Genome editing in wheat with CRISPR/Cas9

Taj Arndell

A thesis submitted to the University of Adelaide

in fulfilment of the requirements for the degree of Doctor of Philosophy

Faculty of Sciences

School of Agriculture, Food and Wine

The University of Adelaide

July 2019

Table of contents

Abstract

Genetically engineered crops have the potential to play a key role in achieving global food security and transitioning to a more sustainable agriculture. In recent years, the CRISPR/Cas9 system has emerged as a powerful tool for crop genome editing. CRISPR/Cas9 enables the targeted and precise modification of plant genomes via the creation and subsequent repair of site-specific DNA double-strand breaks (DSBs). The system consists of the Cas9 endonuclease in complex with a small guide RNA (gRNA) that is designed to target a specific site in the genome. Site-specific DSBs generated by Cas9 are repaired through non-homologous end joining (NHEJ) or homology directed repair (HDR). NHEJ is error-prone and often produces small insertions or deletions (indels) that result in gene knockout. Alternatively, if an exogenous DNA donor template is delivered to the cell, then precise modifications can be made via HDR. The CRISPR/Cas9 system has been successfully applied to many model and crop plants. However, it can be difficult to achieve highly efficient and specific editing in polyploid species. Therefore, the main aim of this PhD project was to develop tools and methods for optimising the CRISPR/Cas9 for efficient and specific genome editing in hexaploid bread wheat (*Triticum aestivum*).

To test the efficacy of the CRISPR/Cas9 system for gene knockout, three gRNAs were designed to target *Ms1*, a male fertility gene that has been proposed for use in hybrid seed production. CRISPR/Cas9 vectors were delivered to immature embryos via *Agrobacterium*-mediated stable transformation, and the regenerated T_0 lines were screened for targeted indels produced via NHEJ. Only one of the three gRNAs was efficacious. Five per cent (2/40) of T_0 lines carrying the active gRNA were edited and male sterile, whereas all unedited lines were fully fertile. The recessive mutations were stably transmitted to the T_1, T_2 and T_3 generations, as was the male sterile phonotype.

Given the observed variability in the efficacy of different gRNAs targeting the same gene, and given that wheat transformation and tissue culture takes months and is laborious, a method was developed for the rapid assessment of gRNA activity and specificity. Seven gRNAs were designed

to target *EPSPS*, a gene involved in aromatic amino acid biosynthesis. CRISPR/Cas9 vectors were then transiently transformed into wheat protoplasts. Three out of the seven gRNAs induced mutations at moderate to high frequencies. gRNA specificity was correlated with the number and distribution of mismatches in the 'seed' region of the gRNA. One of the gRNAs was selected as potentially suitable for the development of non-transgenic herbicide resistant wheat lines.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Taj Arndell

Date: 30 Oct 2018

Acknowledgements

I am forever indebted to my academic supervisors, Dr. Ryan Whitford, Professor Peter Langridge and Dr. Ute Baumann, for their excellent guidance, unending support, and generosity.

Ryan, you taught me so much. Looking back, I can see that the most important things I learned during my PhD came from you. I have benefited immensely from your wisdom and foresight, and for that I am forever grateful. I appreciate the massive contribution of time, ideas and expertise that you gave to my project. Thank you for always being honest, and thank you for sharing your passion for science with me.

Peter, I feel extremely lucky to have had you as one of my supervisors. You are brilliant, yet humble, patient, gentle, and always willing to listen. Thank you for sharing your time and knowledge with me, and thank you for encouraging me throughout my PhD.

Ute, I am extremely grateful for all the help you gave me during my PhD. Thank you for always being willing to take the time to explain things until I was able to understand them. I learnt a great deal from you, and I greatly appreciate the substantial contributions that you made to my project.

To Professor Nikolai Borisjuk, my former supervisor, thank you for taking me on in the very beginning. I still have fond memories of our first meeting, when we sat outside on the balcony on that nice sunny day. Thank you for all the advice and support you gave me. It was a lot of fun to have you around.

To Associate Professor Matthew Tucker, my external advisor, thank you for the invaluable advice that you gave me throughout my PhD, and for being so generous with your time.

To Dr. Nicholas Collins, my mentor, thank you for being there for me whenever I needed help or advice. The pitcher plant you gave me is flowering again!

To Anzu Okada, thank you for all your support. I greatly enjoyed working with you. I don’t know what I would have done without you!

I would like to thank Dr. Niharika Sharma, Dr. Nathan Watson-Haigh and Elise Tucker for their contributions to my project.

I would also like to extend my thanks to the following people at the University of Adelaide who helped me during my PhD: Dr. Larissa Chirkova, Dr. Nataliya Kovalchuk, Dr. Ursula Langridge-Reimold, Dr. Takashi Okada, Dr. Allan Kouidri, Iryna Mazonka, Christine Kastner, Sheridan Morton, and Margaret Parlotta.

I would like to thank all my friends at The University of Adelaide. You know who you are.

I am grateful to the University of Adelaide for supporting me financially during my PhD.

Finally, I would like to thank my family for their love and support.

List of publications

1. Anzu Okada, **Taj Arndell**, Nikolai Borisjuk, Niharika Sharma, Nathan S. Watson-Haigh, Elise J. Tucker, Ute Baumann, Peter Langridge and Ryan Whitford (2019). CRISPR/Cas9-mediated knockout of *Ms1* enables the rapid generation of male-sterile hexaploid wheat lines for use in hybrid seed production. *Plant Biotechnology Journal*. https://doi.org/10.1111/pbi.13106

2. **Taj Arndell**, Niharika Sharma, Peter Langridge, Ute Baumann, Nathan S. Watson-Haigh and Ryan Whitford (2019). gRNA validation for wheat genome editing with the CRISPR-Cas9 system. (submitted for publication)

PREFACE

General introduction

Structure of the thesis

This thesis is presented as a series of three papers. The first paper is a review, which is intended for submission to a peer-reviewed journal. The other two papers are experimental manuscripts, which have been submitted for publication. In the **Preface** (this section), a general introduction is given to set the context for the thesis. **Chapter 1** is the review paper, which provides the background for the experimental manuscripts that follow. The review paper details various strategies for optimising the CRISPR/Cas9 system for different applications, and identifies the main technical challenges that must be overcome in order to realise the full potential of the technology. **Chapter 2** is a published experimental manuscript that reports on the successful application of the CRISPR/Cas9 system in wheat, for the generation of mutant lines that could be used in hybrid seed production. **Chapter 3** is an experimental manuscript that reports on the development of a method for optimising the CRISPR/Cas9 system using wheat protoplasts. Chapter 3 has been submitted for publication and is presented in the format specified by the target journal, with only minor changes to improve readability. Chapters 1-3 are prefaced by statements of authorship describing the authors' contributions. **Chapter 4** is a general discussion covering the overall significance of the work presented in the thesis, the contribution to knowledge, problems encountered, and future research directions.

This thesis is in agreement with The University of Adelaide Higher Degree by Research Specifications for Thesis 2018.

Context of the thesis

A major challenge of the 21st century is to ensure global food security for the growing world population in the face of climate change and dwindling resources. In order to meet this challenge, new crop breeding technologies must be developed to increase the rate of yield gain (Tester and Langridge, 2010). The ideal technology would be targeted, efficient, specific and flexible.

In the last decade, efficient and specific targeted mutagenesis has been achieved in a wide range of organisms (including plants) through the use of genome editing technologies such as engineered meganucleases (MNs) (Arnould *et al.*, 2011), zinc-finger nucleases (ZFNs) (Urnov *et al.*, 2010), and transcription activator-like effector nucleases (TALENs) (Mussolino and Cathomen, 2012). However, MNs, ZFNs and TALENs have not been widely adopted, as their targeting mechanisms are based on complex protein-DNA interactions (Sander and Joung, 2014). In 2012-2013, a revolutionary genome editing technology called CRICPR/Cas9 was introduced (Gasiunas *et al.*, 2012; Jinek *et al.*, 2012, 2013; Cho *et al.*, 2013; Cong *et al.*, 2013; Mali *et al.*, 2013). Compared with older genome editing technologies, the CRISPR/Cas9 system is simpler, much more flexible, and less expensive. Consequently, CRISPR/Cas9 technology was rapidly adopted by plant scientists (Feng *et al.*, 2013; Jiang *et al.*, 2013; Li *et al.*, 2013; Mao *et al.*, 2013; Miao *et al.*, 2013; Nekrasov *et al.*, 2013; Shan *et al.*, 2013; Xie and Yang, 2013).

Wheat (*Triticum spp.*) is the most widely grown cereal in the world with over 220 million ha planted annually. It provides about one-fifth of daily calories and protein for the human population, making it one of the most important staple foods and the world's most important protein source (Shiferaw *et al.*, 2013). Wheat dominates world trade in food crops, and is the largest food import in developing countries (Dixon *et al.*, 2009). Therefore, wheat plays a crucial role in global food security.

At the time this PhD project was started, there were only a few published reports of CRISPR/Cas9-mediated genome editing in wheat (Shan *et al.*, 2013; Upadhyay *et al.*, 2013; Wang *et al.*, 2014). The first two studies (Shan *et al.*, 2013; Upadhyay *et al.*, 2013) focused exclusively on testing the

functionality of the CRISPR/Cas9 system in isolated or cultured wheat cells. The latter of the three studies (Wang *et al.*, 2014) was the first to demonstrate the use of CRISPR/Cas9 for generating wheat plants carrying targeted mutations, although the mutations did not produce the desired phenotype. Moreover, it was not until two years later that the next report on CRISPR/Cas9-mediated genome editing in wheat was published (Zhang *et al.*, 2016). By contrast, reports of CRISPR/Cas9-mediated genome editing in other important crops like rice were abundant (around 30 publications by the end of 2016). Thus, it was (and still is) evident that genome editing in wheat with the CRISPR/Cas9 system can be a challenge. There are a few reasons for this: wheat has a large and complex (hexaploid) genome in containing many functionally redundant genes; until very recently, a high-quality annotated wheat reference genome sequence was not available; wheat is recalcitrant to transformation. Thus, there is a need to develop tools and methods that will improve the efficiency and specificity of the CRISPR/Cas9 system in wheat.

Research objectives

The objectives of this PhD thesis were to:

1. Test the CRISPR/Cas9 system and apply it to wheat for the production of mutant lines.
2. Develop tools and methods for optimising the CRISPR/Cas9 system for genome editing in wheat.

References

Arnould, S. *et al.* (2011) 'The I-CreI meganuclease and its engineered derivatives: applications from cell modification to gene therapy', *Protein Engineering, Design & Selection*, 24(1–2), pp. 27–31. doi: 10.1093/protein/gzq083.

Cho, S. W. *et al.* (2013) 'Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease', *Nature Biotechnology*. Nature Publishing Group, 31(3), pp. 230–232. doi: 10.1038/nbt.2507.

Cong, L. *et al.* (2013) 'Multiplex genome engineering using CRISPR/Cas systems', *Science*, 339, pp. 819–823. doi: 10.1126/science.1231143 RNA-Guided.

Dixon, J. *et al.* (2009) *Wheat Facts and Futures 2009*.

Feng, Z. *et al.* (2013) 'Efficient genome editing in plants using a CRISPR/Cas system', *Cell Research*, 23(10), pp. 1229–1232. doi: 10.1038/cr.2013.114.

Gasiunas, G. *et al.* (2012) 'Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria', *Proceedings of the National Academy of Sciences of the United States of America*, 109(39), pp. E2579-86. doi: 10.1073/pnas.1208507109.

Jiang, W. *et al.* (2013) 'Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice', *Nucleic Acids Research*, 41(20), p. e188. doi: 10.1093/nar/gkt780.

Jinek, M. *et al.* (2012) 'A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity', *Science*, 337, pp. 816–821. doi: 10.1126/science.1225829.

Jinek, M. *et al.* (2013) 'RNA-programmed genome editing in human cells', *eLife*, 2, p. e00471. doi: 10.7554/eLife.00471.

Li, J. *et al.* (2013) 'Multiplex and homologous recombination–mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9', *Nature Biotechnology*, 31(8),

pp. 688–691. doi: 10.1038/nbt.2650.

Mali, P. *et al.* (2013) 'RNA-guided human genome engineering via Cas9', *Science*, 339, pp. 823–826. doi: 10.1126/science.1232033.

Mao, Y. *et al.* (2013) 'Application of the CRISPR-Cas system for efficient genome engineering in plants', *Molecular Plant*, 6(6), pp. 2008–2011.

Miao, J. *et al.* (2013) 'Targeted mutagenesis in rice using CRISPR-Cas system', *Cell Research*, 23(10), pp. 1233–1236. doi: 10.1038/cr.2013.123.

Mussolino, C. and Cathomen, T. (2012) 'TALE nucleases: tailored genome engineering made easy', *Current Opinion in Biotechnology*. Elsevier Ltd, 23(5), pp. 644–650. doi: 10.1016/j.copbio.2012.01.013.

Nekrasov, V. *et al.* (2013) 'Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease', *Nature Biotechnology*, 31(8), pp. 691–693. doi: 10.1038/nbt.2654.

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

Shan, Q. *et al.* (2013) 'Targeted genome modification of crop plants using a CRISPR-Cas system', *Nature Biotechnology*, 31(8), pp. 686–688. doi: 10.1038/nbt.2652.

Shiferaw, B. *et al.* (2013) 'Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security', *Food Security*, 5(3), pp. 291–317. doi: 10.1007/s12571-013-0263-y.

Tester, M. and Langridge, P. (2010) 'Breeding technologies to increase crop production in a changing world', *Science*, 327(5967), pp. 818–822. doi: 10.1126/science.1183700.

Upadhyay, S. K. *et al.* (2013) 'RNA-guided genome editing for target gene mutations in wheat', *G3 Genes Genomes Genetics*, 3(12), pp. 2233–2238. doi: 10.1534/g3.113.008847.

Urnov, F. D. *et al.* (2010) 'Genome editing with engineered zinc finger nucleases', *Nature Reviews. Genetics*. Nature Publishing Group, 11(9), pp. 636–646. doi: 10.1038/nrg2842.

Wang, Y. *et al.* (2014) 'Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew', *Nature Biotechnology*, 32, pp. 947–951. doi: 10.1038/nbt.2969.

Xie, K. and Yang, Y. (2013) 'RNA-guided genome editing in plants using a CRISPR-Cas system', *Molecular Plant*, 6(6), pp. 1975–1983. doi: 10.1093/mp/sst119.

Zhang, Y. *et al.* (2016) 'Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA', *Nature Communications*, 7, p. 12617. doi: 10.1038/ncomms12617.

CHAPTER 1

Crop genome editing with CRISPR/Cas9: optimisation, applications, and challenges

Statement of authorship

Title of Paper	Crop genome editing with CRISPR/Cas9: optimisation, applications and challenges
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☑ Unpublished and Unsubmitted work written in manuscript style
Publication Details	This review manuscript is intended for submission to a peer-reviewed journal

Principal Author

Name of Principal Author (Candidate)	Taj Arndell		
Contribution to the Paper	Researched and wrote the manuscript		
Overall percentage (%)	100		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	29 Oct 2018

Crop genome editing with CRISPR/Cas9: optimisation, applications, and challenges

Taj Arndell

School of Agriculture, Food & Wine, The University of Adelaide, Waite Campus, Urrbrae, SA 5064, Australia

Summary

The CRISPR/Cas9 genome editing system is revolutionising the field of plant biology. Due to its efficiency, specificity and unparalleled flexibility, the technology has been rapidly and widely adopted for a range of applications such as multiplex gene knockout, targeted sequence insertion, and base editing. A variety of strategies have been developed for optimising CRISPR/Cas9 components and delivery systems. This review details the main optimisation strategies and identifies the key challenges that must be overcome in order to realise the full potential of the CRISPR/Cas9 system for genome editing in plants.

Introduction

For a long time, gene functional studies and crop breeding were limited by a reliance on imprecise techniques, namely crossbreeding, physical/chemical mutagenesis, and T-DNA insertion. The inability to target specific sequences in the genome necessitated the screening of large mutant populations in the hope of finding something interesting or potentially useful. Today, the field of plant biology is experiencing a revolution due to the emergence of the CRISPR/Cas9 (**c**lustered **r**egularly **i**nterspaced **s**hort **p**alindromic **r**epeat/CRISPR-associated protein 9) system as a tool for targeted and precise editing of the genome in any species.

CRISPR/Cas systems are prokaryotic adaptive immune systems that confer resistance to invasive nucleic acids such as viral genomes and plasmids [1]. Immunity is conferred by the activity of RNA-guided nucleases (such as Cas9), which target and cleave invading nucleic acids in a sequence-specific manner. CRISPR/Cas systems are divided into two classes, with multi-subunit effector complexes in Class 1 (types I, III and IV) and single protein effector modules in Class 2 (types II, V and VI) [2]. Several Class 2 systems have been repurposed for genome editing, and the CRISPR/Cas9 system (type II) is the most widely used.

In nature, the CRISPR/Cas9 system has three components: the Cas9 endonuclease, a CRISPR RNA (crRNA), and a trans-activating crRNA (tracrRNA). The crRNA, which contains a short 'spacer' (guide) sequence (previously captured from an invading nucleic acid and stored in the host genome) at the 5' end, hybridises to the tracrRNA and then the RNA hybrid molecule forms a complex with Cas9. Recognition of invading viral DNA by Cas9 is dependent on Watson-Crick base pairing between the crRNA guide and the corresponding viral 'protospacer' sequence, as well as the presence of a protospacer adjacent motif (PAM), which is located immediately 3' of the protospacer [3]. Following recognition, Cas9 cleaves the viral DNA and then dissociates.

To repurpose the CRISPR/Cas9 system for genome editing, the crRNA and tracrRNA were fused together into a single synthetic guide RNA (gRNA) that is capable of mediating targeted cleavage of DNA by Cas9 [3]. Subsequently, CRISPR/Cas9 mediated genome editing in human cells was demonstrated [4, 5]. Targeted mutagenesis was achieved by generating Cas9-induced DNA double-strand breaks (DSBs) at specific sites in endogenous genes. Repair of DSBs through error-prone non-homologous end joining (NHEJ) resulted in small insertions/deletions (indels) that disrupted gene function. Some DSBs were repaired though homology directed repair (HDR), which is a more precise DNA repair pathway. In addition, simultaneous editing of multiple loci (multiplex editing) was achieved through combined delivery of different gRNAs. Before long, the CRISPR/Cas9 system was validated in multiple plant species [6–12]. However, it quickly became apparent that optimisation of the CRISPR/Cas9 system would be needed to overcome the many challenges that arose. In this review, the main strategies for optimisation are detailed, and current challenges are identified and discussed with a view to providing guidance for future research.

gRNA design

A number of factors influence gRNA activity and specificity including PAM presence/absence, the PAM sequence, gRNA guide length, the target site sequence, and the presence of off-target sites

in the genome. For Cas9, the gRNA is ~100 nucleotides (nt) in length, and consists of the gRNA scaffold with a guide at the 5' end. The guide is typically 20 nt in length, and its sequence is defined by the target protospacer, which is located immediately upstream of a canonical 5'-NGG-3' PAM. The canonical cut site is 3 bp upstream of the PAM. In human cells, lengthening the guide to 30 nt resulted in lower on-target editing efficiencies, and no increase in specificity [13]. By contrast, truncated gRNAs (tru-gRNAs) containing guides that are 17-18 nt in length can increase specificity without sacrificing on-target editing efficiencies in human cells [14, 15] and *Drosophila* [16]. Efficient targeted mutagenesis has been demonstrated using tru-gRNAs in Arabidopsis [17] and tomato [18]. However, it remains unclear as to whether or not tru-gRNAs have higher specificity compared with their corresponding standard length gRNAs in plants. Shortening the guide to 14-16 nt results in a 'dead gRNA' (dRNA) that can home catalytically active Cas9 to its target site without inducing DSBs [19, 20]. dRNAs and gRNAs have been multiplexed together for simultaneous transcriptional regulation and genome editing using a single Cas9 protein in human cells [19, 20], thereby removing the need to use orthogonal Cas9 proteins from different species for such applications [21]. Thus, standard length guides are recommended for standard applications that require efficient and specific editing, but the use of shorter guides should be considered for applications that require higher specificity, and for advanced applications such as the one described above.

Another important gRNA design consideration is the sequence of the target site. Analyses of 1,841 gRNAs targeting a panel of nine genes in mammalian cells revealed sequence features that correlate with on-target gRNA activity [22]. In particular, it was found that certain nucleotides are favoured/disfavoured at certain positions, especially in and around the PAM [22]. For example, at the position immediately 5' of the PAM, G is strongly favoured and C is strongly disfavoured. The findings were used to develop an online tool called sgRNA Designer that provides gRNA activity scores based on a predictive model, for any input DNA sequence [22]. Subsequently, a number of other online tools for predicting on-target gRNA activity were developed, such as SSC [23], CRISPRscan [24], WU-CRISPR [25] and sgRNA Scorer 2.0 [26]. The sgRNA Designer tool was also

updated with an improved on-target scoring model called Azimuth 2.0 [27]. Some of these tools produce activity scores that take into account not just sequence features but also gRNA structural characteristics, which likely affect gRNA loading into Cas9 and target site recognition/binding. The choice of which tool(s) to use should be based on the nature of the experiment. sgRNA Designer and WU-CRISPR are recommended for gRNAs transcribed *in vivo* using the U6 promoter, whereas CRISPRscan is recommended for gRNAs transcribed *in vitro* using the T7 promoter. CCS provides a unique option to generate activity scores based on the application (genome editing or transcriptional regulation). sgRNA Scorer 2.0 is especially useful for experiments that use orthologous CRISPR systems. However, it should be noted that all of these tools were developed using data generated for animal systems. Position-specific nucleotide preferences that predict gRNA activity in animal systems were not found in a collection of 102 gRNAs that had previously been validated in plants [28]. Nevertheless, the tools are free, accessible, easy to use, and likely to have at least some predictive power for designing gRNAs that induce targeted mutations efficiently in plants.

Cas9 can generate unintended mutations at off-target sites that have sequence similarity to the target site. Several reports based on whole genome sequencing or targeted deep sequencing indicate that off-target mutations are rare in edited plants, especially when compared with pre-existing genetic variation, spontaneous mutations and somaclonal variation [29–33]. However, off-target mutations can sometimes occur at high frequencies [34]. It is desirable to avoid off-target mutations, and this can be achieved by designing gRNAs that are likely to be active only at their intended target sites. The absence of a canonical PAM site greatly reduces or abolishes gRNA activity, and single mismatches and 'bulges' (single-base insertions or deletions) within the PAM-proximal 5-13 nucleotide 'seed' region are generally not tolerated, especially in the presence of other mismatches [35–40]. Based on these and other gRNA/Cas9 binding rules, a number of free online bioinformatics tools have been developed to facilitate both the identification of potential off-target sites in the genome, and the design of gRNAs that have high predicted specificity. Some of these tools are applicable to plants e.g. E-CRISP [41], Breaking-Cas [42], CRISPR-P 2.0 [43],

CRISPR-PLANT v2 [44], and CRISPOR [45]. The key differences between the tools relate to the alignment and scoring algorithms, the adjustable parameters, and the genome sequences available for analyses. Although these bioinformatics tools are certainly useful, some gRNAs that they predict to be highly specific may still induce off-target mutations at high frequencies [34]. CIRCLE-Seq [46] is a highly sensitive *in vitro* method for genome-wide identification of potential off-target sites. The method is simple, rapid, and does not require a reference genome sequence. CIRCLE-Seq was recently used to identify potential off-target sites in edited maize plants [31]. *In vivo* methods for genome-wide identification of potential off-target sites have also been developed, but have yet to be applied to plants. GUIDE-Seq is one such method that could potentially be applied to plant protoplasts [47]. However, given that off-target mutations are rare in plants, and considering that off-target mutations can be removed through backcrossing (assuming they are not linked to the target locus), *in silico* identification of potential off-target sites and/or CIRCLE-seq will usually suffice.

gRNA expression systems

High levels of gRNA expression are needed for efficient genome editing with the CRISPR/Cas9 system. Indeed, early studies reporting on the application of CRISPR/Cas9 in human cells and plant protoplasts indicated that gRNA expression and/or loading into Cas9 may be the limiting factor for Cas9-mediated DNA cleavage [7, 48]. In both animals and plants, the snRNA U6 Pol III promoter is widely used for strong and constitutive gRNA expression. The U6 promoter requires a G at the +1 position for transcription initiation, which means that the nucleotide at the 5’ end of the gRNA must be a G. Therefore, when using the U6 promoter for gRNA expression, it is recommended to either select target sites that have a G at the 5’ end, or else append a mismatched G to the 5’ end of the gRNA [49]. Alternatively, gRNA expression can be driven by the snRNA U3 Pol III promoter, which requires an A for transcription initiation. In either case, the use of an optimised gRNA scaffold in which the internal Pol III pause signal (present in the native

sequence) is mutated may increase transcription and editing efficiencies, as demonstrated in human cells [50].

The CRISPR/Cas9 system lends itself well to multiplexing, but construction and delivery of multiple gRNA vectors or single vectors containing multiple gRNA expression cassettes is not always straightforward. To address this issue, optimised expression systems for multiplex plant genome editing have been developed. One approach is to stack multiple gRNA expression cassettes, each containing a different Poll III promoter, into a single plasmid using the Golden Gate or Gibson Assembly cloning methods [51–53]. Using this approach, up to eight gRNA expression cassettes may be combined into a single vector for efficient delivery to plant cells [51, 53]. However, cloning of six or more gRNA expression cassettes into a single vector can be challenging [53]. Multiple gRNAs can also be expressed from a single polycistronic gene. In this case, the polycistronic transcripts are processed into individual gRNAs by the bacterial CRISPR endoribonuclease Csy4 [54, 55], endogenous tRNA processing enzymes [56], or ribozymes [57]. This approach has significant advantages: it allows a larger number of gRNAs to be cloned into a single vector (up to 12 so far); it allows the use of a constitutive, inducible, or tissue-specific Pol II promoter to drive the expression of all of the gRNAs; in the case of the Cys4 and tRNA systems, it can lead to higher editing efficiencies in plants compared with using constructs containing multiple gRNA expression cassettes [58]. Therefore, the use of a polycistronic construct is recommended for simultaneous expression of multiple gRNAs *in vivo*, especially when more than five gRNAs are multiplexed.

gRNAs can also be expressed *in vitro.* The easiest way to do this is to use a commercial kit. *In vitro* gRNA expression is driven by the T7 promoter, which requires a G or (ideally) GG at the 5’ end of the gRNA to initiate transcription. The purified gRNA transcripts are mixed with purified Cas9 to produce ribonucleoprotein (RNP) complexes that can be delivered directly to animal or plant cells.

Cas9 modifications and Cpf1

Cas9 has been modified in a variety of ways to improve its performance for different applications. Plant codon optimised versions of Cas9 are widely used, and in most cases they show superior performance compared with versions of Cas9 that are optimised for animals. For example, in rice, the use of rice codon optimised Cas9 resulted in editing efficiencies between 87-100%, whereas the efficiency of editing with human codon optimised Cas9 was only 7% [59]. In Arabidopsis protoplasts, plant codon optimised Cas9 was expressed at a substantially higher level than the human codon optimised Cas9 using the same expression vector [7]. Rice codon optimised Cas9 performs well in wheat [6], presumably because they are both monocots.

