Heterologous Expression and Functional Analysis of *Plantago GT61* and *DUF579* Genes in *Arabidopsis thaliana*

By

Lina Herliana

A thesis submitted for the partial fulfilment of the requirements of the Master of Biotechnology (Plant Biotechnology)

The University of Adelaide

Faculty of Sciences

School of Agriculture, Food & Wine

Waite Campus

2017

DECLARATION

I declare that this thesis is a record of original work and contains no material which has been accepted for the award of any other degree or diploma in any university. To the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text.

Lina Herliana
8 November 2017

TABLE OF CONTENTS

DECLARATION i

TABLE OF CONTENTS ii

PREFACE vii

ABSTRACT 1

INTRODUCTION 2

RESULTS 5

Phylogenetic analysis of *GT61* and *DUF579* candidate genes 5

Heterologous expression of *PoGT61_7* driven by a seed-coat promoter in transgenic *Arabidopsis* 12

Mucilage staining with Ruthenium red (RR) 14

Design of single guide RNAs for CRISPR/Cas9 editing of *DUF579* 18

In silico predictions of genome editing 20

In vitro testing of genome editing 22

Evaluating efficiency of guide RNAs (sgRNAs) for targeted genome modification 24

DISCUSSION 26

P. ovata GT61_7 is not directly homologous to gene in *A. thaliana* 26

PoGT61_7 expression controlled by *Arabidopsis* seed coat promoter 27

Overexpression of *PoGT61_7* caused a seed mucilage phenotype 27

AT1G71690 is the only seed specific *DUF579* 28

Low-efficiency protoplast transfection could be the reason for unsuccessful genome editing 29

CONCLUSION 30

MATERIAL AND METHODS 31

Plant material and growth conditions 31

Phylogenetic analyses of *GT61* and *DUF579* genes 31

Constructing an expression vector to overexpress *Plantago GT61_7* in *A. thaliana* 32

Agrobacterium and Plant transformation 32

BASTA selection and transgene detection..32

RNA extraction, cDNA synthesis, and real time qPCR analysis 33

Ruthenium red whole seed staining 33

Selection of sgRNA targets on an *Arabidopsis DUF579* gene (AT1G71690) 33

Construction of AT1G71690 sgRNA and Cas9 expression vectors 34

Transient expression of the CRISPR/Cas9 construct in *Arabidopsis* protoplasts.............35

ACKNOWLEDGEMENTS..36

LITERATURE CITED 37

APPENDIX 1 – SUPPLEMENTARY DATA 41

APPENDIX 2 – PLANT PHYSIOLOGY INSTRUCTIONS FOR AUTHORS......................47

LIST OF FIGURES

Figure 1. Phylogenetic tree of GT61 protein sequences selected from four species. 8

Figure 2. Phylogenetic tree of DUF579 protein sequences from three selected species. 10

Figure 3. Transcript levels of *Arabidopsis DUF579* genes in different tissues and during plant development stages as indicated by publically available RNA-seq data in DRASearch...11

Figure 4. Overexpression of *PoGT61_7* in *A. thaliana*. 13

Figure 5. Wild-type and transgenic seed stained using ruthenium red (RR) 0.01% (w/v) without agitation. 15

Figure 6. Wild-type and transgenic seed stained using ruthenium red 0.01% (w/v) after agitation (400 rpm) for three hours. 16

Figure 7. A preliminary comparison of the thickness of the inner mucilage layer in wild-type versus T_2 seed from three transgenic plants (P1, P18, and P23). 17

Figure 8. Single guide RNA targets were generated using six online tools as indicated by the coloured arrows with a rank order from 1 to 12. 19

Figure 9. *In silico* analysis of potential genomic editing scenarios by the three selected sgRNAs 21

Figure 10. Protoplast isolation and transfection. 23

Figure 11. A pool of *Arabidopsis* protoplast cells treated with Cas9 alone (control) and cells treated with Cas9 and targeting sgRNAs were analysed by TIDE..........................25

LIST OF SUPPLEMENTARY FIGURES

Figure S1. An alignment of GT61 protein from *P. ovata*, *A. thaliana*, *O. sativa,* and *P. trichocarpa* 41

Figure S2. An alignment of DUF579 protein from *A. thaliana*, *P. ovata*, *M. guttatus* 42

Figure S3. *PoGT61_7* with seed-coat promoter in an expression vector 43

Figure S4. An example of expression vector (pDe_Cas9) with sgRNA.....44

Figure S5. An example of CRISPR/Cas9 target from CCTop with the off-target prediction. .45

LIST OF SUPPLEMENTARY TABLES

Table S1. Sequences of all primers and sgRNAs 46

ACKNOWLEDGEMENTS

Alhamdulillah, I can reach this stage, finishing my Master project.

I would like to express my tremendous thanks to Prof. Rachel Burton, who allowed me to learn and explore so many techniques, who taught me to be a good scientist (always include a control and confirm the result!), who connected me with a lot of expertise in this campus, who was ready to give me suggestion anytime especially when my experiment did not work. I will always remember your word “Keep going!”, such powerful words. Thank you for being so patient in guiding me to do research and in correcting the manuscript draft. You are indeed a super supervisor.

Such wonderful experience to do experiments with a lot of support from the University of Adelaide, Plant Cell Walls group and Tucker Lab members. Thank you, Assoc. Prof. Matthew Tucker, Shi Fang (Sandy) Khor, Weng Herng Leong, James Cowley, Dr. Ali Saleh Hasan, Dayton Bird, Jana Phan, Kylie Neumann, Wai Li Lim, and Rosanna Petrella for helping me to troubleshoot my experiments. Thanks to Natalie Kibble, Emma Drew, Chao Ma, and Robyn McBride for a quick system in management and logistics. For the examiners, Caitlin Byrt and Natalie Betts, thank you for valuable suggestions for the manuscript. Special thanks to Kum Foeng (Maple) Ang who assisted me from the very beginning I am in the lab and always be happy to lend me your stuff. To Taj Arndell, I really appreciate your guidance in protoplast experiment, so detail and helpful.

I also want to say thanks to my friend, Elhabti Abdeljalil, who is willing to help me to understand my course during my Master.

To my husband, Alkindi, thank you for your understanding and you always support me to pursue my dream. To my mom, who always pray for my success, thank you for taking care my sweetheart, Aqilah.

Finally, I would like to express my gratitude to Australia Awards Scholarship for generous scholarship and my workplace, BPPT (Badan Pengkajian dan Penerapan Teknologi) to permit me to upgrade my skills.

PREFACE

This research was performed over 10 months as part of a Master of Biotechnology (Plant Biotechnology). In accordance with the requirements of the program, the research is presented in the format of a manuscript for submission to a peer-reviewed scientific journal. I have chosen to follow the format of *Plant Physiology*. My co-author for the manuscript is Prof. Rachel Burton. Prof. Rachel Burton suggested the project, supervised my research and reviewed drafts of the manuscript.

The manuscript in this thesis is intended as the first draft of a manuscript for future publication, once further data have been collected on immunolabeling and monosaccharide analysis of *Arabidopsis* transgenic lines. The word count for the manuscript (excluding references and supplementary material) is 7213 words.

Appendix 1 contains supplementary data. Appendix 2 contains the instructions to authors for the journal *Plant Physiology*.

I have followed these instructions except that I have presented the manuscript in double spacing, Times New Roman 12 on A4 pages, with a 3.5-cm margin on the left-hand side of the page and 1.5-cm margins at the top, bottom and right-hand side of the page, without line numbering, and figures and table are inserted at appropriate places in the text in order to satisfy the thesis guidelines for the Master of Biotechnology (Plant Biotechnology) program.

JOURNAL SUBMISSION REQUIREMENTS:

Short Title:

Analysis of genes related to xylan biosynthesis

Corresponding author details:

Rachel A. Burton*
*ARC Centre of Excellence in Plant Cell Walls, University of Adelaide, Waite Campus

Article title:

Heterologous expression and functional analysis of *Plantago GT61* and *DUF579* genes in *Arabidopsis thaliana*

All authors and affiliations:

Lina Herliana[1] and Rachel A. Burton[1]

[1]ARC Centre of Excellence in Plant Cell Walls, University of Adelaide, Waite Campus

One sentence summary:

Expression of *Plantago GT61_7* in *Arabidopsis* lines produced a phenotype in the inner adherent seed mucilage layer whilst genome editing potential for *DUF579* gene was explored.

List of author contributions:

R.A.B conceived and designed the project and provided critical review of the manuscript; L.H performed the experiments, analysis and wrote the article

Funding information:

All authors are supported by the ARC Centre of Excellence in Plant Cell Walls. L.H is also supported by an Australia Awards Scholarship

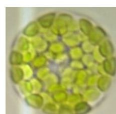
Present addresses:

ARC Centre of Excellence in Plant Cell Walls, School of Agriculture, Food and Wine, University of Adelaide, Waite Campus, Glen Osmond, SA 5064, Australia;

Corresponding author email:

rachel.burton@adelaide.edu.au

Table of contents icon:



ABSTRACT

Mucilage released from *Plantago ovata* seed (psyllium) has been used for maintaining human health as a dietary fibre supplement. Heteroxylan is the main component, and its substitution affects solubility and viscosity of the end product. However, little is known about genes involved in xylan substitution so phylogenetic and transcript information were used to identify candidate genes in the *GT61* and *DUF579* families and their functions were tested in the model plant *Arabidopsis thaliana*. *Plantago GT61*_7, driven by a seed-coat promoter (*Pro$_{DP1}$*) was transformed into *Arabidopsis* using a floral dip and spray method. Ruthenium red staining of wild-type and T_2 seeds from multiple independent transgenic lines showed a significant difference in the thickness of the adherent mucilage layer. The difference in mucilage phenotype suggests that GT61_7 may have a role in xylan substitution that affects seed coat adherence. This preliminary result needs to be examined using immunolabeling and monosaccharide analysis. For the *DUF579* gene AT1G71690, a genome editing approach was adopted. Three single guide RNAs were designed using online tools and *in silico* analysis was performed to predict any changes in coding and protein sequences by each guide RNA. To test them *in vitro*, the CRISPR/Cas9 constructs were successfully delivered to protoplast cells using the Transient Expression in *Arabidopsis* Mesophyll Protoplast (TEAMP) method. However, an analysis using Tracking of Indels by Decomposition(TIDE) showed no evidence of edits in the *DUF569* genomic DNA extracted from the protoplasts. Increasing the transfection efficiency or redesigning the sgRNA could lead to improved CRISPR/Cas9 activity.

