Altschul et al., 1997).

ii. To obtain genomic sequences for the three genes, the EST contig sequences were blasted against the public Wheat_Survey_v2_(IWGSC) sequence data (International Wheat

Genome Sequencing Consortium, 2014 https://wheat-urgi.versailles.inra.fr/Seq-Repository/BLAST) using the Wheat URGI BLAST tool (Altschul et al., 1997). Genomic contigs from chromosome survey sequences with more than 95% homology were downloaded for further analysis. The contigs for *TansLTP9.4* and *TaNFXL1* were retrieved in February 2015; while *TaABCC6* contigs were obtained in July 2016.

iii. ESTs and genomic contigs retrieved from steps i and ii were aligned with MEGA6 software to determine the introns and exons for each gene.

iv. The design of sgRNAs was based on the ESTs from step i. An online program, sgRNA Designer (http://www.broadinstitute.org/rnai/public/analysis-tools/sgrna-design), was used to design sgRNAs for each gene and each sgRNA received a score from 0 to 1. sgRNAs with higher scores were predicted to have higher editing efficiency and were highlighted for further analysis.

v. To have a better understanding of the structure of three genes, including the positions of introns and exons, another BLAST was also performed between the contigs obtained in step i and the *Triticum aestivum* TGACv1 database on EnsemblPlants website (http://plants.ensembl.org/Multi/Tools/Blast?db=core, accessed in July, 2017). On each chromosome, the match with the highest identity was downloaded for further study.

Wheat is a hexaploid plant and most genes have a copy on each of genomes A, B and D. To target the homoeologous genes on the three genomes, we excluded sgRNAs which were not in the conserved regions (Figure 8). According to alignment results in step iii, sgRNAs overlapping with two exons were also ruled out. In addition, the candidate sgRNAs were blasted against the survey sequence on URGI to ensure that they were not homologous to other known sequences, to minimize the risk of off-target editing. For this, the "Wheat_Survey_v2_sequence data" database

was selected for BLAST and the expected threshold was adjusted to 10 under the Advanced Search menu for short input sequences while all other settings were as default. The sgRNAs with matches outside the target regions were eliminated. Since all of the sequence information used to design the sgRNAs was from the cultivar Chinese Spring, while the experiments were carried out in Fielder, the targeted genomic regions from Fielder were amplified and sequenced using primers described in Subsection 2.2.2; sgRNAs that overlapped with sequence polymorphisms between these two cultivars were also excluded. For each gene, the two sgRNAs fitting better the criteria above were selected and used for further testing.

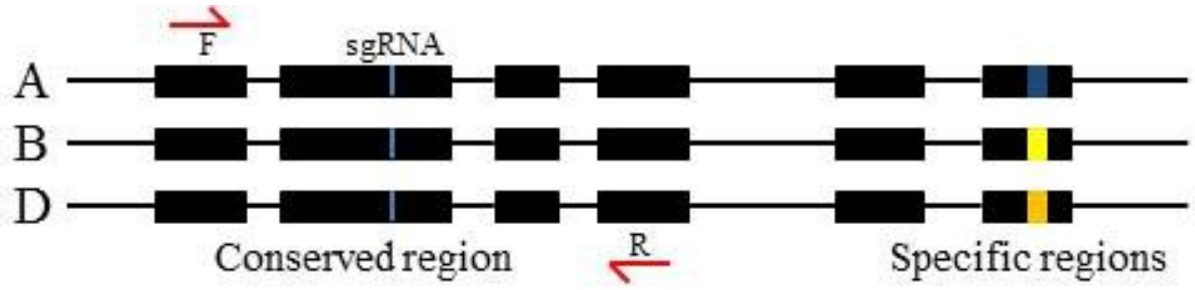


Figure 8. Schematic illustration of a selected sgRNA position on three wheat genomes A, B and D. Introns are indicated by horizontal lines; rectangular boxes represent exons. Red arrows are the forward and reverse primers used to amplify a conserved region including the targeted editing position. The position of the sgRNA is marked with a vertical blue line in each genome. Specific regions on different genomes are marked with boxes in different colors (Modified from Shan et al. 2014).

2.2.1.2. The amplification of sgRNA-encoding templates and *in vitro* transcription

The 'Guide-it Complete sgRNA Screening System' kit (Clontech, USA) was used for sgRNA *in vitro* transcription and screening, following the manufacturer's instructions. First, a 69 nt sgRNA-encoding primer was designed, including a T7 promoter, one of the 20 nt sgRNA targeting sequence selected in Subsection 2.2.1.1 and a sgRNA scaffold sequence (Figure 9). This 69 nt DNA fragment was synthesized by Sigma. A working dilution at a final concentration

of 10 μM was prepared. The fragment was used as a forward primer to amplify a 140bp sgRNA-encoding template by PCR; a reverse primer and a DNA template (20 ng/μl) were already premixed as the Guide-it Scaffold Template (provided with the kit; the ingredients are proprietary). By mixing 5 μl of Guide-it Scaffold Template, 1 μl of forward primer (10 μM) and 19 μl of H_2O with a tube of High Yield PCR EcoDry Premix (also provided with the kit), the PCR reaction was performed as follows: 95 ℃ for 1 min, followed by 33 cycles of 95 ℃ for 30s, 68 ℃ for 1 min; and final elongation is 68 ℃ for 1 min. Proper PCR amplification was confirmed by separation on a 1.5% agarose gel, expecting a PCR product with a size of 140 bp. Amplified products were then purified with NucleoSpin Gel and PCR Clean-Up Kit (provided with the kit), following manufacturer's instructions. The concentration of purified PCR product was measured with a Nanodrop ND-1000 spectrophotometer.

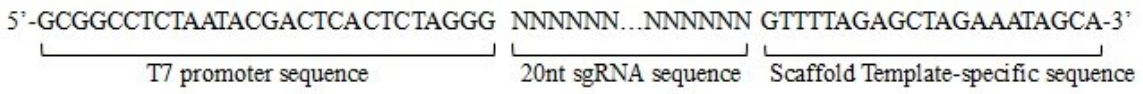


Figure 9. A 69 nt sequence used as a forward primer. The sequence includes a T7 promoter, a 20 nt sgRNA and a sgRNA scaffold sequence.

For *in vitro* transcription of sgRNAs, the reactions were set up as follows: 100 ng of purified PCR product (sgRNA-encoding template), 7 μl of *in vitro* transcription buffer, 3 μl of T7 polymerase mix (33 U/μl) and RNase free water to bring the reaction volume to 20 μl. The reaction was incubated at 42 ℃ for 1 hr. The purification of transcribed sgRNAs was according to the following protocol and all the centrifugations were done at 12000 rpm.

1) Add 2 μl of RNase free DNase I (Thermo Fisher Scientific, Canada) to the entire 20 μl of the *in vitro* transcription mixture and incubate at 37 ℃ for 30 min.
2) Add RNase free water to the reaction mixture to a volume of 100 μl.

3) Add 100 µl of phenol: chloroform: isoamyl alcohol (25:24:1) saturated with 10 mM Tris, pH 8.0, 1 mM EDTA (Sigma, Canada) to the mixture from step 2 and mix well by vortex. Centrifuge for 2 min at room temperature.
4) Transfer the supernatant to a new 1.5 ml tube and add an equal volume of chloroform and vortex to mix. Centrifuge for 2 min at room temperature.
5) Transfer the supernatant to a new 1.5 ml tube and add 1/10 volume of 3 M NaAC and an equal volume of isopropanol. Vortex to mix and incubate at -20 ℃ overnight.
6) Centrifuge the mixture for 30 min at 4 ℃. Discard the supernatant and rinse the pellet with 70% ethanol and centrifuge for 5 min at room temperature.
7) Air dry the pellet for 2 min and resuspend in 20 µl of RNase-free water.
8) Measure the concentration with a Nanodrop ND-1000 spectrophotometer.
9) Store purified sgRNAs at -80 ℃ until future use.

2.2.2. Design primers to amplify and sequence DNA cleavage template

For each gene, primers for amplification of a genomic DNA fragment including the two selected sgRNA sequences were designed with an online program, PrimerQuest Tool (https://www.idtdna.com/Primerquest/Home/Index, Integrated DNA Technologies (IDT), USA). To eventually distinguish on an agarose gel the two smaller fragments generated after cleavage in the *in vitro* test, the primers used for amplification of the target templates were designed to produce asymmetry after the cleavage reaction (Figure 8). For each gene, 3 primer pairs were designed and ordered from Sigma (http://www.sigmaaldrich.com/canada-english.html, Canada), to choose the best amplified fragments for screening sgRNA in the *in vitro* test.

Genomic DNA of wheat cultivar Fielder was isolated with 'Illustra Nucleon Phytopure Genomic DNA Extraction Kit' (GE Life Sciences, Canada) from 12-day-old wheat leaves, following the

manufacturer's instructions. To produce the target DNA templates for sgRNA screening in the *in vitro* test, DNA fragments for each gene, including sgRNA targeting regions, were amplified by PCR using high-fidelity PfuTurbo Cx Hotstart DNA Polymerase (Agilent Technologies, USA) and with the primer pairs selected above. The reactions were set up as described in Appendix 1A. The amplification protocol was as follows: 94 ℃ for 3 min, followed by 35 cycles of 94 ℃ for 30s, 60 ℃ for 30s, 72 ℃ for 1 min and final elongation was 72 ℃ for 10 min. PCR products were firstly identified by separation on a 1% agarose gel followed by purification with PureLink® Quick PCR Purification Kit (Invitrogen, Canada).

To confirm their identity, the PCR products containing targeting regions were sequenced directly with the BigDye® Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific, Canada) by mixing 2 µl of 5x sequencing buffer, 0.5 µl of BigDye® Terminator v3.1 Ready Reaction Mix, 0.5 µl each of forward and reverse primers (10 µM), 1 µl of purified PCR product (~100 ng) and 6 µl of H_2O. The thermal cycler was set up as follows: 95 ℃ for 5 min, followed by 30 cycles of 96 ℃ for 15s, 57 ℃ for 15s and 60 ℃ for 4 min. The sequencing was done by Molecular Microbiology & Next Generation Sequencing Service at ORDC.

2.2.3. Testing of sgRNA efficiency *in vitro*

Cleavage reactions were set up for each sgRNA as described in Appendix 2 using purified sgRNAs from Subsection 2.2.1.2 and templates from Subsection 2.2.2. The reactions were incubated at 37 ℃ for 1 hr followed by 70 ℃ for 10 min to deactivate enzymes. Finally, the products were separated on a 1% agarose gel to evaluate cutting efficiency.

2.3. Testing of sgRNAs in a wheat protoplast system

Figure 10 summarizes the key steps in testing sgRNAs in wheat protoplasts, including gBlocks design and assembly, cloning into expression vectors, protoplasts isolation and transformation. The detailed methodologies are presented in the following subsections.

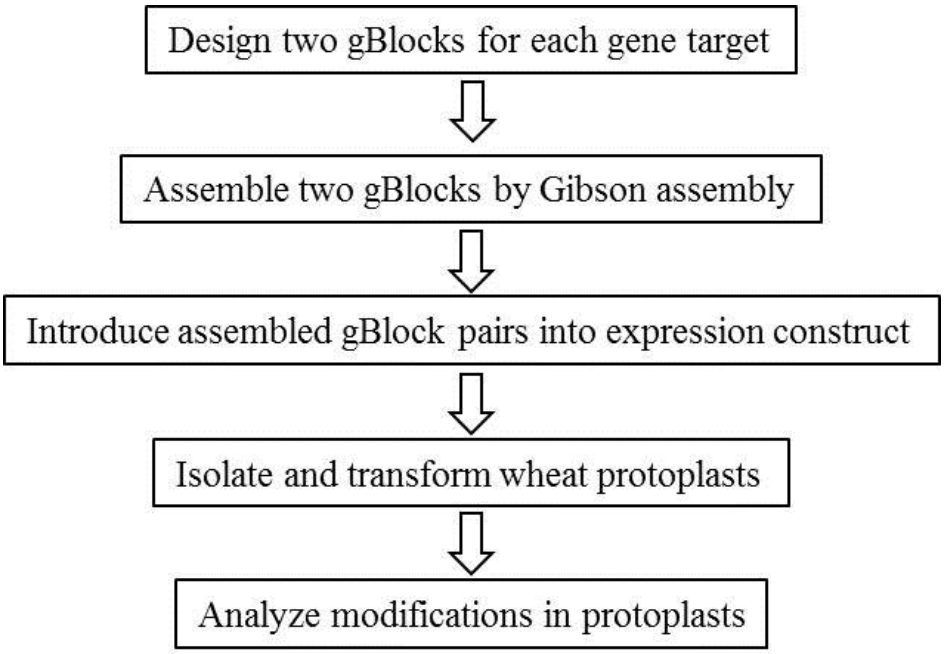


Figure 10. Key steps to test sgRNAs in a wheat protoplast system.

2.3.1. Plasmid construction

2.3.1.1. Design and assembly of gBlocks

For the CRISPR/Cas9 system used in this project, the functional gBlock (sgRNA block) included a wheat U6 promoter, a 20 nt target-specific sgRNA sequence and a sgRNA scaffold (Figure 11). Two 40 bp overlaps at both 5' and 3' were used to assemble two different gBlocks targeting the same gene fragment by Gibson Assembly (Gibson et al., 2009). All gBlocks were designed with DNASTAR Lasergene 10 and ordered directly from IDT (https://www.idtdna.com/pages/products/genes/gblocks-gene-fragments, USA).

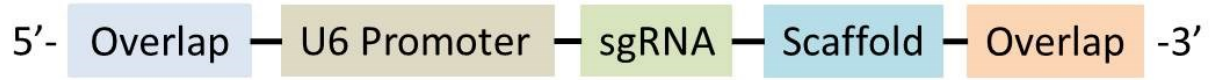


Figure 11. Illustration of the gBlock components. Two overlaps at both 5' and 3' were used to assemble different gBlocks; the U6 promoter (Shan et al., 2013) was used to initiate sgRNA and sgRNA scaffold transcription.

Synthesized gBlocks were firstly dissolved in RNase free water at the concentration of 10 ng/µl. To assemble two gBlocks targeting the same gene fragment, we mixed 5 µl of gBlocks (2.5 µl for each), 10 µl of Gibson Master Mix (NEB, Canada; ingredients are proprietary) and 5 µl of H_2O. The mixture was incubated in a thermocycler at 50 ℃ for 1 hr and stored at -20 ℃ for future use. Primers with EcoRI and KpnI restriction enzyme sites were used to amplify assembled gBlock pairs by PCR (Gib_assem_EcoRI-1F: 5'-CGGAATTCCGCGGTGTCAT-CTATGTTAC-3'; Gib_assem_KpnI_1052R: 5'-GGGGTACCACCCGCCAATATATCCTGTC-3'; restriction enzyme sites are underlined) with PfuTurbo Cx Hotstart DNA Polymerase (Appendix 1A; 94 ℃ for 3 min, followed by 35 cycles of 94 ℃ for 30s, 60 ℃ for 30s, 72 ℃ for 1 min and 72 ℃ for 10 min for final elongation), followed by size verification on an agarose gel and gel purification of the PCR products with the QIAquick Gel Extraction Kit (Qiagen, Canada).

2.3.1.2. Cloning of assembled gBlock pairs into pJet1.2/blunt vector

By mixing 5 µl of 2x reaction buffer, 2 µl of a purified assembled gBlock pair (~50 ng), 0.5 µl of T4 DNA Ligase (5 U/µl), 0.5 µl of linearized pJet 1.2/blunt cloning vector (50 ng/µl, Thermo Scientific, Canada) and 2 µl of H_2O, the EcoRI-KpnI assembled gBlock pairs were ligated into the pJet 1.2 cloning vector by incubating at 4 ℃ overnight. Following ligation, the recombinant vectors and TOP10 competent *E, coli* cells (Thermo Fisher Scientific, Canada) were gently mixed by tapping the tube. The mixture was kept on ice for 30 min followed by heat shock at 42 ℃

for 30s and back on ice for 2 min. Then the cells were transferred to 500 µl of Super Optimal broth with Catabolite repression (S.O.C) medium (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose; Invitrogen, Canada) and incubated at 37 ℃ for 1 hr on a shaker, at 200 rpm. Fifty µl of *E. coli*-containing S.O.C. medium were then spread on a Luria-Bertani (LB) agar (15 g/L agar, 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl; Fisher Scientific, Canada) plate supplemented with 100 µg/ml ampicillin (Sigma, Canada) and incubated at 37 ℃ overnight.

Colony aliquots were picked up from LB agar plates using sterile toothpicks and resuspended in 12 µl of H_2O. Colony PCR reactions were set up using Taq DNA polymerase as described in Appendix 1B (Fisher Scientific, Canada) with forward (Gib_assem_1F: 5'-CGCGGTGTCATCTATGTTAC-3') and reverse (Gib_assem_1052R: 5'-ACCCGCCAATATATCCTGTC-3') primers. The amplification protocol was as follows: 98 ℃ for 2 min, followed by 35 cycles of 98 ℃ for 30s, 60 ℃ for 30s and 72 ℃ for 1 min. The final elongation was 72 ℃ for 5 min. Finally, the PCR products were separated on a 1% agarose gel to identify recombinant clones.

To further confirm if the two restriction enzyme sites had been successfully added to recombinant clones, the remaining 7 µl of cell suspension from last step were firstly mixed with 4 ml of liquid LB medium (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl; BD Company, Canada) and incubated at 37 ℃ overnight on a shaker at 250 rpm. The plasmids were extracted using QIAprep Spin Miniprep Kit (Qiagen, Canada), following manufacturer's instructions. Since the buffers for KpnI and EcoRI are not compatible with each other, we performed two single digestions with KpnI or EcoRI instead of a double enzyme digestion. For each sample, 5 µl of plasmids (~1.5 µg) was mixed with 17 µl of H_2O, 0.5 µl of KpnI or EcoRI (10 U/µl,

Thermo Fisher Scientific, Canada) and 2.5 µl of its unique buffer and incubated at 37 ℃ for 4 hours. The digested and undigested (as control) plasmids were separated on a 1% agarose gel. Finally, a primer pair (pJet_Seq_F: 5'-CGACTCACTATAGGGAGAGCGGC-3'; pJet_Seq_R: 5'-AAGAACATCGATTTTCCATGGCAG-3') was used for sequencing assembled gBlock pairs in recombinant vectors.

2.3.1.3. Transfer of assembled gBlock pairs into an expression vector

A modified pCambia 1302 vector (Appendix 4 and Jiang et al., 2013b) containing a Cas9 nuclease which originated from *Streptococcus pyogenes* (spCas9) and codon optimized for expression in *Chlamydomonas reinhardtii* (crCas9), was kindly provided by Dr. Rajagopal Subramaniam to build expression constructs for delivering assembled gBlock pairs into wheat protoplasts. The pCambia and pJet 1.2-sgRNA recombinant vectors were each digested with EcoRI and KpnI, according to manufacturer's instructions. The digested vectors were firstly separated on a 1% agarose gel and the bands of interest purified using QIAquick Gel Extraction Kit, according to manufacturer's instructions. The assembled gBlock pairs were ligated into the digested pCambia vector by mixing 100 ng of vector DNA, 21 ng of the insert (from recombinant pJet), 1 µl of T4 DNA Ligase (3 U/µl; Promega, USA) and water to a final volume of 10 µl, followed by incubation at 4 ℃ overnight. The ligation products were transformed into TOP 10 competent cells as described in Subsection 2.3.1.2. Transformed cells were then spread on LB agar plates containing 50 µg/ml kanamycin ("LB+Kan", Sigma, Canada) and incubated at 37 ℃ overnight. Colony PCR and sequencing of recombinant plasmids were performed using two sgRNA-specific sequences (Appendix 3A) as primer pairs, to identify correct recombinants plasmids, as described in Subsection 2.3.1.2.

Selected colonies with correct constructs were firstly inoculated in 1 mL of liquid LB+Kan and incubated at 37 ℃ for 6 hours on a shaker at 250 rpm. Then 100 µl of *E. coli* culture was transferred into 100 mL of liquid LB+Kan and incubated in the same conditions overnight. The plasmids were extracted using NucleaBond Xtra Midi kit (Clontech, USA) according to manufacturer's instructions. The DNA concentration was measured by Nanodrop ND-1000 spectrophotometer and was adjusted to about 1000 ng/µl for protoplasts transformation.

Another expression vector for expressing the plant codon optimized spCas9, pFGC-pcoCas9, was a gift from Jen Sheen (Appendix 5, Addgene plasmid # 52256). A different primer pair with EcoRI and Xmal restriction enzyme sites was used to amplify the assembled *TaNFXL1* gBlock pair in order to insert it into the pcoCas9-sgRNA vector (Gib_EcoRI_F: 5'-CGGAATTCCGCGGTGTCATCTATGTTAC-3'; R: 5'-Gib_XmaI_R: TCCCCCCGGGACCC-GCCAATATATCCTGTC-3'; restriction enzyme sites are underlined) with PfuTurbo Cx Hotstart DNA Polymerase, following the same amplification protocol as in subsection 2.2.2. The cloning procedure was as described in subsection 2.3.1.2, including pJet 1.2 vector transformation, restriction enzyme digestion, cloning into expression vector pFGC-pcoCas9 and verification by sequencing.

2.3.2. Protoplast isolation and transformation via polyethylene glycol (PEG) mediated method

A schematic illustration of our optimized protoplast extraction and transformation protocol is shown in Figure 12. This isolation and transformation protocol was inspired from protocols published by Yoo et al. (2007) and Shan et al. (2014). The detailed optimized methods are summarized below. All required solutions are described in Appendix 6. Using our optimized

protocol, the protoplasts were transformed with either one of four sgRNA editing constructs or a pMDC32 vector expressing a green fluorescent protein (GFP) (pMDC32-ZsGreen, Appendix 7, kindly produced and provided by Natalie Labbé and Johann Schernthaner). The fluorescence was observed under a fluorescence microscope.

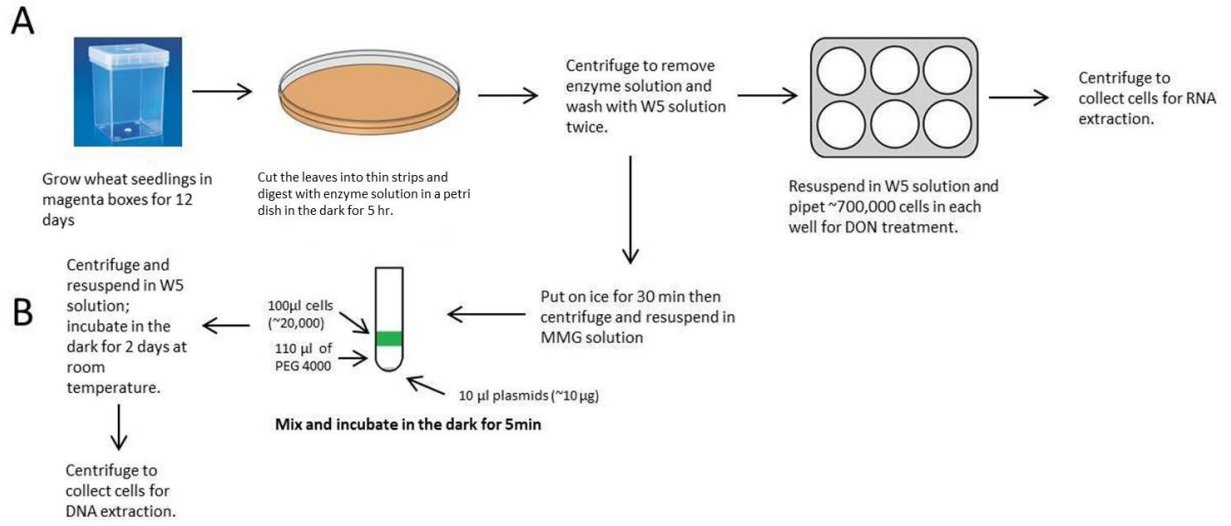


Figure 12. Visual representation of wheat protoplasts isolation (A) and transformation (B).

2.3.2.1. Wheat protoplast isolation

i. Prepare 10 mM MES digestion media containing 0.6 M mannitol, 10 mM KCl, 1.5% cellulose R-10 (w/v, Yakult, Japan) and 0.75% macerozyme R-10 (w/v, Yakult, Japan). Heat the solution at 55 ℃ for 10 min to inactivate DNase and increase enzyme solubility.

ii. Let the solution cool down to room temperature, and add $CaCl_2$ to 10 mM and BSA to 0.1% (w/v).

iii. Cut thin strips (~1 mm) from 12-day-old wheat leaves (20-25 pieces) and put them into a petri dish with 12.5 ml enzyme solution.

iv. Vacuum infiltrate for 40 min while in the dark (covered with aluminum foil).

v. Incubate strips in the dark for 5 hours with gentle shaking (50 rpm) at room temperature.

vi. Place a 70 μm nylon mesh (Greiner Bio-One, Germany) over a 50 ml falcon tube and gently pour liquid through the nylon mesh and wash the petri dish with about 40 ml W5 solution twice (~20 ml each time).

vii. Centrifuge at 100 g for 2 min at room temperature in a swinging bucket rotor, remove supernatant carefully with pipette without losing cells.

viii. Wash the petri dish again with W5 solution and repeat step vi-vii.

ix. Gently resuspend cells in W5 solution (~10 ml) by rotating or tapping the tube.

x. Count the protoplasts using a hemocytometer.

xi. Keep on ice for at least 30 min and centrifuge at 100 g for 1 min to remove supernatant.

xii. Resuspend protoplasts in Mannitol $MgCl_2$ (MMG) solution. Dilute the suspension with MMG until a cell density of $2x10^5$ /ml is reached. For RNA assays, dilute the suspension with W5 at the concentration of $2x10^6$ cells per ml.

2.3.2.2. Protoplast transformation via PEG mediated method

i. Mix 10 μg of recombinant expression plasmids (~1 μg/μl) (described in Subsection 2.3.1.3) with 100 μl (~$2x10^4$ cells) of protoplasts from Subsection 2.3.2.1 in a 2 ml Eppendorf round-bottom tube. Immediately add 110 ul PEG solution and mix gently by tapping the tube.

ii. Leave for 5 min at room temperature in the dark without shaking.

iii. Stop the transfection by adding 440 ul of W5 solution and mix well by inverting the tubes.

iv. Centrifuge at 100 *g* for 2 min and remove PEG using a pipette.

v. Resuspend protoplasts in 2 ml of W5 solution and incubate in a 2 ml tube in the dark at room temperature for 48 hours without shaking.

vi. Centrifuge at 100 g for 2 min and discard the supernatant as much as possible to collect cells.

vii. For protoplasts transformed with one of four sgRNA editing constructs, pool two individual transformations together to extract genomic DNA using "Illustra Nucleon Phytopure Genomic DNA Extraction Kit" immediately or store at -80 ℃ until future use*.

viii. For protoplasts transformed with GFP vector, resuspend cell pellet from step vi in 200 ul of W5 solution and use 10 ul of cell suspension to determine transformation efficiency by counting the number of fluorescent cells using an Axio Scope.A1 (Item No. 430035-9100-000; Carl Zeiss, USA) connected to a Colibri.2 light source (Carl Zeiss, USA). A 505 nm wavelength was selected while the filter No.3 was chosen on the microscope. The protoplasts were observed at 200x magnification. The photos were taken using a Canon EOS 60D camera. The exposure time was set to 15s for dark field photos, while 1/15s was for bright field photos. Other settings were as default.

* Transformed protoplasts were only collected when transformation efficiency was above 60%, as monitored with GFP in step viii.

2.4. Quantification of gene editing in modified protoplasts by high throughput sequencing

2.4.1. PCR amplicons preparation and sequencing

Genomic DNA used in this section was from step vii in Subsection 2.3.2.2. For the first round of amplicon sequencing, for each gene, five DNA samples from protoplasts prepared at different times were extracted (Appendix 8A). The targeted regions were amplified using CloneAmp™ HiFi PCR Premix (Clontech, USA) and the reactions were set up as described in Appendix 1C. The amplification protocol was as follows: 95 ℃ for 3 min, followed by 35 cycles of 98 ℃ for 10s, 60℃ for 15s and 72 ℃ for 10s. PCR products were purified with PureLink® Quick PCR Purification Kit. The first ten amplicon samples (5 for *TansLTP9.4* and 5 for *TaNFXL1*) were sent to the Centre for the Analysis of Genome Evolution & Function (University of Toronto,

Canada) for high throughput sequencing using an Illumina MiSeq system. The amplicon samples were named nsLTP-1~5 and NFXL1-1~5, respectively. Primers used to amplify the target regions for that round of sequencing are listed in Appendix 3B.

Seventeen PCR amplicon samples were sequenced in the second round of high throughput sequencing, using the Molecular Microbiology & Next Generation Sequencing Service at ORDC. Five samples for the *TaABCC6* targeted region were named ABC 1~5, respectively; five PCR products for the *TaNFXL1* targeted region which were amplified from genomic DNA isolated from pcoCas9-transformed protoplasts were named pcoNFXL1-1~5, respectively. Information on protoplast and transformation batches used for sample preparation can be seen in Appendix 8B. Since we used a different sequencing service as well as different primer pairs for amplification, we also re-amplified from genomic DNA and re-sequenced the five *TaNFXL1* samples from the first round of sequencing (NFXL1-1~5), to enable a comparison of efficiency between the two types of Cas9 protein used. In addition, two untransformed samples, for *TaABCC6* (ABC-control) and *TaNFXL1* (NFXL1-control), were also sequenced as controls.

Illumina overhang adapter sequences were added to the 5' of gene specific primers to make PCR products compatible with the protocol used by the second sequencing service. In order to increase base diversity, degenerate and staggered fusion primers were created by adding between 0~3 "N" bases in between the adapter and the gene-specific primers. The primers are listed in Appendix 3C. All the primers were firstly diluted to 200 µM stocks in 10 mM Tris-HCl (pH=8.5). By combining 25 µl of four degenerate and staggered forward or reverse primer stocks for each target region, a primer cocktail was made. Primer cocktail stocks were diluted with 10 mM Tris-HCl (pH=8.5) and the working concentration for each primer in the cocktail was 20 µM. All 17 samples were amplified using KOD Hot Start DNA Polymerase (Appendix 1D,

Novagen, Canada) by two rounds of PCR amplifications. Gene-specific primers (10 µM; ABC: F 5'-TGTCTTCAGCGAAGGTAAGC-3', R 5'-GGTGTGTGCGGTCTTTCT-3'; NFXL1: F 5'-CTTTCCACAGGGCTGGTT-3', R 5'-TGGATGCCAGCATCTCTG-3') were used in the first round of amplification. The PCR products were purified using PureLink® Quick PCR Purification Kit and for each sample, 10 ng of purified PCR product was used as the template for the second round of amplification with forward and reverse primer cocktails prepared above. The amplification protocol was as follows: 95 ℃ for 2 min, followed by 35 cycles of 95 ℃ for 20s, 60 ℃ for 10s and 70 ℃ for 15s for the first round of amplification; for the second round of amplification, the annealing temperature was adjusted to 65 ℃ while all other settings remained the same.

Two µl of each PCR product was loaded on a 1% agarose gel for qualitative verification of amplification. Using a Nanodrop ND-1000 spectrophotometer to quantify the concentration, the PCR product was diluted to a final concentration around 50 ng/µl with 10 mM Tris-HCl (pH=8.5). Amplification products (25 µl) were sent to the service for sequencing. The PCR samples were cleaned up using Agencourt AMPure XP (Beckman Coulter, USA) and sequenced using an Illumina Miseq sequencing system (Illumina, USA).

2.4.2. Sequencing data analyses

The sequencing data of both rounds were analyzed using CLC Genomics Workbench (version 10.0.1, Qiagen). A workflow (Appendix 9) was created to analyze all sequencing reads. For each sample, a quality control (QC) report was firstly created to verify sequencing quality. To remove adapter sequences, 20 bases at both 3' and 5' ends of each read were trimmed off. The sequencing results in Subsection 2.2.2 from cultivar Fielder were used as reference sequences to map the reads for the amplicons analyzed; all other settings were set as default.

Trimmed and paired readings were then used to detect and to quantify targeted mutations using the InDels and Structural Variants tools under Resequencing Analysis in the CLC Genomics Toolbox. “Create breakpoints” was selected under output options while all other settings were set as default (P-value threshold=0.0001; maximum number of mismatch=3; minimum quality score and minimum relative consensus sequence coverage were set to 0; “Ignore broken pairs” was selected). For each sample, four individual files were generated, including InDel, Structural Variants (SV) and Breakpoint (BP) analyses, and a report for Structural Variants (Figure 13). The results of InDel and Breakpoint analyses were exported in excel files and different types of deletions or insertions were mapped to the reference sequences manually.

Figure 13. Examples of output files from InDel, Structural Variant (SV) and breakpoint (BP) analyses, to detect different types of mutations in the regions targeted by gene editing. The examples presented were generated for sample NFXL1-1.

2.4.3. Sequence alignment using the latest wheat reference sequence database

The wheat genome reference sequence was not available until IWGSC Reference Sequence v1.0 was revealed on URGI in January 2017 (https://urgi.versailles.inra.fr/About-us/News/IWGSC-Reference-Sequence-v1.0-assembly-is-now-available). This database is the first version of the reference sequence of Chinses Spring, which provides us a more reliable source for wheat genome sequence analysis. To have a better understanding of six sgRNAs used in this project, we blasted the sgRNAs selected in Subsection 2.2.1.1 and the reference sequences used in the NGS analyses in Subsection 2.4.2 in this database (accessed in March 2017). The default settings

were used to blast reference sequences; while the expected threshold was adjusted to 10 to blast short sgRNA sequences.