Cas9 nickases contain mutations that inactivate one of the two nuclease domains (RuvC or HNH), so that single strand DNA nicks are induced instead of DSBs. The most widely used Cas9 nickase is Cas9-D10A, whose RuvC domain is inactivated. The primary motivation for using Cas9 nickases is to reduce unwanted off-target mutations. This is achieved by using nickases in a paired configuration, such that DSBs can be created at the target site, while any single strand nicks created at off-target sites are repaired faithfully. In Arabidopsis, the use of Cas9-D10A in a non-paired configuration reduced the frequency of DSBs at the target site by at least 740-fold compared with Cas9, yet Cas9-D10A and Cas9 stimulated intra-chromosomal homologous recombination at similar frequencies [60]. Moreover, when Cas9-D10A was used in a paired arrangement, the frequency of DSB induction was comparable to that of Cas9 [61]. In rice, unwanted off-target mutations were completely avoided by using paired Cas9-D10A, but on-target editing efficiencies also decreased [62]. Taken together, these studies suggest that the paired Cas9-D10A method can greatly increase specificity, although highly active gRNAs should be used to ensure efficient induction of DSBs at the target site. Further studies are needed to determine whether non-paired Cas9 nickases can be used to generate single strand nicks that are repaired efficiently via the HDR pathway [63].

CRISPR/Cpf1 is a type V system that has received much attention as an alternative to CRISPR/Cas9. Unlike Cas9, which generates blunt-ended DSBs, Cpf1 produces DSBs with 5 nt overhangs ('sticky ends') [64]. These sticky ends provide more control over targeted insertions. Also, Cpf1 recognises a 5'-TTTN-3' PAM, and therefore expands the options for target site selection. Cpf1 induced mutations at almost 100% efficiency for the four independent target sites assessed in rice [65]. Cpf1 was also delivered as RNPs to soybean and tobacco for efficient and specific editing [66]. These initial results demonstrate that CRISPR/Cpf1 is a powerful tool for genome editing in plants. The next step is to test Cpf1 for targeted sequence insertion. Previously, targeted insertions (exon replacement) were generated in rice using Cas9 and an intron-targeting strategy [67]. However, the efficiency of sequence insertion was low. Therefore, Cpf1 may be a better alternative for such applications.

Delivery of CRISPR/Cas9 components to plant cells

Delivery of CRISPR/Cas9 components has been, and continues to be, a major challenge in plants. This is largely due to the plant cell wall, which acts as a barrier to delivery of exogenous biomolecules. In addition, plant regeneration in tissue culture (which is considered part of the transformation process) is time-consuming, laborious, and inefficient in most species/cultivars. Together, these two limitations currently form the main bottleneck in applying the CRISPR/Cas9 system to plants [68].

Many of the methods used for delivering CRISPR/Cas9 components to plant cells are essentially the same as those used for conventional transgene delivery. PEG-mediated transformation of protoplasts is highly efficient but plant regeneration from protoplasts is currently unachievable in most species. Consequently, protoplasts have mainly been used for the rapid validation of CRISPR/Cas9 system components in transient assays [6, 7, 9, 10, 69], although a notable exception is that edited potato plants have been regenerated from transformed protoplasts [70]. *Agrobacterium*-mediated transformation is the most widely used delivery system for

CRISPR/Cas9-mediated plant genome editing, and the method varies depending on the species. Arabidopsis is one of very few species that can be stably transformed using an *in planta* method, namely floral dip [71], and this method is used routinely to produce edited Arabidopsis plants [10, 12, 60, 72–75]. The main advantage of *in planta* transformation is that there is no tissue culture step. For other species, the process involves co-cultivation of *Agrobacterium* with calli or explants (immature embryos, cotyledons, leaf discs, stem segments etc.) followed by tissue culture and regeneration of stably transformed (and hopefully edited) plants expressing the CRISPR/Cas9 transgenes [10, 12, 76–79]. In tobacco, the CRISPR/Cas9 transgenes can be delivered via leaf Agroinfiltration, and stably transformed plants can be regenerated via tissue culture using explants obtained from the infiltrated leaves [8]. Biolistics (particle bombardment) is also widely used for the delivery of CRISPR/Cas9 transgenes to immature embryos or calli of various crops including wheat [80, 81], rice [67], maize [82, 83] and soybean [84]. Using biolistics in wheat it has been shown that edited plants can be recovered without selecting for stable integration of the CRISPR/Cas9 transgenes during the tissue culture step [85]. The methods, called TECCDNA (Transient Expression of CRISPR/Cas9 DNA) and TECCRNA (Transient Expression of CRISPR/Cas9 RNA), work because transient expression of CRISPR/Cas9 transgenes or *in vitro* transcribed RNA can provide sufficient amounts of Cas9 and gRNA to effect editing. One benefit of using selection-free conditions is that it takes only 6-8 weeks to obtain testable wheat plants, instead of 8-12 weeks. Also, in the case of TECCRNA, the delivery of *in vitro* transcribed RNA instead of DNA greatly reduces the chance of stable integration, although TECCRNA is more laborious (due to the *in vitro* transcription step) and results in lower frequencies of targeted mutagenesis [85]. Typically, stably integrated CRISPR/Cas9 transgenes are removed from the genomes of edited plants by conventional breeding, and the final product is often presented as non-transgenic. However, concerns have been raised about the potential for small insertions of vector DNA to go unnoticed when PCR and/or Southern blot analyses are used to confirm the absence of CRISPR/Cas9 transgenes in the genome [86].

An alternative to delivering CRISPR/Cas9 DNA or RNA is to deliver *in vitro* assembled RNP complexes. Genome editing with RNPs was first demonstrated in human cells [87] and later the method was adapted for plants. In the first demonstration in plants, RNPs were delivered to Arabidopsis, tobacco, rice and lettuce via PEG-mediated transformation of protoplasts, and edited lettuce plants were recovered [88]. RNPs can also be delivered using biolistics, as demonstrated in maize and wheat [82, 89]. These studies showed that delivery of RNPs instead of DNA vectors results in lower frequencies of targeted mutagenesis and much higher specificity. Most importantly, RNP-based genome editing does not require the delivery of a DNA vector, and thus the risk of DNA vector integration is eliminated. Therefore, the edited plants may be exempted from regulatory scrutiny.

With the exception of Arabidopsis floral dip, all of the methods described above include a tissue culture step for plant regeneration. Many attempts have been made to develop *Agrobacterium*-mediated *in planta* transformation methods for species other than Arabidopsis. An example is wheat, which is considered to be one of the more recalcitrant species for transformation. Wheat seed transformation [90, 91] and floral transformation [92] have been reported, but the methods have not proven reliable enough to replace well-established tissue culture-based methods.

A few single-stranded plant RNA viruses have been engineered as vectors for *in planta* and systemic delivery of CRISPR/Cas9 components. The first is *Tobacco rattle virus* (TRV), whose host range under laboratory conditions includes more than 400 species. *Agrobacterium*-mediated delivery of TRV-gRNA vectors to tobacco plants overexpressing Cas9 resulted in highly efficient targeted mutagenesis in inoculated and systemic leaves [93]. Targeted mutations were also detected in 1.7% (2/120) of M1 seedlings derived from early flowers [93]. However, one or both of the M1 seedlings were chimeric, and heritability of the targeted mutations was not demonstrated. Recently, the application of TRV-gRNA vectors was extended to Arabidopsis, and it was shown that *Pea early-browning virus* (PEBV) can be used as a vector for CRISPR/Cas9-mediated genome editing in the same way as TRV [94]. It will be important to determine whether or not TRV and/or PEBV-based CRISPR/Cas9 systems can generate targeted mutations that are

heritable. Plant DNA viruses such as the bipartite begomoviruses (Geminiviridae) and their associated DNA satellite molecules could potentially be engineered as vectors for systemic *in planta* delivery of CRISPR/Cas9 components, and co-delivery of different kinds of viral vectors (e.g. geminivirus vectors and TRV vectors) also remains to be explored [95]. It may even be possible to create synthetic viral vectors built from scratch using genetic elements from different viruses. As well as the need to generate heritable edits, one of the main challenges is to overcome the limited cargo capacity of viral vectors. Ideally, the viral vector would carry all of the CRISPR/Cas9 components (instead of just the gRNAs) and move efficiently between cells for systemic infection. However, due to the large size of the Cas9 gene (>4 kb), this has yet to be achieved. One potential solution might be to use a small Cas9 ortholog such as SaCas9 (3.16 kb) [96] or CjCas9 (2.95 kb) [97]. Depending on the virus used, other issues may arise such as the potential for negative effects on plant growth and reproduction due to systemic viral infection.

The potential of nanoparticles to overcome many of the limitations in conventional plant transformation was recently highlighted [98]. Nanoparticles are promising materials for species-independent *in planta* delivery of diverse cargo including CRISPR/Cas9 transgenes and RNPs. A recent study demonstrated efficient diffusion-based carbon nanotube (CNT)-mediated delivery of DNA reporter constructs to mature leaves of wheat and *Eruca sativa*, which resulted in strong transient protein expression [99]. The same study also demonstrated highly efficient gene silencing in tobacco leaves following CNT-mediated delivery of siRNA. Although much more work will be needed to optimise nanoparticle-based delivery systems for *in planta* germline transformation, these early results are certainly encouraging.

Gene modification via non-homologous end joining (NHEJ)

Non-homologous end joining is the primary DSB repair pathway in somatic plant cells [100], and is therefore relatively easy to harness for genome editing. This repair pathway is error prone and often introduces a small insertion or deletion (indel) at the break site, which can shift the open

reading frame of a coding sequence or create a premature stop codon, resulting in gene knockout. The CRISPR/Cas9 system has been used for gene knockout to study gene function, and to engineer a variety of traits including disease resistance [101], altered development [102], male-sterility [103–105], higher yield [106], and improved nutrition [107]. Multiplex CRISPR/Cas9 systems can be used to knock out multiple genes simultaneously [106, 108], and recently this strategy was used for *de novo* domestication of wild tomato [109]. To achieve gene knockout with the CRISPR/Cas9 system, it is generally recommended to target an early exon because gene activity is much less likely to be abolished if the indel is induced at the 3' end of the coding sequence or in an intron. However, if the target gene encodes multiple proteins as a result of alternative splicing, then the introduction of a frameshift mutation or stop codon in an early exon may not result in gene knockout. In this case, complete knockout can be achieved through gene deletion, using a multiplex CRISPR/Cas9 system targeting the 5' and 3' ends of the gene. Multiplex CRISPR/Cas9 systems can also be used to create large mutations such as chromosomal deletions and translocations. For example, chromosomal deletions (115-245 kb) were created to delete gene clusters in rice [59], and a recent study demonstrated chromosomal translocations in Arabidopsis [110].

Gene knockout in polyploid species is particularly challenging due to gene functional redundancy. For example, in hexaploid wheat, there are six alleles of most genes, and in most cases all six alleles must be knocked out in order to generate the mutant phenotype. This is difficult to achieve in the first generation; even if all six alleles are edited (which is not easy to achieve in itself), some of the targeted mutations may not create a frame shift or premature stop codon. Therefore, additional breeding steps (selfing and/or crossing) are often needed to combine null mutations or make them homozygous [81, 105]. The problem is exacerbated when multiple genes are targeted simultaneously, which probably explains why multiplex editing has rarely been attempted in wheat (but see [111]). The use of highly active gRNAs, combined with efficient delivery of CRISPR/Cas9 components, can help to overcome the problem of gene functional redundancy in polyploids.

Gene modification via homology directed repair (HDR)

Homology directed repair of Cas9-induced DSBs or nicks can occur if a double-stranded DNA (dsDNA) or single-stranded DNA (ssDNA) donor template is delivered to the nucleus. The donor template contains the desired mutation, which is inserted or copied into the genome at the desired location during HDR. Thus, unlike NHEJ, HDR can be harnessed for targeted and precise sequence deletion/insertion/replacement. Unfortunately, however, HDR occurs at very low frequency in plants (it competes with the NHEJ repair pathway), and is therefore difficult to harness for genome editing.

There are a number of ways to increase the frequency of HDR. In plants, the most successful method developed so far is the use of deconstructed mastrevirus (Geminiviridae) vectors for donor template delivery. These vectors lack viral coat and movement proteins (hence 'deconstructed'), and are therefore not capable of cell-to-cell movement. The method was first demonstrated in tobacco using deconstructed bean yellow dwarf virus [112]. In that report, it was shown that the frequency of HDR is dramatically increased due to replication of the donor template in the nucleus (higher donor template concentration) and pleiotropic activity of the mastrevirus Rep/RepA (replication) proteins. The method was then used to precisely insert a strong promoter upstream of a tomato gene (*ANT1*) controlling anthocyanin biosynthesis, resulting in overexpression and accumulation of pigments in foliage, flowers and fruit [113]. Also, point mutations (amino acid substitutions) were introduced into potato *ALS1*, conferring reduced herbicide susceptibility [114]. Later, the method was adapted for wheat [115] and rice [116] using deconstructed wheat dwarf virus vectors, and recently the bean yellow dwarf virus system was further optimised for tomato [117].

A viable alternative to deconstructed mastrevirus vectors is to use biolistics to deliver a large quantity of the donor template. This strategy has been used to generate targeted and precise edits via HDR in maize [82, 118] and rice [119]. In human cells, HDR frequencies were increased by

rationally designing a ssDNA donor template to match the properties of the Cas9-DNA complex [120]. It would be well worth testing this method in plants.

Another way to potentially increase the frequency of HDR in plants is through modulation of DNA repair pathways. In mammalian cells, inhibition of DNA ligase IV (a key enzyme in the NHEJ pathway) by the small molecule Scr7 increased the frequency of CRISPR/Cas9-mediated HDR up to 19-fold [121]. In Arabidopsis, the absence of KU70 (an important DNA repair protein involved in NHEJ) led to a five to 16-fold enhancement in zinc finger nuclease-mediated HDR [122]. In plants, homologous recombination can be induced via heat shock [123] or treatment with UV-B radiation [124].

Recently, it was shown that CRISPR/Cas9-mediated induction of DSBs in somatic cells can result in targeted recombination between homologous chromosomes [125]. In theory, the CRISPR/Cas9 system could also be used for targeted meiotic recombination, which would greatly increase the efficiency of conventional breeding.

Base editing

Tools have been created that enable editing of target nucleotides ('base editing') in plants without the creation of DSBs and without the need to deliver a DNA donor template. Originally developed for base editing in mammalian cells [126], these tools were rapidly adapted for use in plants. Cas9 nickase-cytidine deaminase fusions have been used to generate C to T (or G to A) transitions within a ~7 bp window at the target site in rice, wheat, maize and tomato [127–129]. Recently, a new plant adenine base editor (Cas9 nickase-adenosine deaminase fusion) was developed, enabling A to G (or T to C) transitions [130]. Thus, it is now possible to edit all base pairs. Base editing is highly efficient but currently there are several limitations. First, as base editing occurs within a window, there can be a lack of precision and/or a lack of coverage. Second, the nucleotide transitions are not the same every time. For example, the Cas9 nickase-cytidine deaminase fusion can sometimes generate C to G transversions [129]. Also, the Cas9 nickase

fusions can generate DSBs, albeit at low frequencies. Therefore, further optimisation of base editing tools is needed to improve the precision/coverage and fidelity. Recently, a new base editing tool with a broadened editing window and enhanced fidelity was developed [131]. Moreover, CRISPR/Cas9 base editors can be delivered as RNPs for DNA-free base editing [132]. Thus, it is evident that base editing has great potential for crop improvement.

Discussion

The CRISPR/Cas9 system is a powerful tool for plant genome editing, owing to its simplicity, flexibility, efficiency and specificity. Each of the system components can be optimised in a variety of ways to improve editing outcomes. However, in order to realise the full potential of the CRISPR/Cas9 system, a number of significant technical challenges must be overcome. Currently, the main bottleneck is plant transformation, especially the dependence on tissue culture for plant regeneration. Therefore, the development of simple and reliable methods for *in planta* delivery of CRISPR/Cas9 components would greatly accelerate progress. Virus-mediated delivery and nanoparticle-mediated delivery are potential solutions that should be explored. The other major challenge is to optimise the CRISPR/Cas9 system and other CRISPR systems for highly efficient and precise multiplex editing, especially via HDR and base editing. This would provide a key tool for the rapid development of new crop varieties, which is necessary to meet the much bigger challenge of global food security.

References

1. Horvath P, Barrangou R. CRISPR/Cas, the immune system of Bacteria and Archaea. Science. 2010;327:167–70.

2. Koonin E V, Makarova KS, Zhang F. Diversity, classification and evolution of CRISPR-Cas systems. Curr Opin Microbiol. 2017;37:67–78.

3. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012;337:816–21.

4. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013;339:819–23.

5. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human genome engineering via Cas9. Science. 2013;339:823–6.

6. Shan Q, Wang Y, Li J, Zhang Y, Chen K, Liang Z, et al. Targeted genome modification of crop plants using a CRISPR-Cas system. Nat Biotechnol. 2013;31:686–8.

7. Li J, Norville JE, Aach J, McCormack M, Zhang D, Bush J, et al. Multiplex and homologous recombination–mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. Nat Biotechnol. 2013;31:688–91.

8. Nekrasov V, Staskawicz B, Weigel D, Jones JDG, Kamoun S. Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. Nat Biotechnol. 2013;31:691–3.

9. Xie K, Yang Y. RNA-guided genome editing in plants using a CRISPR-Cas system. Mol Plant. 2013;6:1975–83.

10. Mao Y, Zhang H, Xu N, Zhang B, Gou F, Zhu J-K. Application of the CRISPR-Cas system for efficient genome engineering in plants. Mol Plant. 2013;6:2008–11.

11. Miao J, Guo D, Zhang J, Huang Q, Qin G, Zhang X, et al. Targeted mutagenesis in rice using CRISPR-Cas system. Cell Res. 2013;23:1233–6.

12. Feng Z, Zhang B, Ding W, Liu X, Yang D-L, Wei P, et al. Efficient genome editing in plants using a CRISPR/Cas system. Cell Res. 2013;23:1229–32.

13. Ran FA, Hsu PD, Lin C-Y, Gootenberg JS, Konermann S, Trevino AE, et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell. 2013;154:1380–9.

14. Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat Biotechnol. 2014;32:279–84.

15. Wyvekens N, Topkar V V, Khayter C, Joung JK, Tsai SQ. Dimeric CRISPR RNA-guided FokI-dCas9 nucleases directed by truncated gRNAs for highly specific genome editing. Hum Gene Ther. 2015;26:425–31.

16. Ren X, Yang Z, Xu J, Sun J, Mao D, Hu Y, et al. Enhanced specificity and efficiency of the CRISPR/Cas9 system with optimized sgRNA parameters in *Drosophila*. Cell Rep. 2014;9:1151–62.

17. Osakabe Y, Watanabe T, Sugano SS, Ueta R, Ishihara R, Shinozaki K, et al. Optimization of CRISPR/Cas9 genome editing to modify abiotic stress responses in plants. Sci Rep. 2016;6:26685.

18. Ueta R, Abe C, Watanabe T, Sugano SS, Ishihara R, Ezura H, et al. Rapid breeding of parthenocarpic tomato plants using CRISPR/Cas9. Sci Rep. 2017;7:507.

19. Dahlman JE, Abudayyeh OO, Joung J, Gootenberg JS, Zhang F, Konermann S. Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease. Nat Biotechnol. 2015;33:1159–61.

20. Kiani S, Chavez A, Tuttle M, Hall RN, Chari R, Ter-Ovanesyan D, et al. Cas9 gRNA engineering for genome editing, activation and repression. Nat Methods. 2015;12:1051–4.

21. Esvelt KM, Mali P, Braff JL, Moosburner M, Yaung SJ, Church GM. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. Nat Methods. 2013;10:1116–21.

22. Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, et al. Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. Nat Biotechnol. 2014;32:1262–7.

23. Xu H, Xiao T, Chen C-H, Li W, Meyer CA, Wu Q, et al. Sequence determinants of improved CRISPR sgRNA design. Genome Res. 2015;25:1147–57.

24. Moreno-Mateos MA, Vejnar CE, Beaudoin J-D, Fernandez JP, Mis EK, Khokha MK, et al.

CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. Nat Methods. 2015;12:982–8.

25. Wong N, Liu W, Wang X. WU-CRISPR: characteristics of functional guide RNAs for the CRISPR/Cas9 system. Genome Biol. 2015;16:218.

26. Chari R, Yeo NC, Chavez A, Church GM. sgRNA Scorer 2.0: A species-independent model to predict CRISPR/Cas9 activity. ACS Synth Biol. 2017;6:902–4.

27. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat Biotechnol. 2016;34:184–91.

28. Liang G, Zhang H, Lou D, Yu D. Selection of highly efficient sgRNAs for CRISPR/Cas9-based plant genome editing. Sci Rep. 2016;6:21451.

29. Feng Z, Mao Y, Xu N, Zhang B, Wei P, Yang D-L, et al. Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in *Arabidopsis*. Proc Natl Acad Sci USA. 2014;111:4632–7.

30. Peterson B a., Haak DC, Nishimura MT, Teixeira PJPL, James SR, Dangl JL, et al. Genome-wide assessment of efficiency and specificity in CRISPR/Cas9 mediated multiple site targeting in Arabidopsis. PLoS One. 2016;11:e0162169.

31. Lee K, Zhang Y, Kleinstiver BP, Guo JA, Aryee MJ, Miller J, et al. Activities and specificities of CRISPR/Cas9 and Cas12a nucleases for targeted mutagenesis in maize. Plant Biotechnol J. 2018.

32. Tang X, Liu G, Zhou J, Ren Q, You Q, Tian L, et al. A large-scale whole-genome sequencing analysis reveals highly specific genome editing by both Cas9 and Cpf1 (Cas12a) nucleases in rice. Genome Biol. 2018;19:84.

33. Li J, Manghwar H, Sun L, Wang P, Wang G, Sheng H, et al. Whole genome sequencing reveals rare off-target mutations and considerable inherent genetic or/and somaclonal variations in

CRISPR/Cas9-edited cotton plants. Plant Biotechnol J. 2018.

34. Zhang Q, Xing H-L, Wang Z-P, Zhang H-Y, Yang F, Wang X-C, et al. Potential high-frequency off-target mutagenesis induced by CRISPR/Cas9 in *Arabidopsis* and its prevention. Plant Mol Biol. 2018;96:445–56.

35. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol. 2013;31:827–32.

36. Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature. 2014;507:62–7.

37. Wu X, Scott DA, Kriz AJ, Chiu AC, Hsu PD, Dadon DB, et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Nat Biotechnol. 2014;32:670–6.

38. Lin Y, Cradick TJ, Brown MT, Deshmukh H, Ranjan P, Sarode N, et al. CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. Nucleic Acids Res. 2014;42:7473–85.

39. Cho SW, Kim S, Kim Y, Kweon J, Kim HS, Bae S, et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res. 2014;24:132–41.

40. O'Geen H, Henry IM, Bhakta MS, Meckler JF, Segal DJ. A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. Nucleic Acids Res. 2015;43:3389–404.

41. Heigwer F, Kerr G, Boutros M. E-CRISP: fast CRISPR target site identification. Nature Methods. 2014;11:122–3.

42. Oliveros JC, Franch M, Tabas-Madrid D, San-León D, Montoliu L, Cubas P, et al. Breaking-Cas—interactive design of guide RNAs for CRISPR-Cas experiments for ENSEMBL genomes. Nucleic Acids Res. 2016;44:W267–71.

43. Liu H, Ding Y, Zhou Y, Jin W, Xie K, Chen L-L. CRISPR-P 2.0: an improved CRISPR-Cas9 tool for genome editing in plants. Mol Plant. 2017;10:530–2.

44. Minkenberg B, Zhang J, Xie K, Yang Y. CRISPR-PLANT v2: An online resource for highly specific guide RNA spacers based on improved off-target analysis. Plant Biotechnol J. 2018;:Epub ahead of print.

45. Concordet J-P, Haeussler M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. Nucleic Acids Res. 2018;46:W242–5.

46. Tsai SQ, Nguyen NT, Malagon-Lopez J, Topkar V V, Aryee MJ, Joung JK. CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. Nat Methods. 2017;14:607–14.

47. Shi L, Tang X, Tang G. GUIDE-seq to detect genome-wide double-stranded breaks in plants. Trends Plant Sci. 2016;21:815–8.

48. Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J. RNA-programmed genome editing in human cells. Elife. 2013;2:e00471.

49. Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol. 2014;32:347–55.

50. Dang Y, Jia G, Choi J, Ma H, Anaya E, Ye C, et al. Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. Genome Biol. 2015;16:280.

51. Xing H-L, Dong L, Wang Z-P, Zhang H-Y, Han C-Y, Liu B, et al. A CRISPR/Cas9 toolkit for multiplex genome editing in plants. BMC Plant Biol. 2014;14:327.

52. Ma X, Zhang Q, Zhu Q, Liu W, Chen Y, Qiu R, et al. A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicotplants. Mol Plant. 2015;:1–11.

53. Lowder LG, Zhang D, Baltes NJ, Paul JW, Tang X, Zheng X, et al. A CRISPR/Cas9 toolbox for multiplexed plant genome editing and transcriptional regulation. Plant Physiol. 2015;169:971–85.

54. Nissim L, Perli SD, Fridkin A, Perez-Pinera P, Lu TK. Multiplexed and programmable regulation

of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. Mol Cell. 2014;54:698–710.

55. Tsai SQ, Wyvekens N, Khayter C, Foden JA, Thapar V, Reyon D, et al. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. Nat Biotechnol. 2014;32:569–76.

56. Xie K, Minkenberg B, Yang Y. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. Proc Natl Acad Sci USA. 2015;:3–8.

57. Tang X, Zheng X, Qi Y, Zhang D, Cheng Y, Tang A, et al. A single transcript CRISPR-Cas9 system for efficient genome editing in plants. Mol Plant. 2016.

58. Cermak T, Curtin SJ, Gil-Humanes J, Cegan R, Kono TJY, Konecna E, et al. A multipurpose toolkit to enable advanced genome engineering in plants. Plant Cell. 2017;29:1196–217.

59. Zhou H, Liu B, Weeks DP, Spalding MH, Yang B. Large chromosomal deletions and heritable small genetic changes induced by CRISPR/Cas9 in rice. Nucleic Acids Res. 2014;:Advance access.

60. Fauser F, Schiml S, Puchta H. Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in Arabidopsis thaliana. Plant J. 2014;79:348–59.

61. Schiml S, Fauser F, Puchta H. The CRISPR/Cas system can be used as nuclease for in planta gene targeting and as paired nickases for directed mutagenesis in Arabidopsis resulting in heritable progeny. Plant J. 2014;9.

62. Mikami M, Toki S, Endo M. Precision targeted mutagenesis via Cas9 paired nickases in rice. Plant Cell Physiol. 2016;57:1058–68.

63. Davis L, Maizels N. Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. Proc Natl Acad Sci USA. 2014;111:E924–32.

64. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, et al. Cpf1 Is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell. 2015;163:759–71.

65. Tang X, Lowder LG, Zhang T, Malzahn A a., Zheng X, Voytas DF, et al. A CRISPR–Cpf1 system for

efficient genome editing and transcriptional repression in plants. Nat Plants. 2017;3 February:17018.

66. Kim H, Kim S-T, Ryu J, Kang B-C, Kim J-S, Kim S-G. CRISPR/Cpf1-mediated DNA-free plant genome editing. Nat Commun. 2017;8:14406.

67. Li J, Meng X, Zong Y, Chen K, Zhang H, Liu J, et al. Gene replacements and insertions in rice by intron targeting using CRISPR-Cas9. Nat Plants. 2016;2:16139.

68. Altpeter F, Springer NM, Bartley LE, Blechl A, Brutnell TP, Citovsky V, et al. Advancing crop transformation in the era of genome editing. Plant Cell. 2016;28:tpc.00196.2016.

69. Jiang W, Zhou H, Bi H, Fromm M, Yang B, Weeks DP. Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. Nucleic Acids Res. 2013;41:e188.

70. Andersson M, Turesson H, Nicolia A, Fält AS, Samuelsson M, Hofvander P. Efficient targeted multiallelic mutagenesis in tetraploid potato (*Solanum tuberosum*) by transient CRISPR-Cas9 expression in protoplasts. Plant Cell Rep. 2017;36:117–28.