Keywords: seed mucilage, glycosyltransferase (GT61), Domain of Unknown Function (DUF) 579, xylan, *Plantago ovata*, *Arabidopsis thaliana*, CRISPR/Cas9, protoplast

INTRODUCTION

Many angiosperms produce a viscous and sticky mucilage released from the mature seed upon wetting, which is a trait called myxospermy (Western, 2012; Francoz et al., 2015). This mucilage is thought to have diverse ecophysiological roles such as controlling dormancy and germination, aiding seed dispersal, and protecting seeds from abiotic and biotic stress (Western, 2012). There is particular scientific interest in seed mucilage (SM) because it is an excellent representative of the plant cell wall in an accessible form, is not essential for plant development under laboratory conditions, and is easy to extract (Western et al., 2000; Voiniciuc et al., 2015; Ralet et al., 2016). These characteristics allow observation of mutants and enables the determination of gene function (Haughn and Western, 2012; Jensen et al., 2013). In addition to SM being useful for plant cell wall research, it also has commercial relevance since it can be used as a dietary fibre supplement and as a thickener in processed foods. The model plant *Arabidopsis* is myxospermous and given that it has a fully sequenced genome and many genetic resources available in some ways it is an ideal system to use to study seed mucilage. However, *Arabidopsis* seed mucilage is not industrially valuable and it is comprised of polysaccharides that are different to species that mucilage is used commercially, such as psyllium from *Plantago ovata*. On the other hand, *P. ovata* has far fewer genetic resources available, cannot be transformed and is not yet fully sequenced but recent work at the University of Adelaide is expanding the knowledge base of the *P. ovata* system. Therefore, *Arabidopsis* can be a useful mucilage model for some aspects, and can be transformed with heterologous genes.

SM properties are determined by polysaccharide composition and their molecular structures, particularly substitution. The primary components of SM are pectins and heteroxylans but relative contents differ among species (Western et al., 2000; North et al., 2014). *Arabidopsis* mucilage is composed primarily of pectic unbranched Rhamnogalacturonan I (RG I) (Naran et al., 2008; Arsovski et al., 2010) while *P. ovata* is rich in complex heteroxylans (Fischer et al., 2004; Guo et al., 2008). Heteroxylan is composed of a backbone of xylose residues decorated with a variety of side chains typically arabinose (Ara), xylose

(Xyl), glucuronic acid (GlcA), and traces of other sugars (Fischer et al., 2004; Ebringerová, 2005). Heteroxylan is a key component that defines the end-use of *P. ovata* mucilage in pharmaceutical and food industries (Khaliq et al., 2015). Therefore, a better understanding and potentially manipulation of the heteroxylan composition and structures could improve the value of seed mucilage.

The genes underlying xylan backbone formation belong to two families, which are glycosyltransferase (GT) 43 (*IRREGULAR XYLEM* (*IRX*) *9*, *IRX9-LIKE* (*IRX9-L*), *IRX14*, and *IRX14-L*) and GT47 (*IRX10* and *IRX10-L*) (Jensen et al., 2013; Jensen et al., 2014; Rennie and Scheller, 2014) whose expression is variable among plant species. Transcripts of *IRX10/IRX10-L* are higher in *Plantago* seed integuments than in *Arabidopsis* (Jensen et al., 2013) and in contrast, *Arabidopsis* seeds show higher expression levels of *IRX14/IRX14-L* (Voiniciuc et al., 2015). Xylan backbone formation has been intensively studied, but the substitutions need to be better explored as they affect solubility and viscosity of the polysaccharide (Jensen et al., 2013; Phan et al., 2016; Ralet et al., 2016).

Candidate genes for xylan substitution belong to the *glycosyltransferase* (*GT*) *61* family. Various enzymatic activities have been identified for the GT61 proteins, including as arabinosyltransferases and xylosyltransferases (Anders et al., 2012; Voiniciuc et al., 2015) but many of these genes remain uncharacterised. Seven *GT61* sequences were identified in the mucilage-producing cells of *P. ovata* (Jensen et al., 2013). Phan et al. (2016) identified 10 *GT61* genes that are co-expressed with *IRX10* (that putatively encodes xylan synthases) in seed coat mucilage of *P. ovata* and *P. cunninghamii*. These *Plantago* species have a similar abundance of heteroxylan but are different in structure and *P. ovata* has more duplication of *GT61* genes compared to *P. cunninghamii*. However, they did not determine whether these GT61 proteins act as xylosyltransferases and/or arabinosyltransferases or if they add other, more minor, substituents. Heterologous expression of these genes in other species may reveal their function, for example, expression of wheat and rice *GT61s* in *Arabidopsis* increased arabinose

substitution and provided gain-of-function evidence for arabinosyltransferase activity (Anders et al., 2012).

Another gene family reported to be involved in influencing xylan structure is the Domain of Unknown Function (DUF) 579 (Brown et al., 2011; Jensen et al., 2011). In *Arabidopsis*, ten genes contain this domain and five of them have been characterized, which are *IRX15* and *IRX15L* (Brown et al., 2011; Jensen et al., 2011) and *GXM1*, *GXM2*, and *GXM3* (Lee et al., 2012). Lee et al. (2012) demonstrated that *irx15/15l* and *gxm* mutants have different phenotypes. A specific defect in glucuronic acid (GlcA) methylation on the xylan of *gxm* mutants versus pleiotropic phenotypes of *irx15/15l* indicate that GXMs are likely to be methyltransferases that are essential for GlcA methylation on xylan (Lee et al., 2012) but IRX15/15L proteins are not. The remaining five *DUF579* genes still need to be characterized. According to unpublished data from the Plant Cell Wall Group at the University of Adelaide, there are seven genes in *P. ovata* that contain the *DUF579* motif (J. Phan personal communication) and a mutation in one of these severely affects the solubility of the xylan (Phan et al, unpublished). Therefore, there is an opportunity to explore the function of uncharacterized *DUF579* genes.

The work presented here aimed to identify previously uncharacterised candidate genes involved in xylan substitution and to study their function in determining the final quality of seed mucilage. To do this a *P. ovata GT61* gene was overexpressed using transgenic *Arabidopsis* as a heterologous host whilst preliminary studies were undertaken to design an effective single guide RNA to undertake genome editing of an *Arabidopsis* DUF579 gene via the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR associated 9 (Cas 9) system.

RESULTS

Phylogenetic analysis of *GT61* and *DUF579* candidate genes

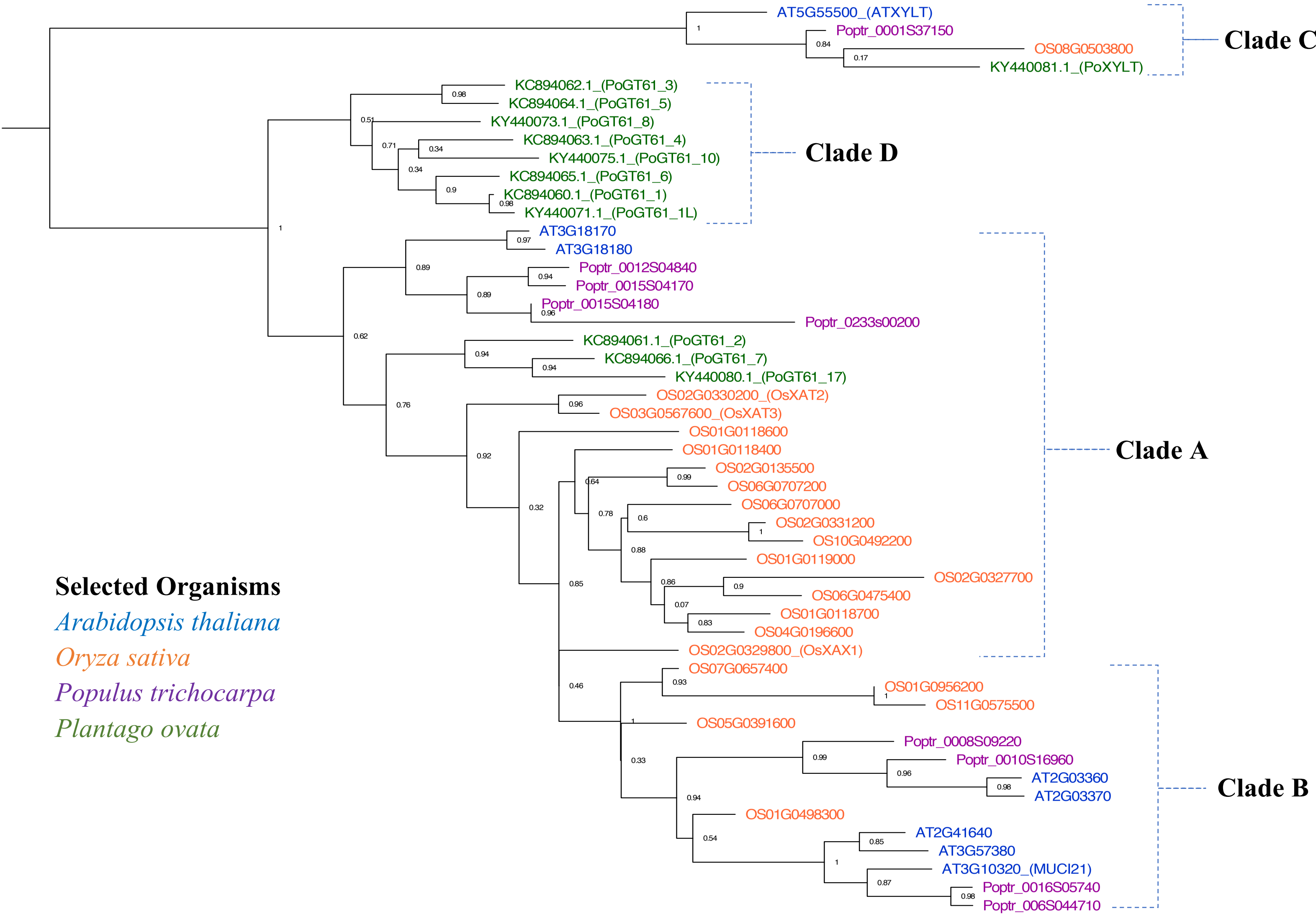
Bioinformatic analysis of *GT61* and *DUF579* genes and proteins identified target genes for functional analysis. After collecting and curating putative sequences, there were 50 *GT61* genes: 12 from *P. ovata*, 8 from *A. thaliana*, 2 from *Oryza sativa* and 9 genes from *Populus trichocarpa* for phylogenetic analysis. For *DUF579*, there were 25 genes, consisting of 10 *A. thaliana*, 7 *P. ovata*, and 8 *Mimulus guttatus* sequences to be sorted. Translation alignments were executed to identify the conserved regions (Figure S1 and S2) and phylogenetic trees were constructed to see the similarity of the sequences among the different species (Figure 1 and 2).