2.5. Expression analysis of nine DON-induced genes in protoplast system by RT-qPCR

Protoplasts were isolated as described from steps i to x in Subsection 2.3.2.1 and suspended in Falcon™ Polystyrene Microplates (6-well, Fisher Scientific, Canada) in W5 solution. Protoplast suspensions were diluted to $2x10^6$ cells per ml with W5 solution supplemented or not with DON, as described in the following two subsections.

2.5.1. Concentration course experiments

For each treatment, about 600,000 protoplasts were suspended in 5 ml of W5 solution supplemented with 0 (control), 5, 10 and 20 ppm of DON, respectively and incubated in 6-well plates in the dark with gentle shaking at 50 rpm for 2 hours at room temperature. Since DON was in methanol solution, 5 µl of methanol was also added to the control treatment. Three biological replicas from protoplasts isolated at different times were also prepared. After treatment, 10 µl of cell suspension from each treatment was stained with 0.4% (w/v) trypan blue (Sigma, Canada) for 5 min and cell viability was assessed under a microscope. The staining was only performed for the first biological replica. The remaining cells were collected in 2 ml round-bottom Eppendorf tubes by centrifuging at 100 g for 2 min. The supernatant was discarded by pipetting and the cells were immediately frozen in liquid nitrogen. RNA was extracted with the TRI Reagent (Molecular Research Center, USA) method followed by RNA clean up using RNeasy Mini Kit (Qiagen, Canada), according to the manufacturer's instructions. The concentration of purified RNA was measured using a Nanodrop ND-1000 spectrophotometer. Two µg of cleaned up RNA was separated on an agarose gel to evaluate quality. cDNA was synthesized with the RETROscript® Reverse Transcription Kit (Ambion, Canada) using 3.5 µg of RNA template per

sample. Cleaned up RNA (3.5 µg) and RNase-free water were combined to a volume of 10 µl and mixed with 2 µl oligo dT primer; the RNA pre-mix was incubated at 65 ℃ for 2 min. Separately, 2 µl of 10x RT buffer, 4 µl of dNTP mix (2.5 mM each), 1 µl of RNase inhibitor (10 U/µl) and 1 µl of reverse transcriptase (100 U/µl) were combined together to form a common mix. Finally, the common mix and RNA pre-mix were combined with quick spin to a total volume of 20 µl and incubated at 42 ℃ for 2 hours to transcribe RNA into cDNA, followed by 92 ℃ for 10 min to denature enzyme activities. Each cDNA sample was diluted 30 times for use in the following RT-qPCR tests using SensiFAST^{TM} SYBR No-ROX mix (Bioline, USA). The RT-qPCR reactions were performed in a PTC-200 Peltier Thermal Cycler (MJ Research, Canada) as follows: 12.5 µl of master mix, 1 µl each of gene-specific forward and reverse primers (at 10 µM each), 5 µl of diluted cDNA template and 5.5 µl of PCR grade water; the amplification protocol was set up with the following conditions: 95 ℃ for 2 min followed by 35 cycles of 95 ℃ for 5s, 60 ℃ for 10s and 72 ℃ for 20s. Gene-specific primer pairs used in this objective are summarized in Appendix 3D. The RT-qPCR reactions were monitored and recorded using Opticon Monitor 3 (BioRad, USA). The relative expression level of each biological replicate was calculated based on two technical repeats using $E^{-\Delta\Delta CT}$ method (Schmittgen & Livak, 2008).

2.5.2. Time course experiments

Based on the result of the concentration course experiments (Subsection 2.5.1), 10 ppm of DON was used as the treatment for this part. Three biological replicates from protoplasts isolated at different times were also prepared. The protoplasts were treated with either water (supplemented with 10 µl of methanol) or DON (diluted in methanol), and incubated in the dark at room temperature for 2, 4, 6, 8, and 12 hours respectively. Cell collection after treatment, RNA

isolation, cDNA synthesis and RT-qPCR analysis were performed as described in Subsection 2.5.1.

2.5.3. Gene functional analysis

The contig sequences for the ten genes investigated in this section were downloaded from the ORDC local wheat database. To explore the possible functions of these genes in wheat, we blasted their nucleotide sequences against an *Arabidopsis* protein sequence database (BLASTx, https://www.arabidopsis.org/Blast/index.jsp, accessed in October, 2016) and all the settings were as default.

3. Results

A detailed methodology has been introduced in Chapter 2. In this chapter, the results of this project are presented.

3.1. *In vitro* test of candidate sgRNAs

As discussed in Subsection 1.1.1, different kinds of software have been developed to design sgRNAs and to predict their efficiency based on designated sgRNA sequence features. This objective aimed at finding sgRNAs with predicted high editing efficiency and no off-target effects using sequences available for the wheat genome survey at URGI. Since all the genome sequences we obtained at URGI were from Chinese Spring, we further performed standard Sanger sequencing to confirm sequence identity between Fielder and Chinese Spring for the proposed sgRNA loci. Selected sgRNAs were tested in an *in vitro* system supplemented with purified Cas9 nuclease and targeted gene fragments amplified by PCR.

3.1.1. sgRNA design

As the genes of interest were identified using Affymetrix microarray, the sequence of the ESTs used to design the corresponding Affymetrix probe set were obtained from GrainGenes and used for the BLASTn search to identify the longest homolog contigs in our ORDC local database of assembled public wheat ESTs. The contig sequences were used for BLASTn search in the wheat genomic survey databases at URGI to obtain the corresponding genomic sequences. Results of three database searches are summarized in Table 2, and the structures of the three genes, including the positions of introns, exons and the two sgRNAs, are presented in Appendix 10.

Table 2. Database searching results to identify predicted coding and genomic sequences for the genes of interest.

Gene	Affymetrix probe set name	Traes name*	Genomic contig**
TaABCC6	Ta.9831.1.S1_at	Traes_2AL_2F9F6DB1A.1	chr2AL_6399751# chr2AL_6435883 chr2BL_8006379 chr2BL_8006380 chr2DL _9821209 chr2DL_9910117
TansLTP9.4	Ta.1282.4.S1_at	Traes_5BL_CE73B6BFF.1	chr5AL_2800777 chr5BL_10897886
TaNFXL1	Ta.11647.2.S1_at	Traes_7DL_ACACA513D.1	chr7DL_3391562

*from the survey sequence gene models, MIPS v2.2, made public in July 2014 (https://wheat-urgi.versailles.inra.fr/Seq-Repository/Genes-annotations)

**from the genomic survey sequence databases at URGI.

#all contig names include the prefix "IWGSC_ab_k71_contigs_longerthan_200" (e.g. IWGSC_chr2AL_ab_k71_contigs_longerthan_200 _6399751).

Based on the database sequences provided above and the methodologies discussed in Subsection 2.2.1.1, two sgRNAs were selected, targeting two different loci on the same gene fragment to enhance editing efficiency (Table 3).

Table 3. Selected sgRNAs for each gene.

sgRNA name	Sequence (5'-3')
ABCC6-sgRNA-1	CACGCCGTCGAGATTACTGG
ABCC6-sgRNA-2	AGTACTCACGGAGATCCAAG
nsLTP9.4-sgRNA-1	GCCGTGCGTGGCGTACGTGA
nsLTP9.4-sgRNA-2	AGTGCTGCTCCGGCGTGCAG
NFXL1-sgRNA-1	TGACTGGCACAACGCAAGGT
NFXL1-sgRNA-2	GATGGAGTTGGTGTGCCGCA

3.1.2. Targeted gene fragment amplification

Three primer pairs listed in Appendix 3E were used to amplify gene fragments including sgRNA target regions by PCR. The amplified fragments were separated on a 1% agarose gel (Figure 14). For each primer pair, only one band with expected size could be clearly observed, indicating that the primers were highly specific and high yielding. PCR products were then column purified and sent for Sanger sequencing. The same amplicons were also used for the sgRNA *in vitro* tests as targeting templates for digestion as described in Subsection 3.1.4.

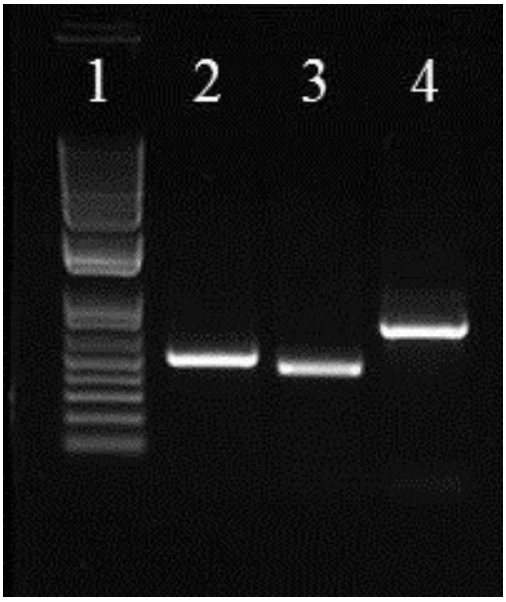


Figure 14. PCR amplification of fragments for three genes. 1: 1kb Plus DNA Ladder; 2: *TaABCC6* gene fragment; 3: *TansLTP9.4* gene fragment; 4: *TaNFXL1* gene fragment.

3.1.3. Sanger sequencing results of PCR amplicons

For each gene, target fragments were amplified from Fielder genomic DNA and sequenced. The sequencing results were aligned as shown in Appendix 11 to Appendix 13. The sequencing results of the six sgRNA regions were checked manually to eliminate sequencing errors. Compared with sequences obtained from Chinese Spring, 100% identity was observed for five of

the six sgRNA target regions in Fielder; the only exception was for the 10th base on ABCC6-sgRNA1 site. The positions of introns and exons for each amplified fragment were also determined using this alignment. Figure 15 summarizes the structure for each gene, including exons and introns, the length of sequenced fragments and the positions of associated sgRNAs.

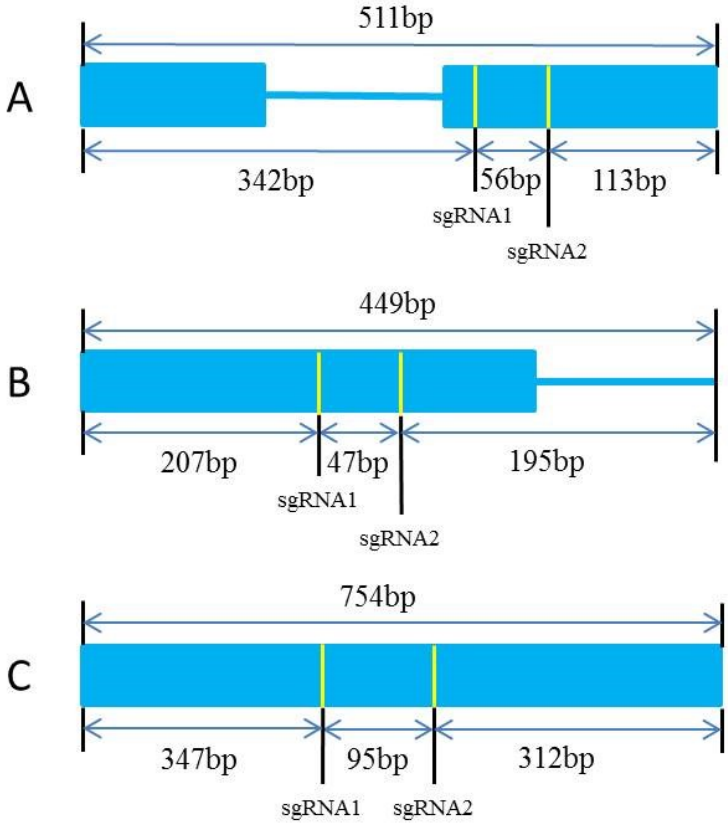


Figure 15. The structure of amplified regions for *TaABCC6* (A), *TansLTP9.4* (B) and *TaNFXL1* (C) genes. Introns are indicated by horizontal lines; rectangular boxes represent of exons. Positions of sgRNAs tested *in vitro* are marked by vertical yellow lines.

3.1.4. *In vitro* test for six sgRNAs

One hundred ng of each purified PCR product amplified with primers in Appendix 3E were used as template to test sgRNAs *in vitro*. The protocol for *in vitro* test can be found in Subsection 2.2.3. As shown in Figure 16, two smaller DNA fragments with expected size were clearly

observed after cleavage with each sgRNA; the uncut PCR products were used as controls. The same PCR amplicons were used to test each of the two sgRNAs targeting the same gene. It can be observed in Figure 16 that no template remains in the digestions with the two sgRNAs for *TaABCC6*, while extra bands representing original PCR products are present for the other four sgRNAs after digestion (Figure 16 B2, B3, C2 and C3). We considered that all six sgRNAs tested *in vitro* were efficient for targeted cleavage driven by sgRNA-Cas9 complex and could be used for editing of the three genes in the protoplast system.

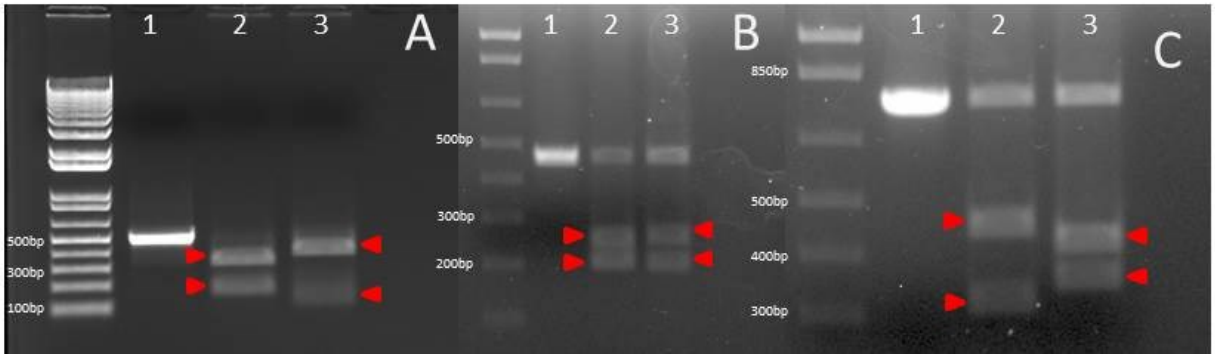


Figure 16. *In vitro* test of six sgRNAs for the *TaABCC6* (A), *TansLTP9.4* (B) and the *TaNFXL1* (C). 1kb Plus DNA Ladder was used in all three gel electrophoresis and digested fragments are marked with red arrowheads. A1： uncleaved template of *TaABCC6* gene fragment; A2 and A3: The fragments after digestion with ABCC6-sgRNA 1 and 2, respectively; B1: uncleaved template of *TansLTP9.4* gene fragment; B2 and B3: The fragments after LTP9.4-sgRNA 1 and 2 digestion, respectively, including some remaining template at the top; C1: uncleaved template of *TaNFXL1* gene fragment; C2 and C3: The fragments after NFXL1-sgRNA 1 and 2 digestion, respectively, including some remaining template at the top.

3.2. Testing of sgRNAs in a wheat protoplast system

The six sgRNAs which have been tested *in vitro* as having sufficient editing efficiency were further tested in a wheat protoplast system. The key steps in this objective included gBlock design, expression vector construction, and protoplast isolation and transformation as discussed in Subsection 2.3. In this section, the results of constructing a pCambia-ABCC6-sgRNA (pCam-

ABC) recombinant expression vector for editing *TaABCC6* gene fragment was used as an example since the methodologies for producing the other three constructs remained similar (pCam-nsLTP9.4, pCam-NFXL1 and pcoCas9-NFXL1).

3.2.1. gBlocks designing and expression vector construction

The sequences for two gBlocks (gBlock 1 and 2) are presented in Appendix 14. The assembled gBlock pairs were firstly amplified by PCR and cloned into pJet 1.2/blunt cloning vectors. Figure 17 shows the results of colony PCR and the expected 1 kb PCR products are marked with red asterisks. Plasmids from clone-1, 4, 7, 10, 13, 14, 16, 17, 18, 20 and 22 were isolated for EcoRI and KpnI restriction enzyme digestion and were renamed ABC-pJet 1~11. Since the buffers for EcoRI and KpnI are unique to its own enzyme, we did single digestion with EcoRI or KpnI for 11 plasmid samples and linearized plasmids migrated more slowly on the agarose gel than undigested supercoiled vectors (Figure 18). Most plasmid samples were successfully digested with both EcoRI and KpnI. The insertion in ABC-pJet 10 was confirmed by sequencing and used for delivering assembled gBlock pairs into pCambia vector.

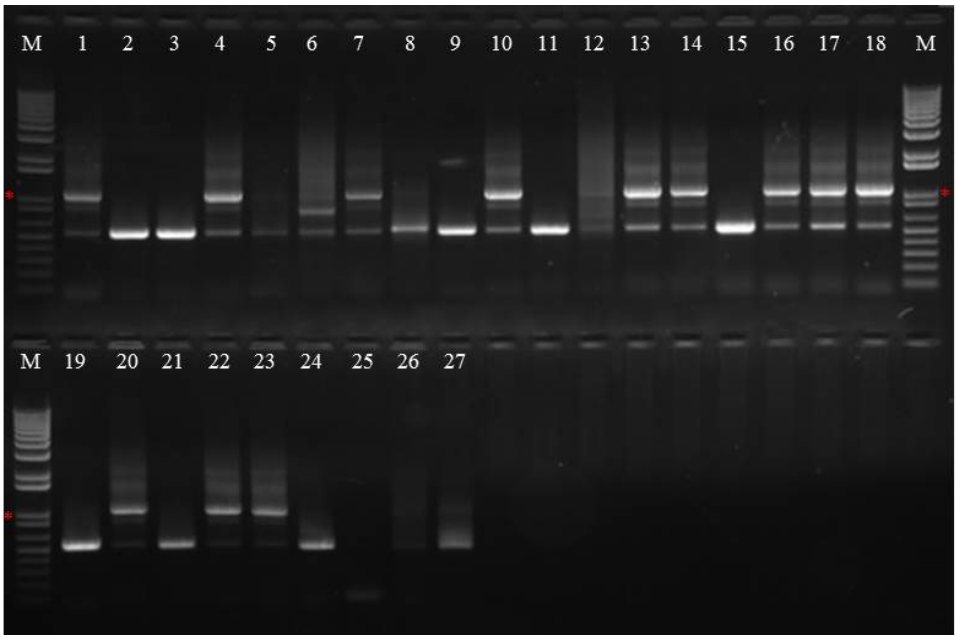


Figure 17. Gel electrophoresis of PCR amplicons from 27 clones for ABC-pJet constructs. The expected 1kb PCR amplicons are marked with red asterisk. M: 1kb Plus DNA Ladder; 1~27: PCR amplicons from 27 clone aliquots.

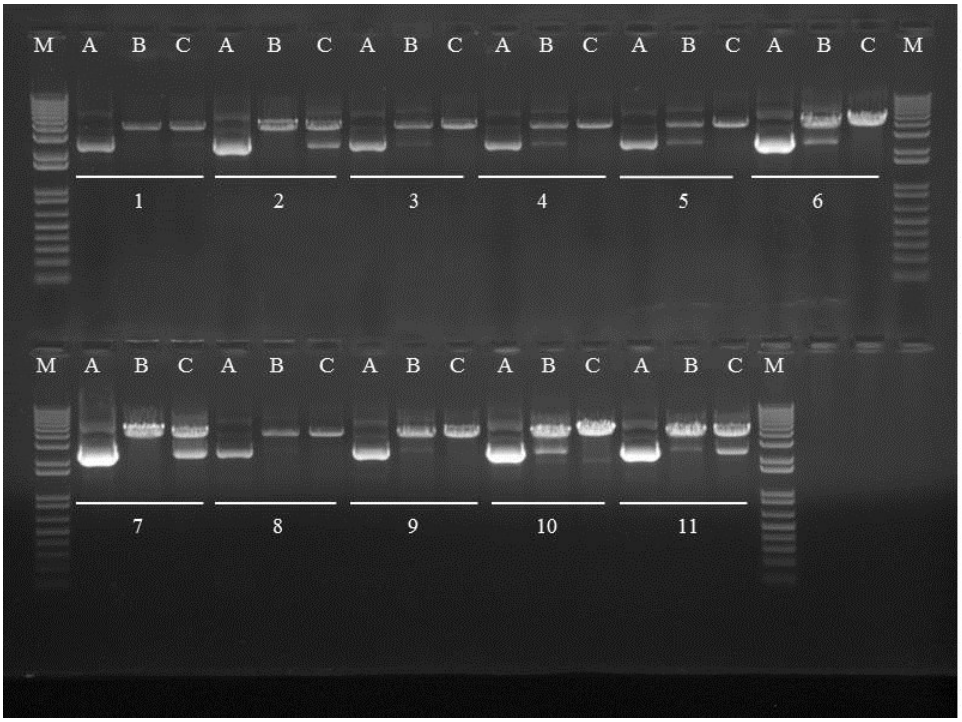


Figure 18. Gel electrophoresis of isolated plasmids ABC-pJet 1~11 before and after restriction enzyme digestion. M: 1kb Plus DNA Ladder; A: undigested plasmids as a control; B: plasmids digested with EcoRI; C: plasmids digested with KpnI.

The ABC-pJet 10 recombinant vector was then digested with EcoRI and KpnI and the recombinant fragment was subcloned into a modified pCambia 1302 vector (Appendix 4, Subsection 2.3.1.3) which was digested with the same enzymes. Two sgRNA-specific sequences for editing *TaABCC6* gene fragment were used as primers (Appendix 3A) for colony PCR (Figure 19). Clone 5 (pCamABC6-5) was verified by sequencing and the plasmid was isolated for protoplast transformation. The same methodology was applied for constructing pCam-nsLTP9.4 and pCam-NFXL1 recombinant expression vectors. For the construction of pcoCas9-NFXL1 expression vector, restriction enzymes EcoRI and XmaI were added to 5' of forward and reverse primers respectively for gBlock pair amplification and the amplified fragments were ligated between EcoRI and XmaI sites on the pFGC-pcoCas9 vector (Appendix 5) while the other steps remained the same.

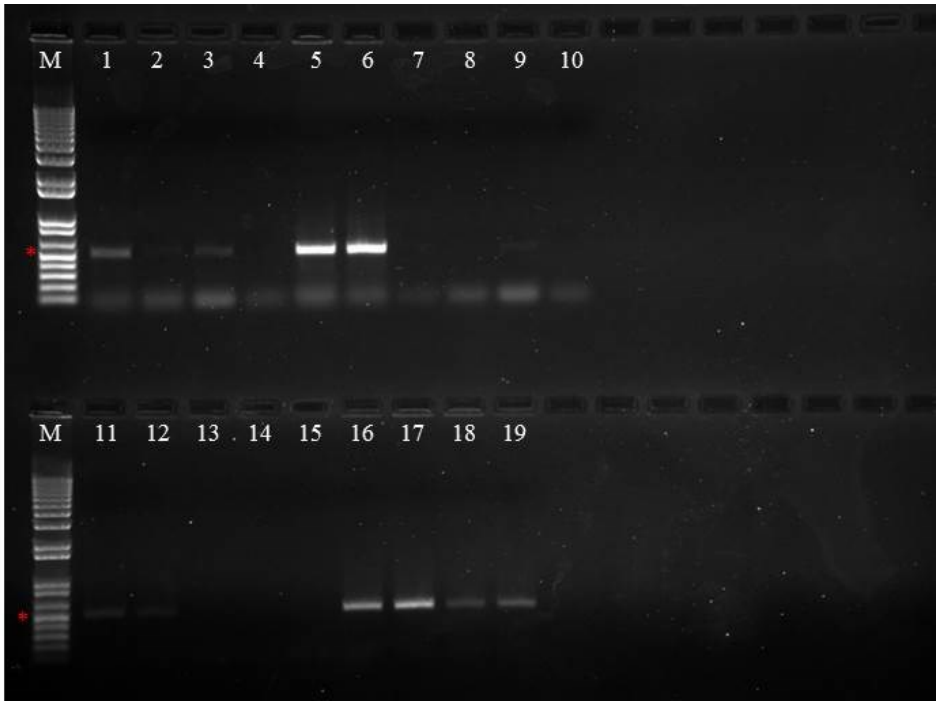


Figure 19. Gel electrophoresis of PCR amplicons from 19 clones for pCambia-ABC constructs. The expected 500 bp amplicons are marked with red asterisk. M: 1 kb Plus DNA Ladder; 1~19: PCR amplicons from 19 *E.coli* clone aliquots.

3.2.2. Wheat protoplast isolation and transformation

Compared with transgenic wheat plants, experiments with protoplasts are time-saving and with lower cost. Transient gene expression assays in protoplast systems have been proven to be powerful and important tools to convey cell-based experiments including gene expression and genomic analyses. So far, a protocol has been well-established for protoplast isolation and transformation from *Arabidopsis* leaves (Yoo et al., 2007). And the use of CRISPR for targeted gene editing in wheat protoplasts has also been reported by Shan et al. (2014). In this section, we firstly tested different conditions which might affect the yield and viability of protoplasts, such as centrifugal speed and time, and the number of cells used for each transformation.

3.2.2.1. The optimization of wheat protoplast isolation and transformation

In comparing the two published protocols mentioned above, some key parameters such as the composition of different media and the time needed for cell wall digestion were common, suggesting that they had been well optimized. In contrast, the major differences between the two protocols were the incubation period post-transformation and the number of cells needed for each transformation. We tested incubation periods of 5 and 30 min (Figure 20) for PEG-mediated transformation (Step ii, Subsection 2.3.2.2). For each transformation, 20,000 and 200,000 cells were also tested (Figure 21). The transformation efficiency was estimated based on the ratio of the number of fluorescent cells counted in the dark field, to the total number of cells counted in the same but bright field by transforming the protoplasts with a vector expressing GFP. The results showed that 5-min incubation was sufficient and worked better than 30-min post PEG-mediated transformation. The estimated transformation efficiency was 10% higher in 5-min incubation while the cell integrity remained the same between the two different incubation

periods (Table 4). However, we did not find a significant difference in transformation efficiency when we compared the number of cells, 20,000 or 200,000, used for each transformation (Table 4).

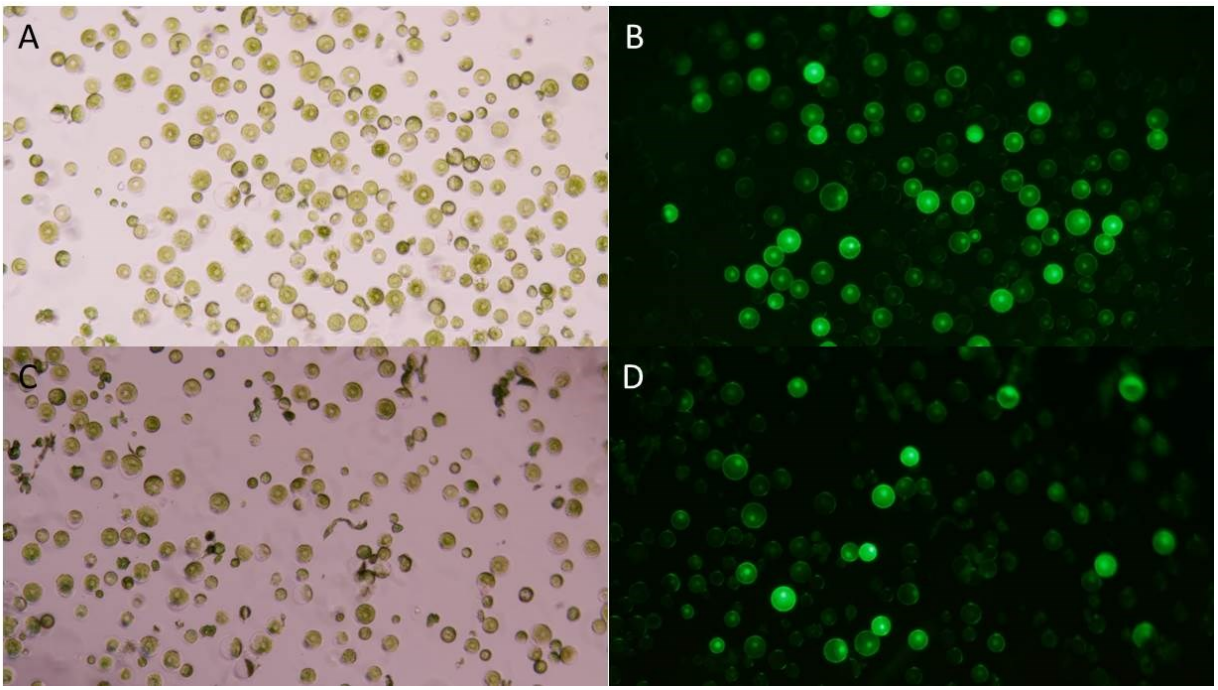


Figure 20. The effect of incubation period on transformation efficiency. 20,000 cells were used per transformation with GFP vector and visualized by microscopy. A, B: 5-min incubation, a bright field and dark field, respectively; C, D: 30-min incubation; A, C: bright field; B, D: dark field fluorescence.

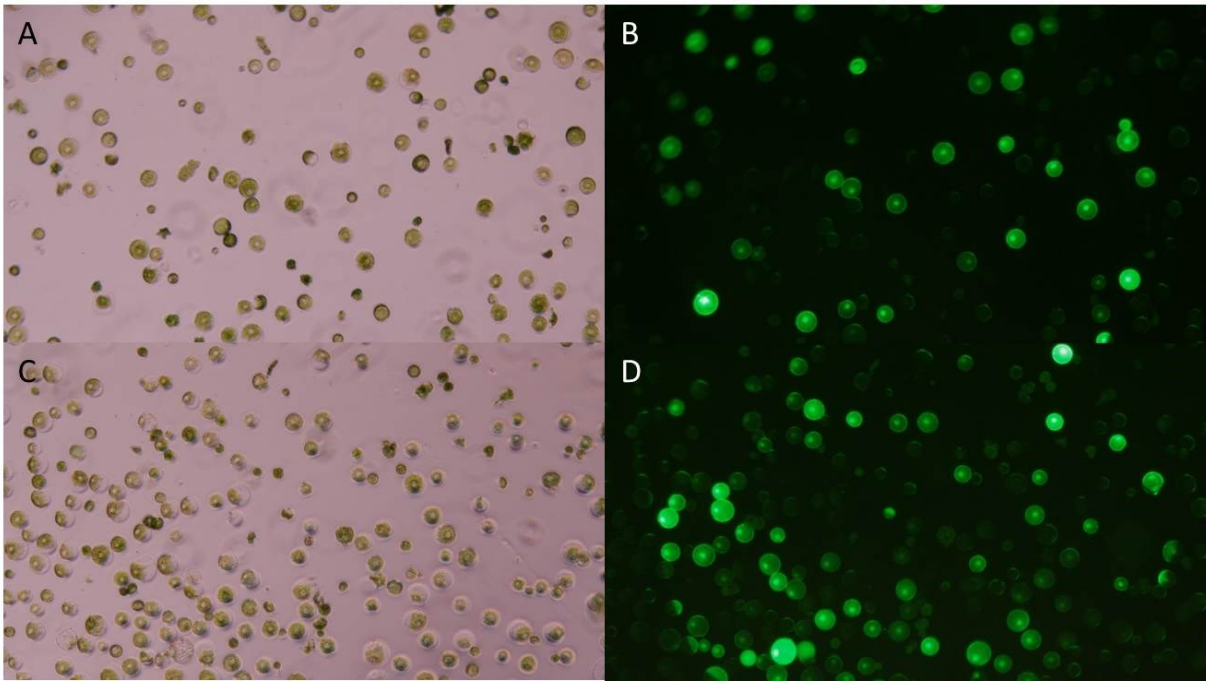


Figure 21. The effect of the cell amounts on transformation efficiency. The cells were incubated for 5 min per transformation with GFP vector and visualized by microscopy. A, B: 20,000 cells used per transformation; C, D: 200,000 cells used per transformation; A, C: bright field; B, D: dark field fluorescence.

Table 4. The estimated transformation efficiency of protoplasts using different number of cells and two incubation periods. The mean value and the standard deviation (SD) were calculated based on three replicas using protoplasts isolated at different times.

Incubation period	The number of cells	Exp. 1	Exp. 2	Exp. 3	Mean	SD
5 min	20,000	58.1%	53.7%	61.4%	57.7%	0.039
30 min		47.6%	45.3%	51.3%	48.1%	0.030
5 min	200,000	54.1%	44.2%	55.9%	51.4%	0.063
	20,000	56.7%	46.9%	54.2%	52.6%	0.051

In addition, we also found that the incubation method had a deep impact on cell viability. In the test with 200,000 cells, we found that most cells were broken after the overnight incubation in an Eppendorf round-bottom 2 ml tube. To minimize cell loss from transformation using 200,000 cells, the incubation was performed in a 6-well plate with gentle shaking at 50 rpm in the dark at

room temperature, while the incubation of 20,000 cells was directly performed in the Eppendorf round-bottom 2 ml tubes (Appendix 15) in the dark without shaking.