71. Clough SJ, Bent AF. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J. 1998;16:735–43.

72. Hyun Y, Kim J, Cho SW, Choi Y, Kim JS, Coupland G. Site-directed mutagenesis in Arabidopsis thaliana using dividing tissue-targeted RGEN of the CRISPR/Cas system to generate heritable null alleles. Planta. 2014;241:271–84.

73. Gao X, Chen J, Dai X, Zhang D, Zhao Y. An effective strategy for reliably isolating heritable and *Cas9*-free Arabidopsis mutants generated by CRISPR/Cas9-mediated genome editing. Plant Physiol. 2016;171:1794–800.

74. Tsutsui H, Higashiyama T. pKAMA-ITACHI vectors for highly efficient CRISPR/Cas9-mediated gene knockout in arabidopsis thaliana. Plant Cell Physiol. 2016;58:46–56.

75. Ryder P, McHale M, Fort A, Spillane C. Generation of stable nulliplex autopolyploid lines of *Arabidopsis thaliana* using CRISPR/Cas9 genome editing. Plant Cell Rep. 2017;36:1005–8.

76. Char SN, Neelakandan AK, Nahampun H, Frame B, Main M, Spalding MH, et al. An Agrobacterium-delivered CRISPR/Cas9 system for high-frequency targeted mutagenesis in maize. Plant Biotechnol J. 2016;:257–68.

77. Brooks C, Nekrasov V, Lippman ZB, Van Eck J. Efficient gene editing in tomato in the first generation using the CRISPR/Cas9 system. Plant Physiol. 2014;166:1292–7.

78. Fan D, Liu T, Li C, Jiao B, Li S, Hou Y, et al. Efficient CRISPR/Cas9-mediated targeted mutagenesis in Populus in the first generation. Sci Rep. 2015;5:12217.

79. Wang S, Zhang S, Wang W, Xiong X, Meng F, Cui X. Efficient targeted mutagenesis in potato by the CRISPR/Cas9 system. Plant Cell Rep. 2015.

80. Wang Y, Cheng X, Shan Q, Zhang Y, Liu J, Gao C, et al. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. Nat Biotechnol. 2014;32:947–51.

81. Zhang Y, Bai Y, Wu G, Zou S, Chen Y, Gao C, et al. Simultaneous modification of three homoeologs of TaEDR1 by genome editing enhances powdery mildew resistance in wheat. Plant J. 2017;:1–11.

82. Svitashev S, Schwartz C, Lenderts B, Young JK, Cigan AM. Genome editing in maize directed by CRISPR-Cas9 ribonucleoprotein complexes. Nat Commun. 2016;7:13274.

83. Shi J, Gao H, Wang H, Lafitte HR, Archibald RL, Yang M, et al. ARGOS8 variants generated by CRISPR-Cas9 improve maize grain yield under field drought stress conditions. Plant Biotechnol J. 2016;:1–10.

84. Jacobs TB, LaFayette PR, Schmitz RJ, Parrott W a. Targeted genome modifications in soybean with CRISPR/Cas9. BMC Biotechnol. 2015;15:16.

85. Zhang Y, Liang Z, Zong Y, Wang Y, Liu J, Chen K, et al. Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. Nat Commun. 2016;7:12617.

86. Kim J, Kim J-S. Bypassing GMO regulations with CRISPR gene editing. Nat Biotechnol. 2016;34:1014–5.

87. Kim S, Kim D, Cho SW, Kim J, Kim J-S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome Res. 2014;24:1012–9.

88. Woo JW, Kim J, Kwon S Il, Corvalán C, Cho SW, Kim H, et al. DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. Nat Biotechnol. 2015;33:1162–4.

89. Liang Z, Chen K, Li T, Zhang Y, Wang Y, Zhao Q, et al. Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. Nat Commun. 2017;8:14261.

90. Supartana P, Shimizu T, Nogawa M, Shioiri H, Nakajima T, Haramoto N, et al. Development of simple and efficient in planta transformation method for wheat (Triticum aestivum L.) using Agrobacterium tumefaciens. J Biosci Bioeng. 2006;102:162–70.

91. Razzaq A, Hafiz IA, Mahmood I, Hussain A. Development of *in planta* transformation protocol for wheat. African J Biotechnol. 2011;10:740–50.

92. Agarwal S, Loar S, Steber C, Zale J. Floral transformation of wheat. In: Jones HD, Shewry PR, editors. Transgenic Wheat, Barley and Oats. Totowa, NJ: Humana Press; 2009. p. 105–13.

93. Ali Z, Abul-Faraj A, Li L, Ghosh N, Piatek M, Mahjoub A, et al. Efficient virus-mediated genome editing in plants using the CRISPR/Cas9 system. Mol Plant. 2015;:1–4.

94. Ali Z, Eid A, Ali S, Mahfouz MM. Pea early-browning virus-mediated genome editing via the CRISPR/Cas9 system in Nicotiana benthamiana and Arabidopsis. Virus Res. 2018;244 July 2017:333–7.

95. Zaidi SS-A, Mansoor S. Viral vectors for plant genome engineering. Front Plant Sci. 2017;8

April:2012–7.

96. Ran FA, Cong L, Yan WX, Scott D a., Gootenberg JS, Kriz AJ, et al. In vivo genome editing using Staphylococcus aureus Cas9. Nature. 2015.

97. Kim E, Koo T, Park SW, Kim D, Kim K, Cho H-Y, et al. In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. Nat Commun. 2017;8:14500.

98. Cunningham FJ, Goh NS, Demirer GS, Matos JL, Landry MP. Nanoparticle-mediated delivery towards advancing plant genetic engineering. Trends Biotechnol. 2018;36:882–97.

99. Demirer GS, Zhang H, Matos J, Chang R, Chio L, Staskawicz B, et al. High aspect ratio nanomaterials enable biomolecule delivery and transgene expression or silencing in mature plants. bioRxiv. 2018;:179549.

100. Puchta H. The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. J Exp Bot. 2005;56:1–14.

101. Wang F, Wang C, Liu P, Lei C, Hao W, Gao Y, et al. Enhanced rice blast resistance by CRISPR/Cas9-targeted mutagenesis of the ERF transcription factor gene *OsERF922*. PLoS One. 2016;11:e0154027.

102. Varkonyi-Gasic E, Wang T, Voogd C, Jeon S, Drummond RSM, Gleave AP, et al. Mutagenesis of kiwifruit CENTRORADIALIS-like genes transforms a climbing woody perennial with long juvenility and axillary flowering into a compact plant with rapid terminal flowering. Plant Biotechnol J. 2018;:Epub ahead of print.

103. Li Q, Zhang D, Chen M, Liang W, Wei J, Qi Y, et al. Development of *japonica* photo-sensitive genic male sterile rice lines by editing *Carbon Starved Anther* using CRISPR/Cas9. J Genet Genomics. 2016;43:415–9.

104. Zhou H, He M, Li J, Chen L, Huang Z, Zheng S, et al. Development of commercial thermo-sensitive genic male sterile rice accelerates hybrid rice breeding using the CRISPR/Cas9-mediated

TMS5 editing system. Sci Rep. 2016;6:37395.

105. Singh M, Kumar M, Albertsen MC, Young JK, Cigan AM. Concurrent modifications in the three homeologs of Ms45 gene with CRISPR-Cas9 lead to rapid generation of male sterile bread wheat (*Triticum aestivum* L.). Plant Mol Biol. 2018;97:371–83.

106. Miao C, Xiao L, Hua K, Zou C, Zhao Y, Bressan RA, et al. Mutations in a subfamily of abscisic acid receptor genes promote rice growth and productivity. Proc Natl Acad Sci. 2018.

107. Sánchez-León S, Gil-Humanes J, Ozuna C V, Giménez MJ, Sousa C, Voytas DF, et al. Low-gluten, nontransgenic wheat engineered with CRISPR/Cas9. Plant Biotechnol J. 2017;16:902–10.

108. Jansing J, Sack M, Augustine SM, Fischer R, Bortesi L. CRISPR/Cas9-mediated knockout of six glycosyltransferase genes in Nicotiana benthamiana for the production of recombinant proteins lacking beta-1,2-xylose and core alpha-1,3-fucose. Plant Biotechnol J. 2018.

109. Zsögön A, Čermák T, Naves ER, Notini MM, Edel KH, Weinl S, et al. *De novo* domestication of wild tomato using genome editing. Nat Biotechnol. 2018; October.

110. Lynagh PG, Inagaki S, Amundson KR, Marimithu MPA, Pike BR, Henry IM, et al. Translocation and duplication from CRISPR-Cas9 editing in *Arabidopsis thaliana*. bioRxiv. 2018.

111. Wang W, Pan Q, He F, Akhunova A, Chao S, Trick H, et al. Transgenerational CRISPR-Cas9 activity facilitates multiplex gene editing in allopolyploid wheat. Cris J. 2018;1:65–74.

112. Baltes NJ, Gil-Humanes J, Cermak T, Atkins PA, Voytas DF. DNA replicons for plant genome engineering. Plant Cell. 2014;26:151–63.

113. Čermák T, Baltes NJ, Čegan R, Zhang Y, Voytas DF. High-frequency, precise modification of the tomato genome. Genome Biol. 2015;16:232.

114. Butler NM, Baltes NJ, Voytas DF, Douches DS. Geminivirus-mediated genome editing in potato (*Solanum tuberosum* L.) using sequence-specific nucleases. Front Plant Sci. 2016;7 July:1045.

115. Gil-Humanes J, Wang Y, Liang Z, Shan Q, Ozuna C V, Sánchez-León S, et al. High efficiency gene targeting in hexaploid wheat using DNA replicons and CRISPR/Cas9. Plant J. 2017;89:1251–62.

116. Wang M, Lu Y, Botella JR, Mao Y, Hua K, Zhu J-K. Gene targeting by homology-directed repair in rice using a geminivirus-based CRISPR/Cas9 system. Mol Plant. 2017;10 July:1007–10.

117. Dahan-Meir T, Filler-Hayut S, Melamed-Bessudo C, Bocobza S, Czosnek H, Aharoni A, et al. Efficient in planta gene targeting in tomato using geminiviral replicons and the CRISPR/Cas9 system. Plant J. 2018;95:5–16.

118. Svitashev S, Young JK, Schwartz C, Gao H, Falco SC, Cigan AM. Targeted mutagenesis, precise gene editing, and site-specific gene insertion in maize using Cas9 and guide RNA. Plant Physiol. 2015;169:931–45.

119. Sun Y, Zhang X, Wu C, He Y, Ma Y, Hou H, et al. Engineering herbicide-resistant rice plants through CRISPR/Cas9-mediated homologous recombination of acetolactate synthase. Mol Plant. 2016;9:628–31.

120. Richardson CD, Ray GJ, DeWitt MA, Curie GL, Corn JE. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. Nat Biotechnol. 2016;34:339–44.

121. Maruyama T, Dougan SK, Truttmann MC, Bilate AM, Ingram JR, Ploegh HL. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. Nat Biotechnol. 2015;33:538.

122. Qi Y, Zhang Y, Zhang F, Baller JA, Cleland SC, Ryu Y, et al. Increasing frequencies of site-specific mutagenesis and gene targeting in Arabidopsis by manipulating DNA repair pathways. Genome Res. 2013;23:547–54.

123. Lebel EG, Masson J, Bogucki A, Paszkowski J. Stress-induced intrachromosomal recombination in plant somatic cells. Proc Natl Acad Sci USA. 1993;90:422–6.

124. Ries G, Buchholz G, Frohnmeyer H, Hohn B. UV-damage-mediated induction of homologous recombination in Arabidopsis is dependent on photosynthetically active radiation. Proc Natl Acad Sci USA. 2000;97:13425–9.

125. Filler Hayut S, Melamed Bessudo C, Levy AA. Targeted recombination between homologous chromosomes for precise breeding in tomato. Nat Commun. 2017;8:15605.

126. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature. 2016;533:420–4.

127. Zong Y, Wang Y, Li C, Zhang R, Chen K, Ran Y, et al. Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. Nat Biotechnol. 2017;35:438–40.

128. Lu Y, Zhu J. Precise editing of a target base in the rice genome using a modified CRISPR/Cas9 system. Mol Plant. 2016; March:523–5.

129. Shimatani Z, Kashojiya S, Takayama M, Terada R, Arazoe T, Ishii H, et al. Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. Nat Biotechnol. 2017;35:441–3.

130. Li C, Zong Y, Wang Y, Jin S, Zhang D, Song Q, et al. Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. Genome Biol. 2018;19:1–9.

131. Jiang W, Feng S, Huang S, Yu W, Li G, Yang G, et al. BE-PLUS: a new base editing tool with broadened editing window and enhanced fidelity. Cell Res. 2018;28:855–61.

132. Rees HA, Komor AC, Yeh W-H, Caetano-Lopes J, Warman M, Edge ASB, et al. Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. Nat Commun. 2017;8:15790.

CHAPTER 2

CRISPR/Cas9-mediated knockout of *Ms1* enables the rapid generation of male-sterile hexaploid wheat lines for use in hybrid seed production

Statement of authorship

Title of Paper	CRISPR/Cas9-mediated knockout of *Ms1* enables the rapid generation of male-sterile hexaploid wheat lines for use in hybrid seed production
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Anzu Okada, Taj Arndell, Nikolai Borisjuk, Niharika Sharma, Nathan S. Watson-Haigh, Elise J. Tucker, Ute Baumann, Peter Langridge and Ryan Whitford (2019). *Plant Biotechnology Journal*. https://doi.org/10.1111/pbi.13106

Principal Author

Name of Principal Author (Candidate)	Taj Arndell	
Contribution to the Paper	Designed the experiments, carried out the experiments, helped with the bioinformatics work, analysed/interpreted the data, and wrote the manuscript	
Overall percentage (%)	40%	
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.	
Signature	Date	17 Oct 2018

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);
ii. permission is granted for the candidate in include the publication in the thesis; and
iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Anzu Okada		
Contribution to the Paper	Designed the experiments, carried out the experiments including greenhouse experiments, helped with the bioinformatics work, and analysed/interpreted the data. Equal contributor with the Principal Author i.e. 40% contribution.		
Signature		Date	18 Oct 2018

Name of Co-Author	Nikolai Borisjuk		
Contribution to the Paper	Initiated the project, supervised designing of the experiments, designed vectors, supervised the experimental flow, and critically reviewed the manuscript		
Signature		Date	17 Oct 2018

Name of Co-Author	Niharika Sharma		
Contribution to the Paper	Designed the bioinformatics work and carried out the bioinformatics work		
Signature		Date	17 Oct 2018

Name of Co-Author	Nathan S. Watson-Haigh		
Contribution to the Paper	Designed the bioinformatics work and carried out the bioinformatics work		
Signature		Date	17 Oct 2018

Name of Co-Author	Elise J. Tucker		
Contribution to the Paper	Designed the gRNAs, designed and tested the capillary separation method, and critically reviewed the manuscript		
Signature		Date	17 Oct 2018

Name of Co-Author	Ute Baumann		
Contribution to the Paper	Supervised designing of the experiments, designed the bioinformatics work, helped with data interpretation, reviewed and critically revised the manuscript		
Signature		Date	17 Oct 2018

Name of Co-Author	Peter Langridge		
Contribution to the Paper	Supervised designing of the experiments, helped with data interpretation, reviewed and critically revised the manuscript		
Signature		Date	17 Oct 2018

Name of Co-Author	Ryan Whitford		
Contribution to the Paper	Supervised designing of the experiments, helped with data interpretation, reviewed and critically revised the manuscript, and acted as corresponding author		
Signature		Date	17 Oct 2018

Plant Biotechnology Journal

Plant Biotechnology Journal (2019), pp. 1–9 doi: 10.1111/pbi.13106

CRISPR/Cas9-mediated knockout of *Ms1* enables the rapid generation of male-sterile hexaploid wheat lines for use in hybrid seed production

Anzu Okada[1,¶], Taj Arndell[1,¶], Nikolai Borisjuk[1,†], Niharika Sharma[1,‡], Nathan S. Watson-Haigh[1], Elise J. Tucker[1,§], Ute Baumann[1], Peter Langridge[1] and Ryan Whitford[1,*]

[1]*School of Agriculture, Food & Wine, The University of Adelaide, Urrbrae, South Australia, Australia*

Received 17 October 2018;
revised 5 February 2019;
accepted 27 February 2019.
**Correspondence* (Tel +61(0) 458 959 973; email ryan.whitford@adelaide.edu.au)
[†]Present address: School of Life Science, Huaiyin Normal University, Huai'an, China.
[‡]Present address: New South Wales Department of Primary Industries, Research Excellence, Orange, New South Wales, Australia
[§]Present address: Commonwealth Scientific and Industrial Research Organisation, Agriculture and Food, Urrbrae, South Australia, Australia
[¶]These authors contributed equally to this work.

Keywords: CRISPR/Cas9, genome editing, hybrid wheat, Ms1, *Triticum aestivum*, male sterility.

Summary

The development and adoption of hybrid seed technology have led to dramatic increases in agricultural productivity. However, it has been a challenge to develop a commercially viable platform for the production of hybrid wheat (*Triticum aestivum*) seed due to wheat's strong inbreeding habit. Recently, a novel platform for commercial hybrid seed production was described. This hybridization platform utilizes nuclear male sterility to force outcrossing and has been applied to maize and rice. With the recent molecular identification of the wheat male fertility gene *Ms1*, it is now possible to extend the use of this novel hybridization platform to wheat. In this report, we used the CRISPR/Cas9 system to generate heritable, targeted mutations in *Ms1*. The introduction of biallelic frameshift mutations into *Ms1* resulted in complete male sterility in wheat cultivars Fielder and Gladius, and several of the selected male-sterile lines were potentially non-transgenic. Our study demonstrates the utility of the CRISPR/Cas9 system for the rapid generation of male sterility in commercial wheat cultivars. This represents an important step towards capturing heterosis to improve wheat yields, through the production and use of hybrid seed on an industrial scale.

Introduction

Global demand for food crops is projected to double between 2005 and 2050 (Tilman *et al.*, 2011). In order to meet future demand and limit the environmental impact associated with doing so, new breeding technologies must be developed to increase crop yields (Tester and Langridge, 2010). Capturing heterosis through hybrid breeding is one of few crop improvement technologies that offers rapid and significant yield gains across diverse production environments. Hybrid seed has long been widely used for the production of major cereal crops such as maize and rice, but it has been a challenge to develop a commercially viable hybridization platform for bread wheat (*Triticum aestivum*) due to wheat's strong inbreeding features, and the absence of a simple, inexpensive and reliable system for hybrid seed production. Heterotic yield gains of more than 10% and enhanced yield stability have been observed in experimental wheat hybrids (Longin *et al.*, 2013; Mühleisen *et al.*, 2014), underscoring the potential of this breeding method. Given that wheat provides approximately one-fifth of dietary calories and protein for the human population (Shiferaw *et al.*, 2013), it is clear that the development of a viable wheat hybridization platform could have a substantial positive impact on global food security.

Commercial hybrid seed production requires efficient cross-pollination between genetically distinct parental inbred lines. In addition, self-pollination of the female parent must be avoided. In wheat, this has been difficult to achieve on a large scale due to a lack of efficient and reliable methods for separating the sexes and forcing outcrossing (Whitford *et al.*, 2013). Recently, a novel hybridization platform that utilizes nuclear male sterility was described for maize (Wu *et al.*, 2016). A key component of the platform is seed production technology (SPT), a process that enables the propagation of non-transgenic nuclear male-sterile inbred lines for use as female parents. This hybridization platform has since been extended to rice, a development that was made possible by the identification and isolation of the rice male fertility gene *OsNP1* (Chang *et al.*, 2016). Recently, we identified the wheat male fertility gene *Ms1* by map-based cloning and demonstrated its function via complementation of the EMS-derived mutation *ms1d* (Tucker *et al.*, 2017). In our previous report, we also described how *Ms1* could be used to establish SPT in wheat, and we highlighted the potential of genome editing for

Please cite this article as: Okada, A., Arndell, T., Borisjuk, N., Sharma, N., Watson-Haigh, N.S., Tucker, E.J., Baumann, U., Langridge, P. and Whitford, R. (2019) CRISPR/Cas9-mediated knockout of *Ms1* enables the rapid generation of male-sterile hexaploid wheat lines for use in hybrid seed production. *Plant Biotechnol. J.*, https://doi.org/10.1111/pbi.13106

rapidly introducing highly penetrant recessive *ms1* alleles into elite wheat lines.

The CRISPR (clustered regularly interspaced short palindromic repeats)/Cas9 (CRISPR-associated protein 9) system is currently the most widely used genome editing technology, largely due to its simplicity and flexibility. The system has two components that together form a ribonucleoprotein complex: the Cas9 endonuclease and a small guide RNA (gRNA). The gRNA contains a 20 nucleotide guide sequence that is designed to target a specific site (protospacer) in the genome via Watson–Crick base pairing. The protospacer must be located immediately 5′ to a protospacer adjacent motif (PAM), whose canonical form is 5′-NGG-3′ (Jinek *et al.*, 2012). Following target site recognition, Cas9 creates a DNA double-strand break (DSB). Repair of the DSB through the error-prone non-homologous end-joining (NHEJ) pathway often leaves a lesion in the form of a small insertion/deletion (indel) mutation. Such mutations can shift the open reading frame of a coding sequence or introduce a pre-mature stop codon, resulting in gene knockout.

Here, we used the CRISPR/Cas9 system to generate *Ms1* knockout wheat lines that exhibit male sterility in the first generation. One of the recessive *ms1* alleles was highly penetrant and stably transmitted to the T_1, T_2 and T_3 generations. Our results demonstrate the utility of the CRISPR/Cas9 system for the rapid generation of nuclear male sterility in hexaploid wheat. We anticipate that this approach will facilitate the development of a commercially viable wheat hybridization platform.

Results

gRNA and vector design

Using partial *Ms1* (chromosome 4BS), *Ms-A1* (4AL homoeolog) and *Ms-D1* (4DS homoeolog) sequences derived from *T. aestivum* cultivars Fielder and Gladius (NCBI GenBank accessions MK039721, MK039722, MK039723, MK039724, MK039725, MK039726), we designed three gRNAs (LTPG1-1, LTPG1-2 and LTPG1-4) that target exon 1 of *Ms1* (Figure 1a). All three gRNAs targeted a predicted signal peptide. LTPG1-1 was mismatched on chromosomes 4AL (positions +8 and +10 upstream of the PAM in the protospacer) and 4DS (position +8). LTPG1-2 was mismatched on chromosomes 4BS (position +20), 4AL (positions +2, +4 and +20) and 4DS (positions +4, +15 and +20). LTPG1-4 was mismatched on chromosomes 4AL (positions +2, +11, +14 and +20) and 4DS (position +20). Given that single mismatches at the PAM-distal end of the guide sequence (i.e. position +20) are well-tolerated unlike mismatches at the PAM-proximal end (Hsu *et al.*, 2013), we expected that LTPG1-1 and LTPG1-2 would specifically target *Ms1* on 4BS, while LTPG1-4 would target both *Ms1* and *Ms-D1*. A schematic of the *Agrobacterium* T-DNA binary vector used for transformation is shown in Figure 1b.

Identification of transgenic T_0 plants with targeted mutations in *Ms1*

We transformed 444 immature embryos of cv. Gladius and 352 immature embryos of cv. Fielder. We produced a total of 74 T_0 transgenic lines (36 of cv. Gladius and 38 of cv. Fielder) carrying gRNA LTPG1-1 (18 lines), LTPG1-2 (40 lines) or LTPG1-4 (16 lines). Transgene copy numbers ranged from 1 to 18 (Table S1). Two edited lines, both carrying LTPG1-2, were identified by capillary separation of fluorescently labelled *Ms1*-specific amplicons (Figure 2a). Line GL353-119 (cv. Gladius, transgene copy number = 12) was a biallelic heterozygous mutant (+1/−3), and FL353-19 (cv. Fielder, transgene copy number = 7) was a biallelic mutant (+1/+1). TIDE (Tracking of Indels by DEcomposition) analysis (Brinkman *et al.*, 2014) of Sanger sequence traces (Data S1-S14) confirmed the presence of targeted mutations in GL353-119 (+1/−3) and FL353-19 (+1/+1; Figure 2b). CRISPResso analysis (Pinello *et al.*, 2016) of NGS reads (NCBI BioProject PRJNA495044) revealed that the targeted mutations in GL353-119 and FL353-19 were located precisely at the canonical cut site for LTPG1-2 (Figure 3). In GL353-119, the +1 insertion was an A nucleotide that created a new *AluI* restriction site, and the −3 deletion removed the Leu7 residue in the predicted signal peptide. In FL353-19, the +1 insertions were A and T nucleotides. The CRISPResso analysis also identified several chimeric T_0 lines in which <4% of genomic DNA was mutated at the target site (Figure 3). Of the three gRNAs, LTPG1-2 induced targeted mutations in *Ms1* at the highest frequency, and it also had the highest WU-CRISPR score (Wong *et al.*, 2015) based on *in silico* predicted on-target gRNA activity (Table 1). By contrast, LTPG1-1 induced targeted mutations at very low frequencies, despite having the highest sgRNA Designer score (Doench *et al.*, 2016; Table 1). None of the T_0 lines contained off-target mutations in *Ms-A1* or *Ms-D1* at a frequency above that observed in the wild-type negative control (0.1%). GL353-119 was partially male-sterile (Figure S1), and FL353-19 was fully male-sterile, whereas all other T_0 lines and the wild-type negative controls were fully fertile.

Inheritance of targeted mutations and male-sterile phenotypes in the T_1, T_2 and T_3 generations

To determine whether the targeted mutations were heritable, we tracked the inheritance of the +1 mutant allele in the progeny of the partially male-sterile line GL353-119 (+1/−3). As the +1 insertion of an A nucleotide created a new *AluI* restriction site, we were able to easily detect the mutant allele via *AluI* restriction enzyme assay.

By crossing GL353-119 (+1/−3) with wild-type cv. Gladius (Figure 4a), we obtained T_1 progeny (+1/WT or −3/WT), all of which were fully fertile (Figure 4b). Line T1-1 (+1/WT) was selfed to produce 94 T_2 seeds that were either DsRed-positive (63 seeds) or DsRed-negative (31 seeds), based on fluorescence microscopy (Figure 4c). The Cas9 transgene was detected by PCR in 59 of the 63 DsRed-positive T_2 plants (94%), while it was detected in only 5 of the 31 DsRed-negative T_2 plants (16%). The 26 DsRed/*Cas9*-negative T_2 plants were genotyped for the +1 allele and phenotyped for fertility. All of the WT/WT and +1/WT plants were fully fertile, whereas all of the +1/+1 plants were male-sterile and produced no seed (Figure 4d and Table 2). The recessive +1 allele was inherited in a Mendelian fashion (Figure 4e and Table 2). Thus, we obtained four male-sterile T_2 plants that were apparently non-transgenic (Table 2). Line T2-14 (+1/WT) was randomly selected and selfed to produce T_3 progeny. Fifty T_3 plants were genotyped and phenotyped. All of the WT/WT and +1/WT plants were fully fertile, whereas all of the +1/+1 plants were male-sterile, and the targeted mutations were inherited in a Mendelian fashion (Table 3). Additionally, the partially male-sterile GL353-119 line produced a T_1 seed through selfing. The T_1 plant grown from this seed had the parental genotype (+1/−3), and it too was partially male-sterile.