GT61 genes can be grouped into 4 clades A, B, C, and D (Figure 1). The naming of clades A-C follows Anders et al. (2012) and Voiniciuc et al. (2015). Clade D is a new clade, not previously identified. The majority of the *Plantago GT61s, PoGT61_1*, *1L*, *3*, *4*, 5, *6*, 8 and *10* are in this clade D, separated from most genes from other plants. Clade B contains no *P. ovata* sequences at all whilst D contains no *Arabidopsis* homologs. Clade A contains three *P. ovata* genes, *PoGT61_2*, *7*, and *17* whilst only one gene belongs to clade C. Based on the *Arabidopsis* sequence, this clade C gene has been annotated as *PoXLYT* with a putative function as a xylosyltransferase. Since there is no candidate gene *GT61* from *Arabidopsis*, *PoGT61_7* was chosen for functional analysis based on protein activity from unpublished data from Zeng, Burton, and Tucker (University of Adelaide).

In contrast, *Plantago* genes with a *DUF579* motif have many homologous genes in *Arabidopsis* and *Mimulus* as shown in Figure 2. The phylogenetic tree for *DUF579* can be divided into four clades, A, B, C, and D. *Plantago* contigs extracted from the Centre of Excellence in Plant Cell Walls proprietary database are found in all clades, and they have *Arabidopsis* homologs. Clades B and D have more members than the other clades. In clade B, two *P. ovata* contigs are close to *Arabidopsis IRX15* and *IRX15L* while three *P. ovata* contigs are close to *Arabidopsis GXMs* in clade D. No members of clades A or C have been annotated meaning that their functions are still unknown. Since all clades contain homologous genes from

both *P. ovata* and *A. thaliana*, further analysis was required to determine which candidate gene might influence the biosynthesis of mucilage in the seed coat.

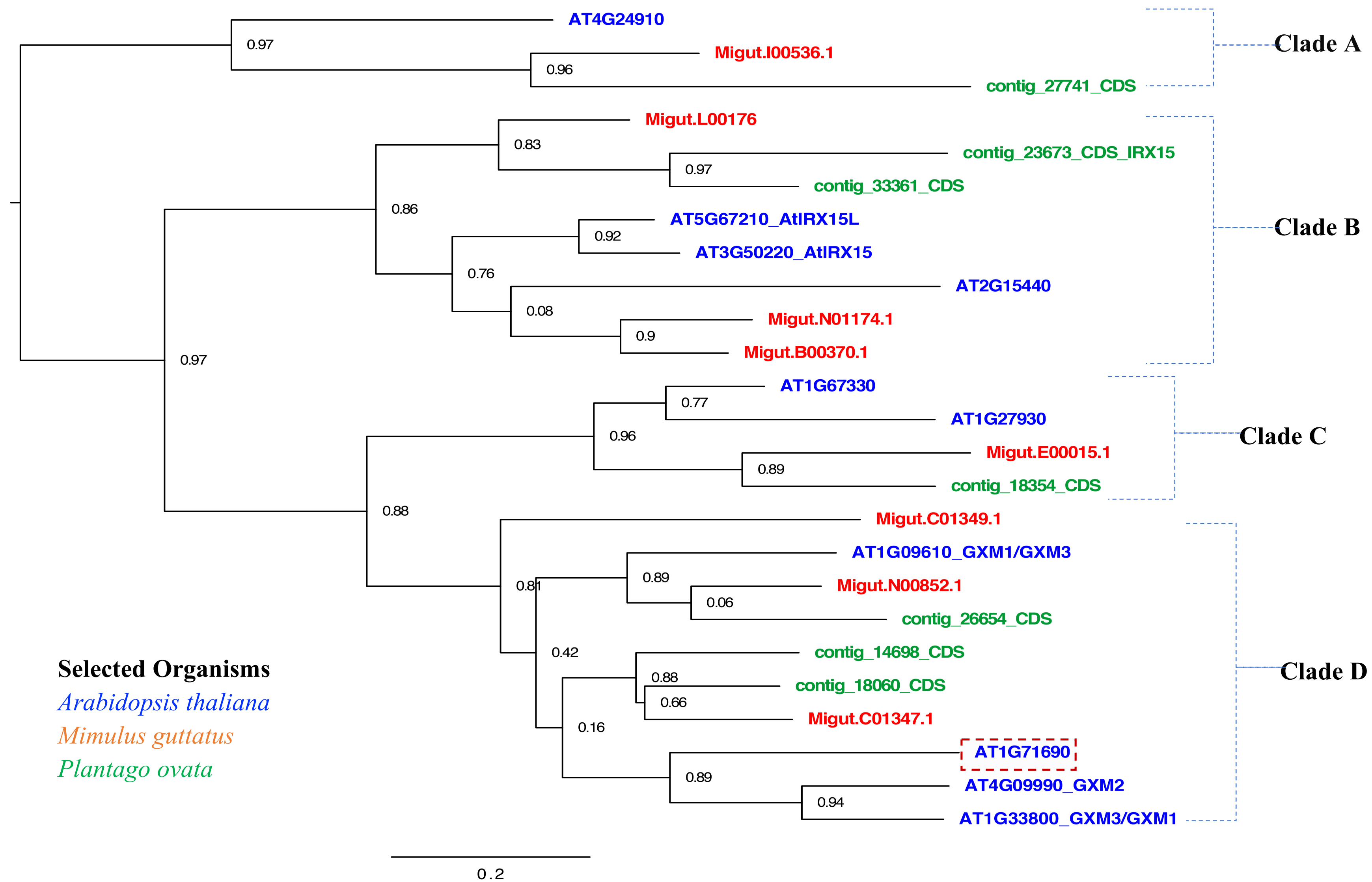
Therefore, the transcript levels of the 10 *Arabidopsis DUF579* genes were evaluated to determine their tissue specificity. Figure 3 shows transcript levels of all *Arabidopsis DUF579* genes during a range of plant developmental stages, as assessed from RNA sequencing data (DRASearch, http://trace.ddbj.nig.ac.jp/DRASearch/). Plant development can be divided into two phases, which are vegetative and generative stages. Vegetative stage occurs before flowering, and includes germinating seed, hypocotyl, root, leaf, and internode tissues, while the generative phase refers to plant tissues at flowering and fruit development stages. According to the transcript profiles across both vegetative and generative periods, these ten *DUF579* genes show six different expression patterns. Firstly, two genes consistently have low transcript levels (below 50 Transcripts per Million (TPM)) during plant development; they are AT2G15440 and AT4G24910. Secondly, only one gene (AT1G27930) is expressed during both vegetative and generative periods even though the expression fluctuates and is at a low level overall (50-100 TPM). Three genes have more transcript in the vegetative stage rather than in the generative stage, namely AT1G09610, AT1G33800, and ATG09990 whilst the fourth group has higher expression during the generative rather than the vegetative stage including AT3G50220 and AT5G67210. However, the third and fourth group have 6 similar peak locations which are in germinating seed, seedling root, internode (stem), root, pod silique, and silique. The last two patterns are represented by only a single gene, each with an entirely different expression time. AT1G67330 is only expressed during the vegetative stage with three prominent peaks in germinating seed at days 2 and 3, seedling root, and root tip with the highest peak around 140 million. Lastly, AT1G71690 is only expressed at the generative stage at a specific time during seed development and so is the only gene which is likely to be involved in mucilage production. This gene also has the highest expression level among all the DUF579 genes; it reaches around 300 TPM. Thus, AT1G7190 was chosen as the best candidate to study for its effect on SM using CRISPR/Cas9 editing.