3.2.3. Mutation detection and quantification by high throughput sequencing

Recombinant vectors prepared in Subsection 3.2.1 were transformed into wheat protoplasts based on an optimized protocol discussed above. To detect and quantify mutations on three target genes, we performed two rounds of amplicon sequencing. Ten samples were sequenced firstly (NFXL1-1~5 and nsLTP-1~5). In the second round of sequencing, amplification and sequencing were performed on 17 samples (ABC-1~5, pcoNFXL1-1~5, resequenced NFXL1-1~5, ABC-control and NFXL1-control, Subsection 2.4). For each sample in both rounds of amplicon sequencing experiments, two individual transformations were pooled to isolate genomic DNA. PCR amplicons were sequenced using an Illumina MiSeq Sequencing System. The sequencing data was analyzed on CLC Genomics Workbench.

3.2.3.1. Sequencing results for pCam-nsLTP and pCam-NFXL1 target regions

The PCR amplicon samples characterized in the first round of amplicon sequencing are listed in Appendix 16. In our first round of sequencing, for each nsLTP sample, ~100,000 reads were mapped to the reference; while for NFXL1, the mapped reads ranged from approximately 150,000 to 410,000. An average of 92.5% reads were mapped to the reference for nsLTP-1~5, while 91.1% of total reads were mapped to the NFXL1 reference sequence (Appendix 17). However, the ratio of broken paired reads for NFXL1 samples (13% to 16%) was higher than those for nsLTP samples (~6%), indicating nsLTP-1~5 were with better mapping quality (Appendix 17). Using the InDels, Structural Variants and Breakpoint Analysis programs under the Resequencing Analysis section on CLC Genomics Workbench, we estimated CRISPR

activity for each sample. The results showed that targeted sites in the samples nsLTP 1~4 and NFXL1-1~5 were successfully edited (Table 5). Summaries of the different modifications observed from all sequenced samples are presented in Figure 22 and Figure 23 for *TansLTP9.4* and *TaNFXL1* targeted regions respectively. Besides insertions and deletions, we also identified three different types of replacements for *TaNFXL1* target regions, the variant sequences ranging from 9 bp to 204 bp (Appendix 18D). In those cases, some nucleotides were firstly deleted from genomic DNA, then a different fragment was inserted at the same locus. In brief, different types of deletions or insertions were discovered by Breakpoint and Indel Analysis and replacements were determined by Structural Variants. The original data exported from CLC Workbench can be found in Appendix 18 (A~D). The estimated editing efficiencies ranged from 0% to 11.9% in 5 samples for nsLTP, while up to 8.4% for editing *TaNFXL1* target regions.

Table 5. A summary of estimated editing efficiency for the 10 sequenced samples in the first round of amplicon sequencing. The calculations were based on the variant ratio of different modifications in Appendix 18.

Sample name	Editing efficiency	Sample name	Editing efficiency
nsLTP-1	9.1%	NFXL1-1	4.0%
nsLTP-2	1.9%	NFXL1-2	4.4%
nsLTP-3	11.3%	NFXL1-3	8.4%
nsLTP-4	11.9%	NFXL1-4	8.1%
nsLTP-5	0.0%*	NFXL1-5	3.0%

* No modification was detected for this sample.

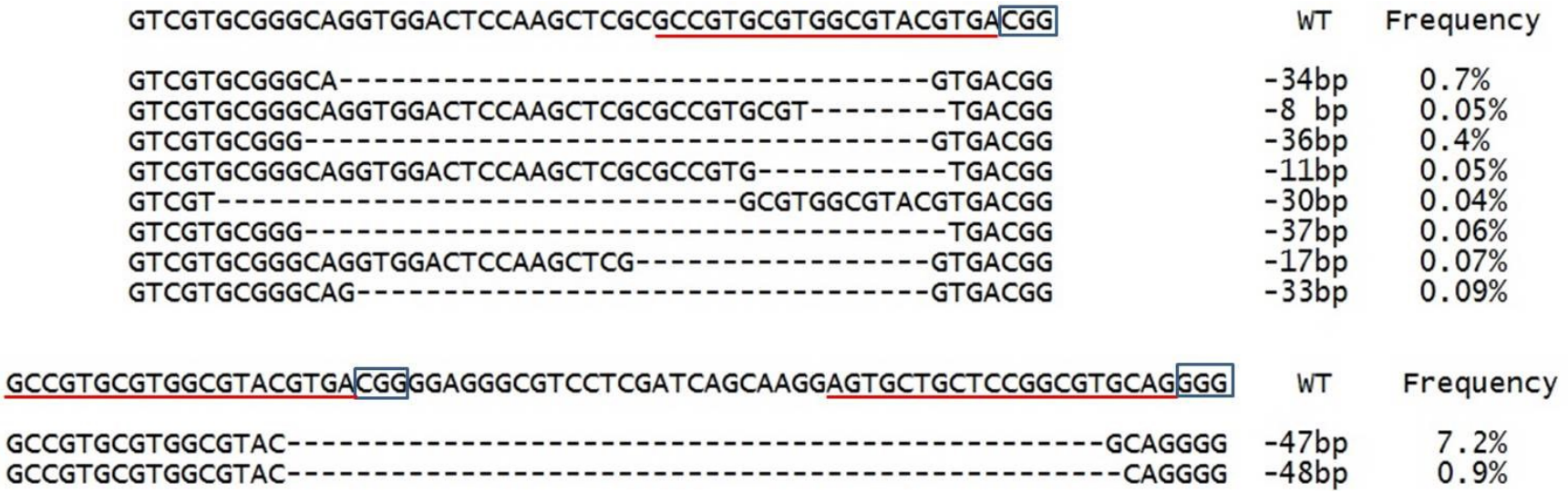


Figure 22. Alignments of the high-throughput sequencing reads with deletion for *TansLTP9*.4 target regions and their respective editing frequency at target regions. The deletions presented in this figure are the cumulative results from samples nsLTP-1~5. The reference sequence is at the top of each panel. The 20 nt sgRNA sequences are underlined in red and the NGG PAM structure is in a blue rectangle.

Figure 23. Alignments of the high-throughput sequencing reads with deletion for *TaNFXL1* target regions and their respective editing frequency at target regions. A: Targeted deletions at the sgRNA1 and 2 sites; B: Insertions at the sgRNA1 site. The modifications presented in this figure are the cumulative results from samples NFXL1 1~5.The reference sequence is at the top of each panel. The 20nt sgRNA sequences are underlined in red and the NGG PAM structure is in a blue rectangle.

To find the origins of three insertions on NFXL1-sgRNA1 site, we blasted the insertion sequences with the pCambia vector itself. The results showed that two insertions (+109 bp and +83 bp) matched with the vector sequence partially and the third, 115 bp insertion, was mapped to a fragment from the *TaNFXL1* gene itself (Figure 24). Interestingly, we found that the last 17 bases of this fragment were perfectly mapped to the NFXL1-sgRNA1 sequence (sgRNA1: TGACTGGCACAACGCAAGGT-*PAM*, the fragment found in 115 bp insertion is underlined). Given the fact that the cutting site of Cas9 nuclease is 3 bases upstream the PAM structure (Wu et al., 2014), this 115 bp insertion is likely a fragment resulting from a cleavage at the sgRNA1 site. In addition, for the other two insertions, the vector was the only foreign DNA introduced into the protoplasts, making it a possible template donor in DSB repairing after cleavage.

Figure 24. The BLAST results for three insertions found at the NFXL1-sgRNA1 site. Sequences which perfectly matched with pCambia vector are underlined and in bold (part of +109 bp and +83 bp insertions); the sequence with a match to a fragment on *TaNFXL1* is in italic (part of +115 bp insertion) and the first 17 bases of NFXL1-sgRNA1 are underlined in red.

3.2.3.2. Sequencing results for ABCC6 and a Cas9 efficiency comparison at NFXL1 target regions

Seventeen samples were sequenced in the second round of sequencing, including 5 for *TaABCC6* targeted regions (ABC 1~5), 5 for *TaNFXL1* targeted regions edited by a plant-codon optimized Cas9 (pcoNFXL1-1~5) and 2 for non-edited *TaABCC6* and *TaNFXL1* gene fragments (ABC-

control, NFXL1-control, respectively). In addition, we also re-amplified and re-sequenced 5 NFXL1 samples from our first round of sequencing (NFXL1-1~5). The PCR amplicons submitted for sequencing are listed in Appendix 19. The sequencing analysis results are summarized below.

The quality check was also carried out by mapping the reads to the reference sequences for both *TaABCC6* and *TaNFXL1* gene fragments (Appendix 20). On average, about 100,000 reads were mapped to the reference sequence for each sample. For ABC-1~5, an average of 84% reads were mapped; while for both NFXL1-1~5 and pcoNFXL1-1~5, about 93% of total reads were mapped. For the ABC and NFXL1 controls, 79% and 93% of total reads were mapped to the reference, respectively. Further, ABC-1~5 had a higher ratio of broken paired reads (~22%) than 10 samples for NFXL1 (~7.4%). Compared with the first-round sequencing, the five resequenced NFXL1 samples were with better mapping quality. However, no considerable difference in the ratio of broken paired reads was observed between edited samples and controls (22% and 24% for ABC, 7.4% and 6.2% for NFXL1, respectively).

InDel and Breakpoint analysis was also performed for this round of sequencing. As expected, we did not find any modifications at targeted regions in the two controls, ABC- and NFXL1-control. In 15 samples transformed with editing constructs, we detected modifications in all samples except for the re-sequenced NFLX1-5. Figure 25 and Figure 26 summarized different modifications on targeted regions for ABC 1~5 and pcoNFXL1-1~5. We also summarized the modifications on 5 resequenced NFXL1-1~5 samples (Figure 27). Compared with the results in the first round of sequencing (Figure 23), we noticed more different types of modifications in this round of editing experiments. The sequencing analyses imported from CLC Workbench are

shown in Appendix 21 and Appendix 22. A summary of the respective editing efficiency for 15 samples is presented in Table 6.

Figure 25. Alignments of the high-throughput sequencing reads with deletion for *TaABCC6* target regions and their respective editing frequency at target sites. The deletions presented in this figure are the cumulative results from samples ABCC6-1~5.The reference sequence is at the top of each panel. The 20 nt sgRNA sequences are underlined in red and the NGG PAM structure is in a blue rectangle.

Figure 26. Alignments of the high-throughput sequencing reads with deletion for the pcoNFXL1 target regions and their respective editing frequency at target sites. The deletions presented in this figure are the cumulative results from samples pcoNFXL1-1~5. The reference sequence is at the top of each panel. The 20 nt sgRNA sequences are underlined in red and the NGG PAM structure is in a blue rectangle.

Figure 27. Alignments of the high-throughput sequencing reads with deletion for the resequenced *TaNFXL1* target regions and their respective editing frequency at target sites. The deletions presented in this figure are the cumulative results from resequenced samples NFXL1-1~5. The reference sequence is at the top of each panel. The 20 nt sgRNA sequences are underlined in red and the NGG PAM structure is in a blue rectangle.

The estimated editing efficiencies ranged from 6.6% to 13% for ABC-1~5. Compared with NFXL1-1~5 in the first round of sequencing, a higher calculated efficiency was observed in this round, up to 42.2% in 5 resequenced samples. For NFXL1-5, we detected an editing efficiency of 3% in the first round of sequencing, while no modification was found in this round. Besides this, we also found that the editing frequencies in NFXL1-1~5 varied significantly from 0% to 42.2% with the crCas9 nuclease (CV (Coefficient of Variation) =87%). However, in samples transformed with the pcoCas9 nuclease (pcoNFXL1-1~5), the estimated efficiencies were relatively consistent among different samples (8.6% ~ 21.2%, CV=38%), suggesting that pcoCas9 is either more consistently expressed, more stable and/or function more efficiently in the wheat protoplasts.

Table 6. A summary of estimated editing efficiency for the 15 sequenced samples in the second round of sequencing. The calculations were based on the variant ratio of different modifications in Appendix 21 and Appendix 22.

Sample name	Editing efficiency	Sample name	Editing efficiency	Sample name	Editing efficiency
ABC-1	9.3%	NFXL1-1	8.4%	pcoNFXL1-1	10.2%
ABC-2	6.6%	NFXL1-2	18.3%	pcoNFXL1-2	8.6%
ABC-3	9.1%	NFXL1-3	42.2%	pcoNFXL1-3	15%
ABC-4	13%	NFXL1-4	22.8%	pcoNFXL1-4	21.2%
ABC-5	9%	NFXL1-5	0%*	pcoNFXL1-5	20.7%

* No modification was detected for this sample.

In addition, we also estimated the editing efficiency for sgRNA1 and sgRNA2 in each sample (Table 7). To have a better comparison of efficiencies between pcoCas9 and crCas9, the results for re-sequenced NFXL1-1~5 sampled were used for this table. In ABC-1~5 and nsLTP-1~4, similar frequencies of editing were observed for sgRNA1 and 2. However, there was a higher

editing efficiency at the NFXL1-sgRNA2 site than at the NFXL1-sgRNA1 site in NFXL1-1~4 and pcoNFXL1-1~5 samples. We also noticed that the editing efficiencies at each of the two sgRNA sites in pcoNFXL1 samples were more consistent than in NFXL1 samples, which correlated with our findings in Table 6. However, there was not a clear relationship between editing frequencies observed and protoplast batches (-1 with -2 or -3 with -4, Appendix 8).

Table 7. The estimated editing efficiency on sgRNA1 or 2 site for each sample from two rounds of sequencing. To have a better comparison between two types of Cas9 nucleases, the re-sequencing data was used to calculate efficiencies for NFXL1-1~5. For modifications between two sgRNA sites, they were counted in both categories.

Sample name	Editing efficiency		Sample name	Editing efficiency	
	sgRNA1	sgRNA2		sgRNA1	sgRNA2
ABC-1	8.4%	8.6%	NFXL1-1	1.8%	6.6%
ABC-2	6.6%	6.5%	NFXL1-2	11.7%	13.8%
ABC-3	8.8%	8.8%	NFXL1-3	10.9%	31.3%
ABC-4	11.1%	11.6%	NFXL1-4	5.2%	17.7%
ABC-5	8.7%	8.9%	NFXL1-5	0.0%*	0.0%*
nsLTP-1	9.1%	8.6%	pcoNFXL1-1	2.9%	7.3%
nsLTP-2	1.9%	1.7%	pcoNFXL1-2	2.2%	6.4%
nsLTP-3	11.3%	9.6%	pcoNFXL1-3	3.4%	11.7%
nsLTP-4	11.9%	9.9%	pcoNFXL1-4	6.5%	18.8%
nsLTP-5	0.0%*	0.0%*	pcoNFXL1-5	9.2%	11.5%

* No modification was detected at this sgRNA site.

3.2.4. Off-target analyses using the latest wheat reference sequence database

A first version of the wheat genome reference sequence was made available in January 2017 with the unveiling of "IWGSC Reference Sequence v1.0" for cultivar Chinese Spring on URGI. We blasted the six sgRNAs, and three reference sequences obtained from cultivar Fielder

(Subsection 3.1.3) to this database. The findings are summarized in Table 8. Two distinct matches, representing closely related genes from the same family, were found on the same chromosome in each genome for each target region, except for *nsLTP9.4* on 5B and 5D. In addition, the search results were consistent with our initial finding that there was no unwanted target (except for the novel second genes) outside the target regions for all six sgRNAs. For the two ABCC6-sgRNAs and nsLTP-sgRNAs and for NFXL1-sgRNA1, we found one to three mismatches on different target sites in the Chinese Spring reference sequence. It is not known at this time if similar mismatches are present in the sequence of the second closely related genes in Fielder.

Table 8. Comparison of the six sgRNAs and target gene fragments used as the reference in NGS to the IWGSC Reference Sequence v1.0 database. The sequences of sgRNAs and target gene fragments were from cultivar Fielder, while the reference sequence database is from Chinese Spring.

Target	Chromosome	Match[A]	PCR[B]	sgRNA1[C]	sgRNA2[C]	RT-qPCR[D]
ABCC6	2A	1 (91%)	Y	0	0	Y
		2 (84%)	Y	2 (7,10)	1 (10)	N
	2B	1 (92%)	Y	1 (10)	1 (8)	Y
		2 (83%)	N	0	1 (8)	N
	2D	1 (94%)	Y	0	0	Y
		2 (84%)	Y	1 (10)	0	N
nsLTP9.4	5A	1 (99%)	Y	0	0	N/A*
		2 (99%)	Y	0	0	N/A
	5B	1 (96%)	Y	0	0	N/A
	5D	1 (93%)	N	1 (10)	1 (12)	N/A
NFXL1	7A	1 (94%)	Y	0	0	Y
		2 (93%)	Y	0	0	Y
	7B	1 (94%)	Y	0	0	N
		2 (93%)	Y	0	0	Y
	7D	1 (94%)	Y	0	0	Y
		2 (93%)	Y	3 (2,7, 14)	0	Y

[A] The percentage refers to the respective identity between the Sanger sequencing results from Fielder in Subsection 3.1.3, which are the consensus of the sequences from the three genomes, and the different homoeologous genes in Chinese Spring genomic DNA.

[B] Indicates if the gene could have been amplified by the primers used for the NGS analysis (Y) or not (N) (Subsection 3.2.3), based on the reference sequence from Chinese Spring.

[C] The number of mismatch (es) between the sgRNA and the potential target site in Chinese Spring, and the relative location (s) of the mismatch(es) on the sgRNA (with position 1 being the farthest base on the sgRNA from the PAM structure; the 20th, the closest: 5'-N (1) N…NN (20)-PAM-3').

[D] Indicates if each gene could be amplified in the RT-qPCR assays (Subsection 3.3), based on the reference sequence from Chinese Spring.

* Not applicable for this gene.

3.3. Expression analyses of nine DON-induced genes in a transient protoplast system

Previous studies in our lab found that some genes can be induced in Roblin heads by direct DON treatment (Balcerzak et al., 2016; T Ouellet, personal communication). In this objective, we explored the expression pattern of these genes in wheat protoplasts treated with DON. Untransformed wheat protoplasts were treated with DON (diluted in methanol) or water (supplied with the same amount of methanol as used in DON solution) and the expression level of these genes was analyzed by RT-qPCR. The appropriate DON treatment was firstly determined by a concentration course experiment with the characterization of four genes; the chosen DON concentration was then applied to a time course experiment for further characterization of gene expression. In this protoplast system, some genes were found to have a similar expression profile as in wheat heads for both concentration and time course experiments. However, possibly due to the difference between cell-based and tissue-based experiments, some genes expressed differentially between these two systems. In addition, comparing to the similar experiments in the wheat heads where up to 100 ppm of DON was used (Balcerzak et al., 2016), and considering that protoplasts are more fragile and the contact with DON is more direct, a

much lower concentration of DON was used in the protoplast experiments than in the wheat heads.

3.3.1. Concentration course experiment

The cell viability of protoplasts when treated with DON was firstly examined with trypan blue staining (Figure 28). Most cells (> 95%) were still alive after 2 hours of treatment with all DON concentrations (0, 5, 10 or 20 ppm). However, we found that when protoplasts were treated with 20 ppm DON, there were a much higher proportion of cells with the chloroplasts gathered together on one side of the cell, often associated with a not perfectly spherical shape of the cell, when compared with protoplasts treated with other DON concentrations or water. This indicated that protoplasts exposed to the highest concentration of DON were stressed.

The RT-qPCR results for four genes known to be induced by DON in wheat heads also showed that a 2-hour treatment with 10 ppm of DON was sufficient to induce their expression (Figure 29). Hence, a treatment with 10 ppm of DON was used for the time-course experiment.

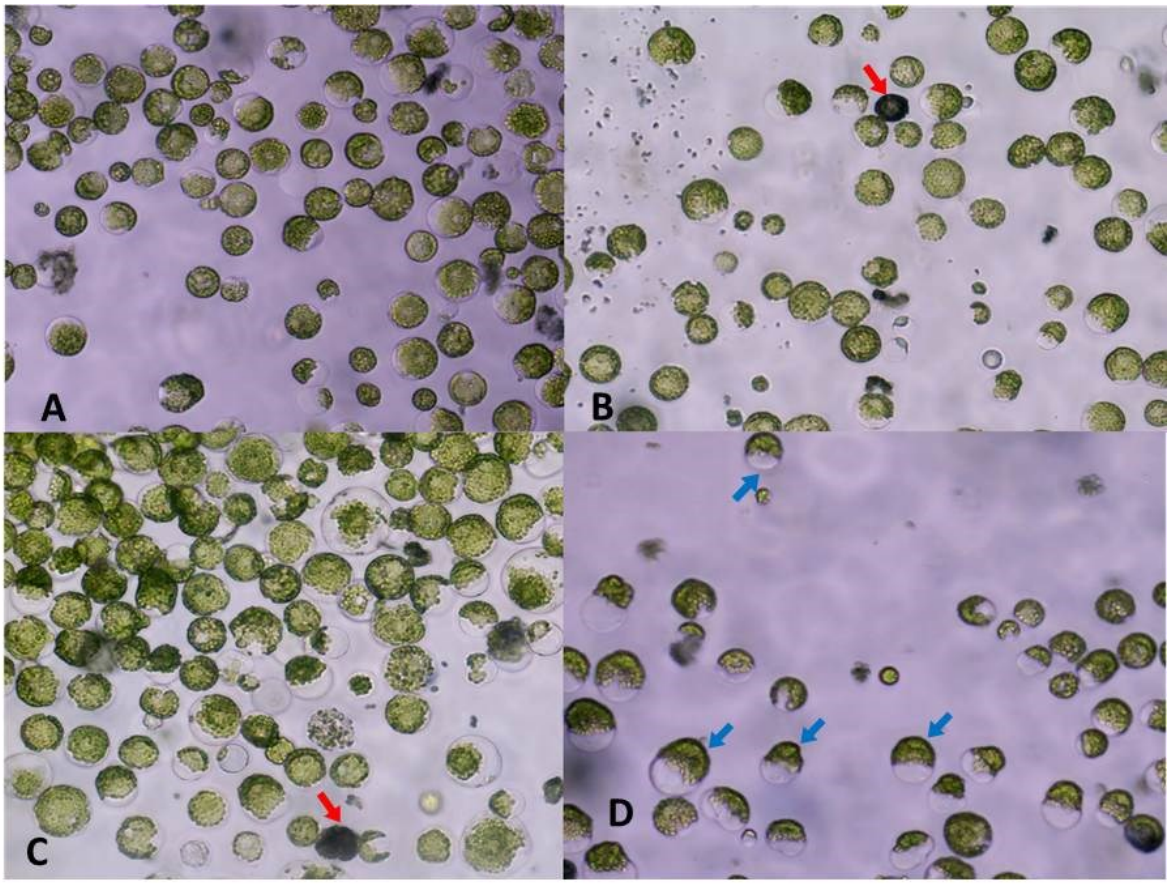


Figure 28. Protoplasts with water or DON treatment were stained with trypan blue. A~D: DON treatment for 2 hours with the concentrations of 0, 5, 10, 20 ppm of DON, respectively. Dead cells are marked with red arrows and some abnormal cells are indicated with blue arrows.

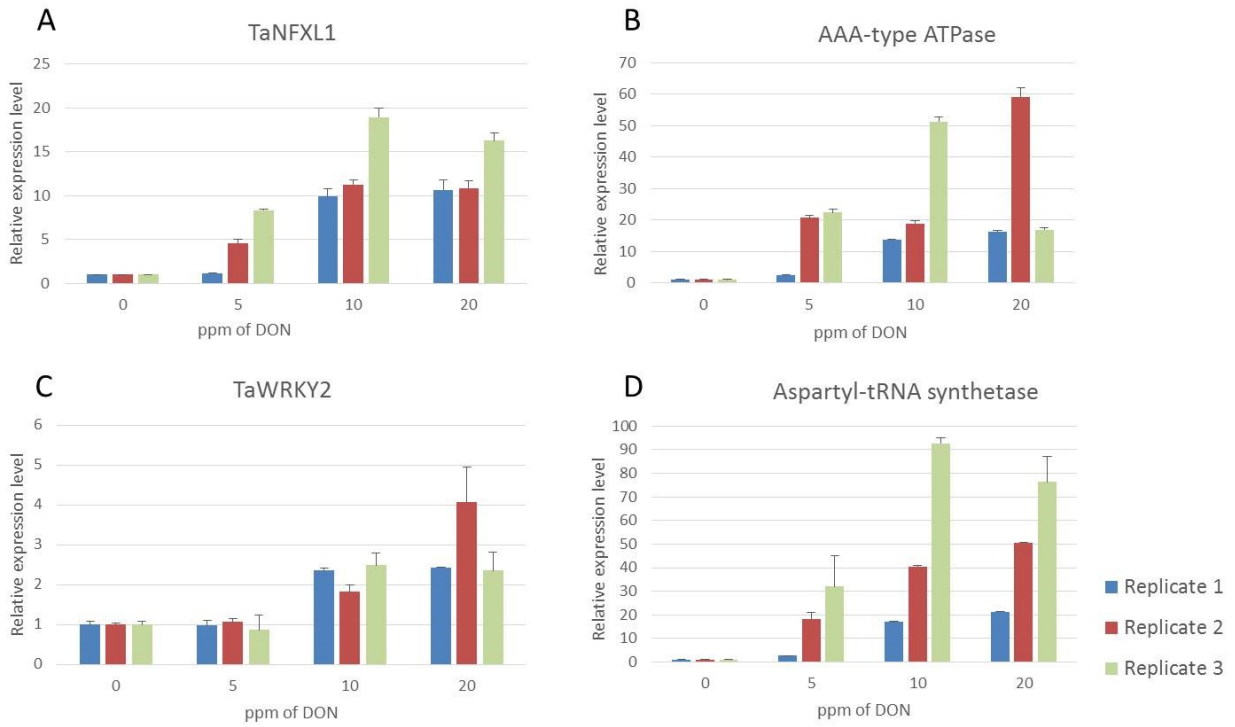


Figure 29. The concentration course experiments of four tested genes. A~D: Relative expression levels of *TaNFXL1*, an AAA-type ATPase, *TaWRKY2* and an Aspartyl-tRNA synthetase in protoplasts treated with different concentrations of DON for 2 hours. Columns with different colors indicate three biological replicates done for each treatment. The replicates were prepared from three different batches of protoplasts. Error bars represent standard deviation between two technical repeats. The calculation was based on $E^{-\Delta\Delta CT}$ method (Schmittgen & Livak, 2008).

3.3.2. Time course experiment

For the time course experiment, the protoplasts were treated with either water or 10 ppm of DON for 2, 4, 6, 8 and 12 hours. The RT-qPCR results showed that the expression levels of six genes, a MRP-like ABC transporter (*TaABCC6*), *TaNFXL1*, an AAA-type ATPase, an aspartyl-tRNA synthetase, Ta.5297 and a glutathione transferase were significantly up-regulated (over 100-fold change) after DON treatment compared with water treatment (Figure 30), which correlate with the expression patterns observed in DON-treated wheat heads (Ouellet et al. 2012). However, we

found a different expression pattern for the other three genes tested: a Cys2/His2 Zin Finger Protein, Ta.22180, and *TaWRKY2*. *TaPR4* was used as a control since this gene can be induced by *Fusarium* infection, but not by DON treatment. While those three genes were highly up-regulated in wheat heads by DON (Ouellet et al., 2012), their expression showed no or only a modest up-regulation in DON-treated protoplasts (Figure 31). These genes exhibited a different expression pattern in the cell-based and tissue-based experiments. A possible explanation is that the expression patterns of these genes are regulated largely by tissue-specific factors, leading to either strong induction or high stability of mRNAs in wheat heads while no or only minimal expression changes are observed in protoplasts from leaf tissues. However, this still needs to be confirmed by further experiments.

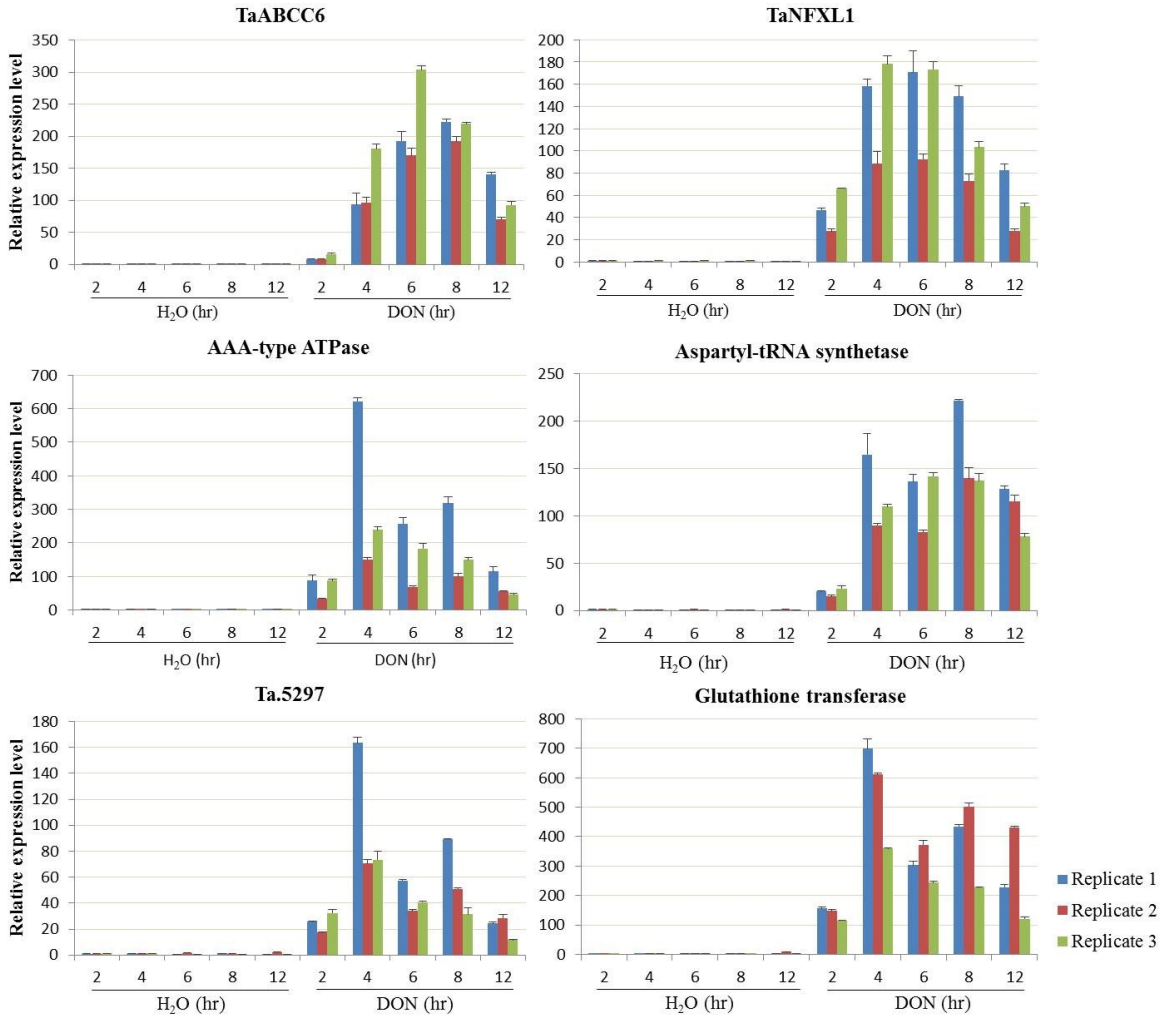


Figure 30. Relative expression levels of six highly up-regulated genes under DON treatment in the time course experiments. Ten ppm of DON was used for treatment. Columns with different colors stand for three biological replicates prepared from different batches of protoplasts. Error bars represent the standard deviation between two technical repeats in the RT-qPCR assay. The calculation was based on $E^{-\Delta\Delta CT}$ method (Schmittgen & Livak, 2008).

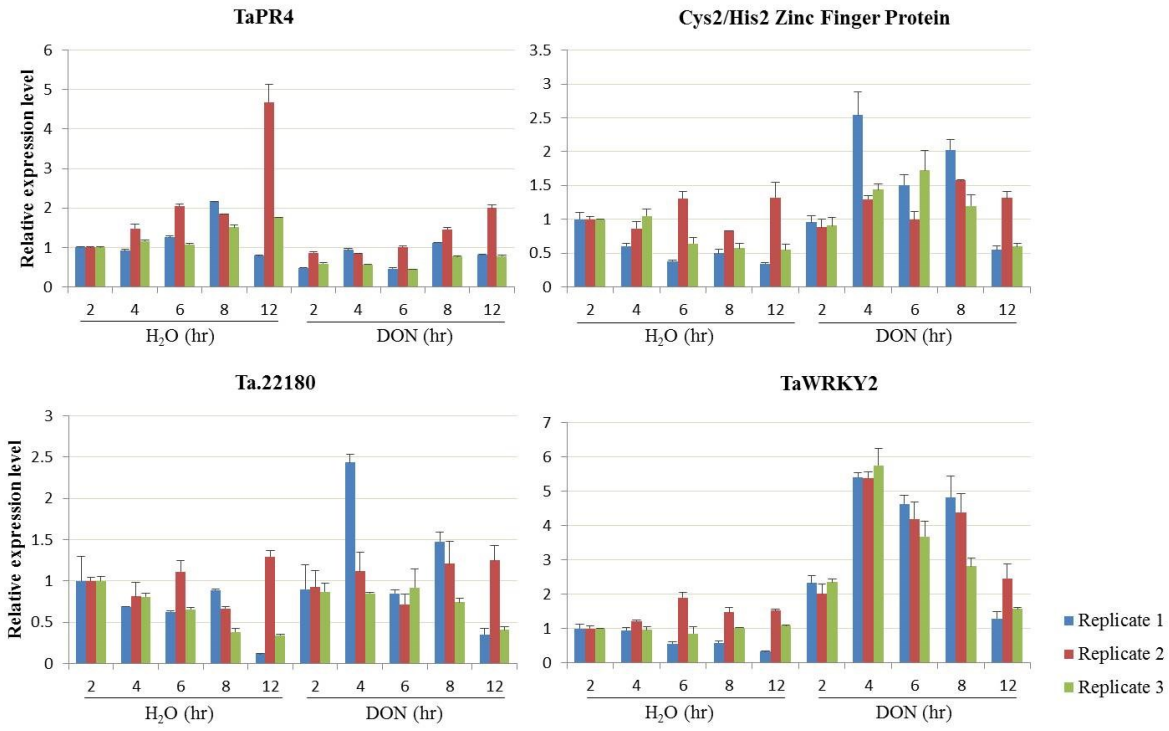


Figure 31. Relative expression levels of four genes with lower expression change under DON treatment in the time course experiments. Ten ppm of DON was used for treatment. Columns with different colors stand for three biological replicates prepared from different batches of protoplasts. Error bars represent the standard deviation between two technical repeats in the RT-qPCR assay. *TaPR4* is used as a control since this gene is induced by FHB but not by DON itself. The calculation was based on $E^{-\Delta\Delta CT}$ method (Schmittgen & Livak, 2008).