Discussion

The CRISPR/Cas9 system is a powerful tool for studying gene function and generating genetic diversity in crops. In this study,

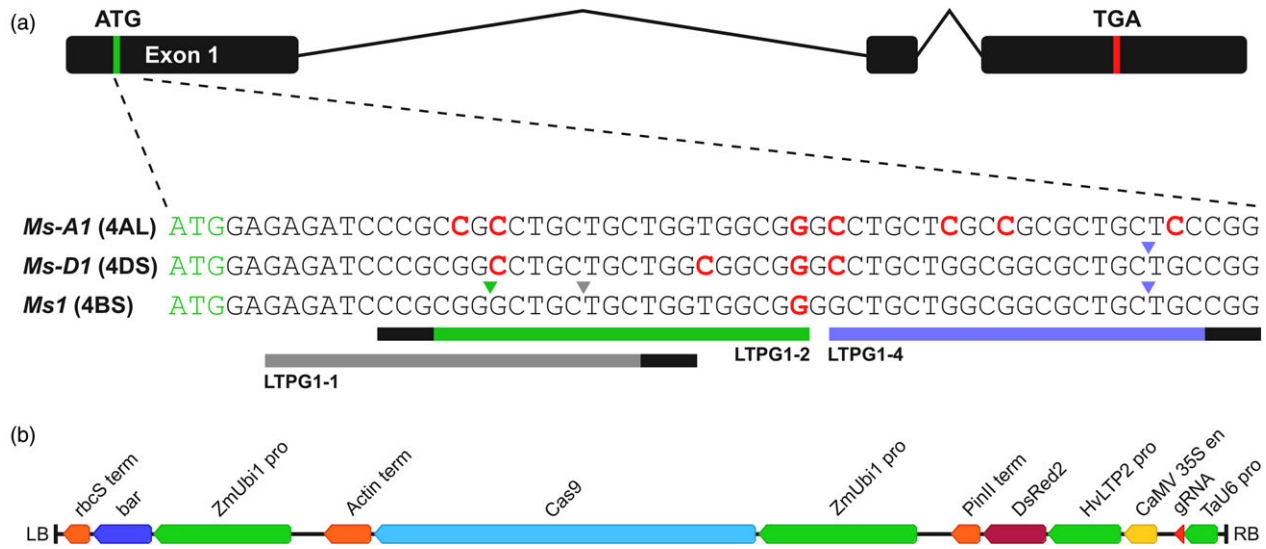


Figure 1 Target sites for gRNAs targeting *Ms1*, and the vector used for transformation. (a) The gene structure of *Ms1* and its homoeologs is shown. The detail underneath shows partial sequences for *Ms1* and its homoeologs, as well as the target sites for the three gRNAs (LTPG1-1, LTPG1-2 and LTPG1-4) targeting exon 1 of *Ms1* on chromosome 4BS. Target sites are indicated by coloured bars (grey, green and purple). PAM sites (5′-NGG-3′) are indicated by black bars at the ends of the coloured bars. Downward-pointing coloured arrow heads indicate the position of the canonical cut site, and the predicted specificity of the gRNA based on the number and distribution of gRNA mismatches (red text). (b) T-DNA (12.8 kb) of the *Agrobacterium* binary vector used for transformation. LB, left border; RB, right border.

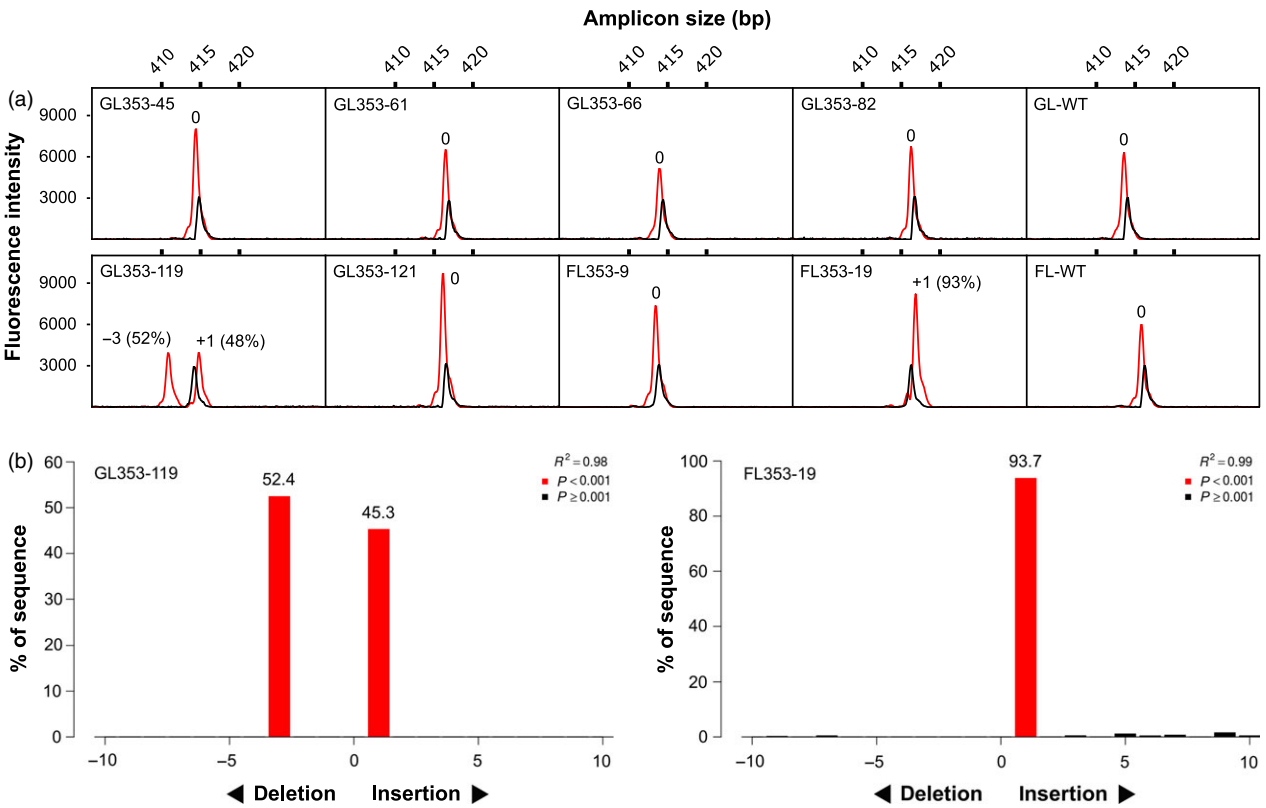


Figure 2 Screening of transgenic T_0 wheat lines for targeted mutations. (a) Capillary separation of fluorescently labelled *Ms1*-specific (chromosome 4BS) amplicons derived from eight representative lines carrying gRNA LTPG1-2. A wild-type control is also shown for each cultivar (GL-WT, cv. Gladius; FL-WT, cv. Fielder). Black peaks, wild-type spike-in (size reference); red peaks, transgenic or wild-type (negative control) line. (b) TIDE analysis of Sanger sequence traces for the two putative biallelic mutants identified in (a).

GL353-119 (total NHEJ = 99.1%)

```
   152   (0.7%) ATGGAGAGATCCCGCGGGCTGCTGCTGGTGGCGG    WT(4BS)
10 806  (48.4%) ATGGAGAGATCCCGCGGAGCTGCTGCTGGTGGCGG   +1
10 303  (46.2%) ATGGAGAGATCCCGCGG---GCTGCTGGTGGCGG   -3
```

FL353-19 (total NHEJ = 99.0%)

```
   242   (0.9%) ATGGAGAGATCCCGCGGGCTGCTGCTGGTGGCGG    WT(4BS)
14 138  (53.2%) ATGGAGAGATCCCGCGGAGCTGCTGCTGGTGGCGG   +1
10 939  (41.2%) ATGGAGAGATCCCGCGGTGCTGCTGCTGGTGGCGG   +1
```

GL353-82 (total NHEJ = 3.8%)

```
16 450  (92.1%) ATGGAGAGATCCCGCGGGCTGCTGCTGGTGGCGG    WT(4BS)
   447   (2.5%) ATGGAGAGATCCCGCGG---GCTGCTGGTGGCGG    -3
    31   (0.2%) ATGGAGAGATCCCGCGGGGCTGCTGCTGGTGGCGG   +1
    27   (0.2%) ATGGAGAGATC----------AGCTGGTGGCGG    -12/+1
    22   (0.1%) ATGGAGAGATCCCGCGG--TGCTGCTGGTGGCGG    -2
    20   (0.1%) ATGGAGAGATCCCGC----------TGGTGGCGG    -10
```

FL353-14 (total NHEJ = 2.3%)

```
28 463  (93.2%) ATGGAGAGATCCCGCGGGCTGCTGCTGGTGGCGG    WT(4BS)
   142   (0.5%) ATGGAGAGATCCCGCGG----CTGCTGGTGGCGG    -4
   109   (0.4%) ATGGAGAGATCCCGCGG----------------     -17
    98   (0.3%) ATGGAGAGATCCCGCGG-------CTGGTGGCGG    -7
    95   (0.3%) ATGGAGAGATCCCGCGG---------GGTGGCGG    -9
    59   (0.2%) ATGGAGAGATCCCGCG---TGCTGCTGGTGGCGG    -3
```

GL353-61 (total NHEJ = 2.1%)

```
19 924  (93.4%) ATGGAGAGATCCCGCGGGCTGCTGCTGGTGGCGG    WT(4BS)
   124   (0.6%) ATGGAGAGATCCCGCGG----------------     -17
    59   (0.3%) ...----------------TGCTGCTGGTGGCGG    -26
    50   (0.2%) ATGGAGAGATCCCGCGG--------------...    -29
    48   (0.2%) ATGGAGAGATCCCGCGG-------CTGGTGGCGG    -7
    38   (0.2%) ATGGAGAGATCCCGCGG----CTGCTGGTGGCGG    -4
```

GL353-66 (total NHEJ = 2.0%)

```
24 243  (93.7%) ATGGAGAGATCCCGCGGGCTGCTGCTGGTGGCGG    WT(4BS)
   396   (1.5%) ATGGAGAGATCCCGCGGGGCTGCTGCTGGTGGCGG   +1
    31   (0.1%) ATGGAGAGATCCCGC----TGCTGCTGGTGGCGG    -4
    25   (0.1%) ATGGAGAGATCCCGCGG-CTGCTGCTGGTGGCGG    -1
    22   (0.1%) ATGGAGAGATCCCGCGG----------------     -17
    16   (0.1%) ATGGAGAGATCCCGCGG----CTGCTGGTGGCGG    -4
```

Figure 3 NGS reads from transgenic T_0 mutant wheat lines carrying gRNA LTPG1-2 targeting *Ms1* on chromosome 4BS. The number of reads is shown in the first column. The per cent of total reads is shown in the second column. Green text, start codon; bold text, PAM; blue text, target site. Downward-pointing arrow heads indicate the position of the canonical cut site. *AluI* restriction sites are highlighted in grey.

our aim was to demonstrate the utility of the CRISPR/Cas9 system for the rapid generation of nuclear male sterility in hexaploid wheat. We hypothesized that knockout of *Ms1* through the introduction of targeted biallelic frameshift mutations would result in male sterility, a trait that has high agronomic value for hybrid seed production.

Five per cent (2 of 40) of T_0 plants carrying gRNA LTPG1-2 were biallelic mutants, while the majority of the plants (26 of 40) were chimeras in which only a small proportion of cells were edited. Similar editing efficiencies have been reported for gRNAs targeting other wheat genes such as *TaMLO-A1* (Wang *et al.*, 2014) and *TaGW2* (Zhang *et al.*, 2016). We observed substantial variation in editing efficiencies between different gRNAs; of the three gRNAs tested, only one (LTPG1-2) had sufficient activity to generate heritable targeted mutations in *Ms1*. *In silico* prediction of gRNA activity was carried out using the sgRNA Designer (Doench *et al.*, 2016) and WU-CRISPR (Wong *et al.*, 2015) tools. These tools use different algorithms to calculate an activity score for each gRNA, where higher scores indicate higher predicted gRNA activity. WU-CRISPR has a score range of 0–100, but only gRNAs with scores ranging from 50 to 100 are displayed in the database as they are considered 'good' candidates (Xiaowei Wang, personal communication). LTPG1-2 was predicted by WU-CRISPR to be a good candidate (activity score = 70), whereas LTPG1-1 and LTPG1-4 (activity scores < 50) were not. Our experimental results were in agreement with these predictions. sgRNA Designer has a score range of 0–1, and there is no threshold for what is considered a 'good' candidate. LTPG1-1 (activity score = 0.4) was predicted by sgRNA Designer to be the most active of the three gRNAs, but our experimental results were not in agreement with this prediction. The lack of correlation between prediction and experiment in this case is not surprising, as we tested only three gRNAs. Given the variability in gRNA efficacy, and considering the time-consuming and laborious nature of wheat transformation and tissue culture, it is recommended that gRNAs be validated through transient expression in wheat protoplasts prior to commencing plant transformation (Shan *et al.*, 2014).

We did not observe any off-target editing of *Ms-A1* or *Ms-D1* in T_0 lines. In the case of LTPG1-2, the presence of multiple mismatches (including at least one mismatch in the critical PAM-proximal 'seed' region of the protospacer) was sufficient to abolish gRNA activity, as expected (Hsu *et al.*, 2013; O'Geen *et al.*, 2015; Sternberg *et al.*, 2014; Wu *et al.*, 2014). Thus, LTPG1-2 shows high specificity for *Ms1* on chromosome 4BS. In the case of LTPG1-1 and LTPG1-4, on-target activity was very low or undetectable, and therefore, the lack of off-target activity was also expected.

We used three different methods for detecting targeted mutations in transgenic T_0 plants: capillary separation of fluorescently labelled amplicons, TIDE analysis of Sanger sequence traces and CRISPResso analysis of NGS reads. Only NGS had sufficient sensitivity to detect the low-frequency targeted mutations in the chimeras, but the three methods produced very similar results for

Table 1 Summary of editing frequencies in T_0 wheat lines carrying gRNAs targeting *Ms1* on 4BS

	Number of transgenic lines				*In silico* predicted gRNA activity	
gRNA	Total	>0.1% NHEJ	≥1.0% NHEJ	≥99% NHEJ	sgRNA Designer score*	WU-CRISPR score†
LTPG1-1	18	4 (22%)	0 (0%)	0 (0%)	0.4	<50
LTPG1-2	40	28 (70%)	10 (25%)	2 (5%)	0.3	70
LTPG1-4	16	0 (0%)	N/A	N/A	–	<50
None	N/A	0 (0%)	N/A	N/A	N/A	N/A

*Potential range of gRNA activity score = 0–1.

†Potential range of gRNA activity score = 0–100 (scores < 50 are not output); –, no score; N/A, not applicable. Editing frequencies (%NHEJ) are based on CRISPResso analysis of NGS reads. Wild-type cv. Gladius was used as the negative control (bottom row).

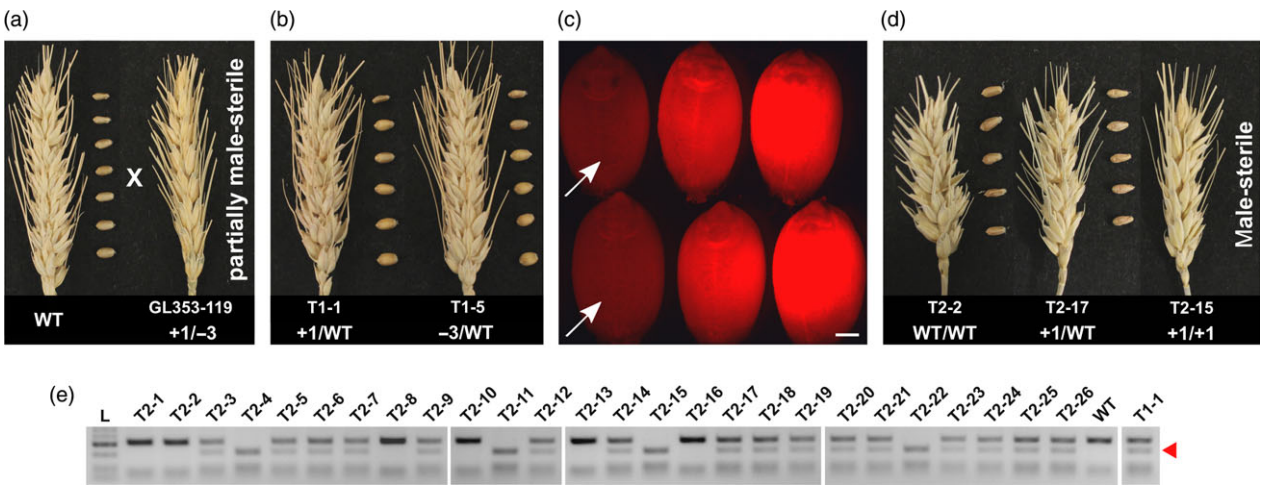


Figure 4 Inheritance and segregation of targeted mutations and male sterility generated with gRNA LTPG1-2. (a) Crossing of the transgenic T_0 mutant line GL353-119 with wild-type cv. Gladius. (b) Representative examples of T_1 progeny derived from the cross shown in (a). The T_1 progeny were selfed to produce T_2 seeds. (c) Selection of DsRed-negative (presumed non-transgenic) T_2 seeds produced by line T1-1. Scale bar = 1 mm. (d) Representative examples of T_2 progeny grown from the selected DsRed-negative seeds. (e) Genotyping of the 26 DsRed/*Cas9*-negative T_2 progeny. The +1 mutant allele contains an *AluI* restriction site that is not present in the WT allele. Cleavage at the *AluI* restriction site results in a 210 bp band (red arrow head).

Table 2 Summary of genotypes and phenotypes of DsRed/*Cas9*-negative T_2 progeny from line T1-1

Genotype	Number of lines	Male-sterile
+1/WT	16 (61.5%)	0 (0%)
+1/+1	4 (15.4%)	4 (100%)
WT/WT	6 (23.1%)	0 (0%)
Chi-squared	1.69*	

*Segregation conforms to a Mendelian 1:2:1 ratio based on the chi-squared test ($P = 0.43$).

Table 3 Summary of genotypes and phenotypes of T_3 progeny from line T2-14

Genotype	Number of lines	Male-sterile
+1/WT	22 (44%)	0 (0%)
+1/+1	11 (22%)	11 (100%)
WT/WT	17 (34%)	0 (0%)
Chi-squared	2.16*	

*Segregation conforms to a Mendelian 1:2:1 ratio based on the chi-squared test ($P = 0.34$).

the two biallelic mutants. Therefore, capillary separation of fluorescently labelled homoeolog-specific amplicons and TIDE analysis of homoeolog-specific Sanger sequence traces are rapid, reliable and cost-effective options for initial screening of T_0 plants for targeted mutations that are likely to be heritable.

As expected, biallelic knockout of *Ms1* in the T_0 line FL353-19 (+1/+1) resulted in complete male sterility. By contrast, incomplete knockout of *Ms1* in the biallelic mutant T_0 line GL353-119 (+1/−3) resulted in partial male sterility. The targeted mutations carried by GL353-119 were inherited in a Mendelian fashion, and completely male-sterile (+1/+1) mutants were recovered in the T_2 and T_3 generations, along with fully fertile +1/WT and WT/WT plants. These results are consistent with previous reports of male sterility in EMS-derived *Ms1* knockout mutants (Tucker *et al.*, 2017; Wang *et al.*, 2017). Furthermore, we obtained four T_2 knockout mutants that were apparently non-transgenic based on fluorescence microscopy and PCR assays. The identification of these mutants was streamlined by employing a fluorescence-based seed sorting strategy, similar to that which has been developed for Arabidopsis (Gao *et al.*, 2016). Interestingly, the partial male-sterile phenotype, observed in GL353-119, was inherited by a +1/−3 T_1 plant. This suggests that the Leu7 residue (deleted in the −3 allele) is required for proper functioning of the signal peptide in Ms1.

Genome editing has been successfully applied for the generation of male-sterile rice and sorghum lines (Chang *et al.*, 2016; Cigan *et al.*, 2017; Li *et al.*, 2016; Zhou *et al.*, 2016). Initial attempts to use (meganuclease-based) genome editing for the generation of male-sterile wheat lines (Cigan *et al.*, 2017) were met with limited success, as editing of only one of the three *Ms26* homoeologs (A, B or D-genome) did not confer male sterility due to functional redundancy of the gene. However, conventional crossing of single-genome biallelic *Ms26* mutants was successfully carried out to produce triple biallelic *Ms26* mutants exhibiting male sterility (Singh *et al.*, 2017). More recently, the CRISPR/Cas9 system was used to edit the wheat male fertility gene *Ms45*, and by selfing a triple monoallelic mutant T_1 plant, male-sterile triple biallelic mutant T_2 plants were recovered (Singh *et al.*, 2018). Thus, gene functional redundancy can slow the process of recovering edited wheat lines with the desired phenotype, as the targeted mutations often need to be combined and/or made homozygous via conventional breeding.

Ms1 is a single copy gene located on chromosome 4BS (Tucker *et al.*, 2017), and the homoeologs *Ms-A1* and *Ms-D1* are epigenetically silenced (Wang *et al.*, 2017). This lack of functional redundancy among homoeologs makes *Ms1* a particularly attractive target for genome editing. Indeed, our results demonstrate that the CRISPR/Cas9 system can be used to generate *Ms1* knockout wheat lines that exhibit male sterility in the first generation. We also report here the sequence for LTPG1-2, an active gRNA that specifically targets *Ms1* in cv. Fielder and cv. Gladius. Further studies will be needed to determine the efficacy of LTPG1-2 in other wheat cultivars, including elite germplasm from different breeding pools. If LTPG1-2 is found to be ineffective in a particular target cultivar, for example due to gene functional redundancy, then a different gRNA or multiple gRNAs may be needed. Furthermore, as we only tested three gRNAs, it is likely that screening of additional gRNAs would lead to the identification of a gRNA(s) that exhibits even higher activity than that observed for LTPG1-2.

Our study provides a methodological foundation and molecular tools for the rapid development of nuclear male-sterile wheat lines. This represents a significant step toward the establishment of a commercially viable hybrid wheat platform. Thus, we anticipate that the adoption of genome editing technologies for precision wheat breeding, together with a better understanding of wheat floral architecture and the flower opening process (Okada *et al.*, 2018), will ultimately lead to increased yield gains through capturing heterosis.

Experimental procedures

gRNA design

Partial *Ms1*, *Ms-A1* and *Ms-D1* sequences derived from *T. aestivum* cultivars Fielder and Gladius were used for gRNA design. The gRNAs were designed to target exon 1 of *Ms1* on chromosome 4BS, in a region immediately downstream of the start codon, based on the presence of the canonical PAM (5′-NGG-3′). Guide sequences were 20 nucleotides in length. To ensure efficient transcription from the TaU6 promoter, all gRNAs had a G nucleotide at position +20 (PAM-distal end) of the guide sequence (Sander and Joung, 2014).

Vector design and construction

All vectors were designed using Vector NTI software. The gRNA expression cassette (Shan *et al.*, 2013) consisting of the TaU6 promoter and a non-targeting (random guide sequence) gRNA was synthesized (GenScript) and cloned into pUC57, resulting in pUC57-gRNA. Annealed oligos containing the *Ms1*-targeting guide sequence (Table S2) were cloned into pUC57-gRNA by simultaneous digestion/ligation with *BbsI* and T4 DNA ligase. Positive clones were identified by diagnostic restriction digest and validated by Sanger sequencing (Australian Genome Research Facility).

The rice codon-optimized *Sp*Cas9 gene with N- and C-terminal nuclear localization signals (Shan *et al.*, 2013) was synthesized (GenScript) and inserted into PHP62407M as an *NcoI–AscI* fragment between the maize *Ubi1* promoter and the *Sorghum bicolor* actin terminator, resulting in the entry vector pCas9-NB.

The *Agrobacterium* T-DNA binary vector pMDC123 (Curtis and Grossniklaus, 2003) was modified by replacing the original selection cassette with an intron-containing *bar* gene regulated by the maize *Ubi1* promoter and the wheat *rbcS* Class II terminator. In addition, an aleurone-specific fluorescent reporter (DsRed2) cassette (Wu *et al.*, 2016) was inserted between the Gateway cassette and the right border, resulting in the destination vector pMDC-Bar-DsRed. The Cas9 expression cassette from pCas9-NB was Gateway cloned into pMDC-Bar-DsRed to produce the intermediate vector pNB1.

Finally, the gRNA expression cassettes from three different pUC57-gRNA vectors pre-loaded with *Ms1*-targeting guide sequences (LTPG1-1, LTPG1-2 and LTPG1-4, respectively) were individually cloned into pNB1 as *AsiSI-PmeI* fragments between the DsRed2 expression cassette and the right border. The resulting T-DNA binary vectors (pNB-LTPG1-1, pNB-LTPG1-2 and pNB-LTPG1-4) were used for *Agrobacterium*-mediated transformation.

Agrobacterium-mediated transformation

Transformation of cv. Fielder and cv. Gladius was carried out as described (Ishida *et al.*, 2015), with minor modifications. Briefly, immature embryos were isolated from spikes harvested at 14 days post-anthesis. Isolated embryos were transferred to WLS-liq solution, centrifuged at 16 000 ***g*** for 10 min, incubated in WLS-inf solution containing *Agrobacterium* (strain AGL1) for 5 min and then transferred to WLS-AS media for 2 days of co-cultivation. After co-cultivation, embryo axes were removed, and then, scutella were transferred to WLS-Res media for 5 days of resting culture. After the resting culture, scutella were transferred to WLS-P5 callus induction media (selection with 5 mg/L phosphinothricin) for 2 weeks, followed by WLS-P10 callus induction media (selection with 10 mg/L phosphinothricin) for 3 weeks. Calli were then transferred to LSZ-P5 regeneration media (selection with 5 mg/L phosphinothricin) for 2 weeks under a cycle of 12 h dark/12 h light (~70 μmol/m^2/s). Regenerants were transferred to LSF-P5 rooting media (selection with 10 mg/L phosphinothricin) for 2 weeks, before being transferred to potted soil in the greenhouse. Timentin was substituted for cefotaxime in all tissue culture media.

Detection of targeted mutations by capillary separation of fluorescently labelled amplicons

Genomic DNA was extracted from the second leaves of transgenic T_0 plants at the vegetative stage, using a freeze-dried method (Kovalchuk, 2014). The target site was amplified by PCR using Phusion High-Fidelity DNA Polymerase (New England BioLabs), Phusion GC Buffer, 5% DMSO, 1 M betaine and a pair of *Ms1*-specific 6-FAM-labelled primers (Table S2). To generate wild-type amplicons for spike-in (size reference), the same primer pair labelled with HEX was used. Touchdown PCR cycling conditions were as follows: initial denaturation at 98 °C for 3 min, denaturation at 98 °C for 15 s, annealing at 70–65 °C for 20 s, extension at 72 °C for 15 s and final extension at 72 °C for 5 min. The starting annealing temperature was decreased by 0.5 °C each cycle for 10 cycles, followed by 25 cycles at the final annealing temperature. A sample of the PCR product was run on an agarose gel to confirm the presence of a single band of the expected size (432 bp). The fluorescently labelled amplicons were diluted and subjected to capillary electrophoresis (Australian Genome Research Facility) on an AB3730 DNA Analyzer (Applied Biosystems, Foster City, CA). The results were analysed using PeakScanner Software 2 (Applied Biosystems). In PeakScanner, the peak for the HEX-labelled wild-type size reference was adjusted to a fluorescence intensity of approximately 3000, to improve visual clarity.

Detection of targeted mutations by Sanger sequencing and TIDE

Genomic DNA was extracted from transgenic T_0 plants as described above. The target site was amplified by PCR using Phusion High-Fidelity DNA Polymerase, Phusion GC Buffer, 5% DMSO, 1 M betaine and a pair of *Ms1*-specific primers (Table S2). Touchdown PCR cycling conditions were as follows: initial denaturation at 95 °C for 8 min, denaturation at 94 °C for 10 s, annealing at 62–57 °C for 30 s, extension at 72 °C for 30 s and final extension at 72 °C for 5 min. The starting annealing temperature was decreased by 0.5 °C each cycle for 10 cycles, followed by 30 cycles at the final annealing temperature. A sample of the PCR product was run on an agarose gel to confirm the presence of a single band of the expected size (577 bp). The amplicons were then column-purified and Sanger sequenced (Australian Genome Research Facility) on a 3730xl DNA Analyzer (Applied Biosystems). Bases were called with KB Basecaller v1.4.1.8, and the AB1 files were uploaded to the online TIDE analysis tool (Brinkman *et al.*, 2014). In TIDE, the indel size range was set at 10, and the other settings were adjusted based on information provided on the online TIDE analysis tool Troubleshooting webpage. The proportion of edited DNA in the sampled tissue was calculated as the sum of all significant indels ($P < 0.001$) detected by TIDE.