AT5G55500_(ATXYLT)
Poptr_0001S37150
OS08G0503800
KY440081.1_(PoXYLT)
Clade C
KC894062.1_(PoGT61_3)
KC894064.1_(PoGT61_5)
KY440073.1_(PoGT61_8)
KC894063.1_(PoGT61_4)
KY440075.1_(PoGT61_10)
KC894065.1_(PoGT61_6)
KC894060.1_(PoGT61_1)
KY440071.1_(PoGT61_1L)
Clade D
AT3G18170
AT3G18180
Poptr_0012S04840
Poptr_0015S04170
Poptr_0015S04180
Poptr_0233s00200
KC894061.1_(PoGT61_2)
KC894066.1_(PoGT61_7)
KY440080.1_(PoGT61_17)
OS02G0330200_(OsXAT2)
OS03G0567600_(OsXAT3)
OS01G0118600
OS01G0118400
OS02G0135500
OS06G0707200
OS06G0707000
OS02G0331200
OS10G0492200
OS01G0119000
OS02G0327700
OS06G0475400
OS01G0118700
OS04G0196600
OS02G0329800_(OsXAX1)
Clade A
OS07G0657400
OS01G0956200
OS11G0575500
OS05G0391600
Poptr_0008S09220
Poptr_0010S16960
AT2G03360
AT2G03370
OS01G0498300
AT2G41640
AT3G57380
AT3G10320_(MUCI21)
Poptr_0016S05740
Poptr_006S044710
Clade B
Selected Organisms
Arabidopsis thaliana
Oryza sativa
Populus trichocarpa
Plantago ovata
0.3

Figure 1. Phylogenetic tree of GT61 protein sequences selected from four species.

Clades A–C are labelled as per Anders et al. (2012) and Voiniciuc et al. (2015). Bootstrap values are indicated and the scale bar shows evolutionary distance in units defined by the FastTree support value. Visualised using Figtree with a midpoint rooting option.



AT4G24910
0.97
Migut.I00536.1
0.96
contig_27741_CDS
Clade A
Migut.L00176
0.83
contig_23673_CDS_IRX15
0.97
contig_33361_CDS
0.86
AT5G67210_AtIRX15L
0.92
AT3G50220_AtIRX15
0.76
AT2G15440
0.08
Migut.N01174.1
0.9
Migut.B00370.1
Clade B
0.97
AT1G67330
0.77
AT1G27930
0.96
Migut.E00015.1
0.89
contig_18354_CDS
Clade C
0.88
Migut.C01349.1
AT1G09610_GXM1/GXM3
0.89
Migut.N00852.1
0.81
0.06
contig_26654_CDS
0.42
contig_14698_CDS
0.88
contig_18060_CDS
0.66
Migut.C01347.1
0.16
AT1G71690
0.89
AT4G09990_GXM2
0.94
AT1G33800_GXM3/GXM1
Clade D
Selected Organisms
Arabidopsis thaliana
Mimulus guttatus
Plantago ovata
0.2

Figure 2. Phylogenetic tree of DUF579 protein sequences from three selected species.

Bootstrap values are indicated and the scale bar shows evolutionary distance in units of FastTree support value. Visualised using Figtree with a midpoint rooting option.

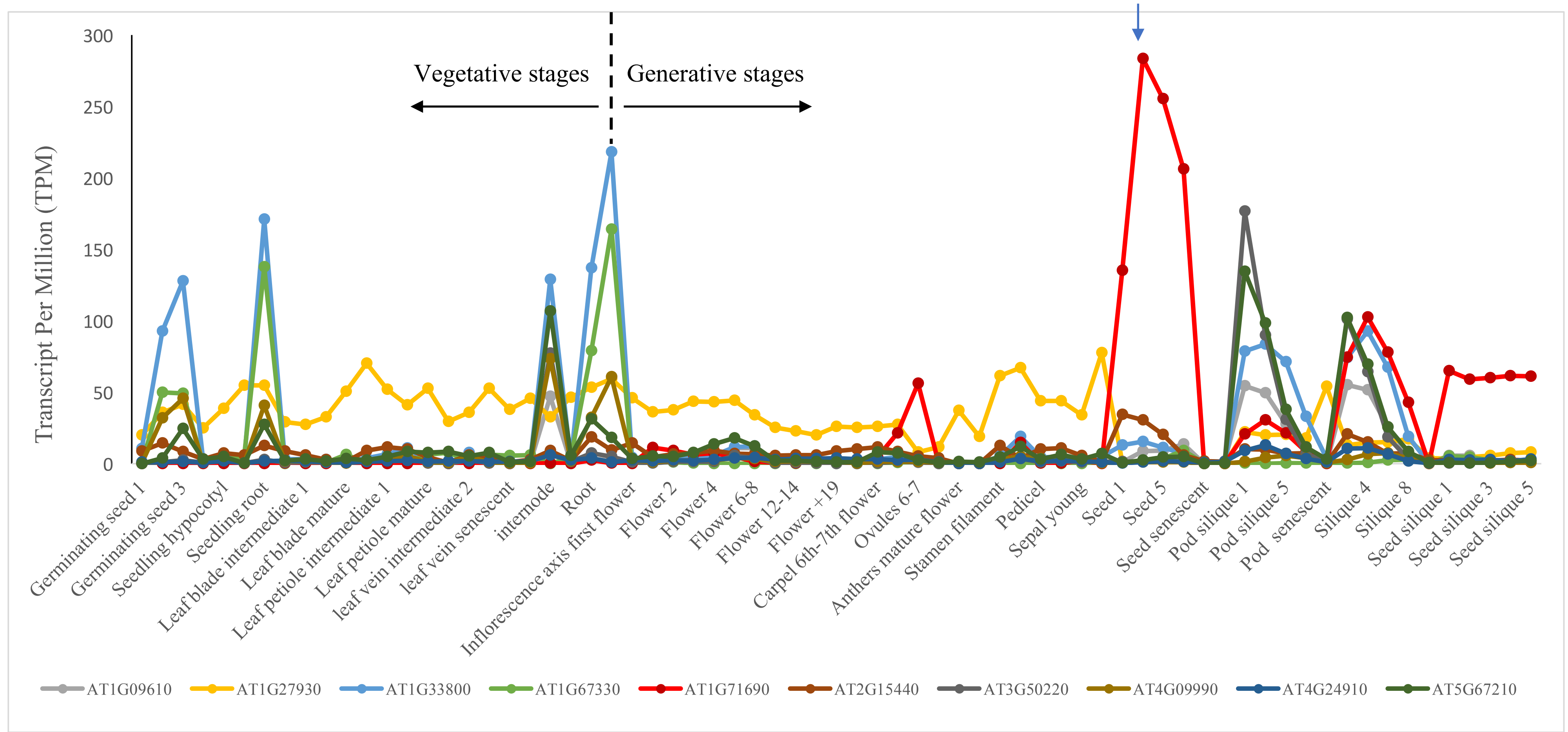


Figure 3. Transcript levels of *Arabidopsis DUF579* genes in different tissues and during plant development stages as indicated by publically available RNA-seq data in DRASearch.

(http://trace.ddbj.nig.ac.jp/DRASearch/). A blue arrow indicates the highest expression of AT1G71690.

Heterologous expression of *PoGT61_7* driven by a seed-coat promoter in transgenic *Arabidopsis*

Candidate gene *Plantago GT61_7* was successfully cloned into a destination vector containing a seed-coat promoter and *bar* resistant gene (Figure S3), transformed into *Agrobacterium* using freeze/thaw, then delivered into plants using a floral dip method. After selecting plants that were resistant to the herbicide Basta, 30 putative transgenic lines were obtained (Figure 4A). These plants were genotyped using the GT61_P1 primers (Table S1) that should amplify a 1.3 kb fragment from the genomic DNA as shown in Figure 4C. The primers also amplified a smaller product at around 500 bp so both products of 1.3 kb and 500 bp were sequenced and mapped to the *PoGT61_7* sequence using Geneious 8.1.9 software. The bigger product (1,302 bp) was confirmed to match the expected sequence *PoGT61_7* while the small PCR product did not match. Using the nucleotide blast (blastn) tool in NCBI, the sequence of the smaller product was aligned against the NCBI database and showed 100% identity with *A. thaliana* chromosome 3 sequence accession CP002686.1. Using a new primer set (GT61_P2, Table S1), two more transgenic lines carrying *PoGT61_7* (P6 and P9) were detected, making a total of 6 independent transformants.

Real-time quantitative PCR (qPCR) analysis was performed for three wild-type plants and six putative transgenic plants using primer set GT61_P2 (Table S1) on cDNA synthesised from RNA extracted from developing siliques. Siliques were harvested from ten-week old plants, as shown in Figure 4B, which were predicted to contain the only tissue, developing seeds, where the seed-coat promoter should be active. Transcript levels of *PoGT61_7* in these 9 plants were compared (Figure 4D). There were background levels of transcript in the wild type samples, potentially from the CP002686.1 gene but three putative transgenic plants, P3, P18 and P23 showed reasonable amounts of *PoGT61_7* transcript.