3.3.3. Functional annotation of the homologs of DON-induced genes in *Arabidopsis*

The genes identified in this objective have not been well studied in wheat. Since the gene expression networks have been relatively well-established in *Arabidopsis*, by predicting the functions of homologs or their similar genes, this will provide us with some clues about the possible roles of these genes in wheat. To further explore the possible functions of these genes, we blasted their coding nucleotide sequences against *Arabidopsis* protein sequence database

(BLASTx, https://www.arabidopsis.org/Blast/index.jsp, accessed in October 2016). The results (Table 9) suggested that the biological roles are different among these genes in *Arabidopsis*.

Table 9. The predicted gene functions in *Arabidopsis*.

Affymetrix probeset name	Contig Number in ORDC local database	Annotation from GeneChip	BLASTx	Amino acid identities	Annotation in *Arabidopsis*
TaAffx.1095.1.A1_at	13A_32559	WRKY2 protein [*Hordeum vulgare subsp. vulgare*]	AtWRKY40 (AT1G80840)	41% (62/149)	Pathogen-induced transcription factor. Defense response to bacterium and fungus; negative regulation of transcription; response to salicylic acid. Cell to cell mobile RNA
TaAffx.12215.1.A1_at	1S2237	AAA-type ATPase [*Oryza sativa* (japonica cultivar-group)]	P-loop containing nucleoside triphosphate hydrolases superfamily protein (AT3G50940)	63% (120/188)	Functions in nucleoside-triphosphatase activity, ATPase activity, nucleotide binding, ATP binding; response to salt stress.
TaAffx.12876.1.S1_at	13A_24426	Cys2/His2 zinc-finger protein [*Oryza sativa* (japonica cultivar-group)]	C_2H_2-type zinc finger family protein (AT3G46070)	45% (45/100)	Regulation of transcription; zinc ion and nucleic acid binding; sequence-specific DNA binding.
Ta.1221.1.A1_at	13B_16082	Aspartyl-tRNA synthetase [*Oryza sativa* (japonica cultivar-group)]	Class II aminoacyl-tRNA and biotin synthetases superfamily protein (AT4G31180)	73% (171/233)	Defense response to fungus, response to cadmium ion; functions in ATP, DNA binding, aspartate-tRNA ligase activity, and metal ion binding.
TaAffx.64766.1.S1_at	1Y2643	Glutathione transferase	Glutathione S-transferase family protein (AT3G62760)	45% (95/207)	Involved in glutathione metabolic process; response to toxic substance; toxic catabolic process
Ta.11647.2.S1_at	13B_12689	TF-like protein [*Oryza sativa* (japonica cultivar-group)]	AtNFXL1 (AT1G10170)	72% (288/465)	Function as a negative regulator of the trichothecene phytotoxin-induced defense response; defense response to bacterium; response to salt stress and salicylic acid biosynthetic process.
Ta.9831.1.S1_s_at	13B_4644	MRP-like ABC transporter [*Oryza sativa* (japonica cultivar-group)]	Multidrug resistance-associated protein 9 (AT3G60160)	59% (242/410)	ABC Transporter; ATP binding; ATPase activity.
Ta.5297.1.A1_at	8S72	unknown			unknown
Ta.22180.1.S1_x_at	wdi1c_pk005_a3 (EST number)	unknown			unknown

4. Discussion

The long-term goal of the research program which includes this project is to elucidate the molecular mechanisms contributing to resistance to FHB in wheat. It includes the characterization of some key genes involved in FHB resistance or susceptibility. Two genes in this work, *TaABCC6* and *TaNFXL1*, contribute to FHB susceptibility, while *TansLTP9.4* is related to FHB resistance. The CRISPR technology has become a promising tool for gene editing, especially in model organisms such as mice, human cells, zebrafish, *Arabidopsis* and rice. However, only a few investigations have been carried out in wheat (Wang et al., 2016; Zhang et al., 2016; Zhang et al., 2017). Since the generation and screening of transgenic plants are extremely time-consuming, many studies have suggested the testing of sgRNA in a protoplast system as a more efficient and an alternative method to evaluate CRISPR activity in plant cells. Therefore, before producing transgenic plants, we conducted experiments to test six sgRNAs both *in vitro* and in a wheat protoplast system. The three target genes were successfully edited in this project. The key findings and learning in this project are summarized below.

In the first objective, two sgRNAs were designed for each of three target genes. Unlike the other four sgRNAs designed to target *TansLTP9.4* and *TaNFXL1* genes, the two ABCC6-sgRNAs were located on the last (11^{th}) exon close to 3' UTR, which means some shorter, yet functional proteins could still possibly be translated. To further optimize the strategy of targeted gene editing, designing sgRNAs targeting conserved domains or motifs could be one of the future directions.

The efficiency of six sgRNAs to guide the Cas9 nuclease complex was shown *in vitro* to be high enough to cleave the templates into two smaller fragments with expected sizes. Due to the

simplicity of templates which only included targeted regions, and the lack of DSB repairing after cleavage, it was not possible to quantify editing efficiency or identify different modification types from the *in vitro* assay.

In the second objective, we firstly optimized a protocol for wheat protoplast isolation and transformation. Transformation efficiencies of about 60% were obtained with different batches of protoplasts, similar to the efficiencies obtained by Shan et al. (2014) for protoplasts from shoot tissues in 1~2 weeks old wheat cultivar Bobwhite seedlings. To quantify editing in protoplasts for the three target genes, we performed PCR amplicon sequencing using an Illumina Miseq system. The application of the NGS allows sequencing of a large number of PCR amplicons directly by avoiding the heavy workload for library preparation. The estimated editing efficiencies that we have obtained were close to the results from Wang et al. (2016) for the edition of four genes in wheat protoplasts; their estimated editing efficiencies ranged from 0% to 44.5% as evaluated by NGS. When a pcoCas9 was used, the estimated editing efficiency was more consistent than when using a crCas9, suggesting that the pcoCas9 nuclease might be more active or have better stability in plant cells. In addition, some features of the expression vector itself could also have an impact on Cas9 activity, including the transcription or translation efficiency. For example, crCas9 in the pCambia vector is driven by a 35S promoter, while pcoCas9 is under the control of a 35SPPDK promoter (constitutive 35S enhancer fused to the maize C4 pyruvate orthophosphate dikinase (C4PPDK) basal promoter). The different promoters used in the two vectors could affect Cas9 expression, leading to a different editing efficiency. The potato *ST-LS1* gene encodes a light inducible, leaf and stem specific protein; the second intron (IV2) of this gene was inserted into the pcoCas9 used in this project, which has been shown by Li et al. (2015) to minimize the adverse effects on bacteria growth of unwanted

expression in *E. coli* during vector cloning. The incorporation of the IV2 intron in a GFP gene also showed an improvement of 1.4-fold on fluoresce intensity in corn and tobacco protoplasts, leading to a better expression efficiency in plants (Pang et al., 1996). A more detailed discussion about additional options to further improve CRISPR efficiency in wheat is presented in the next section.

The estimated editing frequencies of the five resequenced NFXL1 samples differed significantly from the results in the first round of sequencing. The size of PCR amplicons for the first round of NFXL1 sequencing was 754 bp, which exceeded the optimal length for the MiSeq sequencing system (2x250 bp). In contrast, a new primer pair was designed and used in the second round of sequencing with an expected product size of 310 bp, leading to a better sequencing coverage and quality.

The estimated editing efficiencies obtained in this project were close between the two sgRNAs on *TaABCC6* and *TansLTP9.4* target regions, but higher for NFXL1-sgRNA1 than NFXL1-sgRNA2. In addition, the editing efficiencies between the 3 genes were of the same order. This was not fully expected as only one mismatch was identified in Fielder, between ABC-sgRNA1 and its target site, while no mismatch was identified between the other sgRNAs and their respective target sites based on our Sanger sequencing results. According to Anderson et al. (2015), a lower editing frequency was to be expected for the sgRNAs presenting one mismatch to their target sequence. This underscores the importance for the protoplast testing of sgRNAs as predictions of efficiencies are still imperfect.

Still based on Anderson et al. (2015), and assuming that similar mismatches to sgRNAs are present between the Chinese Spring and Fielder genomic sequences for the newly identified

genes, the off-target sites with two or three mismatches (ABC-sgRNA1 on 2A, copy 2 and NFXL1-sgRNA1 on 7D, copy 2) would not be expected to be edited while a lower editing frequency would be expected for the sgRNAs having one mismatch to the off-target sequence (ABC-sgRNA1 on 2D, copy 2; ABC-sgRNA2 on 2A, copy 2). These findings suggested that designing highly specific sgRNAs and off-target analyses in wheat will be difficult until a complete reference sequence is available.

Recently, the use of whole genome sequencing to detect off-target effect was proposed by Schaefer et al. (2017). With this approach, 117 indels and 1397 single nucleotide variants were identified unexpectedly in two CRISPR modified mice compared with a non-edited control. However, considering the potential sequencing error, and the sheer size of the wheat genome, this method is still unpractical in wheat.

Based on the two rounds of amplicon sequencing results, we observed different types of mutations including deletion, insertion and replacement. However, in our analyses, deletions on either target site or between the two sgRNAs on the same gene fragment were more frequent than the other two modification types. In addition, for each gene, some deletions were observed much more frequently than others. In our experiments, only the CRISPR vector DNA was available as a donor DNA template to induce insertion-repairing; it is possible that more insertions would have been observed if a donor DNA template with more extensive homology to the target region had been present. More experiments will be needed to confirm that.

Not all the indels identified in this project could induce frameshift mutation in the predicted resulting protein. Among the indels observed, there were 5/10 for ABCC6, 6/10 for nsLTP, 11/14 for resequenced NFXL1 samples and 13/18 for pcoCas9 edited NFXL1 samples that are

predicted to cause a frameshift mutation. Identifying individual plants with frameshift mutations in the edited gene is important for studies aiming at studying the effect of an inactivated gene. The size of the majority of indels found in this project was longer than what has been reported for other CRISPR experiments in wheat. For example, an enhanced disease resistance 1 gene (*TaEDR1*) was successfully edited in all three homoeolog genes with the deletions ranging from 1 bp to 34 bp (Zhang et al. 2017). Zhang et al. (2016) also reported that deletions between 1 bp and 37 bp were identified when targeting a wheat *TaGASR7*, a gene controlling grain length and weight. In contrast, most of the indels identified in this project were over 50 bp. The difference in indel length between our experiments and those cited above may be associated with the different targeting strategy that we used. In our experiments, the two sgRNAs were targeting a limited region of each gene and were 47 to 95 bp apart; in published experiments in wheat, only one sgRNA was used to target each gene (Zhang et al, 2016, 2017). We cannot exclude the possibility that the close proximity of our targeting sites produced interference or interaction between sgRNA cutting and DSBs repairing mechanisms.

In the last objective, we profiled the expression pattern of nine DON-induced genes in a DON-treated protoplast system; the *TaPR4* gene, which can be induced by FHB and not directly by DON treatment, was used as a control. A ten times lower DON concentration (10 ppm compared to 100 ppm used in wheat heads) was sufficient to induce gene expression in protoplasts when compared to wheat heads, possibly due to the difference in exposure to treatment between individual cells and compact tissues. Among all the tested genes, six of them shared a similar expression pattern as in wheat heads, with up-regulation of over 100 folds under DON stress, while another three genes showed no or only a modest up-regulation in the treated protoplasts. These three genes expressed differentially in cell-based and tissue-based experiments. In addition,

a variation in candidate gene expression level was observed among the three biological replicates. This may be related to the complexity of protoplast preparation steps, as the three biological replicates were prepared from protoplasts isolated at different times from different wheat plants.

One of the long-term goals of the research program is to study the expression networks modulated by DON-induced regulatory genes following their edition by CRISPR in transformed protoplasts. This requires two prerequisites: to have high and consistent editing efficiency between different samples, and having the genes of interest inducible by direct DON treatment. The findings in this work support the second prerequisite and are important for future studies.

The functional annotation of these genes suggests that they might be involved in different biological process in response to DON stress. Homologs or closely related genes to our DON-induced genes have been studied by others with different types of stressors. The key findings are discussed in the following paragraphs, focusing on results in wheat and *Arabidopsis*.

TaNFXL1 was up-regulated by DON treatment in protoplast culture and peaked at 6 hours after DON treatment with a fold change over 100 times. As described in Subsection 1.4.2, this gene was also induced in DON treated wheat heads in two susceptible cultivars, Roblin and Chinese Spring. *AtNFXL1* (*AT1G10170*) is the homolog of *TaNFXL1* in *Arabidopsis*. The expression of *AtNFXL1* was up-regulated by T-2 toxin treatment in 7-day-old *Arabidopsis* seedlings; T-2 toxin is a trichothecene mycotoxin closely related to DON (Masuda et al., 2007). This suggests that at least some aspects of the response to trichothecene compounds are conserved between these two plant species.

In contrast to *TaNFXL1*, *TaWRKY2* was significantly up-regulated by DON treatment only in wheat heads and with a moderate up-regulation in protoplasts. The genes *AtWRKY40*

(*AT1G80840*) and *HvWRKY2* are the most similar to the wheat *TaWRKY2* in *Arabidopsis* and barley, respectively. In barley, the mildew resistance proteins A (MLA) confer resistance to the powdery mildew fungus, *Blumeria graminis*. Using VIGS to lower the expression of *HvWRKY2* in barley seedlings, a lower percentage of germinated *B. graminis* was observed when compared with the seedlings transformed with empty vectors, indicating that *HvWRKY2* functions as a suppressor of defense to *B. graminis*. Yeast two-hybrid assays further showed that *HvWRKY2* interacts with the MLA coiled-coil (CC) domain. This suggests that *HvWRKY2* modulates defense resistance to powdery mildew negatively by interfering with MLA proteins (Shen et al., 2007). In *Atwrky40* and *Atwrky18* double *Arabidopsis* mutants and *Atwrky40*, *Atwrky18*, *Atwrky60* triple mutants, an enhanced resistance to *Pseudomonas syringae* was observed; the same mutants also showed a higher susceptibility to *Botrytis cinerea*. However, little difference was observed in *Atwrky40* single mutants, suggesting that *AtWRKY40* and the other two *WRKY* genes might interact physically or have functional redundancy for their role in defense towards pathogen infection (Xu et al., 2006). In a separate line of investigation aiming at identifying mobile RNAs, *AtWRKY40* mRNA was identified as able to move from root to shoot and vice versa. To determine that, the scientists took advantage of the high frequency of single nucleotide polymorphisms between two *Arabidopsis* ecotypes, Columbia and Pedriza, and carried out a reciprocal hetero-grafting experiment (Thieme et al., 2015). Based on RNA sequencing data analyses, 2006 genes were found to generate mobile RNAs, including At*WRKY40* mRNA. This suggests that this gene is involved in intercellular activities between different tissues. In that context, it is possible that the wheat *TaWRKY2* is expressed differentially in DON-treated wheat heads and the protoplast system.

Using Northern blot analysis, the expression level of *AtWRKY40* was highly up-regulated in 4-week-old *Arabidopsis* plants sprayed with 2 mM salicylic acid (SA), compared with the plants without SA treatment (Dong et al., 2003). In SA-treated transgenic *Arabidopsis* carrying a *β-glucuronidase* (*GUS*) gene driven by an AtNFXL1 promoter, a 40-fold increase in GUS activity was observed when compared with the control without SA treatment, indicating that the expression of *AtNFXL1* was also inducible by SA treatment (Asano et al., 2008). In *F. graminearum* infected wheat spikes, a higher SA accumulation was observed in the FHB-resistant cultivar Sumai 3, in comparison to the SA level in mock-inoculated spikes (Makandar et al., 2012). They also found that in transgenic wheat cultivar Bobwhite overexpressing a bacterial *NahG* gene encoding a salicylate hydroxylase, a lower SA concentration was observed, associated with higher FHB disease severity. It has also been shown that the addition of SA to agar plates strongly inhibited *F. graminearum* mycelial growth, and that the co-inoculation of *F. graminearum* spores and SA on wheat heads resulted in a reduction of disease symptoms in both resistant and susceptible wheat cultivars (Qi et al., 2012). Taken together, SA is shown to have a negative effect on *F. graminearum* growth and FHB disease development in infected wheat heads. *AtNFXL1* and *AtWRKY2* likely function as components in SA-dependent signal pathways.

As *TaWRKY2*, the gene coding a Cys2/His2 zinc finger protein was upregulated by DON treatment only in wheat heads, not in protoplast culture. The *Arabidopsis* gene coding a C_2H_2-type zinc finger protein (*AT3G46070*) shares the highest identity with the wheat C_2H_2-type zinc finger protein identified in this project. *AT3G46070* was up-regulated under H_2O_2 treatment, which represented oxidative stress conditions (Stanley Kim et al., 2004). As *TaNFXL1*, the AAA-type ATPase that we have profiled in this project was upregulated by DON in wheat heads and protoplasts. The *Arabidopsis AT3G50940* gene, which shares the highest identity with our

wheat AAA-type ATPase, was up-regulated under salt stress, but exclusively in root (Ma et al., 2006).

The aspartyl-tRNA synthase and the glutathione S-transferase (GST) genes that we have identified are also upregulated by DON in both tissue types. The general functions of those two genes and their gene families have been well documented. By microarray analyses, the expression of aspartyl-tRNA synthetase (*AT4G31180*) was up-regulated when *Arabidopsis* was infected by the cabbage leaf curl virus (Ascencio-Ibanez et al., 2008). GSTs are believed to play a role in the response to abiotic stress and detoxification. They are also involved in phase II detoxification process in plants, which conjugates xenobiotics and glutathione. The products of phase II detoxification are often nontoxic or less toxic compared with the original compounds (Xu et al., 2005). By overexpressing a tomato GST in *Arabidopsis*, an enhanced tolerance to salt and drought stress was observed (Xu et al., 2015b). The *Arabidopsis* GST (*AT3G62760*) is most similar to the wheat GST analyzed in this project but has not been characterized yet. Of the two last genes that we have further characterized, Ta.5297.1.A1_at and Ta.22180.1.S1_x_at, the first one was up-regulated by DON in both tissues while the later was up-regulated only in the wheat heads. In *Arabidopsis*, no homologs or similar genes were identified for these two wheat genes.

Taken together, the DON-induced genes that we have identified are involved in different biological processes; in *Arabidopsis*, their homologs or closely related genes have been shown to be up-regulated under stress conditions. However, the specific roles of those genes in wheat in the response to DON stress or *F. graminearum* infection are still unclear. Therefore, investigating their possible functions in order to have a better understanding of regulatory network affected by FHB and DON will be a part of future directions.

4.1. Considerations in optimizing CRISPR for future directions

CRISPR system has become a powerful tool for targeted gene editing within only three years, both in prokaryote and eukaryote organisms. One of the major focuses with this technology is to optimize the efficiency and specificity of this system, including the design of sgRNAs with predicted highly efficiency, the utilization of different Cas9 nucleases and delivering systems to transfer the sgRNA-Cas9 complex.

4.1.1. The optimization of sgRNA design

As discussed in Subsection 1.1.1, different software has been developed to design sgRNAs with predicted high efficiency although their predictions might be based on different algorithms. Some studies suggested that sequence features could affect sgRNA efficiency. Wang et al. (2014a) found that in human cell lines transformed with a human-codon optimized spCas9, sgRNAs with low (<20%) or high (>80%) GC content were less effective against their targets. By screening 5183 sgRNAs from two distinct published datasets in human (Wang et al., 2014a) and mouse (Koike-Yusa et al., 2013) cell lines, Xu et al. (2015a) found that guanines at 19^{th} and 20^{th} positions (adjacent to PAM) were strongly preferred in functional sgRNAs, which is also reported by Doench et al. (2014). Xu et al. (2015a) also reported that cytosine was preferred at 18^{th} position on functional sgRNAs. This is the position where the cleavage happens and it might contribute to cutting efficiency, sgRNA-Cas9 binding or repairing after cleavage. By measuring the activities of 1280 sgRNAs targeting 128 zebrafish genes and with a zebrafish codon optimized spCas9, sgRNAs with high activities were often guanine-rich while with fewer adenines, which do help to stabilize sgRNA structure, leading to a higher editing efficiency and more types of mutations (Moreno-Mateos et al., 2015). Compared with the most widely used 20

nt sgRNAs, the utilisation of truncated sgRNAs with complementarity lengths of 17 to 19 nt was shown to enormously decrease mutations by 5000-fold or more at off-target sites while not disturbing on-target activities in human cell lines both with a spCas9 and a pair of Cas9 variants, Cas9-D10A nickases (Fu et al., 2014).

The sgRNAs selected in this project were with the highest scores among the candidates from the "sgRNA designer" program, which indicated they were predicted to have high activity. However, at the time this project was started, only a few programs were available to design sgRNAs and to predict their efficiency and the off-target effects. Following the recent findings discussed above in this subsection, more programs have been developed such as CRISPRscan (Moreno-Mateos et al., 2015), and CRISPOR (Haeussler et al., 2016), which may provide more comprehensive and reliable results for the designing of high-efficient sgRNAs.

Still, some non-functional sgRNAs cannot be predicted only by sequence features. Other factors such as secondary structure, sgRNA transcription efficiency, positions on target sequences or even the conditions of targeted organisms (e.g., cell cultures, the different growth stage of individuals) could also have an impact on CRISPR/Cas9 efficiency and such factors have to be experimentally tested.

4.1.2. The diversity of Cas9 nucleases

As in most published literature, the Cas9 nuclease used in this project originated from *Streptococcus pyogenes* (spCas9); that Cas9 recognizes a NGG PAM structure. Recently, more types of Cas9 nucleases from different organisms have been identified. For example, Bell et al. (2016) reported that two modified Cas9 from *Caenorhabditis elegans* can recognize NGA and NGCG PAM structures, instead of NGG. By modifying PAMs directly without changing sgRNA

sequence, the Cas9 with NGA PAM was found to be as efficient as the NGG PAM Cas9. It was also revealed that a different Cas9 from *Staphylococcus aureus* can induce cleavage with similar efficiency to spCas9 at target sites with a NNGRRT (R stands for A or G) PAM structure (Ran et al., 2015). With these Cas9 orthologues being able to recognize different PAMs, the design of sgRNAs will not be limited to the NGG PAM structure, allowing for a wider selection of sgRNAs, which might lead to higher editing efficiencies.

4.1.3. The delivery system of CRISPR/Cas9 complex

Building an expression construct containing sgRNA and Cas9 sequences followed by transformation has become a conventional method for creating targeted gene editing in different organisms using CRISPR. In this project, six gBlocks targeting different loci on three gene fragments were synthesized and each gBlock cost over 150 CAD. It could be quite expensive if multiplex genome editing was needed. Lowder et al. (2015) created a toolbox which allows researchers to assemble CRISPR/Cas9 constructs quickly and efficiently. Compared with the gBlocks used in this project, the only changeable element in this toolbox is a 20 nt sgRNA sequence, which can be replaced easily by restriction enzyme digestion and ligation using T4 DNA ligase. Using Gateway cloning methods, different sgRNAs with their own promoters were firstly cloned into a series of entry vectors and multi sgRNA assets were assembled into one recipient vector. Finally, a Cas9 gene, a selection marker and sgRNA assets were cloned into one expression construct using Gateway recombination. Gene editing with this construct showed mutation frequencies ranging from 42% to 67% in rice protoplasts, indicating its high activity.

The tRNA processing system precisely cleaves at both ends of tRNA precursors after transcription. By attaching tRNA spacers between different sgRNAs (tRNA-sgRNA1-tRNA-

sgRNA2-…), a construct was made under the control of a single promoter which could target multiple genome sites *in vivo* with high capacity and editing efficiency. With this system, Xie et al. (2015) successfully edited two different loci on *MPK5* gene simultaneously in transgenic rice plants, with efficiency up to 100%. Using the tRNA processing system, gene editing at multiplex target sites can be carried out simultaneously by avoiding the synthesis of expensive gBlocks.

Instead of the construction of plasmids encoding Cas9 and sgRNAs, pre-assembly of Cas9 protein-sgRNA ribonucleoproteins (RNPs) has become a novel way to induce mutations in different organisms. By delivering RNPs directly into protoplasts of *Arabidopsis*, tobacco, lettuce and rice, the editing efficiencies in plants regenerated from these protoplasts can be up to 46% (Woo et al., 2015). The wheat *TaGW2* gene plays a role in grain weight control and has three homoeologs on chromosomes A, B and D. Using the DNA-free approach, Liang et al. (2017) successfully edited this gene in wheat protoplasts with mutagenesis frequencies of 33.4% and 21.8% for genome B and D, respectively and the off-target efficiency was also 5 times lower than the conventional plasmid transformation. Such targeted gene editing using a DNA-free method could bypass the concerns and the regulations about genetically modified plants.

Building on the recent findings with the use of the CRISPR system discussed above, more efforts on optimizing this system will need to be carried out, including the optimization of sgRNA design, the utilization of different types of Cas9 nucleases and delivery systems. In addition, with the recent availability of a fully assembled draft of the wheat genomic sequence (IWGSC Reference Sequence v1.0), it is expected that the design of sgRNAs with high efficiency and the prediction of off-target sites will become easier.

4.2. Conclusion

In conclusion, the major effort of this project aimed at establishing and initiating optimization of the CRISPR/Cas9 system in wheat protoplasts. Two sgRNAs were designed to target each of the genes *TaABCC6*, *TansLTP9.4* and *TaNFXL1*. All three genes were successfully edited both *in vitro* and in a wheat protoplast system. The next step is to produce transgenic plants edited in each of the three genes. These genes are closely related to FHB resistance or susceptibility. In transgenic plants, a higher resistance to FHB should be observed in *TaABCC6* or *TaNFXL1* edited plants, while *TansLTP9.4* edited transgenic plants are expected to be more susceptible to FHB infection.

References

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC146917/pdf/253389.pdf

Anderson, E. M., Haupt, A., Schiel, J. A., Chou, E., Machado, H. B., Strezoska, Ž., … Smith, A. V. B. (2015). Systematic analysis of CRISPR-Cas9 mismatch tolerance reveals low levels of off-target activity. *Journal of Biotechnology*, *211*, 56–65. https://doi.org/10.1016/j.jbiotec.2015.06.427

Asano, T., Masuda, D., Yasuda, M., Nakashita, H., Kudo, T., Kimura, M., … Nishiuchi, T. (2008). *AtNFXL1*, an *Arabidopsis* homologue of the human transcription factor NF-X1, functions as a negative regulator of the trichothecene phytotoxin-induced defense response. *Plant Journal*, *53*(3), 450–464. https://doi.org/10.1111/j.1365-313X.2007.03353.x

Ascencio-Ibanez, J. T., Sozzani, R., Lee, T.-J., Chu, T.-M., Wolfinger, R. D., Cella, R., & Hanley-Bowdoin, L. (2008). Global Analysis of *Arabidopsis* Gene Expression Uncovers a Complex Array of Changes Impacting Pathogen Response and Cell Cycle during Geminivirus Infection. *Plant Physiology*, *148*(1), 436–454. https://doi.org/10.1104/pp.108.121038

Balcerzak, M., Gulden, S., Ouellet, T. (2009). Expression profiling of trichothecene induced genes in wheat. Ottawa, ON: Poster presented at 6th Canadian Workshop on Fusarium Head Blight.

Balcerzak, M., Leung, W., & Ouellet, T. (2016). *Two disease susceptibility genes for FHB in wheat*. Ottawa, ON: Proceedings of the 8th Canadian Workshop on Fusarium Head Blight.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., … Horvath, P. (2007). CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science*, *315*(5819), 1709–1712.

Bell, R. T., Fu, B. X. H., & Fire, A. Z. (2016). Cas9 variants expand the target repertoire in *Caenorhabditis elegans*. *Genetics*, *202*(2), 381–388. https://doi.org/10.1534/genetics.115.185041

Bhati, K.K., Sharma, S., Aggarwal, S., Kaur, M., Shukla, JV., Kaur, J., Mantri, S. & Pandey, A.K. (2015). Genome-wide identification and expression characterization of ABCC-MRP transporters in hexaploid wheat. *Frontiers in Plant science, 6*, Article 488.