Detection of targeted mutations by NGS and CRISPResso analysis

Genomic DNA was extracted from transgenic T_0 plants as described above. To generate amplicons for NGS, two rounds of PCR were carried out. The PCR mixtures contained Phusion High-Fidelity DNA Polymerase, Phusion GC Buffer, 5% DMSO and 1 M betaine. In the first round of PCR, the target site was amplified using conserved primers flanked by 5′ universal tail sequences (Table S2). PCR cycling conditions were as follows: initial denaturation at 98 °C for 3 min, followed by 30 cycles of 98 °C for 20 s, 72 °C for 20 s, with a final extension at 72 °C for 2 min. A sample of the PCR product was run on an agarose gel to confirm the presence of bands of the expected sizes (300 bp, 313 bp, 327 bp). In the second round of PCR, barcodes and adapters were added using Illumina Nextera XT primers that anneal to the tail sequences of the primers used in the first round of PCR. Cycling conditions in the second round of PCR were as follows: initial denaturation at 98 °C for 3 min, followed by six cycles of 98 °C for 30 s, 55 °C for 30 s, 72 °C for 30 s, with a final extension of 72 °C for 5 min. The barcoded PCR products were purified using Agencourt AMPure XP beads (Beckman Coulter), quantified by qPCR, pooled in equimolar amounts, spiked with 10% PhiX Control v3 and then sequenced (Australian Genome Research Facility) on the Illumina MiSeq platform using the MiSeq Reagent Kit v3 300 cycle. The raw reads from each sample were filtered using *Ms1*, *Ms-A1* and *Ms-D1*-specific tag sequences (Table S2) and assigned to their respective homoeologs. FASTQ files containing the homoeolog-specific reads were used as input for the CRISPResso analyses (Pinello *et al.*, 2016). In CRISPResso, the following parameters were used: –w 20 –hide_mutations_outside_window_NHEJ –save_also_png –trim_sequences -q 30 –exclude_bp_from_left 5 –exclude_bp_from_right 5 –ignore_substitutions. Allele frequencies in Figure 3 were calculated by summing the values in the %Reads column of the

CRISPResso allele frequency table, after filtering out aligned sequences that did not contain the partial allele sequence shown. Editing frequencies in Table 1 were calculated using data from the CRISPResso pie charts.

Pollen viability assay

Pollen viability was assessed by Lugol (1% I_3K solution) staining. Pollen grains were mounted on glass microscope slides and imaged using a Nikon Ni-E microscope equipped with a DS-Ri1-U3 camera (Adelaide Microscopy Waite Facility). Images were captured with NIS-Elements software.

DsRed expression assay

Mature seeds were imaged with a Leica MZ FLIII microscope equipped with a Leica DFC450 C digital camera and a DsRed filter set. Images were captured with Leica Application Suite v4.10.0 software.

Detection of the Cas9 transgene in T_2 plants

Genomic DNA was extracted from T_2 plants as described above. The Cas9 transgene was detected by PCR using Phusion High-Fidelity DNA Polymerase, Phusion HF Buffer and a pair of *Cas9*-specific primers (Table S2). Touchdown PCR cycling conditions were as follows: initial denaturation at 98 °C for 2 min, denaturation at 98 °C for 15 s, annealing at 72–69 °C for 20 s, extension at 72 °C for 15 s and final extension at 72 °C for 5 min. The starting annealing temperature was decreased by 0.5 °C each cycle for 6 cycles, followed by 24 cycles at the final annealing temperature. A sample of the PCR product was run on an agarose gel to check for the presence/absence of the expected 281 bp band.

AluI assay for genotyping T_2 and T_3 plants

Genomic DNA was extracted from T_2 and T_3 plants as described above. The target site was amplified by PCR using Phusion High-Fidelity DNA Polymerase, Phusion GC Buffer, 5% DMSO, 1 M betaine and pair of *Ms1*-specific primers (Table S2). Touchdown PCR cycling conditions were as follows: initial denaturation at 95 °C for 8 min, denaturation at 94 °C for 10 s, annealing at 70–65 °C for 30 s, extension at 72 °C for 20 s and final extension at 72 °C for 5 min. The starting annealing temperature was decreased by 0.5 °C each cycle for 10 cycles, followed by 27 cycles at the final annealing temperature. A sample of the PCR product was run on an agarose gel to confirm the presence of a single band of the expected size (432 bp). Five μL of unpurified amplicons was digested with 2 units of *AluI* in a 7 μL reaction and then run on a 2% agarose gel.

Statistics

To test for Mendelian inheritance of edited alleles in the T_2 and T_3 generations, Pearson's chi-squared test was used, as described (Montoliu, 2012).

Post hoc *in silico* prediction of gRNA activity

gRNA on-target activity was predicted using the sgRNA Designer (Doench *et al.*, 2016) and WU-CRISPR (Wong *et al.*, 2015) tools, according to the developers' guidelines.

Acknowledgements

We are grateful to Matthew Tucker (The University of Adelaide) for critical discussions and reading of the manuscript, Adam Croxford (The University of Adelaide) for providing helpful advice and technical assistance with NGS, Yuan Li (The University of Adelaide) for copy number analysis, Ursula Langridge-Reimold and Yuriy Onyskiv (The University of Adelaide) for glasshouse management and plant care, Iryna Mazonka, Sheridan Morton and Christine Kastner (The University of Adelaide) for assistance with plant transformation and tissue culture, and Mathieu Baes (The University of Adelaide) for technical support. This research was funded by DuPont Pioneer Hi-Bred International Inc. We are grateful for the support provided by The University of Adelaide, Australian Research Council, Grains Research and Development Corporation and the South Australian State Government.

Conflict of interest

ET, UB and RW have filed a patent on previous published work on *Ms1*. The remaining authors declare no conflict of interest.

GenBank accession numbers

Partial *Ms1* sequence obtained from cultivar Fielder: MK039721; Partial *Ms-A1* sequence obtained from cultivar Fielder: MK039722; Partial *Ms-D1* sequence obtained from cultivar Fielder: MK039723; Partial *Ms1* sequence obtained from cultivar Gladius: MK039724; Partial *Ms-A1* sequence obtained from cultivar Gladius: MK039725; Partial *Ms-D1* sequence obtained from cultivar Gladius: MK039726.

References

Brinkman, E.K., Chen, T., Amendola, M. and van Steensel, B. (2014) Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, e168.

Chang, Z., Chen, Z., Wang, N., Xie, G., Lu, J., Yan, W., Zhou, J. *et al.* (2016) Construction of a male sterility system for hybrid rice breeding and seed production using a nuclear male sterility gene. *Proc. Natl. Acad. Sci. USA*, **113**, 14145–14150.

Cigan, A.M., Singh, M., Benn, G., Feigenbutz, L., Kumar, M., Cho, M.-J., Svitashev, S. *et al.* (2017) Targeted mutagenesis of a conserved anther-expressed P450 gene confers male sterility in monocots. *Plant Biotechnol. J.* **15**, 379–389.

Curtis, M.D. and Grossniklaus, U. (2003) A Gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol.* **133**, 462–469.

Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I. *et al.* (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191.

Gao, X., Chen, J., Dai, X., Zhang, D. and Zhao, Y. (2016) An effective strategy for reliably isolating heritable and *Cas9*-free Arabidopsis mutants generated by CRISPR/Cas9-mediated genome editing. *Plant Physiol.* **171**, 1794–1800.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y. *et al.* (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832.

Ishida, Y., Tsunashima, M., Hiei, Y. and Komari, T. (2015) Wheat (*Triticum aestivum* L.) transformation using immature embryos. *Methods Mol. Biol.* **1223**, 189–198.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816–821.

Kovalchuk, N. (2014) High-throughput analysis pipeline for achieving simple low-copy wheat and barley transgenics. *Methods Mol. Biol.* **1145**, 239–252.

Li, Q., Zhang, D., Chen, M., Liang, W., Wei, J., Qi, Y. and Yuan, Z. (2016) Development of *japonica* photo-sensitive genic male sterile rice lines by editing *Carbon Starved Anther* using CRISPR/Cas9. *J. Genet. Genom.* **43**, 415–419.

Longin, C., Gowda, M., Mühleisen, J., Ebmeyer, E., Kazman, E., Schachschneider, R., Schacht, J. *et al.* (2013) Hybrid wheat: quantitative genetic parameters and consequences for the design of breeding programs. *Theoret. Appl. Genet.* **126**, 2791–2801.

Montoliu, L. (2012) Mendel: a simple excel workbook to compare the observed and expected distributions of genotypes/phenotypes in transgenic and knockout mouse crosses involving up to three unlinked loci by means of a $\chi 2$ test. *Transgenic Res.* **21**, 677–681.

Mühleisen, J., Piepho, H.-P., Maurer, H.P., Longin, C.F.H. and Reif, J.C. (2014) Yield stability of hybrids versus lines in wheat, barley, and triticale. *Theoret. Appl. Genet.* **127**, 309–316.

O'Geen, H., Henry, I.M., Bhakta, M.S., Meckler, J.F. and Segal, D.J. (2015) A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. *Nucleic Acids Res.* **43**, 3389–3404.

Okada, T., Jayasinghe, J.E.A.R.M., Nansamba, M., Baes, M., Warner, P., Kouidri, A., Correia, D. *et al.* (2018) Unfertilized ovary pushes wheat flower open for cross-pollination. *J. Exp. Bot.* **69**, 399–412.

Pinello, L., Canver, M.C., Hoban, M.D., Orkin, S.H., Kohn, D.B., Bauer, D.E. and Yuan, G.-C. (2016) Analyzing CRISPR genome-editing experiments with CRISPResso. *Nat. Biotechnol.* **34**, 695–697.

Sander, J.D. and Joung, J.K. (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **32**, 347–355.

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K. *et al.* (2013) Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat. Biotechnol.* **31**, 686–688.

Shan, Q., Wang, Y., Li, J. and Gao, C. (2014) Genome editing in rice and wheat using the CRISPR/Cas system. *Nat. Protoc.* Nature Publishing Group **9**, 2395–2410.

Shiferaw, B., Smale, M., Braun, H.-J., Duveiller, E., Reynolds, M. and Muricho, G. (2013) Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. *Food Security*, **5**, 291–317.

Singh, M., Kumar, M., Thilges, K., Cho, M.-J. and Cigan, A.M. (2017) MS26/CYP704B is required for anther and pollen wall development in bread wheat (Triticum aestivum L.) and combining mutations in all three homeologs causes male sterility. *PLoS ONE*, **12**, e0177632.

Singh, M., Kumar, M., Albertsen, M.C., Young, J.K. and Cigan, A.M. (2018) 'Concurrent modifications in the three homeologs of Ms45 gene with CRISPR-Cas9 lead to rapid generation of male sterile bread wheat (*Triticum aestivum* L.). *Plant Mol. Biol.* Springer, Netherlands **97**, 371–383.

Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C. and Doudna, J.A. (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, Nature Publishing Group **507**, 62–67.

Tester, M. and Langridge, P. (2010) Breeding technologies to increase crop production in a changing world. *Science*, **327**, 818–822.

Tilman, D., Balzer, C., Hill, J. and Befort, B.L. (2011) Global food demand and the sustainable intensification of agriculture. *Proc. Natl Acad. Sci. USA*, **108**, 20260–20264.

Tucker, E.J., Baumann, U., Kouidri, A., Suchecki, R., Baes, M., Garcia, M., Okada, T. *et al.* (2017) Molecular identification of the wheat male fertility gene *Ms1* and its prospects for hybrid breeding. *Nat. Commun.* **8**, 869.

Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C. and Qiu, J.-L. (2014) Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat. Biotechnol.* **32**, 947–951.

Wang, Z., Li, J., Chen, S., Heng, Y., Chen, Z., Yang, J., Zhou, K. *et al.* (2017) Poaceae-specific *MS1* encodes a phospholipid-binding protein for male fertility in bread wheat. *Proc. Natl. Acad. Sci. USA*, **114**, 12614–12619.

Whitford, R., Fleury, D., Reif, J.C., Garcia, M., Okada, T., Korzun, V. and Langridge, P. (2013) Hybrid breeding in wheat: technologies to improve hybrid wheat seed production. *J. Exp. Bot.* **64**, 5411–5428.

Wong, N., Liu, W. and Wang, X. (2015) WU-CRISPR: characteristics of functional guide RNAs for the CRISPR/Cas9 system. *Genome Biol.* **16**, 218.

Wu, X., Scott, D.A., Kriz, A.J., Chiu, A.C., Hsu, P.D., Dadon, D.B., Cheng, A.W. *et al.* (2014) Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.* Nature Publishing Group **32**, 670–676.

Wu, Y., Fox, T.W., Trimnell, M.R., Wang, L., Xu, R., Cigan, A.M., Huffman, G.A. *et al.* (2016) Development of a novel recessive genetic male sterility system for hybrid seed production in maize and other cross-pollinating crops. *Plant Biotechnol. J.* **14**, 1046–1054.

Zhang, Y., Liang, Z., Zong, Y., Wang, Y., Liu, J., Chen, K., Qiu, J.-L. *et al.* (2016) Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. *Nat. Commun.* **7**, 12617.

Zhou, H., He, M., Li, J., Chen, L., Huang, Z., Zheng, S., Zhu, L. *et al.* (2016) Development of commercial thermo-sensitive genic male sterile rice accelerates hybrid rice breeding using the CRISPR/Cas9-mediated *TMS5* editing system. *Sci. Rep.* **6**, 37395.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1 Transgene copy numbers for T_0 lines.

Table S2 Primers, oligonucleotides and sequence tags used in this study.

Figure S1 Pollen viability assay based on detection of starch via iodine-potassium iodide staining. Wild type cv. Gladius (left) and the partially male-sterile edited line GL353-119 (right) are shown. Scale bar = 100 μm.

Data S1 *Ms1* Sanger sequence trace for cv. Fielder transgenic T_0 line FL353-9.

Data S2 *Ms1* Sanger sequence trace for cv. Fielder transgenic T_0 line FL353-14.

Data S3 *Ms1* Sanger sequence trace for cv. Fielder transgenic T_0 line FL353-19.

Data S4 *Ms1* Sanger sequence trace for wild type cv. Fielder.

Data S5 *Ms1* Sanger sequence trace for cv. Gladius transgenic T_0 line GL353-45.

Data S6 *Ms1* Sanger sequence trace for cv. Gladius transgenic T_0 line GL353-48.

Data S7 *Ms1* Sanger sequence trace for cv. Gladius transgenic T_0 line GL353-61.

Data S8 *Ms1* Sanger sequence trace for cv. Gladius transgenic T_0 line GL353-69.

Data S9 *Ms1* Sanger sequence trace for cv. Gladius transgenic T_0 line GL353-82.

Data S10 *Ms1* Sanger sequence trace for cv. Gladius transgenic T_0 line GL353-119.

Data S11 *Ms1* Sanger sequence trace for cv. Gladius transgenic T_0 line GL353-121.

Data S12 *Ms1* Sanger sequence trace for cv. Gladius transgenic T_0 line GL353-123.

Data S13 *Ms1* Sanger sequence trace for cv. Gladius transgenic T_0 line GL353-124.

Data S14 *Ms1* Sanger sequence trace for wild type cv. Gladius.

Table S1 Transgene copy numbers for T_0 lines

gRNA	Plant ID	Copy Number
LTPG1-1 (n = 18)	FL352_T0_1	1
	FL352_T0_2	5
	FL352_T0_4	1
	FL352_T0_5	5
	FL352_T0_6	1
	FL352_T0_7	5
	FL352_T0_8	1
	FL352_T0_9	5
	FL352_T0_10	1
	FL352_T0_11	5
	FL352_T0_12	1
	FL352_T0_13	5
	FL352_T0_14	1
	FL352_T0_15	5
	FL352_T0_18	1
	FL352_T0_20	5
	FL352_T0_21	1
	FL352_T0_22a	5
LTPG1-2 (n = 40)	GL353_T0_45	1
	GL353_T0_46	1
	GL353_T0_47	3
	GL353_T0_48	1
	GL353_T0_49	2
	GL353_T0_54	4
	GL353_T0_59	13
	GL353_T0_61	14
	GL353_T0_63	2
	GL353_T0_64	1
	GL353_T0_66	4
	GL353_T0_67	16
	GL353_T0_68	2
	GL353_T0_69	2
	GL353_T0_70	1
	GL353_T0_72	1
	GL353_T0_82	2
	GL353_T0_106	1
	GL353_T0_117	3
	GL353_T0_118	1

	GL353_T0_119	12
	GL353_T0_121	12
	GL353_T0_122	5
	GL353_T0_123	6
	GL353_T0_124	3
	GL353_T0_125	2
	GL353_T0_126	2
	GL353_T0_127	2
	GL353_T0_128	11
	GL353_T0_129	1
	FL353_T0_3	1
	FL353_T0_5	1
	FL353_T0_6	1
	FL353_T0_7	2
	FL353_T0_8	2
	FL353_T0_9	1
	FL353_T0_14	3
	FL353_T0_17	1
	FL353_T0_19	7
	FL353_T0_30	2
LTPG1-4 (n = 16)	GL360_T0_24	18
	GL360_T0_42	12
	GL360_T0_64	10
	GL360_T0_81	8
	GL360_T0_82	6
	GL360_T0_83	1
	FL360_T0_5	1
	FL360_T0_6	2
	FL360_T0_8	4
	FL360_T0_19	2
	FL360_T0_24	1
	FL360_T0_27	8
	FL360_T0_48	2
	FL360_T0_67	1
	FL360_T0_84	3
	FL360_T0_89	1

Table S2 Primers, oligonucleotides and sequence tags used in this study

Oligo name	Sequence 5'-3' (non-binding in red)	Use
TaLTPG1-1_F	CTTGAGATCCCGCGGGCTGCTGC	Sense oligo for gRNA LTPG1-1
TaLTPG1-1_R	AAACGCAGCAGCCCGCGGGATCT	Antisense oligo for gRNA LTPG1-1
TaLTPG1-2_F	CTTGCGCCACCAGCAGCAGCCCG	Sense oligo for gRNA LTPG1-2
TaLTPG1-2_R	AAACCGGGCTGCTGCTGGTGGCG	Antisense oligo for gRNA LTPG1-2
TaLTPG1-4_F	CTTGCTGCTGGCGGCGCTGCTGC	Sense oligo for gRNA LTPG1-4
TaLTPG1-4_R	AAACGCAGCAGCGCCGCCAGCAG	Antisense oligo for gRNA LTPG1-4
ET0495_F5	GCATTCCATTCCGCCACCG	Fwd primer for amplification of target sites on 4BS (capillary separation and AluI assay)
ET0514_R_FAM	[FAM6]GGGAGGAGGCGGACAACGTA	Rev primer (FAM-labelled) for amplification of target sites on 4BS (capillary separation)
ET0514_R_HEX	[HEX]GGGAGGAGGCGGACAACGTA	Rev primer (HEX-labelled) for amplification of target sites on 4BS (capillary separation)
ET0495_F6	CCCCTCCGGCATTCCATTC	Fwd primer for amplification of target site on 4BS (Sanger sequencing)
ET0493_R7	AGATCCGGTAGTACAGCGAGGAAATTAACAG	Rev primer for amplification of target site on 4BS (Sanger sequencing)
ET495_R4	CTACTAGGTAAGCAGGAAAGCGAGCA	Sanger sequencing primer (Gladius lines)
007.0091.7.C1	GGGAGGAGGCGGACAACGTA	Sanger sequencing primer (Fielder lines), Rev primer for amplification of target site on 4BS (AluI assay).
TaLTPG_NGS_F1_P5_3044	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTCCCCCTCCGGCATTCCA	Primer for first round of amplification for NGS. Used with TaLTPG_NGS_R3_P7_3303
TaLTPG_NGS_R3_P7_3303	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGCATGTCGGGCGCGCAGAAG	Primer for first round of amplification for NGS. Used with TaLTPG_NGS_F1_P5_3044
TaLTPG_NGS_F3_P5_3033	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGCATGTCGGGCGCGCAGAAG	Primer for first round of amplification for NGS. Used with TaLTPG_NGS_R2_P7_3044
TaLTPG_NGS_R2_P7_3044	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTCCCCCTCCGGCATTCCA	Primer for first round of amplification for NGS. Used with TaLTPG_NGS_F3_P5_3033
TaLTPG_B genome_1st tag	TTCCGCCACCGCA	Identification of NGS reads derived from 4BS
TaLTPG_B genome_2nd tag	GGCGGCGGCGCAG	Identification of NGS reads derived from 4BS
TaLTPG_D genome_1st tag	CCCCAACACCGCA	Identification of NGS reads derived from 4DS
TaLTPG_D genome_2nd tag	GTTCGGGCAGCAG	Identification of NGS reads derived from 4DS
TaLTPG_A genome_1st tag	TTTCCGTCCCACC	Identification of NGS reads derived from 4AL
TaLTPG_A genome_2st tag	CTTCGGGCAGCAG	Identification of NGS reads derived from 4AL
Cas-F1	GGGACAAGCCTATCAGAGAGC	Fwd primer for detection of Cas9 transgene
Cas-R1	GCGCGCCTAGTTAGTTAGTCAC	Rev primer for detection of Cas9 transgene

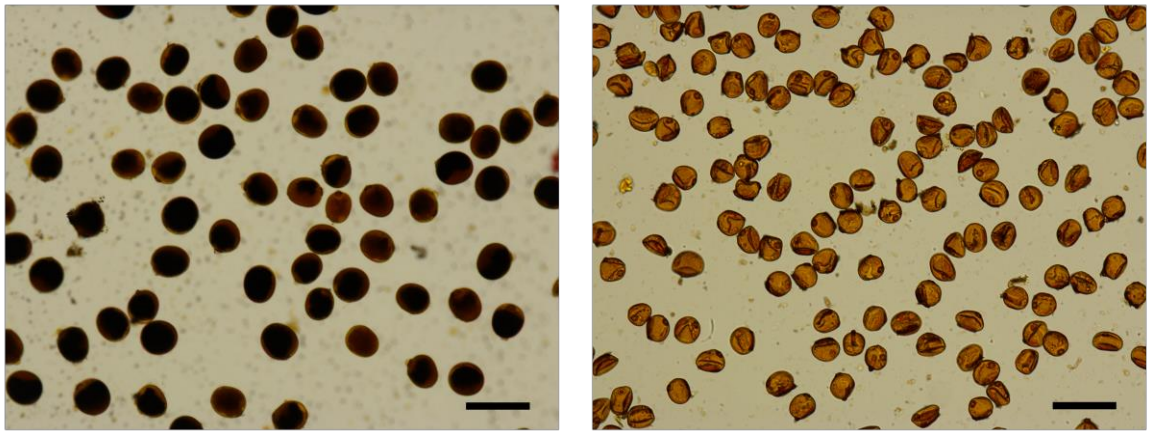


Figure S1 Pollen viability assay based on detection of starch via iodine-potassium iodide staining. Wild type cv. Gladius (left) and the partially male-sterile edited line GL353-119 (right) are shown. Scale bar = 100 µm.

CHAPTER 3

gRNA validation for wheat genome editing with the CRISPR-Cas9 system

Statement of authorship

Title of Paper	gRNA validation for wheat genome editing with the CRISPR-Cas9 system
Publication Status	☐ Published ☐ Accepted for Publication ☑ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	This manuscript has been submitted to a peer-reviewed journal and is currently under review.

Principal Author

Name of Principal Author (Candidate)	Taj Arndell	
Contribution to the Paper	Conceived the project, designed the experiments, carried out the experiments, helped with the bioinformatics work, analysed/interpreted the data, wrote the manuscript, and acted as corresponding author.	
Overall percentage (%)	80%	
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.	
Signature	Date	17 Oct 2018

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

iv. the candidate's stated contribution to the publication is accurate (as detailed above);

v. permission is granted for the candidate in include the publication in the thesis; and

vi. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Niharika Sharma		
Contribution to the Paper	Designed the bioinformatics work, carried out the bioinformatics work, and wrote the Methods section 'Next generation sequencing and CRISPResso analysis'		
Signature		Date	17 Oct 2018

Name of Co-Author	Peter Langridge		
Contribution to the Paper	Supervised designing of the experiments, helped with data interpretation, reviewed and critically revised the manuscript		
Signature		Date	17 Oct 2018

Name of Co-Author	Ute Baumann		
Contribution to the Paper	Supervised designing of the experiments, designed the bioinformatics work, helped with data interpretation, reviewed and critically revised the manuscript		
Signature		Date	17 Oct 2018

Name of Co-Author	Nathan S. Watson-Haigh		
Contribution to the Paper	Designed the bioinformatics work and carried out the bioinformatics work		
Signature		Date	17 Oct 2018

Name of Co-Author	Ryan Whitford		
Contribution to the Paper	Supervised designing of the experiments, helped with data interpretation, reviewed and critically revised the manuscript		
Signature		Date	17 Oct 2018

gRNA validation for wheat genome editing with the CRISPR-Cas9 system

Taj Arndell, Niharika Sharma, Peter Langridge, Ute Baumann, Nathan S. Watson-Haigh and Ryan Whitford

Taj Arndell (corresponding author): taj.arndell@adelaide.edu.au

Niharika Sharma: niharika.sharma@adelaide.edu.au

Peter Langridge: peter.langridge@adelaide.edu.au

Ute Baumann: ute.baumann@adelaide.edu.au

Nathan S. Watson-Haigh: nathan.watson-haigh@adelaide.edu.au

Ryan Whitford: ryan.whitford@adelaide.edu.au

School of Agriculture, Food & Wine, The University of Adelaide, Waite Campus, Urrbrae, SA 5064, Australia

Abstract

Background: The CRISPR-Cas9 system is a powerful and versatile tool for crop genome editing. However, achieving highly efficient and specific editing in polyploid species can be a challenge. The efficiency and specificity of the CRISPR-Cas9 system depends critically on the gRNA used. Here, we assessed the activities and specificities of seven gRNAs targeting 5-enolpyruvylshikimate-3-phosphate synthase (*EPSPS*) in hexaploid wheat protoplasts. EPSPS is the biological target of the widely used herbicide glyphosate.

Results: The seven gRNAs differed substantially in their on-target activities, with mean indel frequencies ranging from 0% to approximately 20%. There was no obvious correlation between experimentally determined and *in silico* predicted on-target gRNA activity. The presence of a single mismatch within the seed region of the guide sequence greatly reduced but did not abolish gRNA activity, whereas the presence of an additional mismatch, or the absence of a PAM, all but abolished gRNA activity. Large insertions (≥20 bp) of DNA vector-derived sequence were detected at frequencies up to 8.5% of total indels. One of the gRNAs exhibited several properties that make it potentially suitable for the development of non-transgenic glyphosate resistant wheat.

Conclusions: We have established a rapid and reliable method for gRNA validation in hexaploid wheat protoplasts. The method can be used to identify gRNAs that have favourable properties. Our approach is particularly suited to polyploid species, but should be applicable to any plant species amenable to protoplast transformation.