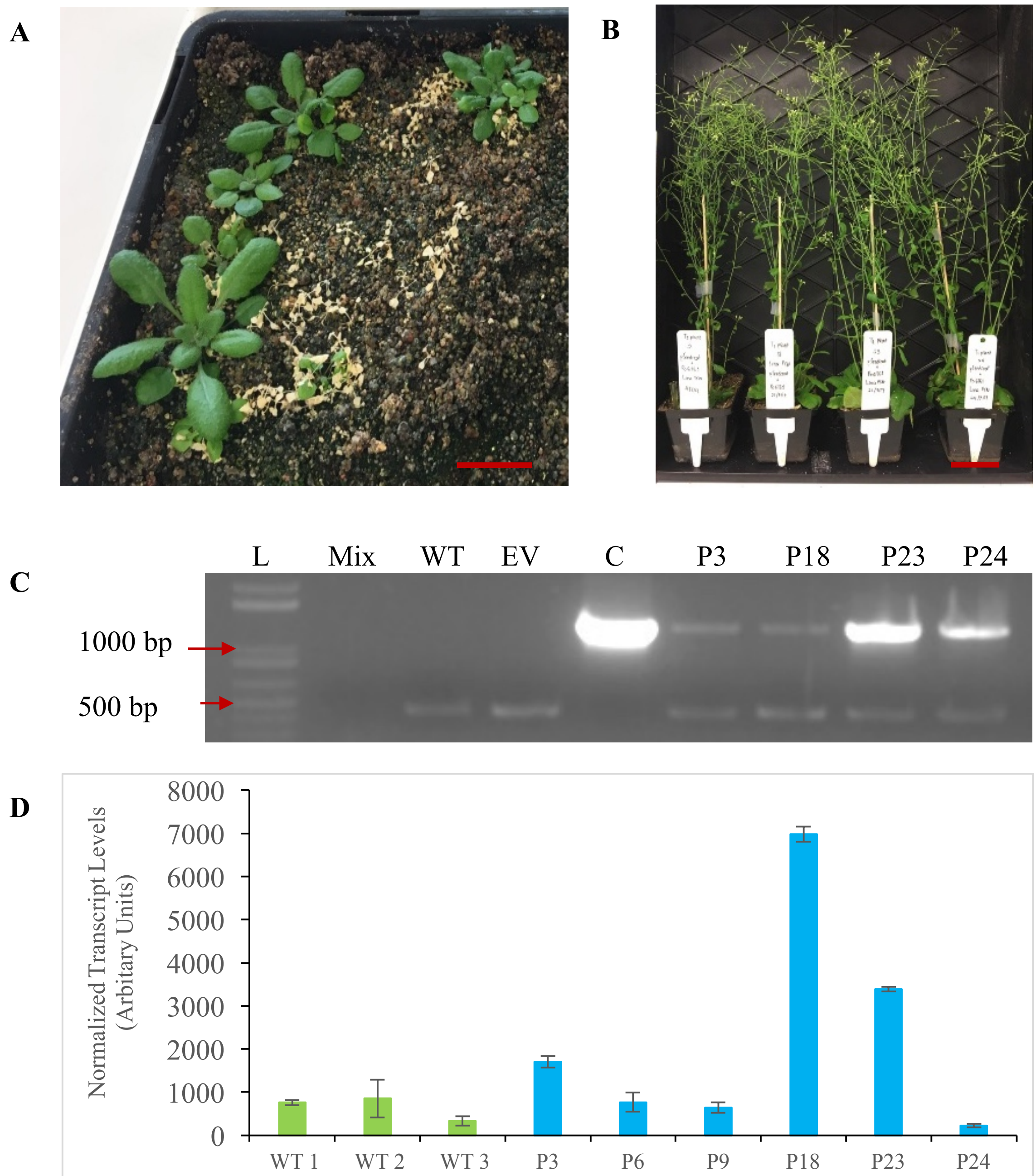


Figure 4. Overexpression of *PoGT61_7* in *A. thaliana*.

(A) Putative T_1 transgenic plants after herbicide selection; (B) Approximately ten-week old putative T_1 plants; (C) PCR fragments amplified from the transgene were detected in the genomic DNA of putative transgenic lines, plus a background product in all lines; (D) Transcript level of *GT61_7* in developing whole siliques of wild-type and putative transgenic plants; L= 1kb ladder; Mix= PCR mixture without DNA; WT= wild-type plant; EV= plant transformed with the empty vector; C= positive control from plasmid DNA; P3, P6, P9, P18, P23 and-P24 transgenic lines; scale bar (A) = 500 mm; scale bar (B) = 5 cm.

Mucilage staining with Ruthenium red (RR)

Mature T_2 seeds from transgenic plants P3, P18 and P23 plus from wild type plants grown at the same time were stained in a solution of ruthenium red (RR) which detects acidic polysaccharides. Figure 5 shows the staining patterns of seeds that were not agitated immediately after adding the RR (Figure 5A, D, G and J), 30 minutes after adding the stain when the inner layer starts to turn pink (Figure 5B, E, H and K) and after 1 hour (Figure 5C, F, I and L). There appears to be a subtle phenotype in the transgenic lines where the inner layer is thicker and more intensely stained on the transgenic seed after an hour (Figure F, I and L) when compared to the wild type (Figure 5C). When the seeds are shaken at 400 rpm for three hours, the inner layer remains attached to the seed coat while the outer layer cannot be observed (Figure 6). In general, the adherent mucilage layers of wild-type without and with agitation are less thick compared to the layers of transgenic seeds, with significant difference between wild-type and P23 without agitation ($P < 0.05$) (Figure 7A). After agitation, the inner layer thickness of wildtype is significant lower than transgenic lines (P3 and P18) (Figure 7B).

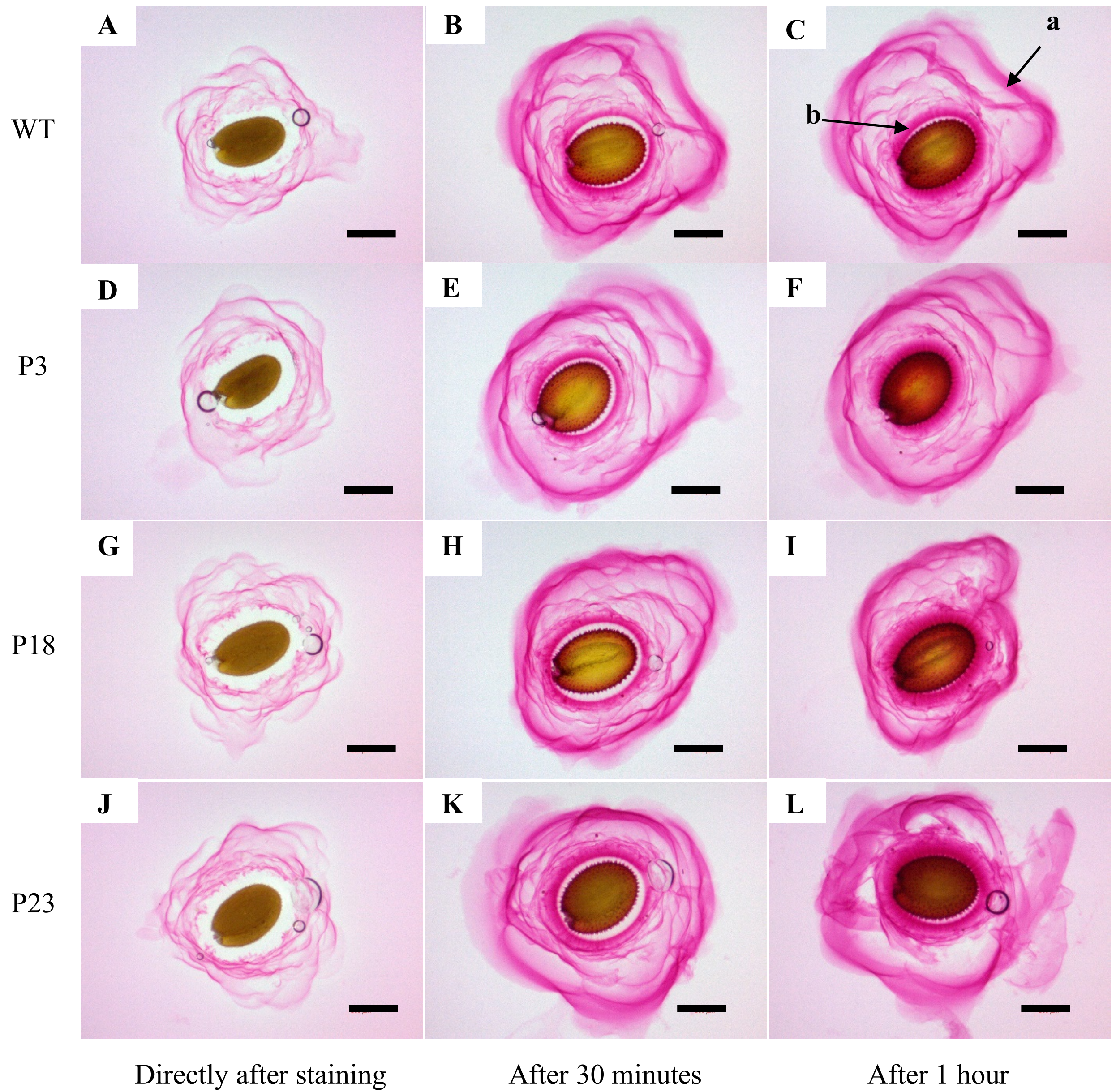


Figure 5. Wild-type and transgenic seed stained using ruthenium red (RR) 0.01% (w/v) without agitation.