Bhaya, D., Davison, M., & Barrangou, R. (2011). CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation. *Annual Review of Genetics*, *45*(1), 273–297. https://doi.org/10.1146/annurev-genet-110410-132430

Borrill, P., Adamski, N., & Uauy, C. (2015). Genomics as the key to unlocking the polyploid potential of wheat. *New Phytologist*, *208*(4), 1008–1022. https://doi.org/10.1111/nph.13533

Brown, D. W., McCormick, S. P., Alexander, N. J., Proctor, R. H., & Desjardins, a E. (2001). A genetic and biochemical approach to study trichothecene diversity in *Fusarium sporotrichioides* and *Fusarium graminearum*. *Fungal Genetics and Biology : FG & B*, *32*(2), 121–133. https://doi.org/10.1006/fgbi.2001.1256

Brown, N. A., Urban, M., van de Meene, A. M. L., & Hammond-Kosack, K. E. (2010). The infection biology of *Fusarium graminearum*: Defining the pathways of spikelet to spikelet colonisation in wheat ears. *Fungal Biology*, *114*(7), 555–571. https://doi.org/10.1016/j.funbio.2010.04.006

Carlson, D. F., Fahrenkrug, S. C., & Hackett, P. B. (2012). Targeting DNA With Fingers and TALENs. *Molecular Therapy - Nucleic Acids*, *1*(1), e3. https://doi.org/10.1038/mtna.2011.5

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., ... & Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*, *339*(6121), 819–823. https://doi.org/10.1126/science.1231143

Daley, J. M., Palmbos, P. L., Wu, D., & Wilson, T. E. (2005). Nonhomologous end joining in yeast. *Annual Review of Genetics*, *39*, 431–451. https://doi.org/10.1146/annurev.genet.39.073003.113340

Davidson, A. L., Dassa, E., Orelle, C., & Chen, J. (2008). Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiology and Molecular Biology Reviews : MMBR*, *72*(2), 317–364. https://doi.org/10.1128/MMBR.00031-07

Ding, L., Xu, H., Yi, H., Yang, L., Kong, Z., Zhang, L., … Ma, Z. (2011). Resistance to hemi-biotrophic *F. graminearum* infection is associated with coordinated and ordered expression of diverse defense signaling pathways. *PLoS ONE*, *6*(4). https://doi.org/10.1371/journal.pone.0019008

Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., Hegde, M., Smith, I., … Root, D. E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. *Nature Biotechnology*, *32*(12), 1262–1267. https://doi.org/10.1038/nbt.3026

Dong, J., Chen, C., & Chen, Z. (2003). Expression profiles of the *Arabidopsis* WRKY gene superfamily during plant defense response. *Plant Molecular Biology*, *51*(1), 21–37. https://doi.org/10.1023/A:1020780022549

Dong, X. (2004). NPR1, all things considered. *Current Opinion in Plant Biology*, *7*(5), 547–552. https://doi.org/10.1016/j.pbi.2004.07.005

Feng, C., Yuan, J., Wang, R., Liu, Y., Birchler, J. A., & Han, F. (2016). Efficient Targeted Genome Modification in Maize Using CRISPR/Cas9 System. *Journal of Genetics and Genomics*, *43*(1), 37–43. https://doi.org/10.1016/j.jgg.2015.10.002

Feng, Z., Zhang, B., Ding, W., Liu, X., Yang, D.-L., Wei, P., … Zhu, J.-K. (2013). Efficient genome editing in plants using a CRISPR/Cas system. *Cell Research*, *23*(10), 1229–1232. https://doi.org/10.1038/cr.2013.114

Foroud, N. A., Ouellet, T., Laroche, A., Oosterveen, B., Jordan, M. C., Ellis, B. E., & Eudes, F. (2012). Differential transcriptome analyses of three wheat genotypes reveal different host response pathways associated with Fusarium head blight and trichothecene resistance. *Plant Pathology*, *61*(2), 296–314. https://doi.org/10.1111/j.1365-3059.2011.02512.x

Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M., & Joung, J. K. (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature Biotechnology*, *32*(3), 279–284. https://doi.org/10.1038/nbt.2808

Gaj, T., Gersbach, C. A., & Barbas, C. F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology*, *31*(7), 397–405.

https://doi.org/10.1016/j.tibtech.2013.04.004

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, *6*(5), 343–345. https://doi.org/10.1038/nmeth.1318

Golkari, S., Gilbert, J., Ban, T., & Procunier, J. D. (2009). QTL-specific microarray gene expression analysis of wheat resistance to Fusarium head blight in Sumai-3 and two susceptible NILs. *Genome*, *52*(5), 409–418. https://doi.org/10.1139/g09-018

Green, R. E., Cornell, S. J., Scharlemann, J. P. W., & Balmford, A. (2005). Farming and the fate of wild nature. *Science*, *307*(5709), 550–555. https://doi.org/10.1126/science.1106049

Güldener, U., Seong, K. Y., Boddu, J., Cho, S., Trail, F., Xu, J. R., … Kistler, H. C. (2006). Development of a *Fusarium graminearum* Affymetrix GeneChip for profiling fungal gene expression *in vitro* and *in planta*. *Fungal Genetics and Biology*, *43*(5), 316–325. https://doi.org/10.1016/j.fgb.2006.01.005

Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J.-B., … Concordet, J.-P. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biology*, *17*(1), 148. https://doi.org/10.1186/s13059-016-1012-2

Harris, L. J., Balcerzak, M., Johnston, A., Schneiderman, D., & Ouellet, T. (2016). Host-preferential *Fusarium graminearum* gene expression during infection of wheat, barley, and maize. *Fungal Biology*, *120*(1), 111–123. https://doi.org/10.1016/j.funbio.2015.10.010

Hattori, J., Ouellet, T., & Tinker, N. A. (2005). Wheat EST sequence assembly facilitates comparison of gene contents among plant species and discovery of novel genes. *Genome*, *48*(2), 197–206. https://doi.org/10.1139/g04-106

International Wheat Genome Sequencing Consortium. (2014). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science*, *345*(6194), 1250088. https://doi.org/10.1126/science.1251788

Jia, Z., Gou, J., Sun, Y., Yuan, L., Tang, Q., Yang, X., … Luo, K. (2010). Enhanced resistance to

fungal pathogens in transgenic *Populus tomentosa Carr.* by overexpression of an nsLTP-like antimicrobial protein gene from motherwort (*Leonurus japonicus*). *Tree Physiology*, *30*(12), 1599–1605. https://doi.org/10.1093/treephys/tpq093

Jiang, W., Bikard, D., Cox, D., Zhang, F., & Marraffini, L. a. (2013a). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology*, *31*(3), 233–239. https://doi.org/10.1038/nbt.2508

Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B., & Weeks, D. P. (2013b). Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in *Arabidopsis*, tobacco, sorghum and rice. *Nucleic Acids Research*, *41*(20), 1–12. https://doi.org/10.1093/nar/gkt780

Jones, P. M., & George, A. M. (2004). The ABC transporter structure and mechanism: Perspectives on recent research. *Cellular and Molecular Life Sciences*, *61*(6), 682–699. https://doi.org/10.1007/s00018-003-3336-9

Kader, J.-C. (1996). Lipid-Transfer Proteins in Plants. *Annual Review of Plant Physiology and Plant Molecular Biology*, *47*(1), 627–654. https://doi.org/10.1146/annurev.arplant.47.1.627

Kang, J., Park, J., Choi, H., Burla, B., Kretzschmar, T., Lee, Y., & Martinoia, E. (2011). Plant ABC Transporters. *The Arabidopsis Book / American Society of Plant Biologists*, *9*, e0153. https://doi.org/10.1199/tab.0153

Karin, M. (1990). Too many transcription factors: positive and negative interactions. *The New Biologist*, *2*(2), 126–31. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2128034

Kazan, K., & Gardiner, D. M. (2017). Transcriptomics of cereal ‐ *Fusarium graminearum* interactions: What we have learned so far. *Molecular Plant Pathology*. DOI: 10.1111/mpp.12561

Koike-Yusa, H., Li, Y., Tan, E.-P., Velasco-Herrera, M. D. C., & Yusa, K. (2013). Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nature Biotechnology*, *32*(3), 267–273. https://doi.org/10.1038/nbt.2800

Koornneef, A., & Pieterse, C. M. J. (2008). Cross Talk in Defense Signaling. *Plant Physiology*, *146*(3), 839–844. https://doi.org/10.1104/pp.107.112029

Krattinger, S. G., Lagudah, E. S., Spielmeyer, W., Singh, R. P., Huerta-espino, J., Mcfadden, H., … Keller, B. (2009). A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. *Science*, *323*(5919), 1360–1363. https://doi.org/10.1126/science.1166453

Kugler, K. G., Siegwart, G., Nussbaumer, T., Ametz, C., Spannagl, M., Steiner, B., … Schweiger, W. (2013). Quantitative trait loci-dependent analysis of a gene co-expression network associated with Fusarium head blight resistance in bread wheat (*Triticum aestivum L.*). *BMC Genomics*, *14*(1), 728. https://doi.org/10.1186/1471-2164-14-728

Kunz, J., Loeschmann, A., Deuter-Reinhard, M., & Hall, M. N. (2000). FAP1, a homologue of human transcription factor NF-X1, competes with rapamycin for binding to FKBP12 in yeast. *Molecular Microbiology*, *37*(6), 1480–1493. https://doi.org/10.1046/j.1365-2958.2000.02105.x

Latchman, D. S. (1997). Transcription factors: An overview. *The International Journal of Biochemistry & Cell Biology*, *29*(12), 1305–1312. https://doi.org/10.1016/S1357-2725(97)00085-X

Li, J.-F., Zhang, D., & Sheen, J. (2015). Targeted Plant Genome Editing via the CRISPR/Cas9 Technology *In Plant Functional Genomics: Methods and Protocols: Second Edition* (pp. 239–255). https://doi.org/10.1007/978-1-4939-2444-8_12

Liang, Z., Chen, K., Li, T., Zhang, Y., Wang, Y., Zhao, Q., … Gao, C. (2017). Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. *Nature Publishing Group*. https://doi.org/10.1038/ncomms14261

Lin, Y., Cradick, T. J., Brown, M. T., Deshmukh, H., Ranjan, P., Sarode, N., ... & Bao, G. (2014). CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic acids research*, gku402. https://doi.org/10.1093/nar/gku402

Lisso, J., Altmann, T., & Müssig, C. (2006). The *AtNFXL1* gene encodes a NF-X1 type zinc finger protein required for growth under salt stress. *FEBS Letters*, *580*(20), 4851–4856. https://doi.org/10.1016/j.febslet.2006.07.079

Lisso, J., Schröder, F., Fisahn, J., & Müssig, C. (2011). NFX1-LIKE2 (NFXL2) suppresses abscisic acid accumulation and stomatal closure in *Arabidopsis thaliana*. *PLoS ONE*, *6*(11). https://doi.org/10.1371/journal.pone.0026982

Lisso, J., Schröder, F., Schippers, J. H. M. M., & Müssig, C. (2012). NFXL2 modifies cuticle properties in *Arabidopsis*. *Plant Signaling & Behavior*, *7*(5), 551–555. https://doi.org/10.4161/psb.19838

Liu, F., Zhang, X., Lu, C., Zeng, X., Li, Y., Fu, D., & Wu, G. (2015). Non-specific lipid transfer proteins in plants: presenting new advances and an integrated functional analysis. *Journal of experimental botany*, 66(19), 5663-5681. https://doi.org/10.1093/jxb/erv313

Lowder, L. G., Zhang, D., Baltes, N. J., Paul, J. W., Tang, X., Zheng, X., … Qi, Y. (2015). A CRISPR/Cas9 Toolbox for Multiplexed Plant Genome Editing and Transcriptional Regulation. *Plant Physiology*, *169*(2), 971–985. https://doi.org/10.1104/pp.15.00636

Ma, S., Gong, Q., & Bohnert, H. J. (2006). Dissecting salt stress pathways. In *Journal of Experimental Botany* (Vol. 57, pp. 1097–1107). https://doi.org/10.1093/jxb/erj098

Mackintosh, C. A., Lewis, J., Radmer, L. E., Shin, S., Heinen, S. J., Smith, L. A., … Muehlbauer, G. J. (2007). Overexpression of defense response genes in transgenic wheat enhances resistance to Fusarium head blight. *Plant Cell Reports*, *26*(4), 479–488. https://doi.org/10.1007/s00299-006-0265-8

Mahfouz, M. M., Li, L., Shamimuzzaman, M., Wibowo, A., Fang, X., & Zhu, J.-K. (2011). De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(6), 2623–2628. https://doi.org/10.1073/pnas.1019533108

Makandar, R., Nalam, V. J., Lee, H., Trick, H. N., Dong, Y., & Shah, J. (2012). Salicylic Acid Regulates Basal Resistance to Fusarium Head Blight in Wheat. *Molecular Plant-Microbe Interactions*, *25*(3), 431–439. https://doi.org/10.1094/MPMI-09-11-0232

Makandar, R., Essig, J. S., Schapaugh, M. a, Trick, H. N., & Shah, J. (2006). Genetically

engineered resistance to Fusarium head blight in wheat by expression of *Arabidopsis* NPR1. *Molecular Plant-Microbe Interactions : MPMI*, *19*(2), 123–129. https://doi.org/10.1094/MPMI-19-0123

Martinoia, E., Grill, E., Tommasini, R., Kreuz, K., & Amrhein, N. (1993). ATP-dependent glutathione S-conjugate "export" pump in the vacuolar membrane of plants. *Nature*, *364*(6434), 247–249. https://doi.org/10.1038/364247a0

Masuda, D., Ishida, M., Yamaguchi, K., Yamaguchi, I., Kimura, M., & Nishiuchi, T. (2007). Phytotoxic effects of trichothecenes on the growth and morphology of *Arabidopsis thaliana*. *Journal of Experimental Botany*, *58*(7), 1617–1626. https://doi.org/10.1093/jxb/erl298

McLaughlin, J. E., Bin-Umer, M. A., Widiez, T., Finn, D., McCormick, S., & Tumer, N. E. (2015). A lipid transfer protein increases the glutathione content and enhances *Arabidopsis* resistance to a trichothecene mycotoxin. *PloS one*, 10(6), e0130204. https://doi.org/10.1371/journal.pone.0130204

McMullen, M., Bergstrom, G., De Wolf, E., Dill-macky, R., Hershman, D., Shaner, G., & Van Sanford, D. (2012). A Unified Effort to Fight an Enemy of Wheat and Barley: Fusarium Head Blight. *Plant Disease*, *96*(12). https://doi.org/http://dx.doi.org/10.1094/PDIS-03-12-0291-FE

Miao, J., Guo, D., Zhang, J., Huang, Q., Qin, G., Zhang, X., … Qu, L.-J. (2013). Targeted mutagenesis in rice using CRISPR-Cas system. *Cell Research*, *23*(10), 1233. https://doi.org/10.1038/cr.2013.123

Moons, A. (2003). *Ospdr9*, which encodes a PDR-type ABC transporter, is induced by heavy metals, hypoxic stress and redox perturbations in rice roots. *FEBS Letters*, *553*(3), 370–376. https://doi.org/10.1016/S0014-5793(03)01060-3

Moreno-Mateos, M. A., Vejnar, C. E., Beaudoin, J.-D., Fernandez, J. P., Mis, E. K., Khokha, M. K., & Giraldez, A. J. (2015). CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting *in vivo*. *Nature Methods*, *12*(10), 982–988. https://doi.org/10.1038/nmeth.3543

Nemudryi, A. A., Valetdinova, K. R., Medvedev, S. P., & Zakian, S. M. (2014). TALEN and CRISPR/Cas genome editing systems: Tools of discovery. *Acta Naturae*, *6*(22), 19–40.

https://doi.org/10.1007/s11103-014-0188-7

Niu, C. F., Wei, W., Zhou, Q. Y., Tian, A. G., Hao, Y. J., Zhang, W. K., … Chen, S. Y. (2012). Wheat WRKY genes *TaWRKY2* and *TaWRKY19* regulate abiotic stress tolerance in transgenic *Arabidopsis* plants. *Plant, Cell and Environment*, *35*(6), 1156–1170. https://doi.org/10.1111/j.1365-3040.2012.02480.x

Ouellet, T., Balcerzak, M., Leung, W., Hattori, J., Martin, T., Gulden, S. 2012. The mycotoxin deoxynivalenol induces genes that increase wheat susceptibility to fusarium head blight in wheat. Proceedings of the 10th International Plant Molecular Biology Conference, Jeju, Korea, October 21-26, 2012. P. 331.

Ouellet, T., Balcerzak, M., Rocheleau, H., Wang, L., Wojcik, P., Dzwinel, W. 2013. Comparison of Fusarium head blight-resistant and -susceptible wheat using global expression profiling. Proceedings of the 7th International Triticeae Symposium, Chengdu, China June 9-13 2013. P 37.

Pang, S. Z., DeBoer, D. L., Wan, Y., Ye, G., Layton, J. G., Neher, M. K., … Fromm, M. E. (1996). An improved green fluorescent protein gene as a vital marker in plants. *Plant Physiology*, *112*(3), 893–900. https://doi.org/10.1104/pp.112.3.893

Pestka, J. J. (2007). Deoxynivalenol: toxicity, mechanisms and animal health risks. *Animal feed science and technology*, 137(3), 283-298. https://doi.org/10.1016/j.anifeedsci.2007.06.006

Petersen, G., Seberg, O., Yde, M., & Berthelsen, K. (2006). Phylogenetic relationships of Triticum and Aegilops and evidence for the origin of the A, B, and D genomes of common wheat (*Triticum aestivum*). *Molecular Phylogenetics and Evolution*, *39*(1), 70–82. https://doi.org/10.1016/j.ympev.2006.01.023

Qi, P. F., Johnston, A., Balcerzak, M., Rocheleau, H., Harris, L. J., Long, X. Y., … Ouellet, T. (2012). Effect of salicylic acid on *Fusarium graminearum*, the major causal agent of fusarium head blight in wheat. *Fungal Biology*, *116*(3), 413–426. https://doi.org/10.1016/j.funbio.2012.01.001

Ran, F. A., Cong, L., Yan, W. X., Scott, D. A., Gootenberg, J. S., Kriz, A. J., … Zhang, F.

(2015). *In vivo* genome editing using *Staphylococcus aureus* Cas9. *Nature*, *520*(7546), 186–191. https://doi.org/10.1038/nature14299

Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., … Zhang, F. (2013a). Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. *Cell*, *154*(6), 1380–1389. https://doi.org/10.1016/j.cell.2013.08.021

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013b). Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, *8*(11), 2281–2308. https://doi.org/10.1038/nprot.2013.143

Rawat, N., Pumphrey, M. O., Liu, S., Zhang, X., Tiwari, V. K., Ando, K., … Gill, B. S. (2016). Wheat *Fhb1* encodes a chimeric lectin with agglutinin domains and a pore-forming toxin-like domain conferring resistance to Fusarium head blight. *Nature Genetics*, *48*(12), 1576–1580. https://doi.org/10.1038/ng.3706

Rice, A. J., Park, A., & Pinkett, H. W. (2014). Diversity in ABC transporters: Type I, II and III importers. *Critical Reviews in Biochemistry and Molecular Biology*, *49*(5), 426–437. https://doi.org/10.3109/10409238.2014.953626

Rueckert, D. G., & Schmidt, K. (1990). Lipid transfer proteins. *Chemistry and Physics of Lipids, 1990, Vol.56(1), pp.1-20*, *56*(1), 1–20. https://doi.org/10.1016/0009-3084(90)90083-4

Sakamoto, H., Maruyama, K., Sakuma, Y., Meshi, T., & Iwabuchi, M. (2004). *Arabidopsis* Cys_2 / His_2-Type Zinc-Finger Proteins Function as Transcription Repressors under Drought. *Society*, *136* (September), 2734–2746. https://doi.org/10.1104/pp.104.046599.2734

Sander, J. D., & Joung, J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature Biotechnology*, *32*(4), 347–355. https://doi.org/10.1038/nbt.2842

Schaefer, K. A., Wu, W. H., Colgan, D. F., Tsang, S. H., Bassuk, A. G., & Mahajan, V. B. (2017). Unexpected mutations after CRISPR-Cas9 editing *in vivo*. *Nature methods*, 14(6), 547-548. doi:10.1038/nmeth.4293

Schmidt-Heydt, M., Parra, R., Geisen, R., & Magan, N. (2011). Modelling the relationship between environmental factors, transcriptional genes and deoxynivalenol mycotoxin

production by strains of two Fusarium species. *Journal of The Royal Society Interface*, *8*(54), 117–126. https://doi.org/10.1098/rsif.2010.0131

Schmittgen, T. D., & Livak, K. J. (2008). Analyzing real-time PCR data by the comparative CT method. *Nature Protocols*, *3*(6), 1101–1108. https://doi.org/10.1038/nprot.2008.73

Schroeder, H. W., & Christensen, J. J. (1963). Factors affecting resistance of Wheat to scab caused by *Gibberella zeae*. *Phytopathology*, *53*(7, 1), 831–838.

Schweiger, W., Steiner, B., Ametz, C., Siegwart, G., Wiesenberger, G., Berthiller, F., … Buerstmayr, H. (2013). Transcriptomic characterization of two major Fusarium resistance quantitative trait loci (QTLs), *Fhb1* and *Qfhs.ifa-5A*, identifies novel candidate genes. *Molecular Plant Pathology*, *14*(8), 772–785. https://doi.org/10.1111/mpp.12048

Shalem, O., Sanjana, E. N., Hartenian, E., & Zhang, F. (2014). Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science*, *343*(6166), 84–88. https://doi.org/10.1038/nbt.2647

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., ... & Gao, C. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. *Nature Biotechnology*, *31*(8), 686–688. https://doi.org/10.1038/nbt.2650

Shan, Q., Wang, Y., Li, J., & Gao, C. (2014). Genome editing in rice and wheat using the CRISPR/Cas system. *Nature Protocols*, *9*(10), 2395–2410. https://doi.org/10.1038/nprot.2014.157

Shang, Y., Xiao, J., Ma, L., Wang, H., Qi, Z., Chen, P., … Wang, X. (2009). Characterization of a PDR type ABC transporter gene from wheat (*Triticum aestivum L.*). *Chinese Science Bulletin*, *54*(18), 3249–3257. https://doi.org/10.1007/s11434-009-0553-0

Shen, Q. H., Saijo, Y., Mauch, S., Biskup, C., Bieri, S., Keller, B., … Schulze-Lefert, P. (2007). Nuclear activity of MLA immune receptors links isolate-specific and basal disease-resistance responses. *Science*, *315*(5815), 1098–1103. https://doi.org/10.1126/science.1136372

Sobrova, P., Adam, V., Vasatkova, A., Beklova, M., Zeman, L., & Kizek, R. (2010). Deoxynivalenol and its toxicity. *Interdisciplinary Toxicology*, *3*(3), 94–99.

https://doi.org/10.2478/v10102-010-0019-x

Song, Z., Krishna, S., Thanos, D., Strominger, J. L., & Ono, S. J. (1994). A Novel Cysteine-rich Sequence-specific DNA-binding Protein Interacts with the Conserved X-box Motif of the Human Major Histocompatibility Complex Class II Genes Via a Repeated Cys-His Domain and Functions as a Transcriptional Repressor. *Journal of Experimental Medicine*, *180*(5), 1763–1774.

Song, W.-Y., Yamaki, T., Yamaji, N., Ko, D., Jung, K.-H., Fujii-Kashino, M., … Ma, J. F. (2014). A rice ABC transporter, *OsABCC1*, reduces arsenic accumulation in the grain. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(44), 15699–15704. https://doi.org/10.1073/pnas.1414968111

Stack, R. W. (2000). Return of an old problem: Fusarium head blight of small grains. *Plant Health Progress*. DOI:10.1094/PHP-2000-0622-01-RV

Stanley Kim, H., Yu, Y., Snesrud, E. C., Moy, L. P., Linford, L. D., Haas, B. J., … Quackenbush, J. (2004). Transcriptional divergence of the duplicated oxidative stress-responsive genes in the *Arabidopsis* genome. *The Plant Journal*, *41*(2), 212–220. https://doi.org/10.1111/j.1365-313X.2004.02295.x

Stephens, A. E., Gardiner, D. M., White, R. G., Munn, A. L., & Manners, J. M. (2008). Phases of infection and gene expression of *Fusarium graminearum* during crown rot disease of wheat. *Molecular Plant-Microbe Interactions : MPMI*, *21*(12), 1571–1581. https://doi.org/10.1094/MPMI-21-12-1571

Sun, J.-Y., Gaudet, D. a, Lu, Z.-X., Frick, M., Puchalski, B., & Laroche, a. (2008). Characterization and antifungal properties of wheat nonspecific lipid transfer proteins. *Molecular Plant-Microbe Interactions*, *21*(3), 346–360. https://doi.org/10.1094/MPMI-21-3-0346

Sun, Y., Xiao, J., Jia, X., Ke, P., He, L., Cao, A., … Wang, X. (2016). The role of wheat jasmonic acid and ethylene pathways in response to *Fusarium graminearum* infection. *Plant Growth Regulation*, *80*(1), 69–77. https://doi.org/10.1007/s10725-016-0147-1

ter Beek, J., Guskov, A., & Slotboom, D. J. (2014). Structural diversity of ABC transporters. *The Journal of General Physiology*, *143*(4), 419–35. https://doi.org/10.1085/jgp.201411164

The Daily — Production of principal field crops, July 2016. Retrieved February 3, 2017, from http://www.statcan.gc.ca/daily-quotidien/160823/dq160823a-eng.htm

Thieme, C. J., Rojas-Triana, M., Stecyk, E., Schudoma, C., Zhang, W., Yang, L., … Kragler, F. (2015). Endogenous *Arabidopsis* messenger RNAs transported to distant tissues. *Nature Plants*, *1*(4), 15025. https://doi.org/10.1038/nplants.2015.25

Thoma, S., Kaneko, Y., & Somerville, C. (1993). A non-specific lipid transfer protein from *Arabidopsis* is a cell wall protein. *Plant Journal*, *3*(3), 427–436. https://doi.org/10.1046/j.1365-313X.1993.t01-25-00999.x

Upadhyay, S. K., Kumar, J., Alok, a., & Tuli, R. (2013). RNA Guided Genome Editing for Target Gene Mutations in Wheat. *G3: Genes|Genomes|Genetics*, *3*(December), 2233–2238. https://doi.org/10.1534/g3.113.008847

Wang, H. W., Hwang, S. G., Karuppanapandian, T., Liu, A., Kim, W., & Jang, C. S. (2012). Insight into the Molecular Evolution of Non-Specific Lipid Transfer Proteins via Comparative Analysis Between Rice and Sorghum. *DNA Research*, *19*, 179–194. https://doi.org/10.1093/dnares/dss003

Wang, T., Wei, J. J., Sabatini, D. M., & Lander, E. S. (2014a). Genetic Screens in Human Cells Using the CRISPR-Cas9 System. *Science*, *343*(6166), 80–84. https://doi.org/10.1126/science.1246981

Wang, W., Akhunova, A., Chao, S., & Akhunov, E. (2016). Optimizing multiplex CRISPR / Cas9-based genome editing for wheat. *bioRxiv*, 51342.

Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C., & Qiu, J.-L. (2014b). Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nature Biotechnology*, *32*(9), 947–951. https://doi.org/10.1038/nbt.2969

Wiedenheft, B., Sternberg, S. H., & Doudna, J. a. (2012). RNA-guided genetic silencing systems in bacteria and archaea. *Nature*, *482*(7385), 331–338. https://doi.org/10.1038/nature10886

Woo, J. W., Kim, J., Kwon, S. Il, Corvalán, C., Cho, S. W., Kim, H., … Kim, J.-S. (2015). DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nature Biotechnology*, *33*(11), 1162–1164. https://doi.org/10.1038/nbt.3389

Wu, X., Kriz, A. J., & Sharp, P. A. (2014). Target specificity of the CRISPR-Cas9 system. *Quantitative biology*, 2(2), 59-70. doi:10.1007/s40484-014-0030-x

Xie, K., Minkenberg, B., & Yang, Y. (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proceedings of the National Academy of Sciences*, *112*(11), 3570–3575. https://doi.org/10.1073/pnas.1420294112

Xing, H.-L., Dong, L., Wang, Z.-P., Zhang, H.-Y., Han, C.-Y., Liu, B., … Chen, Q.-J. (2014). A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biology*, *14*(1), 327. https://doi.org/10.1186/s12870-014-0327-y

Xu, C., Li, C. Y. T., & Kong, A. N. T. (2005). Induction of phase I, II and III drug metabolism/transport by xenobiotics. *Archives of pharmacal research*, 28(3), 249. https://doi.org/10.1007/BF02977789

Xu, H., Xiao, T., Chen, C.-H., Li, W., Meyer, C. A., Wu, Q., … Liu, X. S. (2015a). Sequence determinants of improved CRISPR sgRNA design. *Genome Research*, *25*(8), 1147–1157. https://doi.org/10.1101/gr.191452.115

Xu, J., Xing, X.-J., Tian, Y.-S., Peng, R.-H., Xue, Y., Zhao, W., & Yao, Q.-H. (2015b). Transgenic *Arabidopsis* Plants Expressing Tomato Glutathione S-Transferase Showed Enhanced Resistance to Salt and Drought Stress. *PLOS ONE*, *10*(9), e0136960. https://doi.org/10.1371/journal.pone.0136960

Xu, X., Chen, C., Fan, B., & Chen, Z. (2006). Physical and Functional Interactions between Pathogen-Induced *Arabidopsis WRKY18*, *WRKY40*, and *WRKY60* Transcription Factors. *THE PLANT CELL ONLINE*, *18*(5), 1310–1326. https://doi.org/10.1105/tpc.105.037523

Yoo, S.-D., Cho, Y.-H., & Sheen, J. (2007). *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nature Protocols*, *2*(7), 1565–1572. https://doi.org/10.1038/nprot.2007.199

Zhang, X.-W., Jia, L.-J., Zhang, Y., Jiang, G., Li, X., Zhang, D., & Tang, W.-H. (2012). In planta stage-specific fungal gene profiling elucidates the molecular strategies of *Fusarium graminearum* growing inside wheat coleoptiles. *The Plant Cell*, *24*(12), 5159–5176. https://doi.org/10.1105/tpc.112.105957

Zhang, Y., Bai, Y., Wu, G., Zou, S., Chen, Y., Gao, C., & Tang, D. (2017). Simultaneous modification of three homoeologs of *TaEDR1* by genome editing enhances powdery mildew resistance in wheat. *The Plant Journal*. DOI: 10.1111/tpj.13599

Zhang, Y., Liang, Z., Zong, Y., Wang, Y., Liu, J., Chen, K., ... & Gao, C. (2016). Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. *Nature communications*, 7, 12617. DOI: 10.1038/ncomms12617

Zhou, H., Liu, B., Weeks, D. P., Spalding, M. H., & Yang, B. (2014). Large chromosomal deletions and heritable small genetic changes induced by CRISPR/Cas9 in rice. *Nucleic Acids Research*, *42*(17), 10903–10914. https://doi.org/10.1093/nar/gku806

Zhu, X., Li, Z., Xu, H., Zhou, M., Du, L., & Zhang, Z. (2012). Overexpression of wheat lipid transfer protein gene *TaLTP5* increases resistances to *Cochliobolus sativus* and *Fusarium graminearum* in transgenic wheat. *Functional & integrative genomics*, 12(3), 481-488. DOI 10.1007/s10142-012-0286-z

Appendices

Appendix 1. PCR reactions using PfuTurbo Cx Hotstart DNA Polymerase (A), Taq DNA Polymerase (B), CloneAmp HiFi PCR Premix (C) and KOD Hot Start DNA Polymerase (D). The Taq DNA Polymerase was exclusively used for colony PCR.

A

Ingredients	Volume
PfuTurbo Cx Hotstart DNA Polymerase (2.5 U/µl)	0.5 µl
DNA template (~50 ng/µl)	1 µl
Forward primer (10 µM)	1.25 µl
Reverse primer (10 µM)	1.25 µl
dNTP (10 mM each)	0.5 µl
10x PCR buffer	2.5 µl
H_2O	18 µl

B

Ingredients	Volume
Taq DNA Polymerase (5 U/µl)	0.125 µl
Cell suspension	5 µl
Forward primer (10 µM)	1.25 µl
Reverse primer (10 µM)	1.25 µl
dNTP (10 mM each)	0.5 µl
10x PCR buffer	2.5 µl
$MgCl_2$ (25 mM)	1.5 µl
H_2O	12.875 µl

C

Ingredients	Volume
CloneAmp HiFi PCR Premix	12.5 µl
Forward primer (10 µM)	1 µl
Reverse primer (10 µM)	1 µl
DNA template (~50 ng/µl)	1.5 µl
H_2O	9 µl

D

Ingredients	Volume
KOD Hot Start DNA Polymerase (5 U/µl)	1 µl
DNA template (~50 ng/µl)	2 µl
10x PCR buffer	5 µl
$MgSO_4$ (25 mM)	3 µl
dNTP (2 mM each)	5 µl
Forward primer (20 µM)	0.75 µl
Reverse primer (20 µM)	0.75 µl
H_2O	32.5 µl

Appendix 2. Reactions for sgRNA test *in vitro*.

Ingredients	Volume
Experimental Cleavage Template (100 ng)	1-6 µl
Experimental sgRNA (50 ng)	1-6 µl
Guide-it Recombinant Cas9 Nuclease	1 µl
10× Cas9 Reaction Buffer	1 µl
10× BSA	1 µl
RNase-free water	0-5 µl
Total	10 µl

Appendix 3. Primers used in this project. **A:** sgRNA-specific primers used for colony PCR and sequencing; **B**: Primers used to amplify target regions of *TaLTP9.4* and *TaNFXL1* for the first round of PCR amplicon sequencing; **C**: Twelve degenerate fusion primers and four sequencing primers without degenerate bases for PCR amplification in the second round of sequencing. Gene-specific primers are marked with red letters. Gene-specific primers were designed by Xiucheng Cui while adapter sequences (in black) were provided by Julie Chapados (ORDC, AAFC); **D:** Primers for RT-qPCR analysis. The primer pairs were designed by Margaret Balcerzak (ORDC, AAFC); **E:** Primer pairs used for PCR amplification of three gene fragments in the *in vitro* test.