Keywords: Genome editing, CRISPR-Cas, gRNA, wheat, *Triticum aestivum*, protoplast, EPSPS

Background

Genome editing technologies enable the targeted and precise modification of plant genomes via the creation and subsequent repair of site-specific DNA double-strand breaks (DSBs) [1]. Over the last few years, the field of genome editing has been revolutionised by the introduction of the

CRISPR (clustered regularly-interspaced short palindromic repeats)-Cas9 (CRISPR associated protein) system [2–7]. This system consists of the Cas9 endonuclease in complex with a small guide RNA (gRNA) that is engineered to target a specific site (protospacer) in the genome. The protospacer is defined by a 20 nucleotide guide sequence at the 5' end of the gRNA, making programming of the system relatively straightforward. For the system to function, the protospacer must be located immediately 5' to a protospacer adjacent motif (PAM) whose canonical form is 5'-NGG-3'. Site-specific DSBs generated by Cas9 are repaired through either of two competing pathways: non-homologous end joining (NHEJ) or homology directed repair (HDR). NHEJ, which is the predominant repair pathway in somatic plant cells [8], is error-prone and often produces small insertions/deletions (indels) that result in gene knockout (e.g. through frame shift or the creation of a premature stop codon) [9–12]. Alternatively, if an exogenous DNA donor template is delivered to the cell, then precise modifications (sequence insertion or replacement) can be made through HDR [9, 13–18]. Due to its simplicity, flexibility, and high specificity, the CRISPR-Cas9 system has been and continues to be rapidly adopted by the plant research community for basic research and crop improvement.

Although the CRISPR-Cas9 system has been successfully applied to many model and crop plants, editing efficiencies have varied greatly and in many cases have been lower than one would desire. In particular, gene knockout via NHEJ tends to be relatively inefficient in polyploidy species due to genetic redundancy [19–21]. Furthermore, the inherent low frequency of HDR remains a major challenge in plant genome editing [22]. In addition, despite the high specificity of the CRISPR-Cas9 system, off-target mutations can occur at sites that have sequence similarity to the target site [23–25]. Such off-target sites may be present in non-target genes or non-target alleles. Off-target mutations are undesirable as they may confound results and/or produce impaired phenotypes, in which case they must be removed by backcrossing. Consequently, much effort has been directed toward improving the efficiency and specificity of the CRISPR-Cas9 system in plants and other organisms.

It is well-established that the efficiency and specificity of the CRISPR-Cas9 system depends critically on target site selection, as well as certain sequence features of the gRNA. Thus, an effective strategy for achieving efficient and specific editing is to use gRNAs that exhibit high activity and specificity. A number of bioinformatics tools have been developed for the *in silico* prediction of on-target and/or off-target gRNA activity [26–31]. Some of these tools can provide reliable predictions for potential off-target sites in a limited number of species, and predictions for on-target gRNA activity can reduce the time spent on gRNA screening [32]. However, the predictions are not always accurate, and the development and independent validation of these tools has been based on data obtained from non-plant species. Therefore, it is prudent to carry out experimental validation of gRNAs prior to commencing plant transformation experiments that require substantial investment of time and resources.

Here, we propose and test a strategy for assessing gRNA activity and specificity, using seven gRNAs targeting 5-enolpyruvylshikimate-3-phosphate synthase (*EPSPS*) in hexaploid wheat (*Triticum aestivum*). *EPSPS* is an ideal target for editing via HDR, as several well-characterised amino acid substitutions in this gene are known to confer resistance to the widely used herbicide glyphosate [33]. Phenotypic selection of this agronomically valuable trait is easily achieved through application of the herbicide. The gRNAs were rigorously evaluated through transient expression of CRISPR-Cas9 reagents in wheat mesophyll protoplasts, followed by TIDE (Tracking of Indels by DEcomposition) [34] analysis of Sanger sequence reads, and CRISPResso [35] analysis of NGS reads. On-target activities varied substantially between gRNAs, and there was no obvious correlation between experimentally determined and *in silico* predicted on-target gRNA activity. Off-target mutations in homoeoalleles of *EPSPS* were detected at low frequencies, whereas large insertions (≥20 bp) of DNA vector-derived sequence were detected at surprisingly high frequencies. One of the gRNAs exhibited several properties that make it potentially suitable for the development of non-transgenic glyphosate resistant wheat.

Results

Cloning and sequencing of *EPSPS* in wheat cv. Fielder

A vast number of intervarietal single nucleotide polymorphisms (SNPs) are known to exist in hexaploid wheat [36]. Therefore, for the purpose of designing effective gRNAs, we first obtained sequence information for the three homoeoalleles of *EPSPS* in our target wheat cv. Fielder. For each homoeoallele, two independent partial genomic clones of *EPSPS* were Sanger sequenced. In each case, the sequences of the two independent clones were identical. Thus, the following consensus sequences, covering a region from the 5' end of intron 1 to the middle of exon 5, were obtained: TaEPSPS-FL-7AS [GenBank MG460802], TaEPSPS-FL-4AL [GenBank MG460801], and TaEPSPS-FL-7DS [GenBank MG460803]. These consensus sequences mapped unambiguously to chromosomes 7AS, 4AL and 7DS, as expected [37]. We identified three synonymous, homoeologous SNPs located at the 3' end of exon 2, in close proximity to a universal mutation hotspot for resistance to glyphosate [38]. We exploited these SNPs for the design of homoeoallele-specific gRNAs (see below).

Protoplast transformation

We designed seven gRNAs targeting a region of *EPSPS* that contains the universal mutation hotspot for resistance to glyphosate (Fig. 1). We transiently co-expressed Cas9 and each gRNA in wheat mesophyll protoplasts. To gauge transient transformation efficiencies, we used a positive control in which YFP was substituted for the gRNA. The proportion of fluorescent (YFP expressing) protoplasts in the positive control ranged from 64-72% (mean = 68%) (Additional file 1). We found that the key to achieving high transient transformation efficiencies was to dilute the protoplasts to a concentration of 3.0×10^5 cells/mL (instead of 2.5×10^6 cells/mL as described in another protocol [39]) prior to transformation.

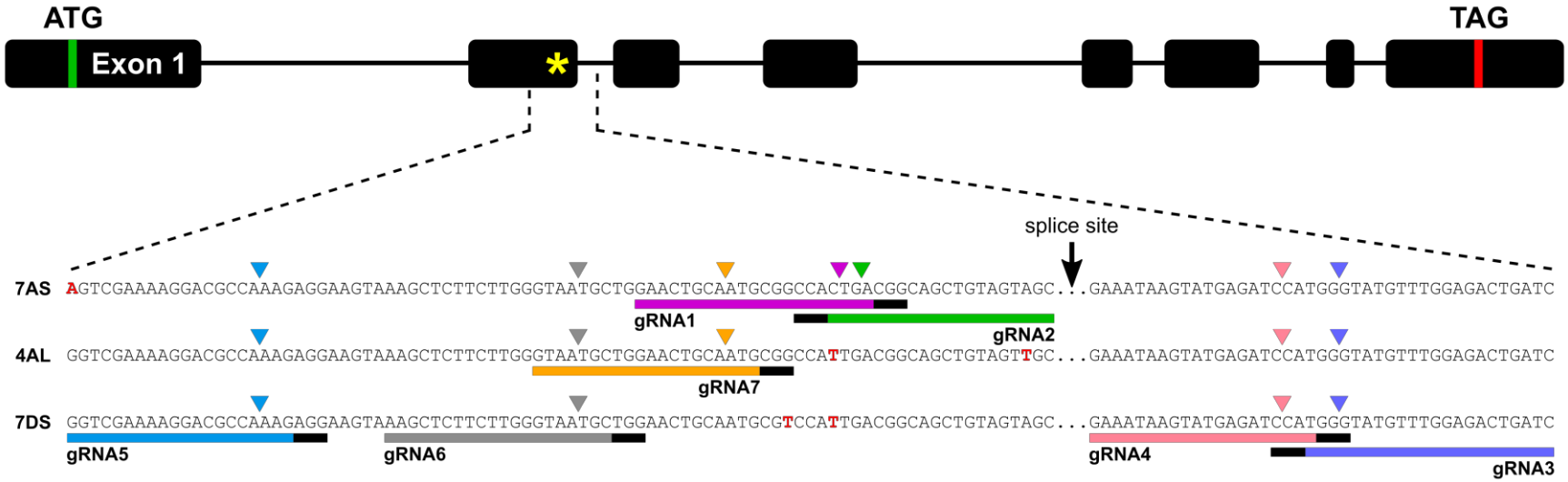


Fig. 1 Target sites for seven gRNAs targeting *EPSPS*. The gene structure of *EPSPS* is shown, annotated with the universal mutation hotspot for glyphosate resistance (yellow asterisk in exon 2). The detail underneath shows partial, abbreviated sequences for the three homoeoalleles of *EPSPS* on chromosomes 7AS, 4AL and 7DS. Target sites are indicated by coloured bars. PAM sites (5'-NGG-3') are indicated by black bars at the ends of the coloured bars. Downward-pointing coloured triangles indicate the position of the canonical cut site and predicted specificity based on the number and distribution of homoeologous SNPs at the corresponding target site.

Assessment of gRNA activity and specificity via TIDE analysis of Sanger sequence traces

We obtained high-quality forward and reverse Sanger sequence reads (Additional files 2-9) of homoeoallele-specific amplicons (Additional file 10) derived from protoplasts treated with each of the seven *EPSPS*-specific gRNAs and one non-targeting (random guide sequence) negative control gRNA. As expected, some of the sequence traces for samples treated with *EPSPS*-specific gRNAs contained mixed peaks downstream of the predicted cut site, and these mixed peaks were detected by TIDE as an increase in the percent of aberrant sequence relative to the negative control (Fig. 2a). There was a strong correlation between the indel spectra/frequencies calculated by TIDE for forward and reverse sequence traces, with each decomposition result having a high goodness of fit ($R^2 \geq 0.93$) (Additional file 11). Individual indels (significant at $p < 0.001$) were detected at frequencies down to approximately 1% (e.g. gRNA2, Rep 3, 7DS in Additional file 11). The mean frequency of significant indels ranged from 0.0-23.3% depending on the gRNA and homoeoallele (Fig. 2b). gRNA5 was the most highly active gRNA on all three homoeoalleles, causing indels at a frequency of 17.2-23.3%. The presence of a single mismatch at the PAM-distal end of the guide sequence (position 20) on 7AS did not reduce the activity of gRNA5. gRNA2 was moderately active on 7AS, causing indels at a frequency of 12.3%, though off-target indels were also detected at low frequency (1.6%) on 7DS in the presence of a single mismatch at the PAM-proximal end of the guide sequence (position 1). In contrast, no off-target indels were detected for gRNA2 on 4AL in the presence of mismatches at positions 1 and 18. gRNA4 was also moderately active on 7AS, causing indels at a frequency of 10.7%, but the frequency of indels appeared to be much lower on 4AL (1.8%) and 7DS (1.9%), even though no mismatches were present. All other gRNAs exhibited low or no activity. gRNA7 caused indels at a frequency of ≤2.3% on 7AS and 4AL, and no off-target activity was detected on 7DS (PAM site absent). gRNA1 caused indels at a frequency of 1.1% on 7AS, and off-target indels were detected at very low frequency (0.2%) on 4AL in the presence of a single mismatch at position 4 in the guide sequence. No off-target activity was detected for gRNA1 on 7DS in the presence of mismatches at positions 4 and 8. gRNA3 caused indels at a frequency of ≤0.6%. No indels were detected for gRNA6.

Assessment of gRNA activity and specificity via CRISPResso analysis of NGS reads

Due to high sequence similarity between NGS reads derived from the three homoeoalleles of *EPSPS*, CRISPResso was unable to accurately map the reads to their respective reference amplicon sequences. Therefore, we used pre-mapped NGS reads [NCBI BioProject PRJNA420019] as input for the CRISPResso analyses. Read mapping resulted in the exclusion of reads containing insertions of >1 bp. In CRISPResso, the total number of aligned (analysed) NGS reads in each subanalysis (i.e. a replicate for a gRNA on a homoeoallele) ranged from 7,067 to 35,668 (mean = 18,110). In general, the CRISPResso results (Fig. 2c and 2d, Additional files 12 and 13) were in agreement with the TIDE results, but there was less variation between replicates in the CRISPResso results, as indicated by smaller standard errors (Fig. 2d). Notably, in the CRISPResso results, the activity of gRNA2 on 7DS (off-target) was only 7% of that on 7AS (on-target), and the difference was statistically significant ($p < 0.05$). Also, the activity of gRNA4 was more consistent across homoeoalleles. The frequency of indels in the negative control was ≤0.1% (mean = 0.005%).

Based on both TIDE and CRISPResso derived data, we determined that gRNA2 is likely to be the most effective gRNA for generating stable homoeoallele-specific edits in *EPSPS*.

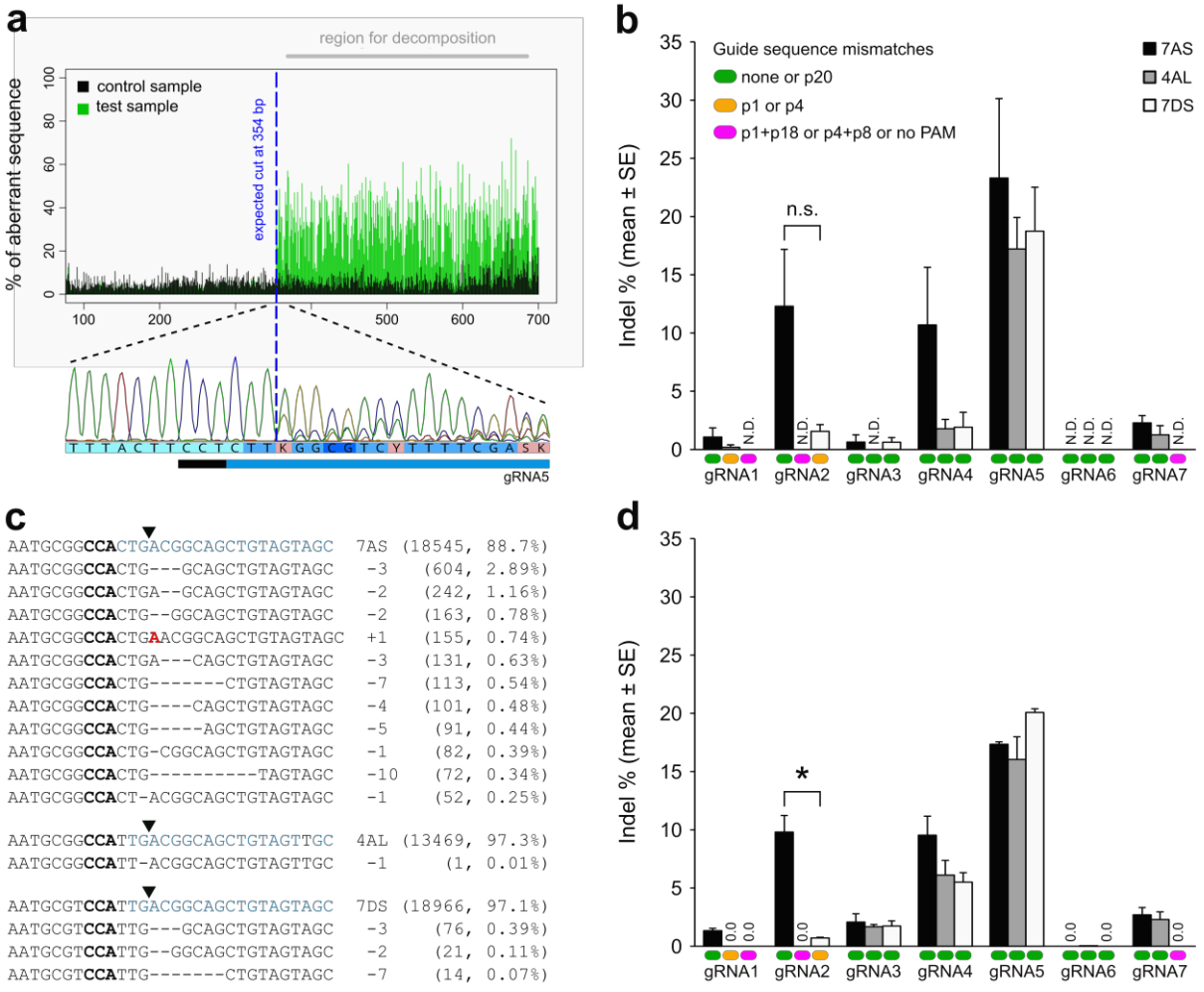


Fig. 2 Mutation detection and summary of editing efficiencies for seven gRNAs targeting *EPSPS* on chromosomes 7AS, 4AL and 7DS. **a** TIDE detection of mixed peaks in the reverse Sanger sequence read for gRNA5 on chromosome 7AS (replicate 1). **b** Summary of TIDE results. N.D., not detected. n.s., not statistically significant. Error bars represent the standard error of the mean. *n* = 3. **c** Alignment of representative mutant NGS reads for gRNA2 on chromosomes 7AS, 4AL and 7DS (replicate 1). Bold black text, PAM; blue text, complementary to gRNA2 guide sequence; red text, inserted nucleotide. Downward-pointing triangles indicate the position of the canonical cut site. The number of reads and percent of total reads is shown in brackets. **d** Summary of CRISPResso results. Error bars represent the standard error of the mean. * statistically significant ($p < 0.05$) based on a two-sample *t*-test assuming unequal variances. *n* = 3. The keys in **b** also apply to **d**. In the key for guide sequence mismatches, p20 means position 20 in the guide sequence, etc.

***In silico* prediction of on-target gRNA activity**

The seven gRNAs differed substantially in their *in silico* predicted on-target activity (Table 1). CRISPRko scores [29] ranged from 0.47-0.85 (potential range = 0-1). WU-CRISPR scores [27] ranged from <50-85 (potential range = 0-100; scores <50 are not output). There was some disagreement between the CRISPRko and WU-CRISPR scores. In particular, the top-scoring gRNA in CRISPRko (gRNA7) had a WU-CRISPR score of <50. There did not appear to be any obvious correlation between experimentally determined and *in silico* predicted on-target gRNA activity (Table 1).

Table 1 Experimentally determined versus *in silico* predicted on-target gRNA activity

	Target homoeoalleles	Indel frequency (%)[a]	CRISPRko score[b]	WU-CRISPR score[c]
gRNA1	7AS	1.3	0.59	58
gRNA2	7AS	9.8	0.58	60
gRNA3	7AS, 4AL, 7DS	1.8	0.64	82
gRNA4	7AS, 4AL, 7DS	7.0	0.52	<50
gRNA5	7AS, 4AL, 7DS	17.8	0.68	85
gRNA6	7AS, 4AL, 7DS	0.0	0.47	<50
gRNA7	7AS, 4AL	2.5	0.85	<50

a Experimentally determined on-target gRNA activity, expressed as the mean proportion of edited NGS reads derived from target homoeoalleles for three replicates. **b** Potential range = 0-1; **c** potential range = 0-100 (scores <50 are not output).

Analysis of large insertions

To detect large insertions (≥20 bp), we used unmapped NGS reads [NCBI BioProject PRJNA420019] as input for a separate CRISPResso analysis. Large insertions were detected in the majority of samples. The third replicate of gRNA2 had the highest frequency of large insertions (8.5% of edited reads), all of which showed 100% sequence identity to components of the DNA vectors used for protoplast transformation (Fig. 3, Additional file 14). Similar frequencies of large insertions were observed for the third replicate of gRNA5 (5.8% of edited reads), and no large insertions were detected in the negative controls (Additional file 14). In some cases, CRISPResso failed to correctly predict the size of the insertion. For example, the +42/-31 mutation (Fig. 3) was listed as a +21 insertion in the CRISPResso allele frequency table (Additional file 14).

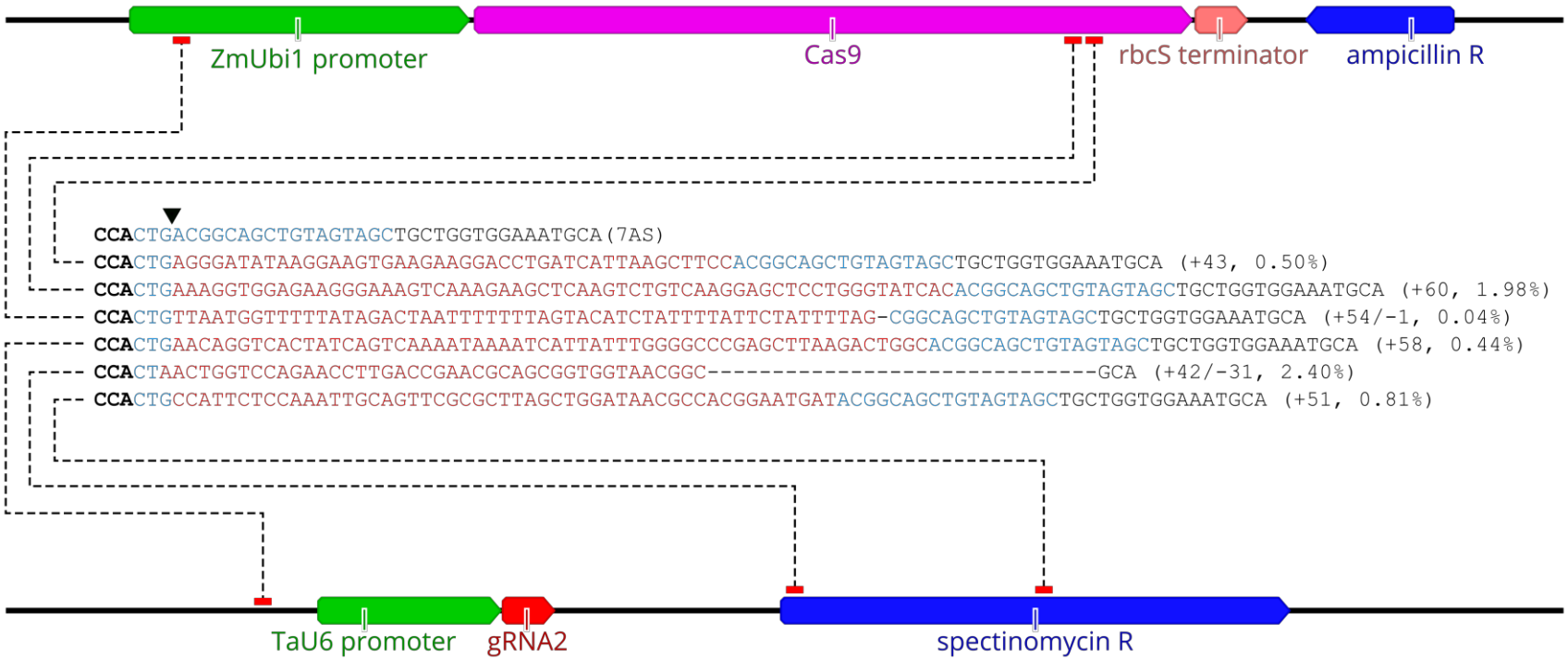


Fig. 3 Representative examples of large insertions (≥20 bp) for gRNA2 (replicate 3). Schematics of pUbi-Cas9-rbcS (top) and pCR8-U6-gRNA2 (bottom) are shown, annotated with links (dotted lines) indicating from where the inserted sequences originate. Bold black text, PAM; blue text, complimentary to gRNA2 guide sequence; red text, inserted sequence. The downward-pointing triangle indicates the position of the canonical cut site. The type/size of the mutation is given in brackets, together with the allele frequency as a percent of edited NGS reads.

Discussion

We assessed the activity and specificity of seven gRNAs targeting *EPSPS* in wheat mesophyll protoplasts. Certain amino acid substitutions in EPSPS are known to confer resistance to the herbicide glyphosate [33], and therefore *EPSPS* is an ideal candidate for editing via HDR given that glyphosate resistance is a robust selectable marker in wheat tissue culture and during plant growth [40, 41]. Furthermore, in wheat there is an opportunity to take advantage of its hexaploid nature by performing homoeoallele-specific editing, thereby potentially avoiding the severe yield penalty associated with homozygous amino acid substitutions and loss-of-function mutations in *EPSPS* [42–44]. However, given the inherent low frequency of HDR, and the potential for simultaneous editing of all three homoeoalleles, this application requires the use of a highly active and (ideally) homoeoallele-specific gRNA. With this in mind, our aim was to identify such a gRNA, and in doing so build on previous work [39] to establish a general approach for validating gRNAs in wheat and other polyploid species.

The wide range of on-target activities observed for the tested gRNAs in this study is consistent with previous reports of CRISPR-based genome editing using hexaploid wheat mesophyll protoplasts. A gRNA targeting *TaMLO-A1* caused indels at a frequency of 29% based on the PCR-restriction enzyme (PCR-RE) assay [9]. In a separate study, the same gRNA caused indels at a frequency of 36% based on a homoeoallele-specific T7E1 assay [45]. Similar editing efficiencies were attained with gRNAs targeting *TaGW2* and *TaGASR7* [46]. gRNAs targeting *TaDEP1*, *TaNAC2*, *TaPIN1* and *TaLOX2* were also evaluated by the PCR-RE assay [19], but the editing efficiencies were not calculated. Nevertheless, it could be seen that editing efficiencies varied substantially between gRNAs and were comparable with those presented here for *EPSPS*. It has been reported that approximately three or four gRNAs out of ten induce indels at a frequency of >20% in wheat protoplasts (whereas in rice protoplasts the number is approximately seven or eight gRNAs out of ten) [39]. Out of seven gRNAs, we found that one induced indels at a frequency of almost 20%, two induced indels at a frequency of 7-10%, and four induced indels at a frequency of <3% (based on data presented in Table 1). Taken together, this limited dataset suggests that gRNAs with high

activity in wheat are likely the exception rather than the rule. Moreover, we did not find any obvious correlation between experimentally determined and *in silico* predicted on-target gRNA activity, which suggests that further improvements to bioinformatics tools for gRNA design are needed. It remains unclear why some of the gRNAs tested in this study have low activity or are completely inactive. A possible explanation is chromatin context, but this seems unlikely considering that all of the gRNAs target a region spanning only 168 bp. In fact, the target site for gRNA5 (the most active gRNA) is very close to the target site for gRNA6 (inactive), so the chromatin context for these two gRNAs is likely to be similar. There are several other possible explanations including inefficient gRNA transcription, inefficient loading of the gRNA into Cas9, and inefficient binding of the guide sequence to the target site. For these reasons, we consider it prudent to carry out gRNA validation prior to commencing experiments for the production of stably edited wheat plants. However, due to there being many additional factors relating to plant transformation and tissue culture that affect overall editing efficiencies in regenerated plants, the use of a highly active gRNA certainly does not guarantee success.

We observed low levels of gRNA activity at off-target homoeoalleles of *EPSPS*. Our results are consistent with established models of gRNA specificity [47–50] in which: a) the absence of a canonical PAM site (5'-NGG-3') greatly reduces or abolishes gRNA activity; b) mismatches within the PAM-proximal 5-13 nucleotide seed region of the guide sequence reduce gRNA activity to a greater degree than mismatches outside the seed region; and c) additional mismatches further reduce gRNA activity. Importantly, a single mismatch within the seed region (at position 4 in gRNA1 and position 1 in gRNA2) greatly reduced but did not abolish gRNA activity. However, when the mismatch within the seed region was accompanied by another mismatch (at position 8 in gRNA1 and position 18 in gRNA2), gRNA activity was further reduced to levels that were undetectable with Sanger sequencing and barely detectable with NGS. Similarly, a gRNA targeting *TaDEP1-B1* caused off-target mutations at reduced frequency in *TaDEP1-A1*, despite the presence of a mismatch at position 7, in T_0 wheat plants produced using the Transient Expression of CRISPR-Cas DNA (TECCDNA) method [19]. Also, despite the presence of a gRNA mismatch at position 8 or

9, off-target mutations in gene copies were detected at reduced frequency in T_1 barley (*Hordeum vulgare* L.) and *Brassica oleracea* plants stably transformed with CRISPR-Cas9 constructs [24]. In the studies cited above, frequencies of off-target mutagenesis were approximately 50-80% lower than frequencies of on-target mutagenesis. By contrast, the gRNA targeting *TaMLO-A1* (mentioned above) did not appear to generate any off-target mutations in homoeoalleles (*TaMLO-B1* and *TaMLO-D1*) in wheat protoplasts or transgenic T_0 plants, due to the presence of a single mismatch at position 2 [45]. The apparent greater reduction in gRNA activity in the presence of a mismatch at position 1, 2 or 4 (compared with a mismatch at position 7, 8 or 9) may be due to the existence of five nucleotide 'core' within the seed region at the PAM-proximal end of the guide sequence [49, 50]. These results suggest that although off-target mutations are substantially reduced in the presence of a single mismatch in the seed region, they are often not eliminated. Therefore, ideally, potential off-target sites should lack a PAM, or else contain multiple mismatches, including at least one in the core of the seed region [23]. Where this is not possible, higher specificity may be achieved through the use of a truncated gRNA [51] and/or high fidelity variant of Cas9 [52].

gRNA2 exhibited several properties that make it potentially suitable for the development of non-transgenic glyphosate resistant wheat. First, the canonical cut site for gRNA2 is adjacent to the universal mutation hotspot for resistance to glyphosate. This is important because the frequency of HDR tends to decrease as the distance between the DSB and the site of the desired mutation increases [53]. Second, gRNA2 was active at its target site on 7AS, although the activity was moderate. Third, gRNA2 was relatively specific for *EPSPS* on 7AS, which is the most highly transcribed copy of *EPSPS* in at least some wheat cultivars [37]. This high specificity would facilitate the creation of an *EPSPS* mutant that is edited on 7AS and wild type on 4AL/7DS. Such a mutant would have the desired trait (glyphosate resistance), and the yield penalty that could otherwise result from simultaneous modification or knockout of *EPSPS* on 4AL/7DS would be avoided.