(A-C) Wild-type seed; (D-F) Transgenic line no. 3; (G-I) Transgenic line no. 18; (J-K) Transgenic line no. 23; a = non-adherent layer; b = adherent layer; Scale bars = 400 µm; n = 16 samples

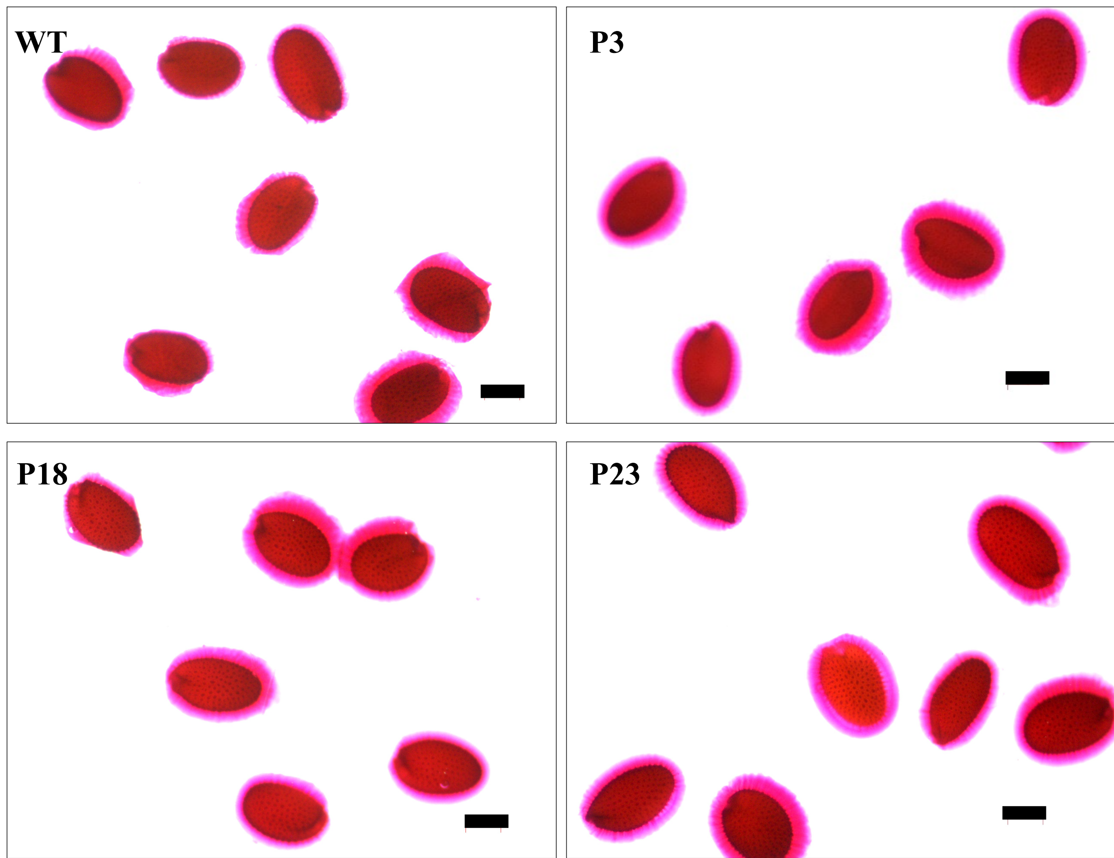


Figure 6. Wild-type and transgenic seed stained using ruthenium red 0.01% (w/v) after agitation (400 rpm) for three hours.

WT = Wild-type plant; P3 = Transgenic plant no. 3; P18 = Transgenic plant no. 18; P24 = Transgenic plant no. 23; Scale bars = 500 μm

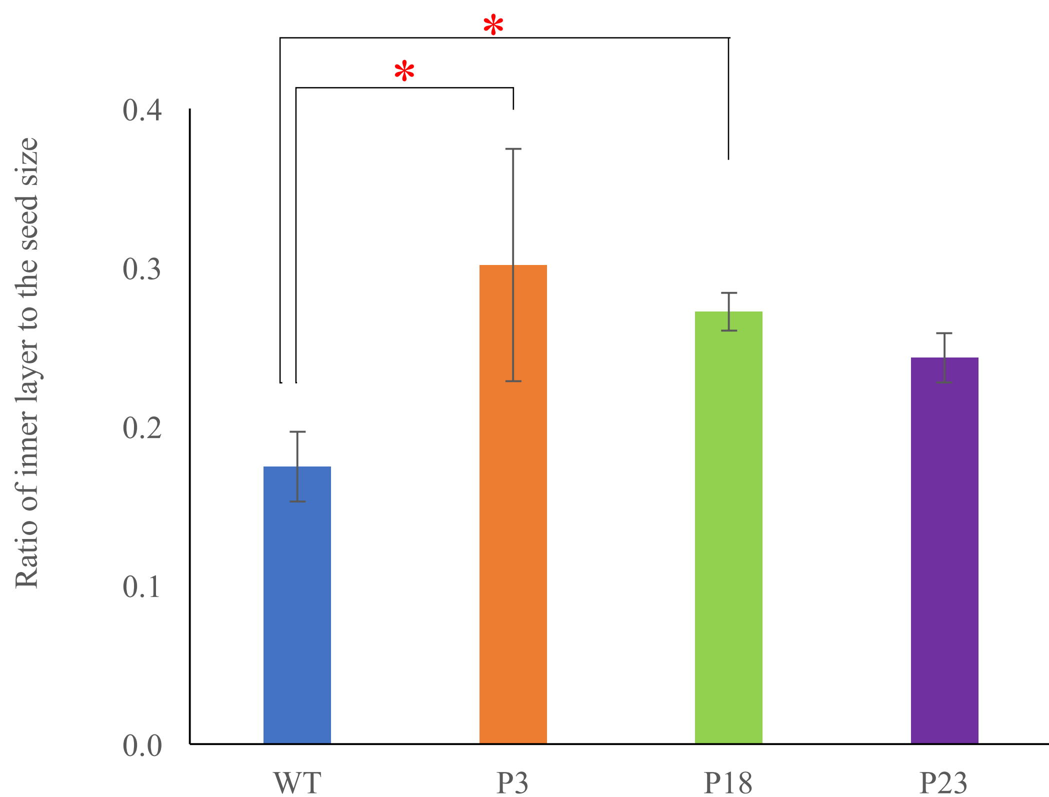


Figure 7. A preliminary comparison of the thickness of the inner mucilage layer in wild-type versus T_2 seed from three transgenic plants (P1, P18, and P23).

Ratio of inner layer width to the seed size with agitation (400 rpm) for 3 hours. * = Significant differences at $P < 0.05$, Tukey's HSD; Error bars show standard deviation from at least 3 biological replicates.

Design of single guide RNAs for CRISPR/Cas9 editing of *DUF579*

In order to ensure disruption of a target, single guide RNAs (sgRNAs) are usually designed to a 5' region of the gene. Figure 8 shows 50 target sequences and three sgRNAs designed using 6 web-tools, which mapped to the first exon of the candidate DUF579 gene (AT1G71690). The target sequences were 23 nucleotides (nts) long and included a 20 nt protospacer for the sgRNA and a three nt (NGG) protospacer for the adjacent motif (PAM) which provides the target for the Cas9 enzyme. The final sgRNAs which are synthesized do not contain the PAM (Liu et al., 2015). The 50 spacers are derived from use of the following websites; CCTop (3), CRISPR-P (3), Cistrome (8), ATUM (10), MIT, (12) and CRISPRSCAN (14). These spacers gave a higher than 0.5 predictive scoring algorithm value so they were included in this selection. Only MIT and CCTop provided information about off targets, and an example can be seen in Figure S5. The spacers were ranked based on the score value from the same websites, and the off-target information from MIT and CCTop. Possible gRNAs were then selected based on the frequency of overlap between the 50 spacers, where at least 3 spacers overlapped each other. Three sgRNAs were selected; sgRNA1 was from CCTop (1), ATUM (6), and CRISPRSCAN (11), and was located 221 – 240 bp from the start codon; sgRNA2 was from CCTop (3), CRISPR-P (2), Cistrome (6), ATUM (1), and MIT (1), located at164 – 183. The third, sgRNA3, was from Cistrome 8, ATUM 2, and MIT 3, located at 161 – 180.

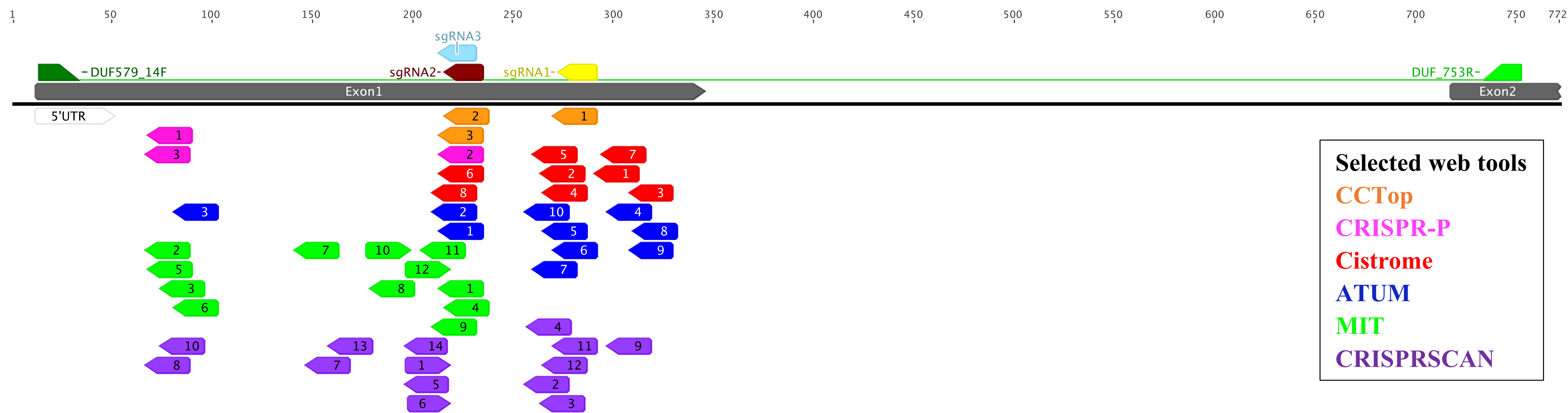


Figure 8. Single guide RNA targets were generated using six online tools as indicated by the coloured arrows with a rank order from 1 to 12.

The three selected sgRNAs target the first exon of AT1G71690.

DUF579_14F and DUF579_753R are the forward and reverse primers that were used in the detection of genomic edits using Tracking of Indels by Decomposition (TIDE).