A

Primer name	Sequence (5'-3')
ABC-sgRNA-F	CACGCCGTCGAGATTACTGG
ABC-sgRNA-R	CTTGGATCTCCGTGAGTACT
nsLTP-sgRNA-F	GCCGTGCGTGGCGTACGTGA
nsLTP-sgRNA-R	CTGCACGCCGGAGCAGCACT
NFXL1-sgRNA-F	TGACTGGCACAACGCAAGGT
NFXL1-sgRNA-R	TGCGGCACACCAACTCCATC

B

Primer name	Sequence (5'-3')
nsLTP-F77-P2	CACACTCATCTGATCACCCATAG
nsLTP-R526-P2	TCAGGAGACCGAAAGGTTAGA
NFXL1-F314-P1	CAGATGGTGGAGTGCTACAA
NFXL1-R1069-P1	GTAGTGATACGGCCACAAGAA

C

Name	Sequence 5` to 3`
Illu_Rd1_0N_ABC-F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGTCTTCAGCGAAGGTAAGC
Illu_Rd1_1N_ABC-F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNTGTCTTCAGCGAAGGTAAGC
Illu_Rd1_2N_ABC-F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNTGTCTTCAGCGAAGGTAAGC
Illu_Rd1_3N_ABC-F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNTGTCTTCAGCGAAGGTAAGC
Illu_Rd1_0N_ABC-R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGTGTGTGCGGTCTTTCT
Illu_Rd1_1N_ABC-R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNGGTGTGTGCGGTCTTTCT
Illu_Rd1_2N_ABC-R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNGGTGTGTGCGGTCTTTCT
Illu_Rd1_3N_ABC-R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNGGTGTGTGCGGTCTTTCT
Illu_Rd1_0N_NFXL-F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTTTCCACAGGGCTGGTT
Illu_Rd1_1N_NFXL-F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNCTTTCCACAGGGCTGGTT
Illu_Rd1_2N_NFXL-F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNCTTTCCACAGGGCTGGTT
Illu_Rd1_3N_NFXL-F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNCTTTCCACAGGGCTGGTT
Illu_Rd1_0N_NFXL-R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGGATGCCAGCATCTCTG
Illu_Rd1_1N_NFXL-R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNTGGATGCCAGCATCTCTG
Illu_Rd1_2N_NFXL-R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNTGGATGCCAGCATCTCTG
Illu_Rd1_3N_NFXL-R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNTGGATGCCAGCATCTCTG

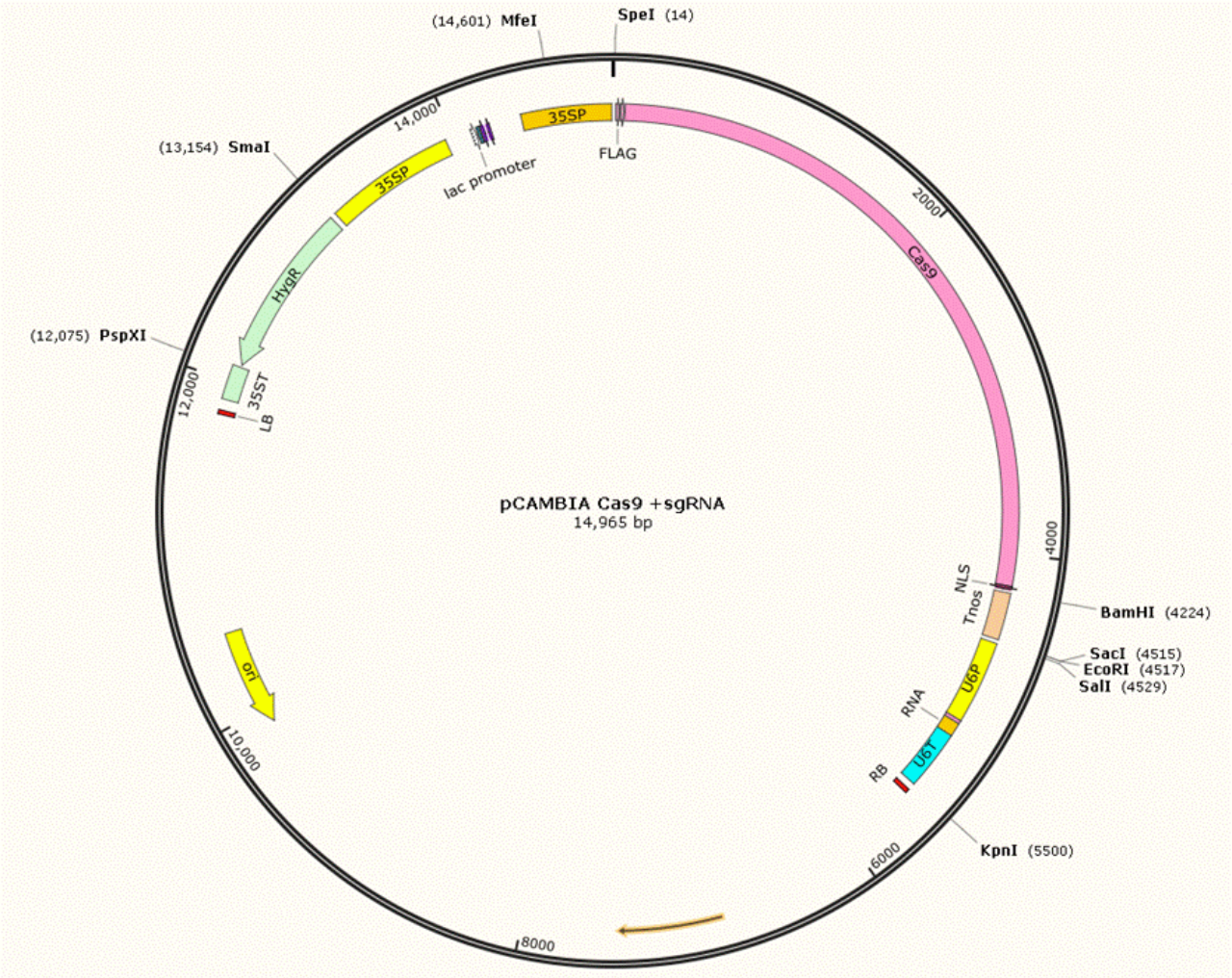
D

Housekeeping Genes	Primer names	Sequence (5'-3')
GAPDH	GAPDH-1F3	AACTGTTCATGCCATCACTGCCAC
	GAPDH-1R1	AGGACATACCAGTGAGCTTGCCAT
AAOx	AAOx-F4	CACAGCAGGATTTAAGCTCTGG
	AAOx-R4	GGGATGGACTAATTTCACAGGC
Genes of Interest	**Primer names**	**Sequence (5'-3')**
NFXL1	TFq-P2-F1089	GTGGAGCCTACAATGGTGATA
	TFq-P2-R1168	TGATTCCACTGGTTGGAGAG
ABCC6	ABC-P6-F1459-qPCR	AAGCACTTTTCCGGATTGTTG
	ABC-P6-R1683-qPCR	GCTTCTTTGGGCTTTGACG
WRKY2	WRKY-qPCR2-F	TCTCGGTAATTAATCAGCAGG
	WRKY-qPCR2-R	TGAGGAGCCCAATAAGCAAACTAC
PR4	PR4-1F1	ACACCGTCTTCACCAAGATCGACA
	PR4-1R1	AGCATGGATCAGTCTCAGTGCTCA
Cys/His2 zinc finger protein	1Q2347-F548P2.1	CATTGCGAGTCGAAGAACAG
	1Q2347-R671P2	CTTCAGGTTTCTTCTGAGCC
AAA-type ATPase	1s2237-F458P1.1	CCTCTATATCTGGATAGGTGGC
	1s2237-R609P1	CAACTACAAGGAACGTCTCG
Aspartyl-tRNA synthetase (Ta.1221)	Ta.1221_F507_P2	TAGCAAGGCATGGTGTAGAA
	Ta.1221_R579_P2	GGTCAGCTTGTTAGGGAGAA
Ta.5297	Ta.5297_F111_P2	GCAGGCTGCTGCTAGAATA
	Ta.5297_R208_P2	GACCAAGAGCCCACCTTT
Ta.22180	Ta.22180_F123_P1	GCAGGATCTCATGACGATAGA
	Ta.22180_R176_P1	CTCAAATACACCTATGCTGGAC
Glutathione transferase (TaAffx.64766)	TaAffx.64766_F173_P2	CGAGCACCGATGTGTATATAGT
	TaAffx.64766_R287_P2	GAAGAAGGTGTTGGATGTGTATG

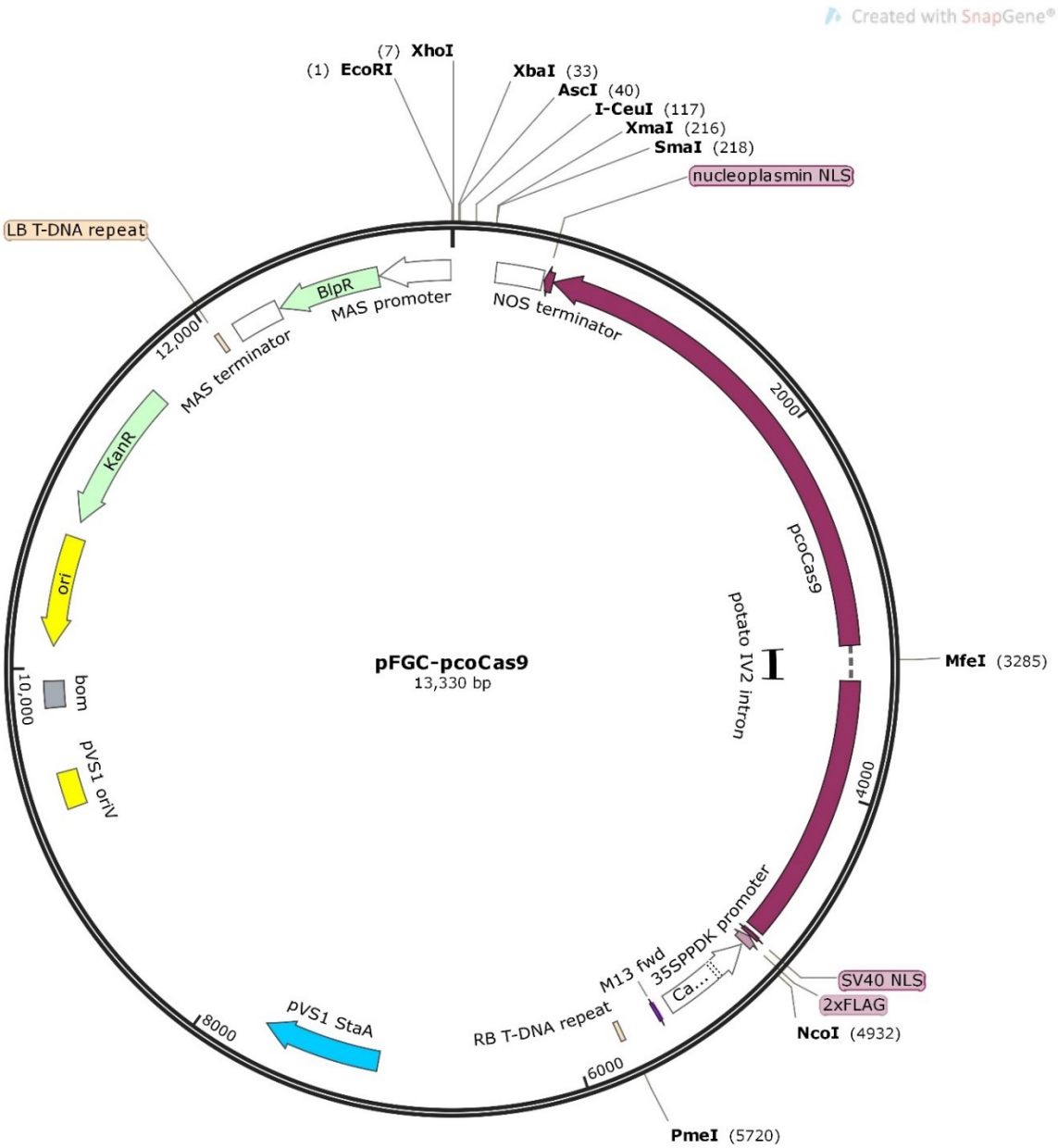
E

Primer name	Sequence (5'-3')	Amplicon size
ABC2-F3398-P2	CGGACAGAGGCAGTTGTTT	511
ABC2-R3913-P2	GAGAAAGGTCGTTGTGGTGT	
nsLTP-F77-P2	CACACTCATCTGATCACCCATAG	449
nsLTP-R526-P2	TCAGGAGACCGAAAGGTTAGA	
NFXL1-F314-P1	CAGATGGTGGAGTGCTACAA	754
NFXL1-R1069-P1	GTAGTGATACGGCCACAAGAA	

Appendix 4. The map of modified pCambia 1302 vector. The crCas9 nuclease is in pink. The sequence between KpnI and EcoRI sites was removed and replaced by assembled gBlock pairs. Reproduced from Jiang et al. (2013b).



Appendix 5. The map of pFGC-RCS vector with plant codon optimized spCas9 (in purple, Addgene plasmid # 52256). Restriction enzymes EcoRI and XmaI in multiple cloning site sequence were used for the cloning of the assembled gBlock pair editing the *TaNFXL1* gene. Modified from Li et al. (2015) using SnapGene and the plasmid sequence was retrieved from Addgene website.



Appendix 6. Required solutions for protoplast isolation and transformation.

W5 Solution (300 mL)

Compounds	Stock	Working Concentration	Final Volume
NaCl	5 M	154 mM	9.24 mL
$CaCl_2$	1 M	125 mM	37.5 mL
MES, pH=5.8	100 mM	2 mM	6 mL
KCl	0.5 M	5 mM	3 mL
Sterile H_2O			To 300 mL

Digestion Media (25 mL, freshly prepared)

Compounds	Stock	Working Concentration	Final Volume
Mannitol	1 M	0.6 M	15 mL
$CaCl_2$	1 M	10 mM	0.25 mL
MES (pH 5.8)	100 mM	20 mM	5 mL
KCl	0.5 M	10 mM	0.5 mL
Sterile H_2O			To 25 mL

Add fresh: 0.1% BSA (w/v, 0.025 g), 1.5% cellulase R10 (0.375 g) and 0.75% macerozyme R10 (0.188 g)

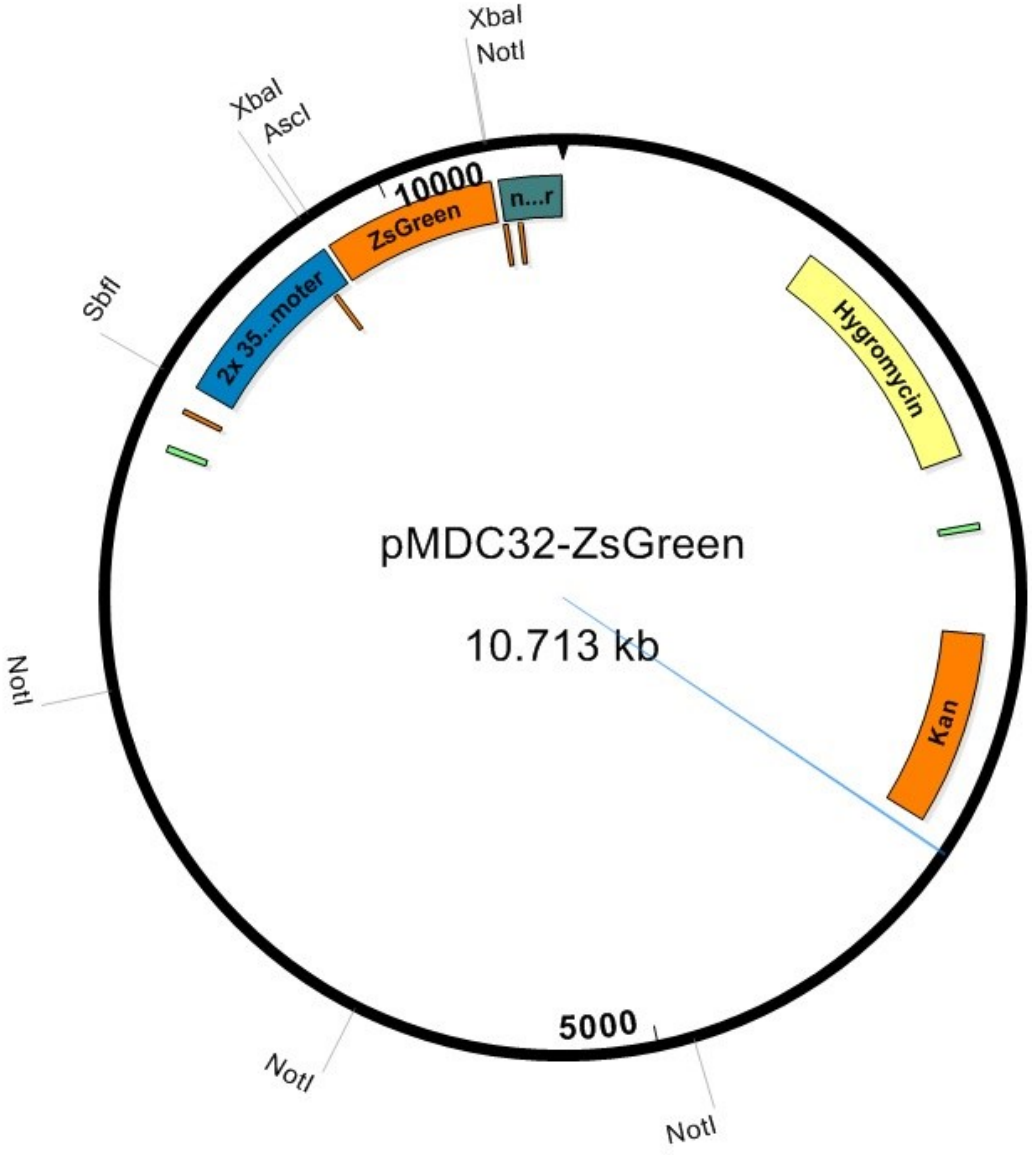
MMG Solution (100 mL)

Compounds	Stock	Working Concentration	Final Volume
Mannitol	1 M	0.4 M	40 mL
$MgCl_2$	1 M	15 mM	1.5 mL
MES pH 5.8	100 mM	4 mM	4 mL
Sterile H_2O			To 100 mL

PEG (1 mL)

Compounds	Stock	Working Concentration	Final Volume
Polyethylene glycol (PEG)	MW=4000	40% (w/v)	0.4 g
Mannitol	1 M	200 mM	0.2 mL
$CaCl_2$	1 M	100 mM	0.1 mL
Sterile H2O			To 1 mL

Appendix 7. The map of the pMDC32-ZsGreen vector (provided by Natalie Labbé and Johann Schernthaner, ORDC).



Appendix 8. Protoplast and transformation batches used for the preparation of PCR amplicons for the first round (A) and the second round (B) of sequencing. In the second round of sequencing, the five samples of NFXL1-1~5 from the first round were also reamplified from genomic DNA and resequenced. For each sample, two individual transformations were pooled together for DNA extraction; samples on the same preparation date were from the same batch of protoplasts.

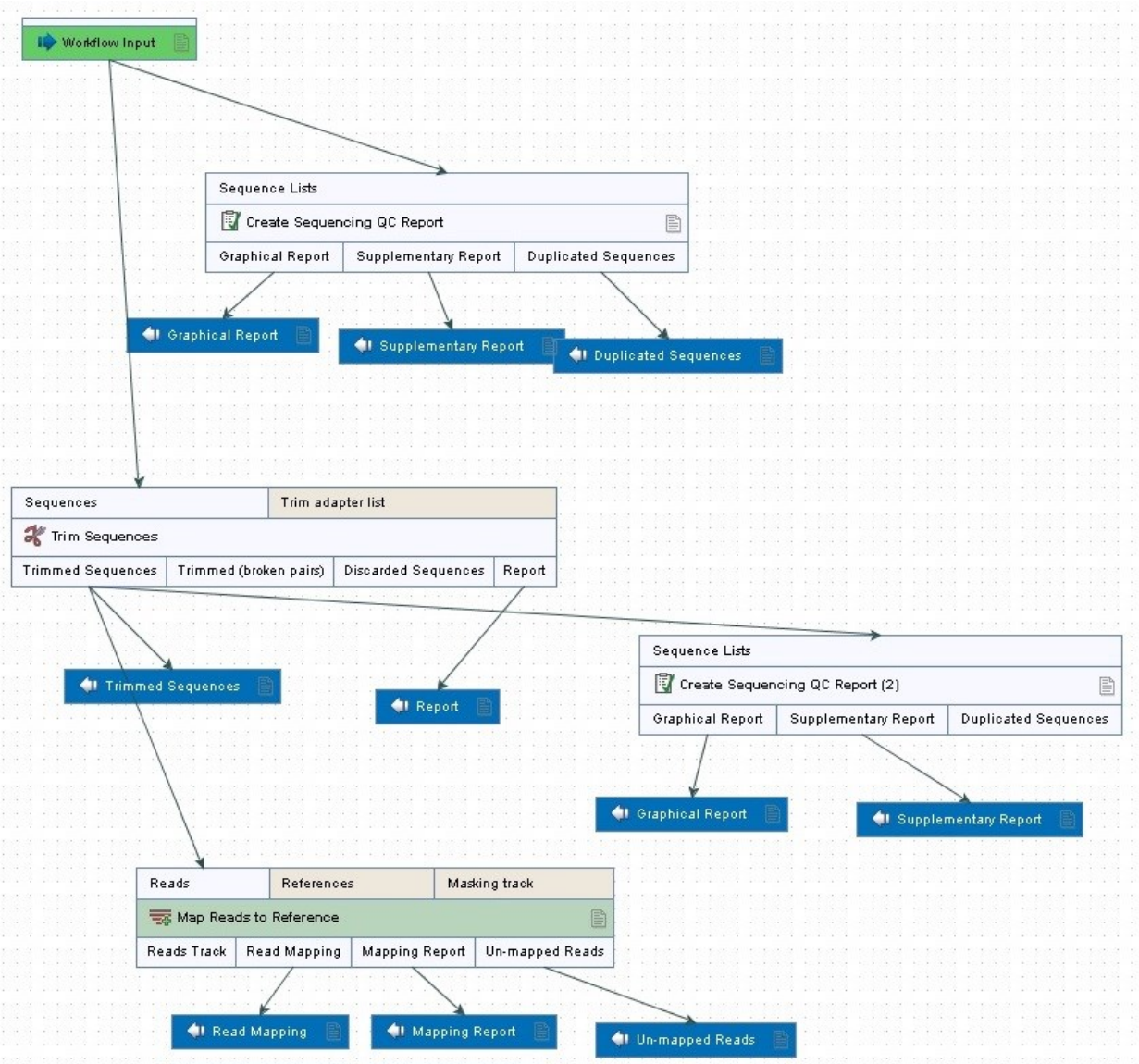
A

Preparation date	Prepared DNA sample name	
04-May-16	nsLTP-1,-2	NFXL1-1,-2
18-May-16	nsLTP-3,-4	NFXL1-3,-4
23-May-16	nsLTP-5	NFXL1-5

B

Preparation date	Prepared DNA sample name	
07-Dec-16	ABC-1,-2	pcoNFXL1-1,-2
12-Dec-16	ABC-3,-4	pcoNFXL1-3,-4
15-Dec-16	ABC-5,-control	pcoNFXL1-5, -control

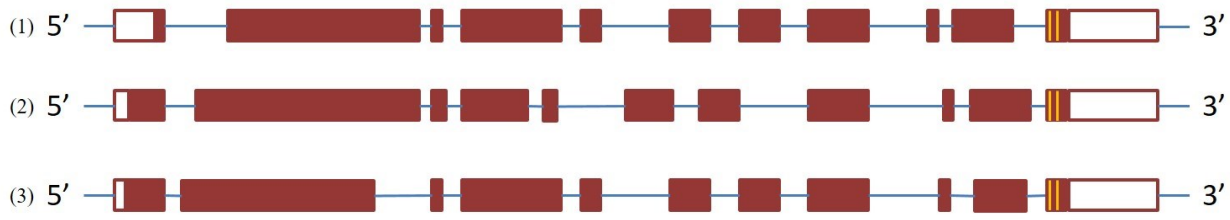
Appendix 9. A workflow created on CLC Genomics Workbench to analyze sequencing results. The first QC report was for the quality control of raw data and the QC report (2) was used for quality control of sequences after trimming.



Appendix 10. Structure of the three genes to be edited. A~C: The structure of *TaABCC6*, *TansLTP9.4* and *TaNFXL1*, respectively. Each figure includes the transcript ID from each chromosome, the respective transcript length, the number of amino acids (# aa), and the number of exons (# exons) for each transcript. The solid red rectangular boxes represent coding regions; the white-filled red boxes indicate 5' or 3' untranslated regions (UTRs); the introns are marked with blue lines while the sgRNAs are presented with yellow vertical lines. *TansLTP9.4* has four splicing transcript variants on chromosome 5AL and two on 5DL, respectively.

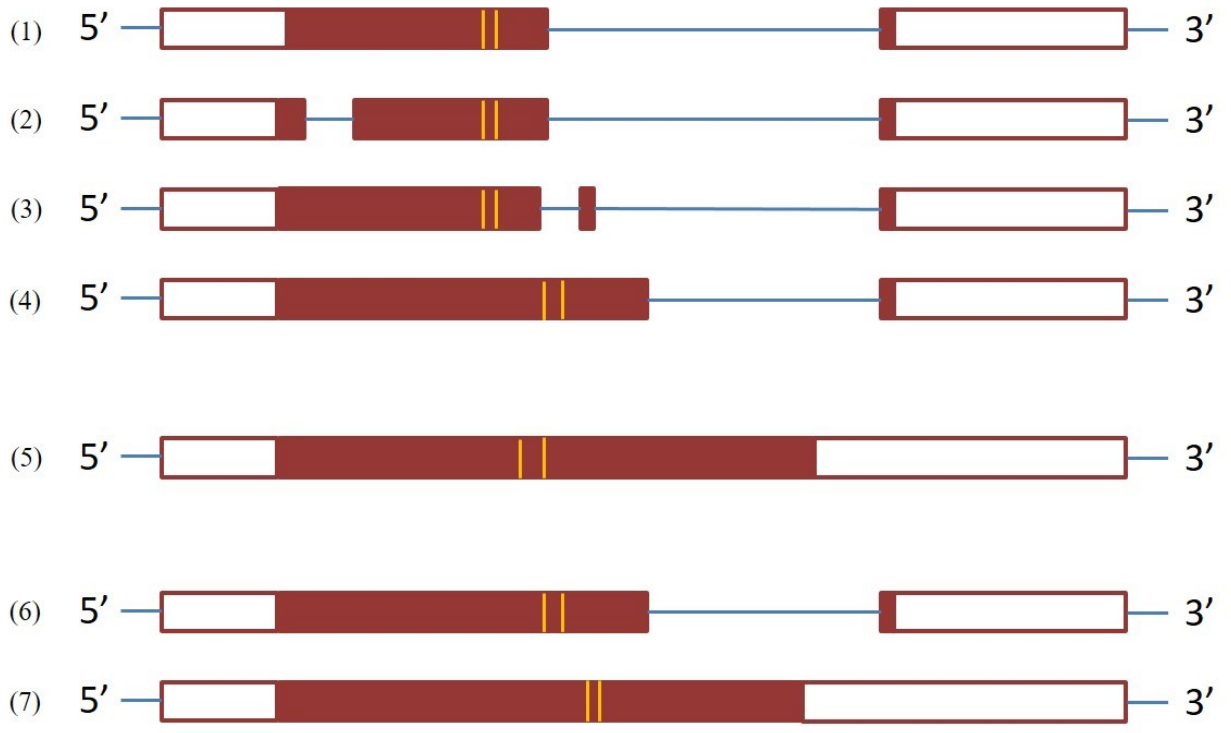
A

Gene	Transcript ID	length	# aa	# exons
TaABCC6	(1) TRIAE_CS42_2AL_TGACv1_095357_AA0309910	5236	1537	11
	(2) TRIAE_CS42_2BL_TGACv1_131165_AA0423600	5425	1661	11
	(3) TRIAE_CS42_2DL_TGACv1_159782_AA0542700	5291	1626	11



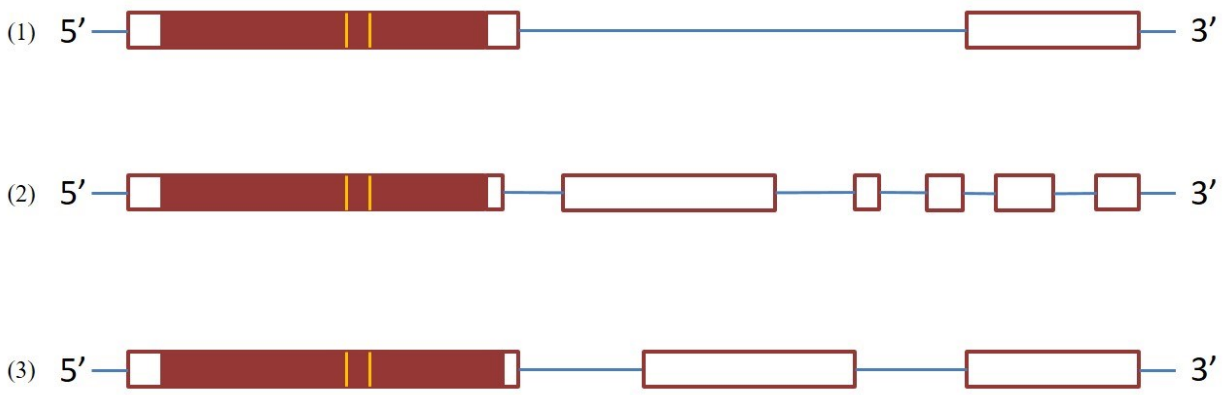
B

Gene	Transcript ID	length	# aa	# exons
	(1) TRIAE_CS42_5AL_TGACv1_375169_AA1217160.1	1005	122	2
	(2) TRIAE_CS42_5AL_TGACv1_375169_AA1217160.2	954	105	3
	(3) TRIAE_CS42_5AL_TGACv1_375169_AA1217160.3	933	98	3
TansLTP9.4	(4) TRIAE_CS42_5AL_TGACv1_375169_AA1217160.4	1276	182	2
	(5) TRIAE_CS42_5BL_TGACv1_404733_AA1309500	1169	155	1
	(6) TRIAE_CS42_5DL_TGACv1_432952_AA1395900.1	954	121	2
	(7) TRIAE_CS42_5DL_TGACv1_432952_AA1395900.2	1240	154	1

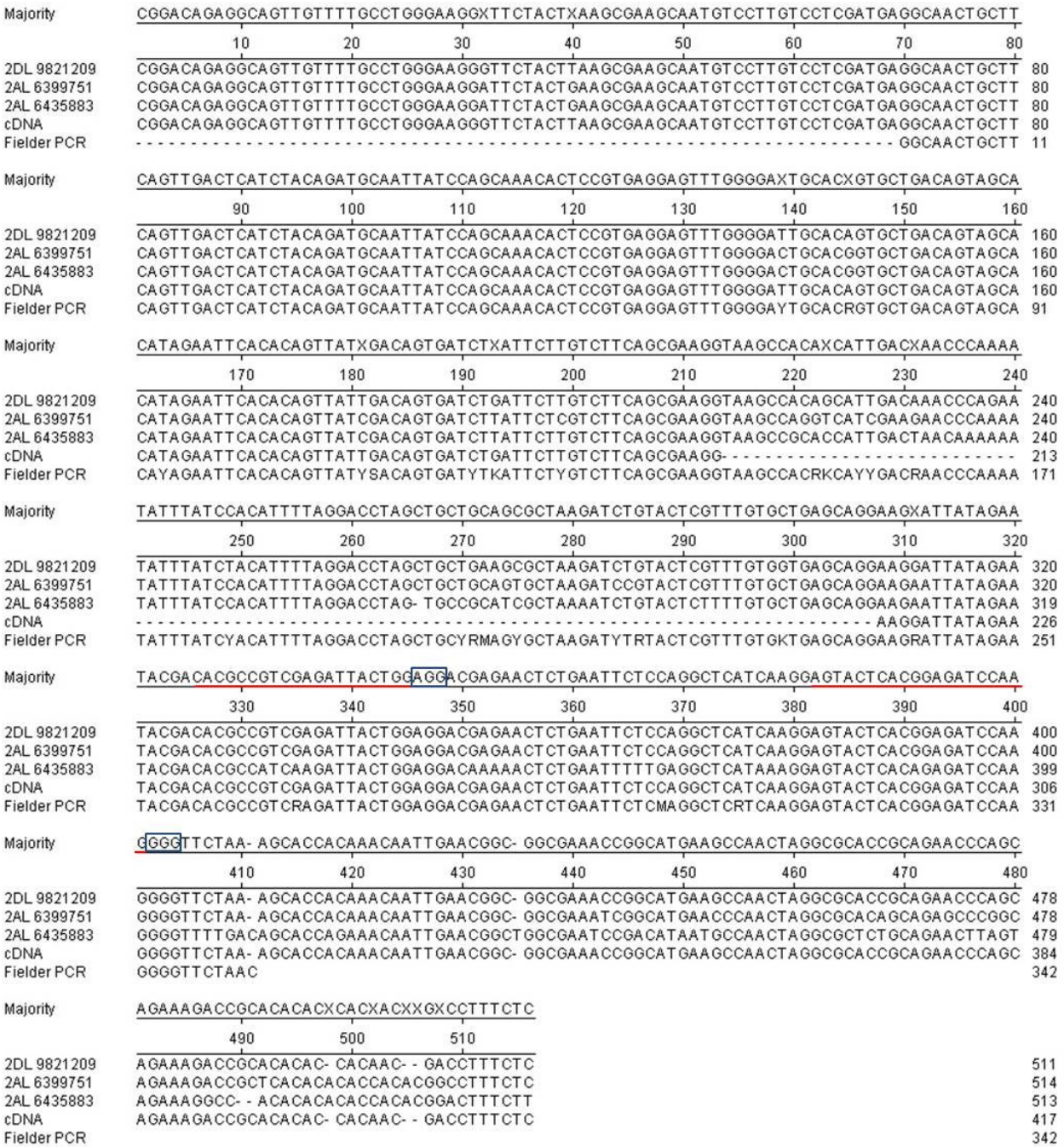


C

Gene	Transcript ID	Length	# aa	# exons
TaNFXL1	(1) TRIAE_CS42_7AL_TGACv1_557614_AA1784060	3790	1124	2
	(2) TRIAE_CS42_7BL_TGACv1_579256_AA1905850	5877	1126	6
	(3) TRIAE_CS42_7DL_TGACv1_603815_AA1989690.1	6732	974	3

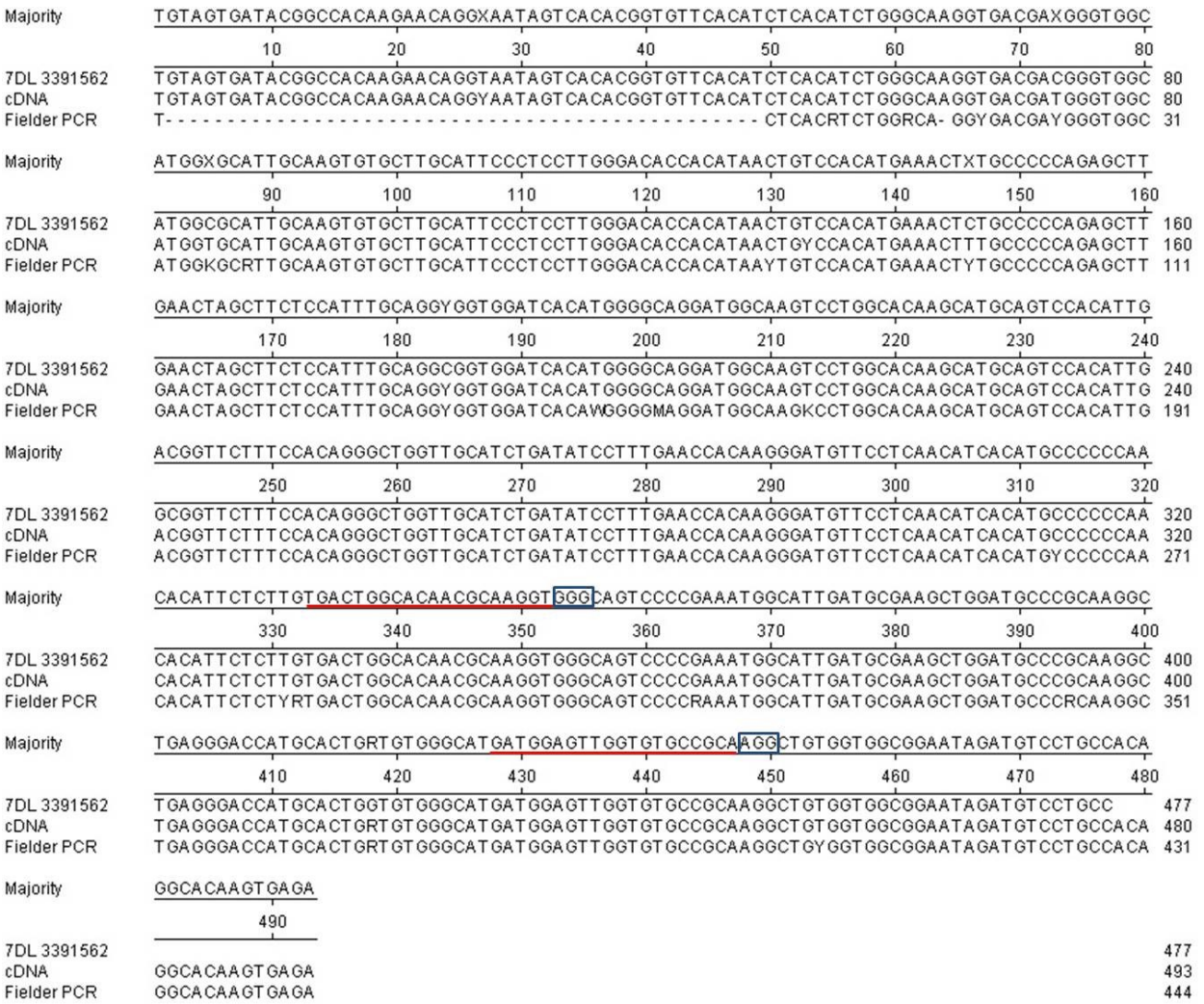


Appendix 11. Alignment of sequences for *TaABCC6* fragment, including three Chinese Spring (CS) genomic sequences obtained from URGI, cDNA sequence from ORDC local EST contig database, and PCR product sequencing results in Fielder. The two sgRNAs are underlined in red and PAMs are in blue rectangles. Majority on the very top represents the consensus sequence.