One somewhat surprising finding in this study was the relatively high frequency of insertions (up to 8.5% of edited NGS reads) that show 100% sequence identity to components of the DNA vectors used for transformation. These insertions are almost certainly vector-derived, and some (e.g. synthetic Cas9 sequences) are undoubtedly vector-derived. Recently, it was reported that DNA vector-derived insertions occur at very low frequencies (0.06-0.14% of edited NGS reads) in *Arabidopsis* protoplasts transiently transformed with CRISPR-Cas9 vectors [54], although the authors state that the frequencies were likely underestimated because insertions of >50 bp were excluded from the analysis. In addition to experimental differences, species-specific differences in NHEJ [55] may help to explain the much higher frequencies of DNA vector-derived insertions in wheat. If so, then sequence knockin via microhomology-mediated end joining may be a particularly effective genome editing strategy in wheat [45, 56]. On a related note, if DNA vector-derived sequences were to be found in an unregulated genome-edited crop (e.g. at an uncharacterised off-target site), it would set back both the regulatory authority and the industry [54]. Fortunately, recent advances in crop genome editing with *in vitro* transcribed RNA [19] and Cas9 ribonucleoproteins [46, 57, 58] suggest that the potential for unintended insertions of DNA vector-derived sequences can be eliminated in many cases. Where these methods cannot be used, unwanted DNA-vector derived insertions could be detected by whole genome sequencing and, if not linked to the target modification, removed by backcrossing.

Finally, this study demonstrates the utility of two methods for detecting and quantifying on-target and off-target indels: TIDE analysis of Sanger sequence traces, and CRISPResso analysis of NGS reads. Although restriction enzyme-based assays such as PCR-RE and T7E1 are rapid, cheap and widely used for mutation detection, they suffer from several drawbacks and may not be suitable in some circumstances. For example, if the target site lacks a restriction site at the canonical cut site (as was the case with gRNA6 in this study), then the PCR-RE assay is impractical. On the other hand, the T7E1 enzyme recognises and cleaves mismatched DNA produced through denaturation and re-annealing of wild type and mutant PCR amplicons, so it does not require a restriction site. However, the T7E1 assay may produce results that are difficult to interpret if the wild type

amplicons are polymorphic (a potential problem in polyploids) [59]. Moreover, restriction enzyme-based assays provide essentially no information about the indel spectrum or sequences of mutant alleles. Such information can be useful for the purpose of selecting gRNAs for particular applications [60, 61]. The sequencing-based mutation detection methods used in this study overcome the limitations described above, and come with the added benefit of greater sensitivity. Sanger sequencing combined with TIDE analysis takes only a few days and can be cost-effective for low/medium throughput screening, considering that sequencing of the opposite strand is unnecessary if the forward sequence trace is high-quality. NGS combined with CRISPResso analysis taken longer and is only cost-effective for high throughput screening. In some cases, significant expertise in bioinformatics may be required for the analysis of NGS reads derived from polyploid species. A number of other mutation detection methods have been established [62, 63]. Ultimately, the choice of mutation detection method should be made on a case-by-case basis, taking into consideration the goals of the experiment and the available resources [59].

Conclusions

In summary, this study demonstrates that gRNA validation is an essential step in the application of the CRISPR-Cas9 system in wheat. gRNA validation should be carried out prior to commencing transformation and tissue culture experiments for the production of stably edited wheat plants. We have established a reliable method for assessing gRNA activity and specificity in hexaploid wheat. The method was used to identify a gRNA that could potentially be used for the production of non-transgenic glyphosate resistant wheat lines. Our approach is applicable to any plant species amenable to protoplast transformation, and should facilitate the adoption of CRISPR-Cas9 technology for genome editing in wheat and other polyploid crops.

Methods

Cloning and sequencing of *EPSPS* in wheat cv. Fielder

A full-length *Triticum aestivum EPSPS* cDNA consensus sequence (1789 bp) was retrieved from GenBank [EU977181] and used as the query for a BLASTN search against the *T. aestivum* EST database. The returned ESTs were assembled *de novo* into contigs using the Geneious Assembler in Geneious v9. The cDNA consensus sequence and EST-derived contigs were then aligned with genomic sequences from the TGACv1 wheat genome assembly (scaffold_569503_7AS:9611-10115, scaffold_290435_4AL:41770-42544 and scaffold_623048_7DS:39649-41774), using the MUSCLE Alignment tool in Geneious. Based on this multiple sequence alignment, primers were designed (Additional file 15) to amplify a >2 kb region of the three homoeologous copies of *EPSPS* in cv. Fielder. Amplicons were TOPO cloned into pCR8 (Invitrogen), and two independent pCR8-TaEPSPS-FL clones derived from each homoeoallele (based on diagnostic restriction enzyme digest) were validated by Sanger sequencing (Australian Genome Research Facility). The Sanger sequence reads were aligned to produce a consensus sequence for each homoeoallele. The consensus sequences were then incorporated into the multiple sequence alignment and used as the basis for gRNA design.

gRNA design

Seven gRNAs were manually designed to target *EPSPS*. Protospacer sequences were 20-22 nucleotides in length, and were located immediately 5' of a PAM sequence (5'-NGG-3'). An extra G nucleotide was appended to the 5' end of gRNA6 in order to ensure efficient transcription of the gRNA [64].

***In silico* prediction of on-target gRNA activity**

On-target gRNA activity was predicted using the CRISPRko [29, 65] and WU-CRISPR [27, 66] tools, according to the developers' guidelines.

Vector design and construction

All vectors were designed using Geneious software. To construct the gRNA vector, the gRNA expression cassette [9] consisting of the TaU6 promoter and a non-targeting gRNA was synthesised (GenScript) and TOPO cloned into pCR8 (Invitrogen). The *BbsI* site in the pCR8 backbone was then removed by digestion with *NheI* and self-ligation, resulting in pCR8-U6-gRNA (negative control for editing). To insert guide sequences into pCR8-U6-gRNA, the guide sequence oligos (Additional file 15) were first annealed by combining 1 µL of each oligo (100 µM) with 1X T4 DNA ligase buffer (Invitrogen) in a total reaction volume of 10 µL. The reaction was heated to 95°C for 5 min and then left at room temperature for 30 min. Annealed oligos were inserted into pCR8-U6-gRNA by simultaneous digestion/ligation using 1 µL annealed oligos, 50 ng pCR8-U6-gRNA, 1X NEBuffer 2.1, 2 units *BbsI* (New England Biolabs), 1X T4 DNA ligase buffer, and 0.5 units T4 DNA ligase (Invitrogen) in a total reaction volume of 10 µL. Cycling conditions were as follows: 37°C for 1 h, 15°C for 1 min, 20°C for 10 min (2 cycles), and finally 37°C for 1 h. Positive clones of pCR8-U6-gRNA loaded with the guide sequence were identified by diagnostic double digest with *BbsI* and *EcoRI-HF* (New England Biolabs), and validated by Sanger sequencing (Australian Genome Research Facility).

To construct the Cas9 vector, the rice codon-optimised SpCas9 gene with N- and C-terminal nuclear localisation signals [9] was synthesised (GenScript) and inserted into the generic vector pUbi-rbcS as an *NcoI–AscI* fragment between the maize Ubiquitin 1 promoter [67, 68] and the wheat rbcS Class II terminator [69], resulting in pUbi-Cas9-rbcS.

To construct the YFP vector, the EYFP gene was inserted into pUbi-rbcS in the same manner as above, resulting in pUbi-YFP-rbcS.

Protoplast isolation and transformation

Protoplast isolation and transformation was carried out as described [39], with several modifications. Seedlings of *T. aestivum* cv. Fielder were grown in potted soil within a growth

chamber at 24°C with a photoperiod of 12 h light (~100 µmol m^{-2} s^{-1}) and 12 h dark, for 7-8 days. Only vigorous seedlings (five to eight in total) were used for protoplast isolation. A razor blade was used to make a shallow cut across the adaxial surface of the primary leaf, from which the abaxial epidermis was peeled off. Leaf peels were placed abaxial side down in a petri dish containing 0.6 M mannitol for 15 min. Leaf peels were then placed abaxial side down in a petri dish containing 10 mL of cell wall-dissolving enzyme solution [20 mM MES-KOH (pH 5.7), 1.5% (wt/vol) cellulase Onozuka RS, 0.75% (wt/vol) macerozyme R10, 0.6 M mannitol, 10 mM KCl, 10 mM $CaCl_2$, 0.1% (wt/vol) BSA] for 3-4 h with very gentle agitation. After addition of one volume of W5 solution [2 mM MES-KOH (pH 5.7), 154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl], protoplasts were filtered through a 100 µm nylon mesh into a petri dish and then carefully transferred to a 30 mL round-bottom tube (Sarstedt 55.517). Protoplasts were centrifuged for 3 min at 80 x *g*, resuspended in 15 mL of W5 solution, and incubated on ice for 30 min. The W5 solution was removed, and the protoplasts were resuspended in 500 µL MMG solution [4 mM MES-KOH (pH 5.7), 0.4 M mannitol, 15 mM $MgCl_2$]. The protoplast concentration was determined by cell counting on a hemocytometer, and subsequently adjusted to 3.0 x 10^5 cells/mL using MMG solution.

In an empty 2 mL tube, Ubi-Cas9-rbcS (20 µg, 3.5 pmol) was mixed with either pCR8-U6-gRNA (loaded with an *EPSPS*-specific guide sequence) (20 µg, 10.5 pmol), pCR8-U6-gRNA (negative control for editing) (20 µg, 10.5 pmol), or pUbi-YFP-rbcS (positive control for transformation) (20 µg, 5.7 pmol). Transformation was carried out by adding (in quick succession) 200 µL of protoplasts and then 200 µL of PEG solution [40% (wt/vol) PEG-4000, 0.2 M mannitol, 100 mM $CaCl_2$] to the tube containing pre-mixed DNA. The DNA/protoplast/PEG mixture was homogenised by gently flicking the tube, and then incubated for 15 min at room temperature. The transformation reaction was stopped by adding 840 µL of W5 solution and gently inverting the tube three times. The protoplasts were centrifuged for 2 min at 100 x *g*. The supernatant was removed and the protoplasts were resuspended in 500 µL W5 solution. The protoplasts were then

transferred to 12-well plates (Sarstedt 83.3921.500) coated with 5% vol/vol fetal bovine serum (Sigma-Aldrich F4135), and incubated at 23°C in the dark for 48 h.

The experiment was repeated twice, from the seed planting step. Thus, there were three biological replicates for each treatment and control.

Microscopy

After 16-24 h of incubation, protoplasts co-transformed with pUbi-Cas9-rbcS and pUbi-YFP-rbcS (positive control for transformation) were imaged using a Nikon Ni-E microscope equipped with a 490-500 nm excitation filter and a 520-560 nm emission filter (Adelaide Microscopy Waite Facility). Transformation efficiencies were calculated as the proportion of spherical protoplasts (*n* = 100, bright field image) that emitted yellow fluorescence (dark field image).

gDNA extraction

At the end of the 48 h incubation period, protoplasts were transferred to 2 mL tubes and centrifuged for 2 min at 100 x *g*. The supernatant was removed and gDNA was extracted from the protoplast pellet using the DNeasy Plant Mini Kit (QIAGEN) according to manufacturer's instructions. DNA was eluted from the spin column with 28 µL of elution buffer.

Target site amplification

To obtain amplicons for Sanger sequencing, a genomic region containing all seven target sites was amplified by PCR using homoeoallele-specific primers (Additional file 15). PCR was performed using 30-40 ng gDNA template, 0.8 µM primers, 200 µM dNTPs, 1X Phusion HF buffer, and 0.6 units Phusion Hot Start Flex DNA Polymerase (New England Biolabs) in a total reaction volume of 50 µL. gDNA obtained from nulli-tetrasomic lines of *T. aestivum* cv. Chinese Spring was used as template in control PCR reactions to confirm that amplification was homoeoallele-specific. Cycling

conditions for touchdown PCR were as follows: initial denaturation at 98°C for 1 min, denaturation at 98°C for 5 sec, annealing at 68-63°C (7AS and 7DS) or 66-61°C (4AL) for 15 sec, extension at 72°C for 55 sec, and final extension at 72°C for 5 min. The starting annealing temperature was decreased by 0.5°C each cycle for 10 cycles, followed by 30 cycles at the final annealing temperature. The PCR product was run on 1% agarose gel, from which amplicons were extracted using the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel) according to the manufacturer's instructions. DNA was eluted from the spin column with 15 µL of diluted (1 in 40) elution buffer and quantified using a NanoDrop 1000 spectrophotometer.

To obtain amplicons for NGS, two rounds of PCR were carried out. In the first round of PCR, a genomic region containing all seven target sites was amplified using conserved primers containing 5' universal tail sequences (Additional file 15) to which Illumina index primers anneal in the second round of PCR. PCR was performed using 20-40 ng gDNA template, 0.25 µM primers, 200 µM dNTPs, 1X Phusion HF buffer, and 0.2 units Phusion Hot Start Flex DNA Polymerase in a total reaction volume of 20 µL. Cycling conditions for touchdown PCR were as follows: initial denaturation at 98°C for 1 min, denaturation at 98°C for 5 sec, annealing at 62-57°C for 15 sec, extension at 72°C for 10 sec, and final extension at 72°C for 2 min. The starting annealing temperature was decreased by 0.5°C each cycle for 10 cycles, followed by 25 cycles at the final annealing temperature. The PCR product was purified using Agencourt AMPure XP beads (Beckman Coulter) according to the manufacturer's instructions. The second round of PCR was performed using 10 ng DNA template (purified amplicons from the first round of PCR), 0.3 µM primers (Illumina Nextera XT), 200 µM dNTPs, 1X Phusion HF buffer, and 0.2 units Phusion Hot Start Flex DNA Polymerase in a total reaction volume of 10 µL. Cycling conditions were as follows: initial denaturation at 98°C for 1 min, denaturation at 98°C for 5 sec, annealing at 60°C for 15 sec, extension at 72°C for 6 sec, and final extension at 72°C for 2 min (7 cycles in total). The indexed PCR products were purified using Agencourt AMPure XP beads.

Sanger sequencing and TIDE analysis

To detect targeted indels produced via NHEJ, homoeoallele-specific amplicons from each PCR reaction were subjected to Sanger sequencing (Australian Genome Research Facility) in the forward and reverse directions with nested homoeoallele-specific primers (Additional file 15). The 3730xl DNA Analyzer (Applied Biosystems) was used for sequencing, and bases were called with KB Basecaller v1.4.1.8. Output AB1 files were uploaded to the online TIDE analysis tool [34]. In TIDE, minor adjustments to the decomposition window were made based on information provided on the online TIDE analysis tool Troubleshooting webpage. All other TIDE settings were the default. The indel frequency for each gRNA/homoeoallele/replicate was calculated as the mean percent of sequences containing significant indels ($p < 0.001$) for the forward and reverse reads.

Next generation sequencing and CRISPResso analysis

To detect targeted indels produced via NHEJ, indexed PCR products were quantified by qPCR, diluted to 4 nM, pooled in equal volumes, spiked with 10% PhiX Control v3, and then sequenced on the Illumina MiSeq platform using the MiSeq Reagent Kit v3 300 cycle (Australian Genome Research Facility). The raw reads from each sample were mapped to the three homoeologous amplicon reference sequences in two phases using Bowtie 2 [70]. The aim of the first phase was to map unedited reads using the following parameters: --end-to-end --very-sensitive --np 0 --mp 6,1 --rdg 999,999 --rfg 999,999 --score-min L,-6,0. Unmapped reads from the first phase were used as input for the second phase, where reads with indels and some low quality mismatches were mapped using the following parameters: --end-to-end --very-sensitive --np 0 --mp 76,1 --rdg 24,1 --rfg 9,14 --score-min L,-75,0. Next, the resulting two BAM files were sliced for the reads that mapped to the respective three amplicons and merged together using SAMtools [71]. An in-house bash script was used to extract the mapped unedited/edited reads from the merged BAM files, and these sequence files in FASTQ format were used as input for CRISPResso [35] analysis using the following parameters: --w 20 --hide_mutations_outside_window_NHEJ --save_also_png --

trim_sequences -q 30 --exclude_bp_from_left 5 --exclude_bp_from_right 5 --ignore_substitutions. Allele frequencies were calculated by summing the values in the %Reads column of the CRISPResso allele frequency table, after filtering out all aligned sequences that did not contain the allele sequence shown in Fig. 2c. Indel frequencies were calculated as the %NHEJ for the gRNA minus the %NHEJ for the negative control, based on data from the CRISPResso pie charts.

Analysis of large insertions

To detect large insertions (≥20 bp), a separate CRISPResso analysis was carried out using unmapped NGS reads as input, with the same CRISPResso settings as above. Data in CRISPResso allele frequency tables were sorted based on insertion size (largest to smallest), and then filtered to exclude aligned sequences containing insertions of <20 bp. Reads containing insertions of ≥20 bp were aligned to the cv. Fielder consensus sequences in Geneious using the MUSCLE alignment tool. The sequences of the insertions were then searched for in the sequences of pUbi-Cas9-rbcS and pCR8-U6-gRNA. Allele frequencies were calculated by summing the values in the %Reads column of the CRISPResso allele frequency table, after filtering out all aligned sequences that did not contain the allele sequence shown in Fig. 3.

Abbreviations

CRISPR, clustered regularly-interspaced short palindromic repeats; Cas9, CRISPR-associated protein 9; gRNA, guide RNA; PAM, protospacer adjacent motif; DSB, double-strand break; NHEJ, non-homologous end joining; HDR, homology directed repair; EPSPS, 5-enolpyruvylshikimate-3-phosphate synthase; TIDE, Tracking of Indels by DEcomposition; SNP, single nucleotide polymorphism; PCR-RE, PCR-restriction enzyme.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets supporting the conclusions of this article are included in the article and its additional files. The next generation sequencing data have been deposited in the NCBI SRA repository [BioProject PRJNA420019; http://www.ncbi.nlm.nih.gov/bioproject/420019].

Competing interests

The authors declare that they have no competing interests.

Funding

This work was funded by the Australian Centre for Plant Functional Genomics (ACPFG) Pty Ltd. The ACPFG had no role in the study other than providing financial support. TA acknowledges financial support received through an Australian Government Research Training Program Scholarship.

Authors' contributions

TA conceived the project. TA, PL, UB and RW designed the experimental work. TA carried out the experimental work and analysed the data. TA, NS, UB and NW-H designed the bioinformatics

work. TA, NS and NW-H carried out the bioinformatics work. TA and NS wrote the paper. PL, UB and RW critically revised the paper. All authors read and approved the final manuscript.

Acknowledgements

We are grateful to Matthew Tucker (The University of Adelaide) for critical discussions and reading of the manuscript, Adam Croxford (The University of Adelaide) for providing helpful advice and technical assistance with NGS, Anzu Okada (The University of Adelaide) for providing technical assistance with qPCR and NGS, Gwen Mayo (Adelaide Microscopy Waite Facility) for providing microscopy assistance, and Nikolai Borisjuk (Huaiyin Normal University) for providing support in the beginning stages of the project. We are grateful for the support provided by The University of Adelaide, Australian Research Council, Grains Research and Development Corporation, and the Government of South Australia.

References

1. Puchta H, Fauser F. Gene targeting in plants: 25 years later. Int J Dev Biol. 2013;57:629–37.

2. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012;337:816–21.

3. Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci USA. 2012;109:E2579-86.

4. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013;339:819–23.

5. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human genome engineering via Cas9. Science. 2013;339:823–6.

6. Cho SW, Kim S, Kim JM, Kim J-S. Targeted genome engineering in human cells with the Cas9

RNA-guided endonuclease. Nat Biotechnol. 2013;31:230–2.

7. Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J. RNA-programmed genome editing in human cells. Elife. 2013;2:e00471.

8. Puchta H. The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. J Exp Bot. 2005;56:1–14.

9. Shan Q, Wang Y, Li J, Zhang Y, Chen K, Liang Z, et al. Targeted genome modification of crop plants using a CRISPR-Cas system. Nat Biotechnol. 2013;31:686–8.

10. Feng Z, Zhang B, Ding W, Liu X, Yang D-L, Wei P, et al. Efficient genome editing in plants using a CRISPR/Cas system. Cell Res. 2013;23:1229–32.

11. Mao Y, Zhang H, Xu N, Zhang B, Gou F, Zhu J. Application of the CRISPR-Cas system for efficient genome engineering in plants. Mol Plant. 2013;6:2008–11.

12. Feng Z, Mao Y, Xu N, Zhang B, Wei P, Yang D-L, et al. Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in *Arabidopsis*. Proc Natl Acad Sci USA. 2014;111:4632–7.

13. Li J, Norville JE, Aach J, McCormack M, Zhang D, Bush J, et al. Multiplex and homologous recombination–mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. Nat Biotechnol. 2013;31:688–91.

14. Baltes NJ, Gil-Humanes J, Cermak T, Atkins PA, Voytas DF. DNA replicons for plant genome engineering. Plant Cell. 2014;26:151–63.

15. Svitashev S, Young JK, Schwartz C, Gao H, Falco SC, Cigan AM. Targeted mutagenesis, precise gene editing, and site-specific gene insertion in maize using Cas9 and guide RNA. Plant Physiol. 2015;169:931–45.

16. Li Z, Liu Z-B, Xing A, Moon BP, Koellhoffer JP, Huang L, et al. Cas9-guide RNA directed genome editing in soybean. Plant Physiol. 2015;169:960–70.

17. Sun Y, Zhang X, Wu C, He Y, Ma Y, Hou H, et al. Engineering herbicide-resistant rice plants

through CRISPR/Cas9-mediated homologous recombination of acetolactate synthase. Mol Plant. 2016;9:628–31.

18. Endo M, Mikami M, Toki S. Biallelic gene targeting in rice. Plant Physiol. 2016;170:667–77.

19. Zhang Y, Liang Z, Zong Y, Wang Y, Liu J, Chen K, et al. Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. Nat Commun. 2016;7:12617.

20. Andersson M, Turesson H, Nicolia A, Fält AS, Samuelsson M, Hofvander P. Efficient targeted multiallelic mutagenesis in tetraploid potato (*Solanum tuberosum*) by transient CRISPR-Cas9 expression in protoplasts. Plant Cell Rep. 2017;36:117–28.

21. Ryder P, McHale M, Fort A, Spillane C. Generation of stable nulliplex autopolyploid lines of *Arabidopsis thaliana* using CRISPR/Cas9 genome editing. Plant Cell Rep. 2017;36:1005–8.

22. Baltes NJ, Gil-Humanes J, Voytas DF. Genome engineering and agriculture: opportunities and challenges. Prog Mol Biol Transl Sci. 2017;149:1–26.

23. Cho SW, Kim S, Kim Y, Kweon J, Kim HS, Bae S, et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res. 2014;24:132–41.

24. Lawrenson T, Shorinola O, Stacey N, Li C, Østergaard L, Patron N, et al. Induction of targeted, heritable mutations in barley and *Brassica oleracea* using RNA-guided Cas9 nuclease. Genome Res. 2015;16:258.

25. Wolt JD, Wang K, Sashital D, Lawrence-Dill CJ. Achieving plant CRISPR targeting that limits off-target effects. Plant Genome. 2016;9.

26. Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, et al. Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. Nat Biotechnol. 2014;32:1262–7.

27. Wong N, Liu W, Wang X. WU-CRISPR: characteristics of functional guide RNAs for the CRISPR/Cas9 system. Genome Biol. 2015;16:218.

28. Graham DB, Root DE. Resources for the design of CRISPR gene editing experiments. Genome Biol. 2015;16:260.

29. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat Biotechnol. 2016;34:184–91.

30. Oliveros JC, Franch M, Tabas-Madrid D, San-León D, Montoliu L, Cubas P, et al. Breaking-Cas—interactive design of guide RNAs for CRISPR-Cas experiments for ENSEMBL genomes. Nucleic Acids Res. 2016;44:W267–71.

31. Liu H, Ding Y, Zhou Y, Jin W, Xie K, Chen L-L. CRISPR-P 2.0: an improved CRISPR-Cas9 tool for genome editing in plants. Mol Plant. 2017;10:530–2.

32. Haeussler M, Schönig K, Eckert H, Eschstruth A, Mianné J, Renaud J-B, et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biol. 2016;17:148.

33. Sammons RD, Gaines TA. Glyphosate resistance: state of knowledge. Pest Manag Sci. 2014;70:1367–77.

34. Brinkman EK, Chen T, Amendola M, van Steensel B. Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Res. 2014;42:e168.

35. Pinello L, Canver MC, Hoban MD, Orkin SH, Kohn DB, Bauer DE, et al. Analyzing CRISPR genome-editing experiments with CRISPResso. Nat Biotechnol. 2016;34:695–7.

36. Lai K, Lorenc MT, Lee HC, Berkman PJ, Bayer PE, Visendi P, et al. Identification and characterization of more than 4 million intervarietal SNPs across the group 7 chromosomes of bread wheat. Plant Biotechnol J. 2015;13:97–104.

37. Aramrak A, Kidwell KK, Steber CM, Burke IC. Molecular and phylogenetic characterization of the homoeologous EPSP Synthase genes of allohexaploid wheat, *Triticum aestivum* (L.). BMC Genomics. 2015;16:844.

38. Zhou M, Xu H, Wei X, Ye Z, Wei L, Gong W, et al. Identification of a glyphosate-resistant mutant of rice 5-enolpyruvylshikimate 3-phosphate synthase using a directed evolution strategy. Plant Physiol. 2006;140:184–95.

39. Shan Q, Wang Y, Li J, Gao C. Genome editing in rice and wheat using the CRISPR/Cas system. Nat Protoc. 2014;9:2395–410.

40. Zhou H, Arrowsmith JW, Fromm ME, Hironaka CM, Taylor ML, Rodriguez D, et al. Glyphosate-tolerant CP4 and GOX genes as a selectable marker in wheat transformation. Plant Cell Rep. 1995;15:159–63.

41. Hu T, Metz S, Chay C, Zhou HP, Biest N, Chen G, et al. Agrobacterium-mediated large-scale transformation of wheat (*Triticum aestivum* L.) using glyphosate selection. Plant Cell Rep. 2003;21:1010–9.