In silico predictions of genome editing

The products and potential effects of genome editing by particular sgRNAs can be tested *in silico*. This is demonstrated in Figure 9 for the three sgRNAs designed for the DUF579 gene. The AT1G71690 CDS is 888 bp long to produce a protein of 296 amino acids. The double stranded DNA break is assumed to occur 3-4 bp after the PAM (Brinkman et al., 2014), potentially throwing the protein out of frame after this point. For sgRNA 1 (Figure 9A and B), a single nucleotide deletion is predicted to result in amino acid changes starting at position 75 causing early termination at a stop codon at position 100, while a two bp deletion would change amino acid 76 and the protein would stop at 116. A nucleotide deletion caused by sgRNA1 could replace alanine to valine at 75 and deletion of leucine at position 76. For sgRNA 2, one nucleotide deletion may cause amino acid sequence changes starting at 56 and terminating at 100. Replacement of glycine to glutamic acid at 56 and stop codon at 57 could happen if two nts are deleted. For sgRNA 3, one bp deletion would lead to amino acid changes after position 55 and termination at 100. Early termination could happen if there is a two-base pair deletion. Overall, all three guide RNAs could induce changes in protein sequences.

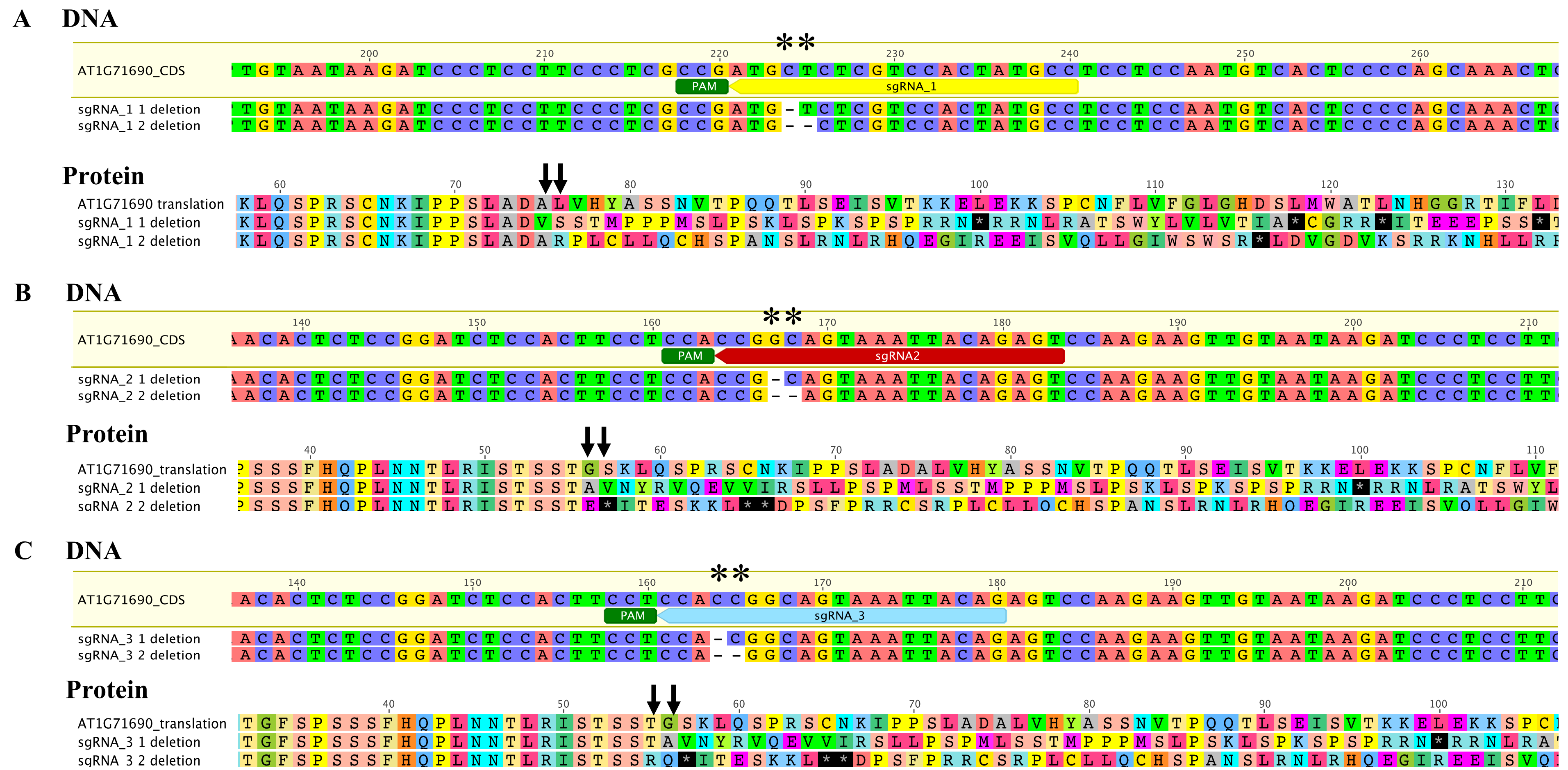


Figure 9. *In silico* analysis of potential genomic editing scenarios by the three selected sgRNAs

Edit location and amino changes generated by (A) sgRNA1; (B) sgRNA2; (C) sgRNA3 are shown; Only partial sequences of the DNA (888 bp) and protein (296 bp) are presented, close to the edits which are marked by asterisks (DNA) and arrows (protein) at the first amino acid modification.

In vitro testing of genome editing

Healthy 4-week-old *Arabidopsis* leaves were used for protoplast isolation and transfection (Figure 10A) experiments initially with a fluorescent marker and subsequently with the sgRNAs. Several experiments were carried out to establish a robust protocol for protoplast isolation and transfection at a high enough efficiency to be likely to yield detectable results. Intact protoplasts of varying sizes were isolated using a cutting technique (Figure 10B and C) and numbers were estimated using a hemocytometer, to be around 3.5×10^5 protoplasts per 40 leaves. For each transfection, 35000 protoplasts were used. The treatments were protoplasts transfected with PEG4000 only or PEG-4000 plus the empty construct (pDe-Cas9) as the negative controls, PEG-4000 with both pDe-Cas9 and sgRNA_1, pDe-Cas9 and sgRNA_2 or pDe-Cas9 and sgRNA_3 (Figure S3), and a YFP construct as a positive control to monitor successful transfection rates. YFP was found to be more suitable for detecting protoplast transfection in this experiment (Figure 10I) since GFP-treated cells were not easily distinguished from untreated protoplasts (data not shown). Protoplasts in both the bright field and DAPI channels were counted for calculating the efficiency of transfection. In the DAPI channel, protoplasts overlapped each other making it challenging to calculate how many cells were present (Figure 10E and H). As a result, the comparison was made between the numbers of cells in the turbo-YFP channel to the total cell number in the bright field channel. As an example, around 16 cells are visible in the turbo-YFP channel (Figure 10I) while about 60 cells are shown in Figure 10G, so transfection efficiency is estimated at about 26%, which was the average efficiency over multiple experiments.

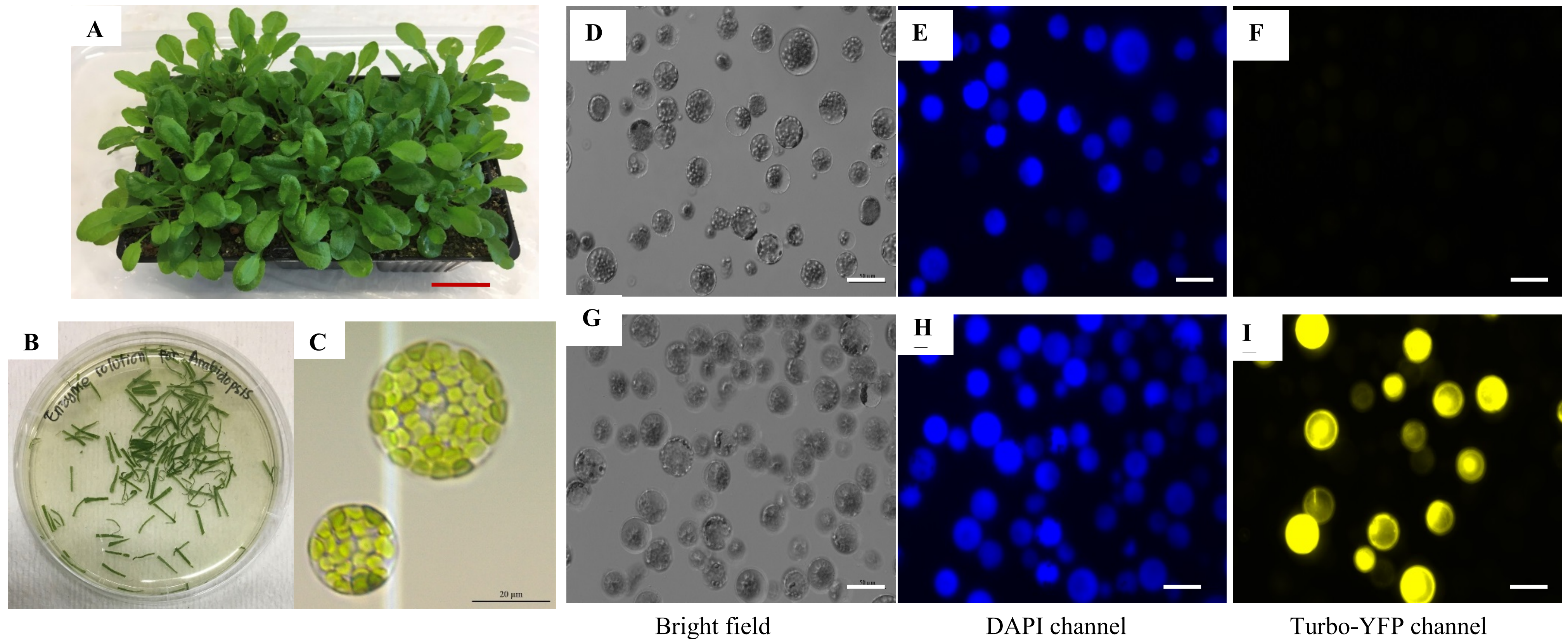


Figure 10. Protoplast isolation and transfection.