Appendix 12. Alignment of sequences for *TansLTP9.4* fragment, including two Chinese Spring (CS) genomic sequences obtained from URGI, cDNA sequence from ORDC local EST contig database, and PCR product sequencing results in Fielder. The two sgRNAs are underlined in red and PAMs are in blue rectangles. Majority on the very top represents the consensus sequence.

Appendix 13. Alignment of sequences for *TaNFXL1* fragment, including one Chinese Spring (CS) genomic sequence obtained from URGI, cDNA sequence from ORDC local EST contig database, and PCR product sequencing results in Fielder. The two sgRNAs are underlined in red and PAMs are in blue rectangles. Majority on the very top represents the consensus sequence.

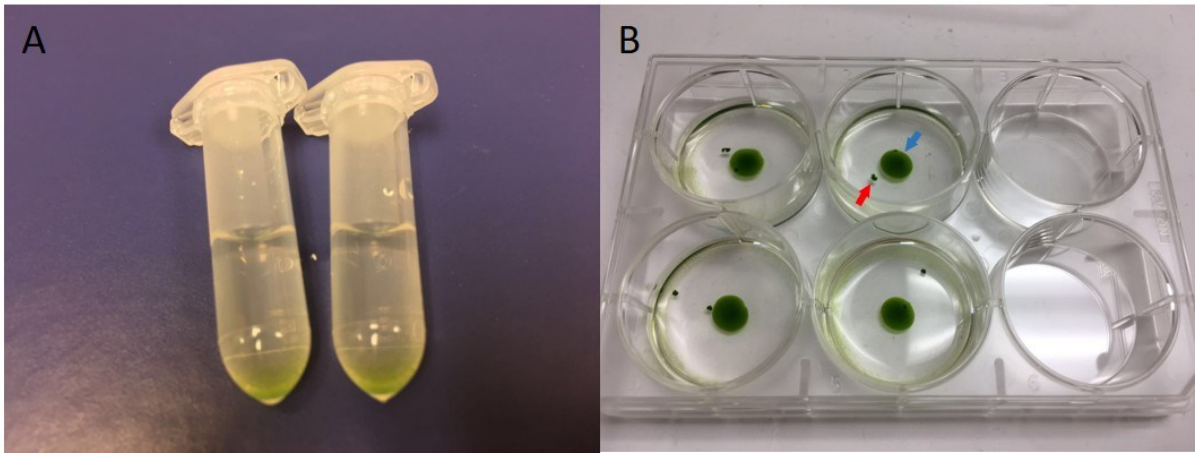


Appendix 14. Full sequences for two gBlocks (gBlocks 1 and 2; 546 bp for one block). Letters in italic: 40 bp overlap with pCambia vector; letters underlined: wheat U6 promoter (Shan et al., 2013); 20 bp in blue: sgRNA sequence; letters in green: sgRNA scaffold and terminator; letters in red: 40 bp overlapping sequence between the two gBlocks required for Gibson assembly.

gBlock1: *CGCGGTGTCATCTATGTTACTAGATCGGGGAGCTCGAATT*GACCAAGCCCGTT
ATTCTGACAGTTCTGGTGCTCAACACATTTATATTTATCAAGGAGCACATTGTTACTC
ACTGCTAGGAGGGAATCGAACTAGGAATATTGATCAGAGGAACTACGAGAGAGCTG
AAGATAACTGCCCTCTAGCTCTCACTGATCTGGGTCGCATAGTGAGATGCAGCCCAC
GTGAGTTCAGCAACGGTCTAGCGCTGGGCTTTTAGGCCCGCATGATCGGGCTTTTGT
CGGGTGGTCGACGTGTTCACGATTGGGGAGAGCAACGCAGCAGTTCCTCTTAGTTTA
GTCCCACCTCGCCTGTCCAGCAGAGTTCTGACCGGTTTATAAACTCGCTTGCTGCAT
CAGACTTGNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAA
AATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTA
GATGGGGCGACGAGTTACTGGCCCTGATTTCTCCGCTTC

gBlock2: AGATGGGGCGACGAGTTACTGGCCCTGATTTCTCCGCTTCGACCAAGCCCG
TTATTCTGACAGTTCTGGTGCTCAACACATTTATATTTATCAAGGAGCACATTGTTAC
TCACTGCTAGGAGGGAATCGAACTAGGAATATTGATCAGAGGAACTACGAGAGAGC
TGAAGATAACTGCCCTCTAGCTCTCACTGATCTGGGTCGCATAGTGAGATGCAGCCC
ACGTGAGTTCAGCAACGGTCTAGCGCTGGGCTTTTAGGCCCGCATGATCGGGCTTTT
GTCGGGTGGTCGACGTGTTCACGATTGGGGAGAGCAACGCAGCAGTTCCTCTTAGTT
TAGTCCCACCTCGCCTGTCCAGCAGAGTTCTGACCGGTTTATAAACTCGCTTGCTGC
ATCAGACTTGNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTT
AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTT
T*CAATTAAACTATCAGTGTTTGACAGGATATATTGGCGGGT*

Appendix 15. Different methods to incubate (A) 20,000 cells in 2 ml round-bottom tubes and (B) 200,000 cells in a 6-well plate after transformation. The incubation of 20,000 cells was without shaking while 200,000 cells were incubated on a shaker at 50 rpm. The live cell clump is marked with a blue arrow while cell debris is marked with a red arrow in B.



Appendix 16. Ten samples submitted for the first round of amplicon sequencing. For each sample, the concentration and the OD value ratio at 260/280 were measured using a Nanodrop ND-1000 spectrophotometer.

Sample ID	Concentration (ng/µl)	OD 260/280
NFXL1-1	57.9	1.80
NFXL1-2	64.2	1.82
NFXL1-3	64.6	1.85
NFXL1-4	61.0	1.79
NFXL1-5	44.9	1.82
nsLTP-1	46.9	1.85
nsLTP-2	27.9	1.81
nsLTP-3	51.0	1.86
nsLTP-4	43.6	1.83
nsLTP-5	27.4	1.73

Appendix 17. A summary of mapping results for nsLTP-1~5 and NFXL1-1~5 target regions. The percentage of mapped reads is highlighted in yellow and includes reads in pairs and broken paired reads.

nsLTP-1	Count	Percentage of reads	Average length	Number of bases	Percentage of bases
References	1	-	449	449	-
Mapped reads	81121	92.14%	131.91	10700338	98.5%
Not mapped reads	6921	7.86%	23.5	162637	1.5%
Reads in pairs	75410	85.65%	400.53	10122064	93.18%
Broken paired reads	5711	6.49%	101.26	578274	5.32%
Total reads	88042	100%	123.38	10862975	100%
nsLTP-2	Count	Percentage of reads	Average length	Number of bases	Percentage of bases
References	1	-	449	449	-
Mapped reads	97790	87.67%	129.29	12642896	97.48%
Not mapped reads	13748	12.33%	23.77	326851	2.52%
Reads in pairs	90620	81.25%	400.63	11947058	92.11%
Broken paired reads	7170	6.43%	97.05	695838	5.37%
Total reads	111538	100%	116.28	12969747	100%
nsLTP-3	Count	Percentage of reads	Average length	Number of bases	Percentage of bases
References	1	-	449	449	-
Mapped reads	106607	94.66%	132.95	14173035	99.37%
Not mapped reads	6019	5.34%	15.04	90546	0.63%
Reads in pairs	100500	89.23%	402.23	13603494	95.37%
Broken paired reads	6107	5.42%	93.26	569541	3.99%
Total reads	112626	100%	126.65	14263581	100%
nsLTP-4	Count	Percentage of reads	Average length	Number of bases	Percentage of bases
References	1	-	449	449	-
Mapped reads	70481	94.04%	131.94	9299174	98.93%
Not mapped reads	4467	5.96%	22.62	101040	1.07%
Reads in pairs	66130	88.23%	401.81	8883275	94.5%
Broken paired reads	4351	5.81%	95.59	415899	4.42%
Total reads	74948	100%	125.42	9400214	100%
nsLTP-5	Count	Percentage of reads	Average length	Number of bases	Percentage of bases
References	1	-	449	449	-
Mapped reads	75370	93.83%	130.61	9844036	99.03%
Not mapped reads	4952	6.17%	19.53	96730	0.97%
Reads in pairs	70158	87.35%	400.82	9334707	93.90%
Broken paired reads	5212	6.49%	97.72	509329	5.12%
Total reads	80322	100.00%	123.76	9940766	100.00%

NFXL1-1	Count	Percentage of reads	Average length	Number of bases	Percentage of bases
References	1	-	756	756	-
Mapped reads	137384	92.12%	124.52	17106490	98.79%
Not mapped reads	11756	7.88%	17.84	209779	1.21%
Reads in pairs	118986	79.78%	156.84	14730776	85.07%
Broken paired reads	18398	12.34%	129.13	2375714	13.72%
Total reads	149140	100.00%	116.11	17316269	100.00%
NFXL1-2	Count	Percentage of reads	Average length	Number of bases	Percentage of bases
References	1	-	756	756	-
Mapped reads	308098	91.51%	114.27	35206882	98.46%
Not mapped reads	28586	8.49%	19.23	549580	1.54%
Reads in pairs	256076	76.06%	126.54	28151441	78.73%
Broken paired reads	52022	15.45%	135.62	7055441	19.73%
Total reads	336684	100.00%	106.2	35756462	100.00%
NFXL1-3	Count	Percentage of reads	Average length	Number of bases	Percentage of bases
References	1	-	756	756	-
Mapped reads	374318	91.13%	116.08	43450131	98.34%
Not mapped reads	36424	8.87%	20.17	734828	1.66%
Reads in pairs	308216	75.04%	128.41	34344240	77.73%
Broken paired reads	66102	16.09%	137.76	9105891	20.61%
Total reads	410742	100.00%	107.57	44184959	100.00%
NFXL1-4	Count	Percentage of reads	Average length	Number of bases	Percentage of bases
References	1	-	756	756	-
Mapped reads	247257	90.42%	107.2	26505387	98.1%
Not mapped reads	26211	9.58%	19.55	512514	1.9%
Reads in pairs	210420	76.95%	116.8	21627974	80.05%
Broken paired reads	36837	13.47%	132.41	4877413	18.05%
Total reads	273468	100%	98.8	27017901	100%
NFXL1-5	Count	Percentage of reads	Average length	Number of bases	Percentage of bases
References	1	-	756	756	-
Mapped reads	333038	90.08%	109.33	36410523	97.38%
Not mapped reads	36682	9.92%	26.71	979726	2.62%
Reads in pairs	283634	76.72%	121.9	29957013	80.12%
Broken paired reads	49404	13.36%	130.63	6453510	17.26%
Total reads	369720	100.00%	101.13	37390249	100.00%

Appendix 18A. Modifications by Breakpoint Analysis on nsLTP-1~4 target regions. Perfect mapped refers to reads that aligned to the reference sequences perfectly; "Number of reads" indicates the reads with the specified unaligned sequences; the variant ratio= (Number of reads)/ (Number of reads + Perfect mapped). We didn't find any modifications on nsLTP-5 target regions.

Sample name	Evidence	p-value	Unaligned	Unaligned length	Mapped to self	Perfect mapped	Number of reads	Variant ratio
nsLTP-1	Left breakpoint*	0	GGTGGACTCCAAGCTCGCGCCGTGCGTGGCG TAC	34	Deletion	16519	55	0.003318
nsLTP-1	Left breakpoint	2.15E-06	GGCGTACG	8	Deletion	12939	4	0.000309
nsLTP-1	Left breakpoint	0	CAGGTGGACTCCAAGCTCGCGCCGTGCGTGG CGTAC	36	Deletion	16614	25	0.001502
nsLTP-2	Left breakpoint	0	GGTGGACTCCAAGCTCGCGCCGTGCGTGGCG TAC	34	Deletion	19650	28	0.001423
nsLTP-2	Left breakpoint	0	CAGGTGGACTCCAAGCTCGCGCCGTGCGTGG CGTAC	36	Deletion	19735	14	0.000709
nsLTP-3	Left breakpoint	3.49E-14	GGCGTACG	8	Deletion	17370	13	0.000748
nsLTP-3	Left breakpoint	5.12E-09	CGTGGCGTACG	11	Deletion	18210	9	0.000494
nsLTP-3	Left breakpoint	3E-07	GCGGGCAGGTGGACTCCAAGCTCGCGCCGT	30	Deletion	22021	8	0.000363
nsLTP-3	Left breakpoint	0	CAGGTGGACTCCAAGCTCGCGCCGTGCGTGG CGTAC	36	Deletion	22577	299	0.01307
nsLTP-3	Left breakpoint	0	CAGGTGGACTCCAAGCTCGCGCCGTGCGTGG CGTAC	36	Deletion	22720	60	0.002634
nsLTP-3	Left breakpoint	6.65E-14	CAGGTGGACTCCAAGCTCGCGCCGTGCGTGG CGTACG	37	Deletion	22846	14	0.000612
nsLTP-4	Left breakpoint	2.48E-10	CGCCGTGCGTGGCGTAC	17	Deletion	12211	9	0.000736
nsLTP-4	Left breakpoint	6E-15	GTGGACTCCAAGCTCGCGCCGTGCGTGGCGT AC	33	Deletion	14391	13	0.000903
nsLTP-4	Left breakpoint	0	GGTGGACTCCAAGCTCGCGCCGTGCGTGGCG TAC	34	Deletion	14457	235	0.015995
nsLTP-4	Left breakpoint	0	CAGGTGGACTCCAAGCTCGCGCCGTGCGTGG CGTAC	36	Deletion	14538	39	0.002675

* Left breakpoint: If an unaligned end started at the left of a mapped read, then the "Left breakpoint" was assigned.

Appendix 18B. Modifications by Breakpoint Analysis on NFXL1-2~4 target regions. Perfect mapped refers to reads that aligned to the reference sequences perfectly; "Number of reads" indicates the reads with the specified unaligned sequences; the variant ratio= Number of reads/ (Number of reads + perfect mapped). The "Right breakpoint" is the mappings with an excess of reads with right unaligned ends. We didn't find any modifications on NFXL1-1 target regions.

Sample name	Name	p-value	Unaligned	Unaligned length	Mapped to self	Perfect mapped	Number of reads	Variant ratio
NFXL1-2	Right breakpoint**	0	CGCAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGGCACAAGTGAGATCGTGG AATATGGTCTCA	68	Deletion	9846	19	0.001926
NFXL1-2	Right breakpoint	9.99E-16	CAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGGCACAAGTGAGATCGTGGAA TATGGTC	63	Deletion	9800	12	0.001223
NFXL1-2	Right breakpoint	6.81E-05	CTGTGGTGGCGGAATAGATGTCCTGCCACAGGC	33	Deletion	9769	4	0.000409
NFXL1-3	Right breakpoint	0	CGCAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGGCACAAGTGAGATCATGG AATATGGTCTCGCGGCAAGGGTCACAGTGACCGCTGTGGCAGAGATGCTGGCAACC ATGCTGTCC	121	Deletion	11878	27	0.002268
NFXL1-3	Left breakpoint	4.2E-05	CAACGCACTCTCTCATGACTGGCACAACGCTCGGCTTCTGACGTTCAGTGCAGCCGT CTTCTGAAAACGACATGTCGCACAAGTCCTAAGTTACGCGACAGGCTGCCGC	109	Insertion	11807	5	0.000423
NFXL1-3	Left breakpoint	2.62E-07	TGTGACTGGCACAACGCAACAATGTACGGGTTCCGGTTCCCAATGTACGGCTTTGGG TTCCCAATGTACGTGCTATCCACAGG	83	Insertion	11550	7	0.000606
NFXL1-3	Right breakpoint	0	CGCAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGGCACAAGTGAGATCATGG AATATGGTCTCGCGGCAAGGGTCACAGTGACCGCTGTGGCAGAGATGCTGGCAACC ATGCTGTCC	121	Deletion	11550	169	0.014421
NFXL1-3	Right breakpoint	0	CAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGGCACAAGTGAGATCATGGAA TATGGTCTCGCGGCAAGGGTCACAGTGACCGCTGTGGCAGAGATGCTGGCAACCAT GCTGTCC	119	Deletion	11513	69	0.005958
NFXL1-3	Left breakpoint	0	CATGCCCCCCAACACATTCTCTTGTGACTGGCACAACGCAA	41	Deletion	3800	24	0.006276
NFXL1-3	Left breakpoint	9.34E-06	CCCAACACATTCTCTTGTGACTGGCACAACGCAA	34	Deletion	3804	6	0.001575
NFXL1-4	Right breakpoint	0	CGCAAGGCTGTGGTGGTGGAATGGATGTCCTGCCACAGGCACAAGTGAGATCATGG AATATGGTCTC	67	Deletion	8629	15	0.001735
NFXL1-4	Left breakpoint	1.9E-13	AAGGGATGTTCCTCAACATCACATGCCCCCCAACACATTCTCTTGTGACTGGCACAA CGCAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAACGGCCG CG	115	Insertion	8391	11	0.001309
NFXL1-4	Right breakpoint	0	CGCAAGGCTGTGGTGGTGGAATGGATGTCCTGCCACAGGCACAAGTGAGATCATGG AATATGGTCTCACGGCAAGGGTCACAGTGACCGCTGTGGCAGAGATGCTGGCAACC ATGCTGTCC	121	Deletion	8391	112	0.013172
NFXL1-4	Right breakpoint	0	CAAGGCTGTGGTGGTGGAATGGATGTCCTGCCACAGGCACAAGTGAGATCATGGAA TATGGTCTCACGGCAAGGGTCACAGTGACCGCTGT	91	Deletion	8372	43	0.00511
NFXL1-4	Right breakpoint	2.94E-09	CTGTGGTGGTGGAATGGATGTCCTGCCACAGGCACAAGTGAGATCATGGAATATGG	56	Deletion	8337	9	0.001078
NFXL1-4	Left breakpoint	0	CAACACATTCTCTTGTGACTGGCACAACGCAA	32	Deletion	2964	16	0.005369

** Right breakpoint: If an unaligned end started at the right of a mapped read, then the "Right breakpoint" was assigned.

Appendix 18C. Deletions in nsLTP-1~4 and NFXL1-4 samples by InDel Analysis. The "Reference" in the table refers to the deleted sequences between two sgRNA sites.

Sample name	Type	Reference	Length	Evidence	Variant ratio
nsLTP-1	Deletion	GTGACGGGGAGGGCGTCCTCGATCAGCAAGGAGTGCTG CTCCGGCGT	47	Paired breakpoint[#]	0.086013
nsLTP-2	Deletion	GTGACGGGGAGGGCGTCCTCGATCAGCAAGGAGTGCTG CTCCGGCGT	47	Self-mapped[##]	0.008538
nsLTP-2	Deletion	GTGACGGGGAGGGCGTCCTCGATCAGCAAGGAGTGCTG CTCCGGCGTG	48	Self-mapped	0.008501
nsLTP-3	Deletion	GTGACGGGGAGGGCGTCCTCGATCAGCAAGGAGTGCTG CTCCGGCGT	47	Paired breakpoint	0.095866
nsLTP-4	Deletion	GTGACGGGGAGGGCGTCCTCGATCAGCAAGGAGTGCTG CTCCGGCGT	47	Paired breakpoint	0.099019
NFXL1-4	Deletion	GGTGGGCAGTCCCCGAAATGGCATTGATGCGAAGCTGG ATGCCCGCAAGGCTGAGGGACCATGCACTGRTGTGGGC ATGATGGAGTTGGTGTGC	94	Self-mapped	0.054819
NFXL1-4	Deletion	GGTGGGCAGTCCCCGAAATGGCATTGATGCGAAGCTGG ATGCCCGCAAGGCTGAGGGACCATGCACTGRTGTGGGC ATGATGGAGTTGGTGTGC	94	Self-mapped	0.063118

Paird breakpoint: Right and left breakpoints were both detected at the unaligned ends for the same reference sequence.

Self-mapped: An unaligned end at the breakpoint was found to map back to the reference sequence in the vicinity of breakpoint itself.

Appendix 18D. Replacements in NFXL1-1~5 samples by Structural Variant Analysis.

Sample Name	Name	Evidence	Length	Reference	Variant sequence	Variant ratio
NFXL1-1	Replacement[#]	Paired breakpoint	9	TGTGACTGG	CATTACTGA	0.04
NFXL1-2	Replacement	Paired breakpoint	20	CCCCCCAACACATTCTCTTG	TCCCCCAACGCACTCTCTCA	0.04
NFXL1-3	Replacement	Paired breakpoint	204	GGTGGGCAGTCCCCGAAATGGCATTGATGCGAAGC TGGATGCCCGCAAGGCTGAGGGACCATGCACTGRT GTGGGCATGATGGAGTTGGTGTGCCGCAAGGCTGT GGTGGCGGAATAGATGTCCTGCCACAGGCACAAGT GAGATCRTGGAATATGGTCTCACGGCAAGGGTCGC AATGACCGCTGTGGCAGAGATGCTGGCAT	CGCAAGGCTGTGGTGGCGG AATAGATGTCCTGCCACAG GCACAAGTGAGATCATGGA ATATGGTCTCGCGGCAAGG GTCACAGTGACCGCTGTGG CAGAGATGCTGGCAA	0.06
NFXL1-5	Replacement	Paired breakpoint	20	CCCCCCAACACATTCTCTTG	TCCCCCAACGCACTCTCTCA	0.03

Replacement: An original sequence (Reference) was deleted before a different sequence (Variant sequence) was inserted at the same locus.

Appendix 19. Seventeen samples submitted for the second round of amplicon sequencing. For each sample, the concentration and the OD value ratio at 260/280 were measured using a Nanodrop ND-1000 spectrophotometer.

Sample ID	Concentration (ng/µl)	OD 260/280
NFXL1-1	38.2	1.93
NFXL1-2	65.2	1.88
NFXL1-3	63.0	1.81
NFXL1-4	62.5	1.89
NFXL1-5	64.2	1.85
NFXL1-control	68.1	1.82
pcoNFXL1-1	65.3	1.91
pcoNFXL1-2	48.4	1.84
pcoNFXL1-3	81.3	1.91
pcoNFXL1-4	60.3	1.83
pcoNFXL1-5	61.1	1.85
ABC-1	61.9	1.82
ABC-2	63.0	1.82
ABC-3	63.4	1.77
ABC-4	63.6	1.82
ABC-5	63.2	1.83
ABC-control	60.1	1.80

Appendix 20. A summary of mapping results for ABC 1~5, resequenced NFXL1-1~5 and pcoNFXL1-1~5 targeting regions. The percentage of mapped bases is highlighted in yellow and includes reads in pairs and broken paired reads.

ABC-1	Count	Percentage of reads	Average length	Number of bases	Percentage of bases
References	1	-	298	298	-
Mapped reads	105447	79.24%	146.52	15450024	93.13%
Not mapped reads	27625	20.76%	41.26	1139685	6.87%
Reads in pairs	67662	50.85%	254.53	11208393	67.56%
Broken paired reads	37785	28.39%	112.26	4241631	25.57%
Total reads	133072	100.00%	124.67	16589709	100.00%
ABC-2	Count	Percentage of reads	Average length	Number of bases	Percentage of bases
References	1	-	298	298	-
Mapped reads	107943	84.72%	157.18	16966289	94.76%
Not mapped reads	19471	15.28%	48.23	939023	5.24%
Reads in pairs	80162	62.91%	256.66	13912049	77.70%
Broken paired reads	27781	21.80%	109.94	3054240	17.06%
Total reads	127414	100.00%	140.53	17905312	100.00%
ABC-3	Count	Percentage of reads	Average length	Number of bases	Percentage of bases
References	1	-	298	298	-
Mapped reads	107672	85.16%	159.64	17188521	94.92%
Not mapped reads	18760	14.84%	49.07	920498	5.08%
Reads in pairs	81230	64.25%	256.83	14300416	78.97%
Broken paired reads	26442	20.91%	109.22	2888105	15.95%
Total reads	126432	100.00%	143.23	18109019	100.00%
ABC-4	Count	Percentage of reads	Average length	Number of bases	Percentage of bases
References	1	-	298	298	-
Mapped reads	111052	85.51%	155.19	17234246	95.82%
Not mapped reads	18812	14.49%	39.99	752257	4.18%
Reads in pairs	85438	65.79%	254.14	14488207	80.55%
Broken paired reads	25614	19.72%	107.21	2746039	15.27%
Total reads	129864	100.00%	138.5	17986503	100.00%
ABC-5	Count	Percentage of reads	Average length	Number of bases	Percentage of bases
References	1	-	298	298	-
Mapped reads	115833	86.95%	158.08	18310867	96.32%
Not mapped reads	17391	13.05%	40.26	700095	3.68%
Reads in pairs	90884	68.22%	254.78	15658966	82.37%
Broken paired reads	24949	18.73%	106.29	2651901	13.95%
Total reads	133224	100.00%	142.7	19010962	100.00%

ABC-control	Count	Percentage of reads	Average length	Number of bases	Percentage of bases
References	1	-	298	298	-
Mapped reads	51466	79.07%	139.76	7192707	94.77%
Not mapped reads	13626	20.93%	29.14	397010	5.23%
Reads in pairs	35900	55.15%	247.42	5276153	69.52%
Broken paired reads	15566	23.91%	123.12	1916554	25.25%
Total reads	65092	100.00%	116.6	7589717	100.00%

NFXL1-1	Count	Percentage of reads	Average length	Number of bases	Percentage of bases
References	1	-	310	310	-
Mapped reads	60665	93.06%	150.28	9116848	99.23%
Not mapped reads	4525	6.94%	15.67	70902	0.77%
Reads in pairs	56578	86.79%	264.36	8699278	94.68%
Broken paired reads	4087	6.27%	102.17	417570	4.54%
Total reads	65190	100.00%	140.94	9187750	100.00%
NFXL1-2	Count	Percentage of reads	Average length	Number of bases	Percentage of bases
References	1	-	310	310	-
Mapped reads	103988	93.77%	163.59	17011280	99.42%
Not mapped reads	6908	6.23%	14.46	99863	0.58%
Reads in pairs	97602	88.01%	263.31	16332542	95.45%
Broken paired reads	6386	5.76%	106.29	678738	3.97%
Total reads	110896	100.00%	154.3	17111143	100.00%
NFXL1-3	Count	Percentage of reads	Average length	Number of bases	Percentage of bases
References	1	-	310	310	-
Mapped reads	118369	94.12%	160.05	18944582	99.35%
Not mapped reads	7395	5.88%	16.72	123666	0.65%
Reads in pairs	111392	88.57%	261.48	18188118	95.38%
Broken paired reads	6977	5.55%	108.42	756464	3.97%
Total reads	125764	100.00%	151.62	19068248	100.00%
NFXL1-4	Count	Percentage of reads	Average length	Number of bases	Percentage of bases
References	1	-	310	310	-
Mapped reads	127103	90.91%	149.11	18952877	99.05%
Not mapped reads	12703	9.09%	14.38	182698	0.95%
Reads in pairs	101128	72.33%	268.77	16217465	84.75%
Broken paired reads	25975	18.58%	105.31	2735412	14.29%
Total reads	139806	100.00%	136.87	19135575	100.00%

NFXL1-5	Count	Percentage of reads	Average length	Number of bases	Percentage of bases
References	1	-	310	310	-
Mapped reads	110987	94.25%	162.27	18010258	99.54%
Not mapped reads	6771	5.75%	12.2	82639	0.46%
Reads in pairs	103964	88.29%	265.62	17279791	95.51%
Broken paired reads	7023	5.96%	104.01	730467	4.04%
Total reads	117758	100.00%	153.64	18092897	100.00%
NFXL1-control	Count	Percentage of reads	Average length	Number of bases	Percentage of bases
References	1	-	310	310	-
Mapped reads	102014	93.33%	166.44	16979612	99.33%
Not mapped reads	7292	6.67%	15.75	114876	0.67%
Reads in pairs	95146	87.05%	266.02	16227535	94.93%
Broken paired reads	6868	6.28%	109.5	752077	4.40%
Total reads	109306	100.00%	156.39	17094488	100.00%

pcoNFXL1-1	Count	Percentage of reads	Average length	Number of bases	Percentage of bases
References	1	-	310	310	-
Mapped reads	115517	93.55%	161.83	18693999	99.38%
Not mapped reads	7959	6.45%	14.68	116826	0.62%
Reads in pairs	107956	87.43%	263.96	17883058	95.07%
Broken paired reads	7561	6.12%	107.25	810941	4.31%
Total reads	123476	100.00%	152.34	18810825	100.00%
pcoNFXL1-2	Count	Percentage of reads	Average length	Number of bases	Percentage of bases
References	1	-	310	310	-
Mapped reads	140816	93.45%	158.24	22282730	99.41%
Not mapped reads	9874	6.55%	13.41	132380	0.59%
Reads in pairs	131464	87.24%	263.98	21302951	95.04%
Broken paired reads	9352	6.21%	104.77	979779	4.37%
Total reads	150690	100.00%	148.75	22415110	100.00%
pcoNFXL1-3	Count	Percentage of reads	Average length	Number of bases	Percentage of bases
References	1	-	310	310	-
Mapped reads	100001	92.79%	151.63	15163616	99.19%
Not mapped reads	7775	7.21%	15.84	123123	0.81%
Reads in pairs	84750	78.64%	269.12	13670974	89.43%
Broken paired reads	15251	14.15%	97.87	1492642	9.76%
Total reads	107776	100.00%	141.84	15286739	100.00%

pcoNFXL1-4	Count	Percentage of reads	Average length	Number of bases	Percentage of bases
References	1	-	310	310	-
Mapped reads	157879	93.43%	163.29	25780392	99.36%
Not mapped reads	11101	6.57%	14.91	165544	0.64%
Reads in pairs	148378	87.81%	260.22	24698385	95.19%
Broken paired reads	9501	5.62%	113.88	1082007	4.17%
Total reads	168980	100.00%	153.54	25945936	100.00%
pcoNFXL1-5	Count	Percentage of reads	Average length	Number of bases	Percentage of bases
References	1	-	310	310	-
Mapped reads	106400	93.72%	161.98	17234818	99.41%
Not mapped reads	7126	6.28%	14.32	102027	0.59%
Reads in pairs	99894	87.99%	261.4	16499700	95.17%
Broken paired reads	6506	5.73%	112.99	735118	4.24%
Total reads	113526	100.00%	152.71	17336845	100.00%

Appendix 21A. Modifications by Breakpoint Analysis on ABC-1~5 target regions. Perfect mapped refers to reads that aligned to the reference sequences perfectly; "Number of reads" indicates the reads with the specified unaligned sequences; the variant ratio= (Number of reads)/ (Number of reads + perfect mapped).