42. Yu Q, Jalaludin A, Han H, Chen M, Sammons RD, Powles SB. Evolution of a double amino acid substitution in the 5-enolpyruvylshikimate-3-phosphate synthase in *Eleusine indica* conferring high-level glyphosate resistance. Plant Physiol. 2015;167:1440–7.

43. Li J, Meng X, Zong Y, Chen K, Zhang H, Liu J, et al. Gene replacements and insertions in rice by intron targeting using CRISPR-Cas9. Nat Plants. 2016;2:16139.

44. Zhang HH, Zhang J, Wei P, Zhang B, Gou F, Feng Z, et al. The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. Plant Biotechnol J. 2014;12:797–807.

45. Wang Y, Cheng X, Shan Q, Zhang Y, Liu J, Gao C, et al. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. Nat Biotechnol. 2014;32:947–51.

46. Liang Z, Chen K, Li T, Zhang Y, Wang Y, Zhao Q, et al. Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. Nat Commun. 2017;8:14261.

47. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, et al. DNA targeting

specificity of RNA-guided Cas9 nucleases. Nat Biotechnol. 2013;31:827–32.

48. Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature. 2014;507:62–7.

49. Wu X, Scott DA, Kriz AJ, Chiu AC, Hsu PD, Dadon DB, et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Nat Biotechnol. 2014;32:670–6.

50. O'Geen H, Henry IM, Bhakta MS, Meckler JF, Segal DJ. A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. Nucleic Acids Res. 2015;43:3389–404.

51. Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat Biotechnol. 2014;32:279–84.

52. Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, et al. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. Nature. 2016;529:490–5.

53. Paquet D, Kwart D, Chen A, Sproul A, Jacob S, Teo S, et al. Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. Nature. 2016;533:125–9.

54. Kim J, Kim J-S. Bypassing GMO regulations with CRISPR gene editing. Nat Biotechnol. 2016;34:1014–5.

55. Vu GTH, Cao HX, Reiss B, Schubert I. Deletion-bias in DNA double-strand break repair differentially contributes to plant genome shrinkage. New Phytol. 2017;214:1712–21.

56. Nakade S, Tsubota T, Sakane Y, Kume S, Sakamoto N, Obara M, et al. Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. Nat Commun. 2014;5:5560.

57. Woo JW, Kim J, Kwon S Il, Corvalán C, Cho SW, Kim H, et al. DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. Nat Biotechnol. 2015;33:1162–4.

58. Svitashev S, Schwartz C, Lenderts B, Young JK, Cigan AM. Genome editing in maize directed by CRISPR-Cas9 ribonucleoprotein complexes. Nat Commun. 2016;7:13274.

59. Zischewski J, Fischer R, Bortesi L. Detection of on-target and off-target mutations generated by CRISPR/Cas9 and other sequence-specific nucleases. Biotechnol Adv. 2017;35:95–104.

60. Bae S, Kweon J, Kim HS, Kim J-S. Microhomology-based choice of Cas9 nuclease target sites. Nat Methods. 2014;11:705–6.

61. Vu GTH, Cao HX, Fauser F, Reiss B, Puchta H, Schubert I. Endogenous sequence patterns predispose the repair modes of CRISPR/Cas9-induced DNA double strand breaks in *Arabidopsis thaliana*. Plant J. 2017.

62. Yang Z, Steentoft C, Hauge C, Hansen L, Thomsen AL, Niola F, et al. Fast and sensitive detection of indels induced by precise gene targeting. Nucleic Acids Res. 2015;43:e59.

63. Mock U, Hauber I, Fehse B. Digital PCR to assess gene-editing frequencies (GEF-dPCR) mediated by designer nucleases. Nat Protoc. 2016;11:598–615.

64. Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol. 2014;32:347–55.

65. CRISPRko. https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design. Accessed 16 Feb 2017.

66. WU-CRISPR. http://crispr.wustl.edu/. Accessed 16 Feb 2017.

67. Christensen AH, Sharrock RA, Quail PH. Maize polyubiquitin genes: structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation. Plant Mol Biol. 1992;18:675–89.

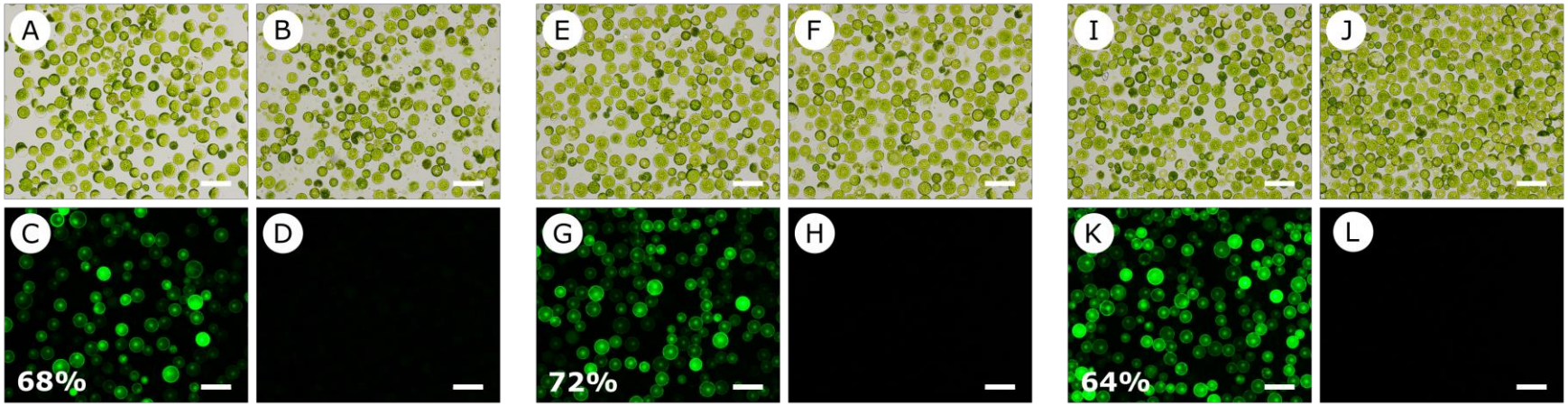
68. Christensen AH, Quail PH. Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants. Transgenic Res. 1996;5:213–8.

69. Sasanuma T. Characterization of the rbcS multigene family in wheat: subfamily classification, determination of chromosomal location and evolutionary analysis. Mol Genet Genomics. 2001;265:161–71.

70. Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.

71. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–2079.

Additional files



Additional file 1. Bright field and dark field microscopy of protoplasts co-transformed with pUbi-Cas9-rbcS and pUbi-YFP-rbcS (A, C, E, G, I, K), and untransformed protoplasts (B, D, F, H, J, L). Three replicates (A-D, E-H, and I-L) for each treatment are shown, along with the transformation efficiency for each replicate (% of protoplasts transformed). Scale bar = 100 µm.

Additional file 2. Forward and reverse Sanger sequence reads of homoeoallele-specific amplicons derived from protoplasts treated with gRNA1. (ZIP/AB1 2.8 MB)

Additional file 3. Forward and reverse Sanger sequence reads of homoeoallele-specific amplicons derived from protoplasts treated with gRNA2. (ZIP/AB1 2.8 MB)

Additional file 4. Forward and reverse Sanger sequence reads of homoeoallele-specific amplicons derived from protoplasts treated with gRNA3. (ZIP/AB1 2.8 MB)

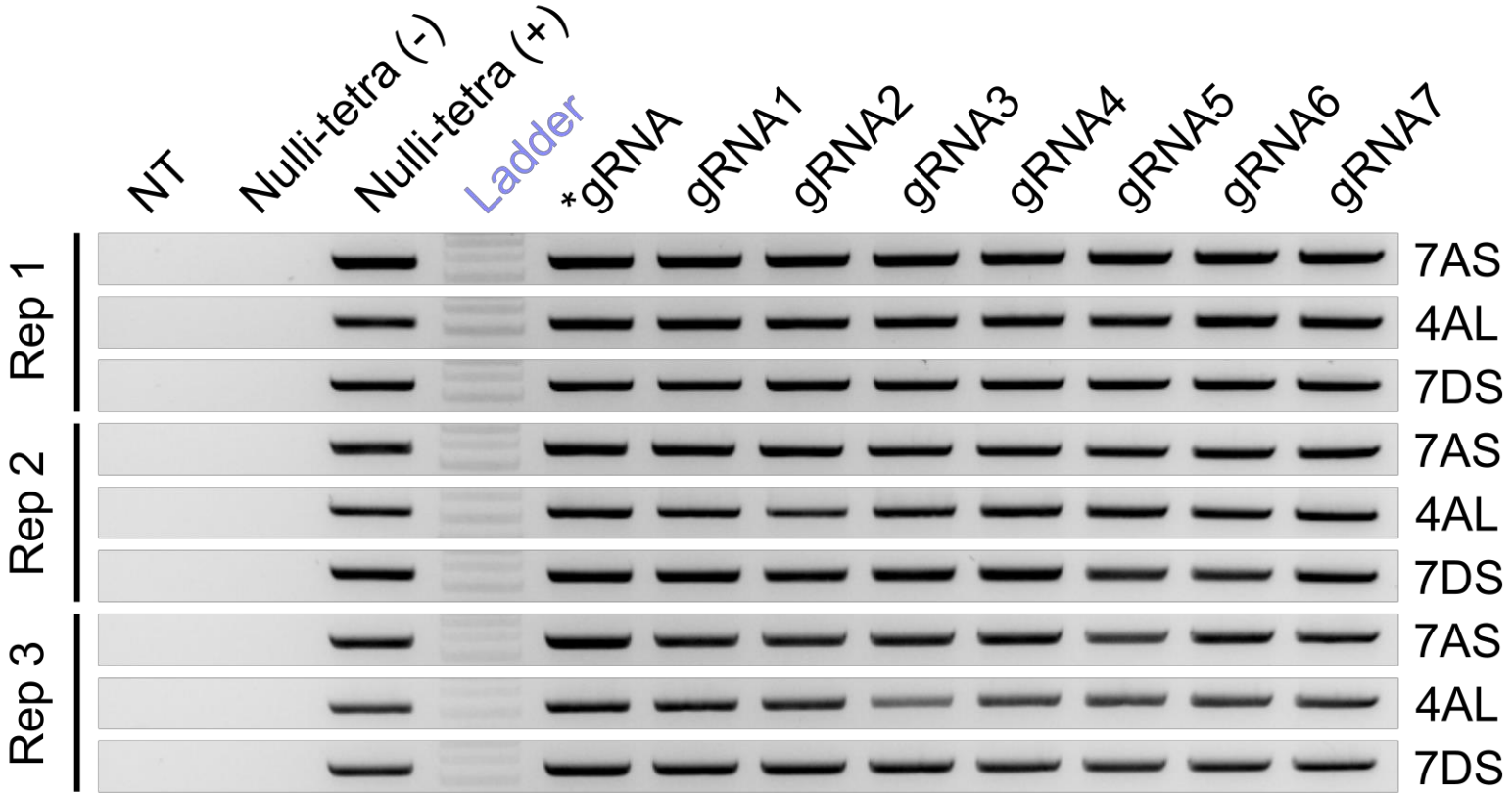
Additional file 5. Forward and reverse Sanger sequence reads of homoeoallele-specific amplicons derived from protoplasts treated with gRNA4. (ZIP/AB1 2.9 MB)

Additional file 6. Forward and reverse Sanger sequence reads of homoeoallele-specific amplicons derived from protoplasts treated with gRNA5. (ZIP/AB1 2.9 MB)

Additional file 7. Forward and reverse Sanger sequence reads of homoeoallele-specific amplicons derived from protoplasts treated with gRNA6. (ZIP/AB1 2.8 MB)

Additional file 8. Forward and reverse Sanger sequence reads of homoeoallele-specific amplicons derived from protoplasts treated with gRNA7. (ZIP/AB1 2.8 MB)

Additional file 9. Forward and reverse Sanger sequence reads of homoeoallele-specific amplicons derived from protoplasts treated with a non-targeting (random guide sequence) negative control gRNA. (ZIP/AB1 2.8 MB)



Additional file 10. Homoeoallele-specific amplification of *EPSPS* on chromosomes 7AS, 4AL and 7DS following transient co-expression of Cas9 and gRNA in wheat protoplasts. Nulli-tetra (-) genomic DNA template does not contain the target chromosome. Nulli-tetra (+) genomic DNA template does contain the target chromosome. NT, no template; *gRNA, non-targeting (random guide sequence) gRNA. Three replicates were performed. All bands to the right of the ladder were gel purified and Sanger sequenced in the forward and reverse directions.

Additional file 11. TIDE indel spectra/frequencies for gRNAs 1-7 targeting *EPSPS* in wheat protoplasts. Shown are results for forward and reverse Sanger sequence reads of homoeoallele-specific amplicons derived from chromosomes 7AS, 4AL, and 7DS. Three replicates were performed. (PDF 9.8 MB)

Additional file 12. CRISPResso allele frequency tables used for analysis of indels induced by gRNA2 in replicate 1. (ZIP/XLS 224 KB)

Additional file 13. CRISPResso NHEJ pie charts. (PDF 1.3 MB)

Additional file 14. CRISPResso allele frequency tables used for analysis of large insertions (≥20 bp) induced by gRNA2 and gRNA5 in replicate 3. (ZIP/XSL 4.4 MB)

Additional file 15. Primers and oligonucleotides used in this study.

Oligo name	Sequence 5'-3' (non-binding in red)	Use
TaEPSPS_Ex1_3'_F2	GCTCCAAGTCGCTCTCCAAC	Cloning of EPSPS on 7AS/4AL
TaEPSPS_Ex5_R2	TTCGACAGTCACAGTCCCTCC	Cloning of EPSPS on 7AS/4AL and Sanger seq of pCR8-TaEPSPS-FL clones
TaEPSPS_D_Int6_R	CGCACGTGCTAGATTGC	Cloning of EPSPS on 7DS
TaEPSPS_Ex2_F	CTGTGGTTGTTGGCTGTGG	Sanger seq of pCR8-TaEPSPS-FL clones
TaEPSPS_Ex3_R2	CGCTCCCTCATTCTTGGTAC	Sanger seq of pCR8-TaEPSPS-FL clones
TaEPSPS-gRNA1_F	CTTGAACTGCAATGCGGCCACTGA	Sense oligo for gRNA1
TaEPSPS-gRNA1_R	AAACTCAGTGGCCGCATTGCAGTT	Antisense oligo for gRNA1
TaEPSPS-gRNA2_F	CTTGCTACTACAGCTGCCGTCAG	Sense oligo for gRNA2
TaEPSPS-gRNA2_R	AAACCTGACGGCAGCTGTAGTAG	Antisense oligo for gRNA2
TaEPSPS-gRNA3_F	CTTGATCAGTCTCCAAACATACCCA	Sense oligo for gRNA3
TaEPSPS-gRNA3_R	AAACTGGGTATGTTTGGAGACTGAT	Antisense oligo for gRNA3
TaEPSPS-gRNA4_F	CTTGAAATAAGTATGAGATCCAT	Sense oligo for gRNA4
TaEPSPS-gRNA4_R	AAACATGGATCTCATACTTATTT	Antisense oligo for gRNA4
TaEPSPS-gRNA5_F	CTTGGTCGAAAAGGACGCCAAAG	Sense oligo for gRNA5
TaEPSPS-gRNA5_R	AAACCTTTGGCGTCCTTTTCGAC	Antisense oligo for gRNA5
TaEPSPS-gRNA6_F	CTTGAAGCTCTTCTTGGGTAATGC	Sense oligo for gRNA6
TaEPSPS-gRNA6_R	AAACGCATTACCCAAGAAGAGCTT	Antisense oligo for gRNA6
TaEPSPS-gRNA7_F	CTTGTAATGCTGGAACTGCAATG	Sense oligo for gRNA7
TaEPSPS-gRNA7_R	AAACCATTGCAGTTCCAGCATTA	Antisense oligo for gRNA7
TaEPSPS_A_Int1_F	AGCTTGTATTTGGATCAAGATATC	Amplification of EPSPS on 7AS
TaEPSPS_A_Int4_R	TCAGTAAAAGCGTAGATGTGAC	Amplification of EPSPS on 7AS
TaEPSPS_B_Int1_F	TTCTGATGGACCCTTTATGAG	Amplification of EPSPS on 4AL
TaEPSPS_B_Int4_R	CAAACAACTGCCGATGTT	Amplification of EPSPS on 4AL
TaEPSPS_D_Int1_F	CATTGGATTGTGTAGCTTGTAC	Amplification/cloning of EPSPS on 7DS
TaEPSPS_D_Int4_R	GCCAAACAACTGCCGATATC	Amplification of EPSPS on 7DS
TaEPSPS_A_Int1_F2	GATCCGTAGCTTGTTTCTGG	Sanger seq of amplicons from 7AS
TaEPSPS_A_Int3_R	ACAGAAGTGTGCACAAAGTAG	Sanger seq of amplicons from 7AS
TaEPSPS_B_Int1_F2	TCACTGTCATATCTAGCTGATCT	Sanger seq of amplicons from 4AL
TaEPSPS_B_Int3_R	GTACATAACATGCAATTTCACG	Sanger seq of amplicons from 4AL
TaEPSPS_D_Int1_F2	TAGCTGATCCGTAGCTTGTC	Sanger seq of amplicons from 7DS
TaEPSPS_D_Int3_R	GTACATAACATGCAAGTTCACA	Sanger seq of amplicons from 7DS
TaEPSPS_NGS_F_P5	TCGTCGGCAGCGTCAGATGTGTATAAG AGACAGGTGGAAGCAGATAAAGTTGC	First round of amplification for NGS
TaEPSPS_NGS_R_P7	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGCCATCAAGCACATAACTGAAG	First round of amplification for NGS

CHAPTER 4

General discussion

This PhD project was designed to:

1. Test the CRISPR/Cas9 system and apply it to wheat for the production of mutant lines.
2. Develop tools and methods for optimising the CRISPR/Cas9 system for genome editing in wheat.

The main findings are chapter specific and presented in the experimental manuscripts (Chapters 2 and 3), but in short they are:

I. The CRISPR/Cas9 system can be used for targeted gene knockout in wheat, and mutant lines can be regenerated.
II. The choice of gRNA can have a significant impact on the efficiency and specificity of the CRISPR/Cas9 system.
III. gRNA activity and specificity can be rapidly assessed in wheat protoplasts.

Significance of the work and contribution to knowledge

The CRISPR/Cas9 system is a powerful tool for crop improvement. However, the application of this technology to wheat is particularly challenging due to wheat being hexaploid and recalcitrant to transformation. In addition, a high-quality annotated reference genome sequence for wheat has only recently become available. In diploid crops such as rice, the CRISPR/Cas9 system can be used to generate biallelic mutations (complete gene knockout) in the first generation with high efficiency (Zhang *et al.*, 2014). In wheat, complete gene knockout can be difficult to achieve in the first generation due to gene functional redundancy (Zhang *et al.*, 2017; Singh *et al.*, 2018). In this PhD project, we circumvented this problem by selecting a target gene (*Ms1*) whose homoeologs are non-functional. In doing so, we successfully generated knockout lines in the first generation, thereby avoiding the need to carry out additional conventional breeding steps to recover the mutant phenotype (male sterility). This strategy could potentially be used for other wheat genes whose homoeologs are non-functional. Although it was already known that recessive mutations in

Ms1 result in male sterility (Tucker *et al.*, 2017), there had been no demonstration of a method for rapidly introducing recessive *Ms1* alleles into different wheat varieties. We showed that the CRISPR/Cas9 system can be used for this purpose. As a result, *Ms1* can now be deployed for commercial hybrid seed production.

We found that the choice of gRNA can have a significant impact on the efficiency of targeted mutagenesis. Only one out the three gRNAs targeting *Ms1* worked. Therefore, we established a method for the rapid assessment of gRNA activity and specificity in wheat protoplasts. Although we built on previous work (Shan *et al.*, 2013, 2014), our method is more comprehensive and we also discovered that *in silico* prediction of gRNA activity is currently not very accurate, at least for small sample sizes. Furthermore, we identified a gRNA that could be used in future experiments to develop non-transgenic glyphosate-resistant wheat lines.

Problems encountered

The main problems encountered during this project were related to transformation and mutation detection. Agrobacterium-mediated transformation was carried out by following a published protocol (with minor modifications) that is reported to give transformation efficiencies of 50-90% (Ishida *et al.*, 2015). However, the transformation efficiencies in this project were only 8% for cultivar Gladius and 11% for cultivar Fielder. The lower transformation efficiencies could be due to a variety of factors, including minor modifications to the published protocol, such as the use of Agrobacterium strain AGL1 instead of strain EHA105, and the use of a different T-DNA binary vector. However, such minor modifications are unlikely to account for such a large reduction in transformation efficiencies. The most likely explanation is differences in plant growth conditions. As stated in the original protocol (Ishida *et al.*, 2015), the quality of the immature embryos is one of the most important factors for achieving highly efficient wheat transformation, and the production of high-quality embryos depends on the use of optimised greenhouse conditions. Unfortunately, greenhouse conditions can be very difficult to control year-round, especially when

there is seasonal variation, as is the case in the location where this project was carried out. A suitable growth chamber was not available for this project. Therefore, options for dealing with this problem were limited.

Protoplast transformation was also challenging in the beginning. In the first experiment, DNA delivery to protoplasts failed due to the way in which the samples were processed. Specifically, protoplasts were added to each tube and mixed with the DNA, and then PEG was added to each tube. As there were many samples (> 30) in the first experiment (because the aim was to optimise several variables), this processing method resulted in prolonged incubation of DNA with protoplasts prior to the addition of PEG. This resulted in a transformation efficiency of close to 0%, presumably because the DNA was rapidly degraded by nucleases released from broken protoplasts. The importance of avoiding prolonged incubation of DNA with protoplasts prior to the addition of PEG was realised after two samples of leftover wildtype protoplasts from the first experiment were successfully transformed (albeit at low frequencies of around 20%) the following day. Fresh protoplasts should show higher transformation efficiencies, so this result was unexpected. However, as only two samples were processed, prolonged incubation of DNA with protoplasts prior to the addition of PEG was unintentionally avoided. Follow up experiments confirmed that protoplast transformation efficiencies decrease as DNA/protoplast incubation time increases. Even if the DNA is incubated with protoplasts for only five minutes, transformation efficiencies are significantly reduced. This is consistent with a previous report on oil palm protoplast transformation (Masani *et al.*, 2014). In that study, increasing the incubation time from 10 minutes to 15 minutes or 30 minutes reduced the transformation efficiency from 2.5% to 1.42% and 0.65%, respectively (Masani *et al.*, 2014). Unfortunately, the authors did not test the effect of reducing the incubation time, and therefore it seems likely that the low transformation efficiency of 2.5% was due to the DNA being incubated with protoplasts for 10 minutes. It is hoped that the method reported in Chapter 3 of this thesis will help researchers to avoid this technical pitfall in the future.

Despite solving the above problem, protoplast transformation efficiencies were still relatively low.

Results from optimisation experiments showed that the low transformation efficiencies were due to the use of a high protoplast-to-DNA ratio. Protoplast transformation was carried out by following a published protocol (Shan *et al.*, 2014), with modifications. In the original protocol, it is recommended to mix 500,000 protoplasts (in 200 µL) with 20 µg of each DNA construct (Shan *et al.*, 2014). Despite many attempts in this project, high transformation efficiencies were never achieved using that protoplast-to-DNA ratio. However, reducing the number of protoplast to 60,000 boosted transformation efficiencies dramatically. High transformation efficiencies (typically around 70%) were consistently achieved by using the lower protoplast-to-DNA ratio and avoiding prolonged incubation of DNA with protoplasts prior to the addition of PEG. In addition, due to the extreme fragility of protoplasts, handling them very gently at all times was critical to success.

Mutation detection was challenging due to wheat being hexaploid. Since homoeologous SNPs and indels can produce false positives, it was necessary to design and test several sets of homoeoallele-specific primers. Unfortunately, the best set of primers produced amplicons that were larger than the ideal size and this may have caused problems downstream. Initially, the T7E1 enzyme was used to detect mutations in DNA extracted from transformed protoplasts, but this method produced high background noise on the agarose gel, possibly due the to large amplicon size. The individual strands of larger amplicons could be expected to form more complex secondary structures following the denaturation and re-annealing step of the T7E1 protocol. Such secondary structures might result in DNA mismatches that are cleaved by the T7E1 enzyme, resulting in high background noise. To deal with this problem, the T7E1 assay was not used again, and instead the amplicons were subjected to Sanger sequencing. The Sanger sequence traces were then analysed using the TIDE tool. This approach worked well, as reported in Chapter 3.

Amplicon deep sequencing was also used for mutation detection, and this approach presented significant challenges during the bioinformatics analysis. In particular, the high sequence similarity between homoeoalleles made it difficult to assign individual reads to the correct homoeoallele. Several online tools were evaluated to see if they could reliably assign reads to the correct

homoeoallele, but none of them produced satisfactory results. Therefore, a custom script had to be written, and this was not straightforward.

Future research directions

Research on CRISPR/Cas9 and other CRISRP systems is advancing at a rapid pace and this is likely to continue. Plant transformation and regeneration is still a major bottleneck in the genome editing pipeline, but exciting new technologies such as nanoparticle-mediated transformation are starting to be developed. New viral vectors, including synthetic viral vectors, could potentially help to reduce the current reliance on time-consuming and laborious conventional transformation methods that use tissue culture. In addition, the low efficiency of editing via HDR is a significant limitation. The use of Cas9 nickases in combination with rationally designed ssDNA donor templates needs to be tested, especially for small sequence changes. Cpf1 could potentially be used for making large insertions, as it generates DSBs with sticky ends. The development and optimisation of CRISPR/Cas9 base editing systems is another high priority, particularly as this would in many cases obviate the need to harness HDR. In addition, improved tools are needed for *in silico* prediction of gRNA activity and identification of potential off-target sites. At the same time, CRISPR/Cas9 should be applied for crop improvement, and this will require identification and prioritisation of gene targets. The recent *de novo* domestication of wild tomato is exemplary. The future looks bright for plant genome editing.

References

Ishida, Y. *et al.* (2015) 'Wheat (Triticum aestivum L.) transformation using immature embryos', *Methods in Molecular Biology*, 1223, pp. 189–198. doi: 10.1007/978-1-4939-1695-5.

Masani, M. Y. A. *et al.* (2014) 'Efficient transformation of oil palm protoplasts by PEG-mediated transfection and DNA microinjection', *PLOS ONE*, 9(5), pp. 1–11. doi:

10.1371/journal.pone.0096831.

Shan, Q. *et al.* (2013) 'Targeted genome modification of crop plants using a CRISPR-Cas system', *Nature Biotechnology*, 31(8), pp. 686–688. doi: 10.1038/nbt.2652.

Shan, Q. *et al.* (2014) 'Genome editing in rice and wheat using the CRISPR/Cas system', *Nature Protocols*. Nature Publishing Group, 9(10), pp. 2395–410. doi: 10.1038/nprot.2014.157.

Singh, M. *et al.* (2018) 'Concurrent modifications in the three homeologs of Ms45 gene with CRISPR-Cas9 lead to rapid generation of male sterile bread wheat (Triticum aestivum L.)', *Plant Molecular Biology*, 97, pp. 371–383. doi: 10.1007/s11103-018-0749-2.

Tucker, E. J. *et al.* (2017) 'Molecular identification of the wheat male fertility gene *Ms1* and its prospects for hybrid breeding', *Nature Communications*, 8(1), p. 869. doi: 10.1038/s41467-017-00945-2.

Zhang, H. H. *et al.* (2014) 'The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation', *Plant Biotechnology Journal*, 12(6), pp. 797–807. doi: 10.1111/pbi.12200.

Zhang, Y. *et al.* (2017) 'Simultaneous modification of three homoeologs of TaEDR1 by genome editing enhances powdery mildew resistance in wheat', *The Plant Journal*, pp. 1–11. doi: 10.1111/tpj.13599.