Sample name	Name	p-value	Unaligned	Unaligned length	Mapped to self	Perfect mapped	Number of reads	Variant ratio
ABC-1	Right breakpoint	1.55E-07	AGGGGTTCTAAAGCACCACAAACAATTGAACGGCGGCGAAACCGG CATGAAGCCAACTAGGCGCACCGCAGAACCCAGC	79	Deletion	3686	8	0.002166
ABC-1	Right breakpoint	0	AAGGGGTTCTAAAGCACCACAAACAATTGAACGGCGGCGAAACCG GCATGAAGCCAACTAGGCGCACCGCAGAACCCAGCA	81	Deletion	3696	24	0.006452
ABC-1	Left breakpoint	0	AGCTGCTGAAGCGCTAAGATCTGTACTCGTTTGTGGTGAGCAGGAA GGATTATAGAATACGACACGCCGTCGAGATTA	78	Deletion	3965	21	0.005268
ABC-1	Left breakpoint	1.04E-05	AGCTGCTGAAGCGCTAAGATCTGTACTCGTTTGTGGTGAGCAGGAA GGATTATAGAATACGACACGCCGTCGAGATT	77	Deletion	3960	7	0.001765
ABC-2	Left breakpoint	2.07E-06	TGAAGCGCTAAGATCTGTACTCGTTTGTGGTGAGCAGGAAGGATTA TAGAATACGACACGCCGGCGAGATTA	72	Deletion	6267	8	0.001275
ABC-3	Right breakpoint	9.45E-14	AAGGGGTTCTAAAGCACCACAAACAATTGAACGGCGGCGAAACCG GCATGAAGCCAACTAGGCGCACCGCAGAACCCAGCA	81	Deletion	4494	14	0.003106
ABC-3	Left breakpoint	3.13E-12	CCTAGCTGCTGAAGCGCTAAGATCTGTACTCGTTTGTGGTGAGCAG GAAGGATTATAGAATACGACACGCCGTCGAGATTA	81	Deletion	4661	14	0.002995
ABC-4	Right breakpoint	0	AAGGGGTTCTAAAGCACCACAAACAATTGAACGGCGGCGAAACCG GCATGAAGCCAACTAGGCGCACCGCAGAACCCAGCA	81	Deletion	4794	95	0.019431
ABC-4	Left breakpoint	0	GCCACAGCACCGACGAACCCAAAATATTTATCCACATTTTAGGACC TAGCTGCTGCAGTGCTAAGATCCGTACTCGTTTGTGCTGAGCAGGA AGAATTATAGAATACGACACGCCGTCGAGATTA	125	Deletion	5052	56	0.010963
ABC-4	Left breakpoint	2.41E-10	CTACATTTTAGGACCTAGCTGCTGAAGCGCTAAGATCTGTACTCGT TTGTGGTGAGCAGGAAGGATTATA	70	Deletion	5046	14	0.002767
ABC-5	Left breakpoint	1.44E-15	AGCCAGGTCATCGAAGAACCCAAAATATTTATCCACATTTTAGGAC CTAGCTGCTGCAGTGCTAAGATCCGTACTCGTTTGTGCTGAGCAGG AAGAATTATAGAATACGACACGCCGTCGAGATTA	126	Deletion	6131	19	0.003089

Appendix 21B. Modifications by Breakpoint Analysis on pcoNFXL1-1~5 target regions. Perfect mapped refers to reads that aligned to the reference sequences perfectly; "Number of reads" indicates the reads with the specified unaligned sequences; the variant ratio= (Number of reads)/ (Number of reads + perfect mapped).

Sample name	Name	p-value	Unaligned	Unaligned length	Mapped to self	Perfect mapped	Number of reads	Variant ratio
pcoNFXL1-1	Right breakpoint	0	CGCAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGGCACAAG TGAGATCGTGGAATATGGTCTCACGGCAAGGGTCGCAATGACCGC TGTGGC	96	Deletion	23577	255	0.0107
pcoNFXL1-1	Right breakpoint	0	CAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGGCACAAGTG AGATCGTGGAATATGGTCTCACGGCAAGGGTCGCAATGACCGCTG TGGC	94	Deletion	23609	491	0.020373
pcoNFXL1-1	Left breakpoint	0	TGCATCTGATATCCTTCGAACCACAGGGGATGTTCCTTAGCATCAC ATGCCCTCCAACGCACTCTCTCATGACTGGCACAACGCAA	86	Deletion	29071	472	0.015977
pcoNFXL1-1	Left breakpoint	0	TGCATCTGATATCCTTTGAACCACAAAGGATGTTCCTCAACATCAC ATGCCCCCCAACACATTCTCTTGTGACTGGCACAACGCAA	86	Deletion	29082	384	0.013032
pcoNFXL1-2	Right breakpoint	0	GCAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGGCACAAGT GAGATCGTGGAATATGGTCTCACGGCAAGGGTCGCAATGACCGCT GTGG	94	Deletion	32051	608	0.018617
pcoNFXL1-2	Right breakpoint	0	CGCAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGGCACAAG TGAGATCGTGGAATATGGTCTCACGGCAAGGGTCGCAATGACCGC TGTGGC	96	Deletion	31928	218	0.006782
pcoNFXL1-2	Left breakpoint	0	TGCATCTGATATCCTTTGAACCACAAGGGATGTTCCTCAACATCAC ATGCCCCCCAACACATTCTCTTGTGACTGGCACAAC	82	Deletion	40024	381	0.00943
pcoNFXL1-2	Left breakpoint	0	CATCTGATATCCTTTGAACCACAAGGAATGTTCCTCAACATCACAT GCCCCCCAACACATTCTCTTGTGACTGGCACAACG	81	Deletion	40440	504	0.012309
pcoNFXL1-3	Right breakpoint	0	GCAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGGCACAAGT GAGATCGTGGAATATGGTC TCACGGCAAGGGTCGCAATGACCGCTGTGGCAGAGATGCTGG	106	Deletion	17277	429	0.024229
pcoNFXL1-3	Right breakpoint	0	CGCAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGGCACAAG TGAGATCGTGGAATATGGTCTCACGGCAAGGGTCGCAGTGACCGC TGTGGCAGAG	100	Deletion	17176	235	0.013497
pcoNFXL1-3	Right breakpoint	0	CTGTGGTGGCGGAATAGATGTCCTGCCACAGGCACAAGTGAGATC GTGGAATATGGTCTCACGGCAAGGGTCGCAATGACCGCTGTGGC	89	Deletion	17226	345	0.019635
pcoNFXL1-3	Left breakpoint	0	TCCTTTGAACCACAAGGGATGTTCCTCAACATCACATGCCCCCCAA CACATTCTCTTGTGACTGGCACAACG	72	Deletion	20925	726	0.033532
pcoNFXL1-4	Right breakpoint	0	TTTGTGCCGCAAGGCTGTGGCGGCGGAATAGATGTCCTGCCACAGG CACAAGTGAGATCGTGGAATATGGTCTCACGGCAAGGGTCGCAAT GACCGCTGTGGC	105	Deletion	28601	107	0.003727

Sample name	Name	p-value	Unaligned	Unaligned length	Mapped to self	Perfect mapped	Number of reads	Variant ratio
pcoNFXL1-4	Right breakpoint	0	GCAAGGCTGCGGTGGCGGAATAGATGACCTGCCACACGCACACGT GAGATCATTGAATATGGTCTCACGGCAAGGGTCGCAGTGACCGCT GTGGC	95	Deletion	29750	155	0.005183
pcoNFXL1-4	Right breakpoint	8.37E-08	GCAAGGCTGTGGTGGTGGAATGGATGTCCTGCCACAGGCACAAGT GAGATCATGGAATATGGTCTCACGGCAAGGGTCACAGTGACCGCT GTGGC	95	Deletion	29785	16	0.000537
pcoNFXL1-4	Right breakpoint	3.66E-15	GGCAAGGCTGCGGTGGCGGAATAGATGACCTGCCACACGCACACG TGAGATCATTGAATATGGTCTCACGGCAAGGGTCGCAGTGACCGCT GTGG	95	Deletion	29961	25	0.000834
pcoNFXL1-4	Right breakpoint	0	CGCAAGGCTGTGGTGGTGGAATGGATGTCCTGCCACAGGCACAAG TGAGATCATGGAATATGGTCTCACGGCAAGGGTCACAGTGACCGC TGTGGC	96	Deletion	29945	255	0.008444
pcoNFXL1-4	Right breakpoint	0	CAAGGCTGTGGTGGTGGAATGGATGTCCTGCCACAGGCACAAGTG AGATCATGGAATATGGTCTCACGGCAAGGGTCACAGTGACCGCTG TGGCAGAGATGCTGGCATCCA	111	Deletion	30020	265	0.00875
pcoNFXL1-4	Left breakpoint	0	TGCATCTGATATCCTTTGAACCACAAGGGATGTTCCTCAACATCAC ATGCCCCCCAACACATTCTCTT	68	Deletion	37027	776	0.020527
pcoNFXL1-4	Left breakpoint	0	TGCATCTGATATCCTTTGAACCACAAGGGATGTTCCTTAACATCAC ATGTCCCCCAACGCACTCTCTCATGACTGGCACAACGCAA	86	Deletion	37482	101	0.002687
pcoNFXL1-4	Left breakpoint	0	CCTCAACATCACATGCCCCCCAACACATTCTCTTGTGACTGGCACA ACGCAA	52	Deletion	37471	33	0.00088
pcoNFXL1-5	Right breakpoint	0	CGCAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGGCACAAG TGAGATCGTGGAATATGGTCTCACGGCAAGGGTCGCAATGACCGC TGTGGCAGAG	100	Deletion	20455	714	0.033729
pcoNFXL1-5	Right breakpoint	0	CGCAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGGCACAAG TGAGATCGTGGAATATGGTCTCACGGCAAGGGTCGCAGTGACCGC TGTGGCAGAGATGCTGGCATCCA	113	Deletion	20372	1587	0.072271
pcoNFXL1-5	Right breakpoint	0	CAAGGCTGTGGTGGTGGAATGGATGTCCTGCCACAGGCACAAGTG AGATCATGGAATATGGTCTCACGGCAAGGGTCACAGTGACCGCTG TGGC	94	Deletion	20421	189	0.00917
pcoNFXL1-5	Left breakpoint	0	TGCATCTGATATCCTTTGAACCACAAGGGATGTTCCTCAACATCAC ATGCCCCCCAACACATTCTCTTGTGACTGGCACAAC	82	Deletion	25164	1120	0.042611
pcoNFXL1-5	Left breakpoint	0	TGCATCTGATATCCTTTGAACCACAAGGGATGTTCCTCAACATCAC ATGCCCCCCAACACATTCTCTTGTGACTGGCACAACGCAA	86	Deletion	25421	1325	0.04954

Appendix 21C. Modifications by Breakpoint Analysis on resequenced NFXL1-1~5 target regions. Perfect mapped refers to reads that aligned to the reference sequences perfectly; "Number of reads" indicates the reads with the specified unaligned sequences; the variant ratio= (Number of reads)/ (Number of reads + perfect mapped). We didn't find any modifications for NFXL1-5.

Sample name	Name	p-value	Unaligned	Unaligned length	Mapped to self	Perfect mapped	Number of reads	Variant ratio
NFXL1-1	Right breakpoint	0	CGCAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGGCACAAGTGA GATCGTGGAATATGGTCTCACGGCAAGGGTCGCAATGACCGCTGTGGC	96	Deletion	9702	138	0.014024
NFXL1-1	Right breakpoint	0	CAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGGCACAAGTGAGA TCGTGGAATATGGTCTCACGGCAAGGGTCGCAATGACCGCTGTGGC	94	Deletion	9706	124	0.012614
NFXL1-1	Left breakpoint	0	GCATCTGATATCCTTTGAACCACAGGGGATGTTCCTCAACATCACATGC CCCCCAACACATTCTCTTGTGACTGGCACAACGCAA	85	Deletion	11242	91	0.00803
NFXL1-1	Left breakpoint	0	CATCTGATATCCTTTGAACCACAAGGGATGTTCCTCAACATCACATGCC CCCCAACACATTCTCTTGTGACTGGCACAACGCAA	84	Deletion	11261	111	0.009761
NFXL1-2	Right breakpoint	3.32E-11	GCAAGGCTGTGGTGGTGGAATGGATGTCCTGCCACAGGCACAAGTGAG ATCATGGAATATGGTCTCACGGCAAGGGTCGCAATGACCGCTGTGG	94	Deletion	18600	27	0.00145
NFXL1-2	Right breakpoint	0	CGCAAGGCTGCGGTGGCGGAATAGATGACCTGCCACACGCACAAGAGA GATCATTGAATATGGTCTCACGGCAAGGGTCGCAGTGACCGCTGTGGCA GAGATGCTGGCATCCACCTGTCTCTTATA	126	Deletion	18537	625	0.032617
NFXL1-2	Right breakpoint	0	CAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGGCACAAGTGAGA TCGTGGAATATGGTCTCACGGCAAGGGTCGCAATGACCGCTGTGGC	94	Deletion	18556	619	0.032282
NFXL1-2	Left breakpoint	0	TGCATCTGATATCCTTTGAACCACAAGGGACGTTCCTCAACATCACATG CCCCCCAACACATTCTCTTGTGACTGGCACAACGCAA	86	Deletion	22890	363	0.015611
NFXL1-2	Left breakpoint	0	TATAAGAGGACAGAGCTTTCCACAGGGCTGGTTGCATCTGATATCCTTT GAACCACAGGGGATGTTCCTTAGCATCACATGTCCTCCAACGCACTCTC TCATGACTGGCACAACGCAA	118	Deletion	22903	528	0.022534
NFXL1-2	Left breakpoint	0	TGCATCTGATATCCTTTGAACCACAAGGGATGTTCCTCAACATCACATG CCCCCCAACACATTCTCTTGTGACTGGC	77	Deletion	22414	148	0.00656
NFXL1-3	Right breakpoint	1.48E-12	GGCAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGGCACAAGTGA GATCGTGGAATATGGTCTCACGGCAAGGGTCGCAATGACCGCTGTGGC	96	Deletion	31229	62	0.001981
NFXL1-3	Right breakpoint	0	GCAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGGCACAAGTGAG ATCGTGGAATATGGTCTCACGGCAAGGGTCGCAATGACCGCTGTGGC	95	Deletion	31308	545	0.01711
NFXL1-3	Right breakpoint	0	CGCAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGGCACAAGTGA GATCGTGGAATATGGTCTCACGGCAAGGGTCGCAATGACCGCTGTGGC	96	Deletion	31263	145	0.004617
NFXL1-3	Right breakpoint	0	CGCAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGGCACAAGTGA GATCGTGGAATATGGTCTCACGGCAAGGGTCGCAATGACCGCTGTGGC AG	98	Deletion	31193	3291	0.095436
NFXL1-3	Right breakpoint	0	CAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGGCACAAGTGAGA TCGTGGAATATGGTCTCACGGCAAGGGTCGCAATGACCGCTGTGGC	94	Deletion	31230	1587	0.048359

Sample name	Name	p-value	Unaligned	Unaligned length	Mapped to self	Perfect mapped	# of Reads	Variant ratio
NFXL1-3	Left breakpoint	0	GCATCTGATATCCTTTGAACCACAAGGGATGTTTCTTAACATCACATGT CCCCCAACACATTCTCTTGTGACTGGCACAAC	81	Deletion	39007	195	0.004974
NFXL1-3	Left breakpoint	0	TGCATCTGATATCCTTTGAACCACAAGGGATGTTCCTCAACATCACATG CCCCCCAACACATTCTCTTGTGACTGGCACAACGCAA	86	Deletion	39284	3097	0.073075
NFXL1-3	Left breakpoint	0	TGCATCTGATATCCTTTGAACCACAAGGGATGTTCCTCAACATCACATG CCCCCCAACACATTCTCTTGTGACTGGCACAACGCAA	86	Deletion	39401	1193	0.029389
NFXL1-3	Left breakpoint	1.16E-10	CATCTGATATCCTTTGAACCACAAGGGATGCTCCTCAACATCACATGCC CCCCAACACATTCTCTTGTGACTGGCACAACGCAA	84	Deletion	38886	56	0.001438
NFXL1-4	Right breakpoint	0	CGCAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGGCACAAGTGA GATCGTGGAATATGGTCTCACGGCAAGGGTCGCAATGACCGCTGTGGC	96	Deletion	21141	130	0.006112
NFXL1-4	Right breakpoint	0	CGCAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGGCACAAGTGA GATCGTGGAATATGGTCTCACGGCAAGGGTCGCAGTGACCGCTGTGGC AGAGATGCTGGCATCCACCGCTGTCTCTA	125	Deletion	21108	1181	0.052986
NFXL1-4	Right breakpoint	0	CAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGGCACAAGTGAGA TCGTGGAATATGGTCTCACGGCAAGGGTCGCAATGACCGCTGTGGC	94	Deletion	21134	597	0.027472
NFXL1-4	Left breakpoint	0	GATATCCTTTGAACCACAAGGGATGTTCCTCAACATCACATGCCCCCCA ACACATTCTCTTGTGACTGGCACAACGCAA	79	Deletion	27826	966	0.033551
NFXL1-4	Left breakpoint	0	ATCCTTTGAACCACAAGGGATGTTCCTCAACATCACATGCCCCCCAACA CATTCTCTTGTGACTGGCACAACGCAA	76	Deletion	27914	523	0.018392

Appendix 22A. Deletions on ABC-1~5 and pcoNFXL1 samples by InDel Analysis. The "Reference" in the table refers to the deleted sequences between two sgRNA sites.

Sample name	Type	Reference	Length	Evidence	Variant ratio
ABC-1	Deletion	CTGGAGGACGAGAACTCTGAATTCTCCAGGCTCATCAAGGAGTAC TCACGGAGATC	56	Paired breakpoint	0.077099
ABC-2	Deletion	CTGGAGGACGAGAACTCTGAATTCTCCAGGCTCATCAAGGAGTAC TCACGGAGATC	56	Self-mapped	0.064906
ABC-3	Deletion	CTGGAGGACGAGAACTCTGAATTCTCCAGGCTCATCAAGGAGTAC TCACGGAGATC	56	Paired breakpoint	0.084866
ABC-4	Deletion	CTGGAGGACGAGAACTCTGAATTCTCCAGGCTCATCAAGGAGTAC TCACGGAGATC	56	Paired breakpoint	0.097029
ABC-5	Deletion	CTGGAGGACGAGAACTCTGAATTCTCCAGGCTCATCAAGGAGTAC TCACGGAGATC	56	Self-mapped	0.086561
pcoNFXL1-4	Deletion	GTGACTGGCACAACGCAAGGTGGGCAGTCCCCGAAATGGCATTGA TGCGAAGCTGGATGCCCGCAAGGCTGAGGGACCATGCACTGRTGT GGGCATGATGGAGTTGG	107	Self-mapped	0.041357

Appendix 22B. Replacements in NFXL1-1~4 and pcoNFXL1-1~4 samples by Structural Variant Analysis.

Chromosome	Name	Evidence	Length	Reference sequence	Variant sequence	Variant ratio
NFXL1-1	Replacement	Paired breakpoint	147	GGTGGGCAGTCCCCGAAATGGCATTGATGCGAAGCTGGAT GCCCGCAAGGCTGAGGGACCATGCACTGRTGTGGGCATGA TGGAGTTGGTGTGCCGCAAGGCTGTGGTGGCGGAATAGAT GTCCTGCCACAGGCACAAGTGAGATCR	CGCAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGG CACAAGTGAGATCG	0.039408
NFXL1-2	Replacement	Paired breakpoint	244	GGTGGGCAGTCCCCGAAATGGCATTGATGCGAAGCTGGAT GCCCGCAAGGCTGAGGGACCATGCACTGRTGTGGGCATGA TGGAGTTGGTGTGCCGCAA	CGCAAGGCTGCGGTGGCGGAATAGATGACCTGCCACACG CACAAGAGAGATCATTGAATATGGTCTCACGGCAAGGGT CGCAGTGACCGCTGTGGCAGAGATGCTGGCATCCACCTG TCTCTTATAAGAGGACAGAGCTTTCCACAGGGCTGGTTG CATCTGATATCCTTTGAACCACAGGGGATGTTCCTTAGCA TCACATGTCCTCCAACGCACTCTCTCATGACTGGCACAAC GCAAGCAA	0.072017
NFXL1-3	Replacement	Paired breakpoint	147	GGTGGGCAGTCCCCGAAATGGCATTGATGCGAAGCTGGAT GCCCGCAAGGCTGAGGGACCATGCACTGRTGTGGGCATGA TGGAGTTGGTGTGCCGCAAGGCTGTGGTGGCGGAATAGAT GTCCTGCCACAGGCACAAGTGAGATCR	CGCAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGG CACAAGTGAGATCG	0.145537
NFXL1-4	Replacement	Paired breakpoint	219	GGTGGGCAGTCCCCGAAATGGCATTGATGCGAAGCTGGAT GCCCGCAAGGCTGAGGGACCATGCACTGRTGTGGGCATGA TGGAGTTGGTGTGCCGCAAGGCTGTGGTGGCGGAATAGAT GTCCTGCCACAGGCACAAGTGAGATCRTGGAATATGGTCT CACGGCAAGGGTCGCAATGACCGCTGTGGCAGAGATGCTG GCATCCANNNNNNNNNNNN	CGCAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGG CACAAGTGAGATCGTGGAATATGGTCTCACGGCAAGGGT CGCAGTGACCGCTGTGGCAGAGATGCTGGCATCCACCGC TGTCTCTA	0.089956
pcoNFXL1-1	Replacement	Paired breakpoint	147	GGTGGGCAGTCCCCGAAATGGCATTGATGCGAAGCTGGAT GCCCGCAAGGCTGAGGGACCATGCACTGRTGTGGGCATGA TGGAGTTGGTGTGCCGCAAGGCTGTGGTGGCGGAATAGAT GTCCTGCCACAGGCACAAGTGAGATCR	CGCAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGG CACAAGTGAGATCG	0.042286
pcoNFXL1-2	Replacement	Paired breakpoint	149	AAGGTGGGCAGTCCCCGAAATGGCATTGATGCGAAGCTGG ATGCCCGCAAGGCTGAGGGACCATGCACTGRTGTGGGCAT GATGGAGTTGGTGTGCCGCAAGGCTGTGGTGGCGGAATAG ATGTCCTGCCACAGGCACAAGTGAGATCR	GCAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGGC ACAAGTGAGATCG	0.038925
pcoNFXL1-3	Replacement	Paired breakpoint	149	AAGGTGGGCAGTCCCCGAAATGGCATTGATGCGAAGCTGG ATGCCCGCAAGGCTGAGGGACCATGCACTGRTGTGGGCAT GATGGAGTTGGTGTGCCGCAAGGCTGTGGTGGCGGAATAG ATGTCCTGCCACAGGCACAAGTGAGATCR	GCAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGGC ACAAGTGAGATCG	0.059226
pcoNFXL1-4	Replacement	Paired breakpoint	221	GGTGGGCAGTCCCCGAAATGGCATTGATGCGAAGCTGGAT GCCCGCAAGGCTGAGGGACCATGCACTGRTGTGGGCATGA TGGAGTTGGTGTGCCGCAAGGCTGTGGTGGCGGAATAGAT GTCCTGCCACAGGCACAAGTGAGATCRTGGAATATGGTCT CACGGCAAGGGTCGCAATGACCGCTGTGGCAGAGATGCTG GCATCCANNNNNNNNNNNNNN	CGCAAGGCTGTGGTGGCGGAATAGATGTCCTGCCACAGG CACAAGTGAGATCGTGGAATATGGTCTCACGGCAAGGGT CGCAGTGACCGCTGTGGCAGAGATGCTGGCATCCAACTG TCTCTTATAC	0.118976

Appendix 23. The information of extracted protoplast RNA for concentration course (A) and time course (B) experiments. Each table includes the sample ID, the concentration and the OD value ratio at 260/280 measured by a Nanodrop ND-1000 spectrophotometer, and the amount of RNA and H_2O was used for producing cDNA samples. For each reaction, 3.5 µg of RNA was used and H_2O was added to bring the total volume to 10 µl. For each treatment, 3 biological replicates were indicated as -1, -2 or -3. Table A: protoplasts were treated with 0, 5, 10 and 20 ppm of DON for 2 hours, respectively. Table B: 10 ppm of DON was used for all the DON treatments.

A

Sample ID	Concentration (ng/µl)	OD 260/280	To 3.5 µg (µl)	H_2O added to 10 µl
0DON-2 hr-1	675	2.13	5.19	4.81
5DON-2 hr-1	408	2.1	8.58	1.42
10DON-2 hr-1	640	2.14	5.47	4.53
20DON-2 hr-1	839	2.14	4.17	5.83
0DON-2 hr-2	891	2.13	3.93	6.07
5DON-2 hr-2	815	2.13	4.29	5.71
10DON-2 hr-2	480	2.11	7.29	2.71
20DON-2 hr-2	830	2.17	4.22	5.78
0DON-2 hr-3	693	2.17	5.05	4.95
5DON-2 hr-3	557	2.16	6.28	3.72
10DON-2 hr-3	405	2.08	8.64	1.36
20DON-2 hr-3	429	2.12	8.16	1.84

B

Samle ID	Concentration (ng/µl)	OD 260/280	To 3.5 µg (µl)	H_2O added to 10 µl
H_2O-2 hr-1	938	2.15	3.73	6.27
DON-2 hr-1	817	2.15	4.28	5.72
H_2O-4 hr-1	905	2.15	3.87	6.13
DON-4 hr-1	813	2.16	4.31	5.69
H_2O-6 hr-1	833	2.17	4.20	5.80
DON-6 hr-1	861	2.15	4.07	5.93
H_2O-8 hr-1	1013	2.15	3.46	6.54
DON-8 hr-1	1317	2.14	2.66	7.34
H_2O-12 hr-1	808	2.16	4.33	5.67
DON-12 hr-1	857	2.15	4.08	5.92
H_2O-2 hr-2	716	2.14	4.89	5.11
DON-2 hr-2	773	2.19	4.53	5.47
H_2O-4 hr-2	461	2.05	7.59	2.41
DON-4 hr-2	864	2.16	4.05	5.95
H_2O-6 hr-2	632	2.23	5.54	4.46
DON-6 hr-2	828	2.19	4.23	5.77
H_2O-8 hr-2	640	2.21	5.47	4.53
DON-8 hr-2	675	2.2	5.19	4.81
H_2O-12 hr-2	646	2.21	5.42	4.58
DON-12 hr-2	923	2.17	3.79	6.21
H_2O-2 hr-3	917	2.18	3.82	6.18
DON-2 hr-3	718	2.18	4.87	5.13
H_2O-4 hr-3	701	2.2	4.99	5.01
DON-4 hr-3	836	2.19	4.19	5.81
H_2O-6 hr-3	785	2.18	4.46	5.54
DON-6 hr-3	922	2.17	3.80	6.20
H_2O-8 hr-3	654	2.2	5.35	4.65
DON-8 hr-3	778	2.18	4.50	5.50
H_2O-12 hr-3	565	2.21	6.19	3.81
DON-12 hr-3	688	2.22	5.09	4.91

Appendix 24. Permissions to use figures in this thesis.

Figure 1: by Sander, J. D., & Joung, J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature biotechnology*, *32*(4), 347-355.

NATURE PUBLISHING GROUP LICENSE TERMS AND CONDITIONS

Jul 19, 2017

This Agreement between Mr. Xiucheng Cui ("You") and Nature Publishing Group ("Nature Publishing Group") consists of your license details and the terms and conditions provided by Nature Publishing Group and Copyright Clearance Center.

License Number	4152731016411
License date	Jul 19, 2017
Licensed Content Publisher	Nature Publishing Group
Licensed Content Publication	Nature Biotechnology
Licensed Content Title	CRISPR-Cas systems for editing, regulating and targeting genomes
Licensed Content Author	Jeffry D Sander, J Keith Joung
Licensed Content Date	Mar 2, 2014
Licensed Content Volume	32
Licensed Content Issue	4
Type of Use	reuse in a dissertation / thesis
Requestor type	academic/educational
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Figures	Figure 3: Naturally occurring and engineered CRISPR-Cas systems.
Author of this NPG article	no
Your reference number	
Title of your thesis / dissertation	Targeted Gene Editing Using CRISPR/Cas9 in a Wheat Protoplast System

Figure 2: by Shan, Q., Wang, Y., Li, J., & Gao, C. (2014). Genome editing in rice and wheat using the CRISPR/Cas system. *Nature protocols*, *9*(10), 2395-2410.

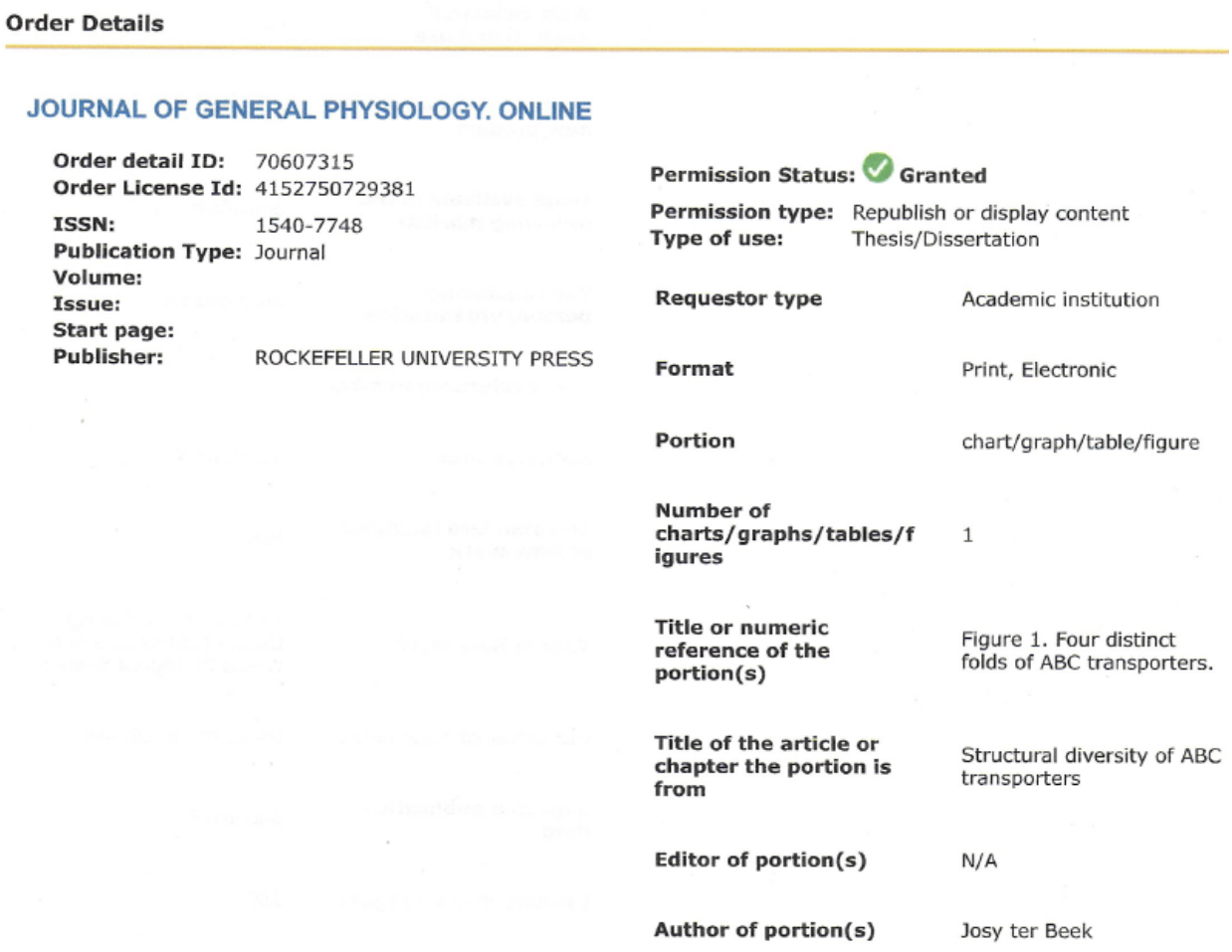
NATURE PUBLISHING GROUP LICENSE TERMS AND CONDITIONS

Jul 19, 2017

This Agreement between Mr. Xiucheng Cui ("You") and Nature Publishing Group ("Nature Publishing Group") consists of your license details and the terms and conditions provided by Nature Publishing Group and Copyright Clearance Center.

License Number	4152731271670
License date	Jul 19, 2017
Licensed Content Publisher	Nature Publishing Group
Licensed Content Publication	Nature Protocols
Licensed Content Title	Genome editing in rice and wheat using the CRISPR/Cas system
Licensed Content Author	Qiwei Shan, Yanpeng Wang, Jun Li, Caixia Gao
Licensed Content Date	Sep 18, 2014
Licensed Content Volume	9
Licensed Content Issue	10
Type of Use	reuse in a dissertation / thesis
Requestor type	academic/educational
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Figures	Figure 1: Schematic description of RNA-guided genome editing using the CRISPR/Cas system.
Author of this NPG article	no
Your reference number	
Title of your thesis / dissertation	Targeted Gene Editing Using CRISPR/Cas9 in a Wheat Protoplast System

Figure 3: by ter Beek, J., Guskov, A., & Slotboom, D. J. (2014). Structural diversity of ABC transporters. *The Journal of general physiology*, jgp-201411164.



Order Details

JOURNAL OF GENERAL PHYSIOLOGY. ONLINE

Order detail ID:	70607315
Order License Id:	4152750729381
ISSN:	1540-7748
Publication Type:	Journal
Volume:	
Issue:	
Start page:	
Publisher:	ROCKEFELLER UNIVERSITY PRESS

Permission Status: Granted

Permission type: Republish or display content
Type of use: Thesis/Dissertation

Requestor type	Academic institution
Format	Print, Electronic
Portion	chart/graph/table/figure
Number of charts/graphs/tables/figures	1
Title or numeric reference of the portion(s)	Figure 1. Four distinct folds of ABC transporters.
Title of the article or chapter the portion is from	Structural diversity of ABC transporters
Editor of portion(s)	N/A
Author of portion(s)	Josy ter Beek

Appendix 4: by Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B., & Weeks, D. P. (2013). Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic acids research*, *41*(20), e188-e188.

RightsLink®

Title: Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice

Author: Jiang, Wenzhi; Zhou, Huanbin

Publication: Nucleic Acids Research

Publisher: Oxford University Press

Date: 2013-08-31

Creative Commons