(A) Four-week old *A. thaliana* Col-0; (B) Leaves digested for releasing protoplast into solution; (C) Protoplast (400X) under microscope; (D-F) Protoplasts treated only with PEG-4000; (G-I) Protoplasts treated with PEG-4000 and plasmid DNA (YFP construct); Scale bar (A) = 2 cm; Scale bar (D-I) = 50 μm

Evaluating efficiency of guide RNAs (sgRNAs) for targeted genome modification

Three CRISPRS/Cas9 constructs did not generate edits in transfected protoplasts, as shown in Figure 11. A method called Tracking of Indels by Decomposition (TIDE) based on analysis of sequencing chromatograms assumes that a double stranded DNA break will be induced between nucleotides 17 and 18 of the sgRNA, 3-4 bp upstream of PAM (Brinkman et al., 2014). Two panels are produced, one panel shows the expected site of the Cas9 cut in the targeted genomic region (Figure 11A, B, and C) and the other panel demonstrates the efficiency of genome editing and type of insertion or deletion (indels) in the pooled DNA samples (Figure 11D, E, and F). The predicted cuts for the three guide RNAs were at 236 bp, 179 bp, and 176 bp from the start codon, respectively. There is no aberrant nucleotide detected directly after the expected cut. The aberrant nucleotides were detected at 300 bp in the repetitive region (Figure 11A, B and C) and in Figure 11A, the abnormal green signal is caused by sequencing error. The efficiency of three guide RNAs are 0% with R^2 = 0.96, 0.97, and 0.98, respectively (Figure 11D, E, and F).

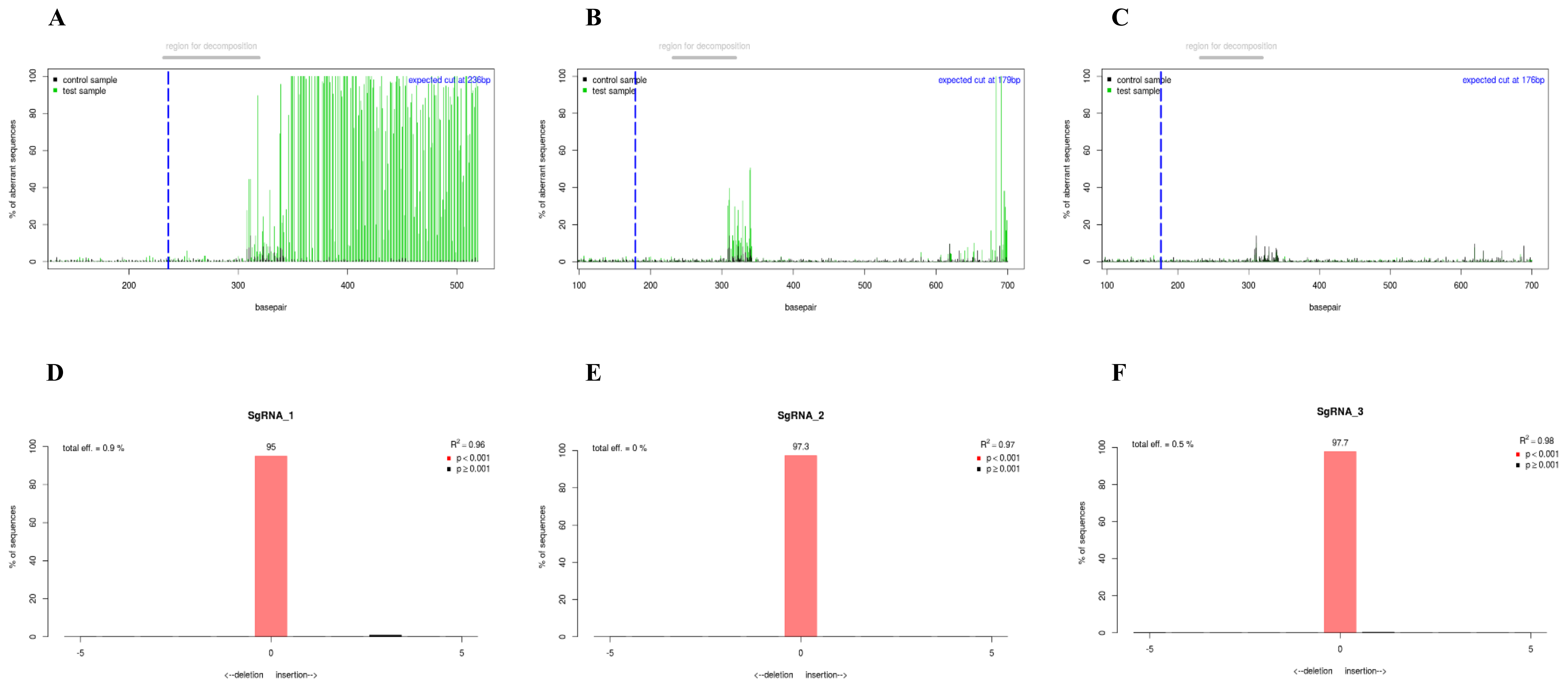


Figure 11. A pool of *Arabidopsis* protoplast cells treated with Cas9 alone (control) and cells treated with Cas9 and targeting sgRNAs were analysed by TIDE. (A, B, C) Aberrant nucleotide signal of the sample (green) compared to that of the control (black); (D, E, F) Indel spectrum determined by TIDE.

DISCUSSION

P. ovata GT61_7 is not directly homologous to gene in *A. thaliana*

Out of 12 *P. ovata GT61s* included in the analysis (Figure 1), only four genes show close homology to *A. thaliana* genes, which are *PoGT61_2*, *7*, and *17* (clade A) and *PoXYLT* (clade C). *Plantago* GT61 numbers 2 and 7 are also shown to be homologs of AT3G18170 and AT3G18180 in phylogenetic trees from Jensen et al. (2013) and Phan et al. (2016). *Plantago GT61_17* was not included in Jensen et al. (2013) since they only analysed *PoGT61*_1 to 7. *PoGT61_17* was not included in the GT61 phylogenetic tree by Phan et al. (2016), however, they mentioned that *PoGT61_7* and *PoGT61_17* sequences are highly similar.

Although AT3G18180 and AT3G18170 have not been annotated, the indirect function of these two *Arabidopsis GT61* genes have been speculated on (Anders et al., 2012; Voiniciuc et al., 2015). They are members of clade A (Figure 1) (Anders et al., 2012; Voiniciuc et al., 2015) and several members of this clade, *OsXAX1*, *OsXAT2*, *OsXAT3*, *TaXATI*, and *TaXAT2* have been annotated (Figure 1). XAX is a xylan β-1,2-xylosyltransferase and XAT is an α-1,3-arabinosyltransferase. By knocking out a *GT61* gene (*TaXAT1*) in wheat endosperm, Anders et al. (2012) found that this gene is responsible for monosubstitution of arabinoxylan. They also found that heterologous expression of *TaXAT2*, *OsXAT2*, and *OsXAT3* in *A. thaliana* leads to arabinosylation of *Arabidopsis* xylan. Hence, they suggested that GT61 family members of clade A are responsible for arabinosylation.

However, *OsXAX1* is also grouped in clade A on the GT61 phylogenetic tree (Figure 1) more recently annotated by (Voiniciuc et al., 2015). By knocking out *OsXAX1* (OS02G22380), the protein was found to be a xylosyltransferase (Chiniquy et al., 2012). Chiniquy et al. (2012) divided clade A, which was clade C in their paper, into subgroups I, II, III, and IV. Clade A.IV, which contains *XAX1*, has no published dicot ortholog or grass equivalent. AT3G18180 and AT3G18170 are grouped in clade A.II with *TaXAT2*. Chiniquy et al. (2012) proposed that members of clade A contain *GT61s* for both arabinose and xylose substitution. Protein activity

assay of *PoGT61_7* (clade A, Figure 1) by the Plant Cell Wall group, University of Adelaide revealed that this protein act as xylosyltransferase (unpublished data). PoGT61_7 was the only one out of 10 GT61 proteins to be active, but Jensen et al. (2011) suggested that xylan is likely to be synthesised by a protein complex with many components. This could explain why no activity was found for the other nine *PoGT61* genes. Since there are two predictive functions for clade A members and only *PoGT61_7* showed protein activity, this gene was chosen for heterologous expression in *Arabidopsis* plants.

***PoGT61_7* expression controlled by *Arabidopsis* seed coat promoter**

The *PoGT61_7* transgene was successfully integrated into the *A. thaliana* genome as indicated by modest transcript levels in qPCR experiments. Lower transcripts were not surprising since the material for this analysis was taken from mixed developing siliques and the seed specific promoter is only expressed 7 days after fertilisation in the epidermal layer of the seed coat. Esfandiari et al. (2013) reported that expression level of GUS driven by the seed-coat promoter (Pro_{DP1}) was up to six times lower than levels controlled by the constitutive promoter (Pro_{35S}). Specific expression is needed to prevent deleterious effects (Esfandiari et al., 2013) since expressing the transgene in all parts of the plant could lead to pleiotropic phenotypes (Jensen et al., 2011). As a result, further examination of the right time and tissue is needed to measure transcript levels accurately.

Overexpression of *PoGT61_7* caused a seed mucilage phenotype

Arabidopsis seeds produce mucilage upon wetting and form an inner and outer layer (Figure 5) (Western et al., 2001; Naran et al., 2008; Arsovski et al., 2010). Comparison of the thickness of the inner layer between wild-type and transgenic plants with or without shaking (Figure 7) shows that the transgenic lines have a thicker adherent layer than the wild-type which may be caused by expression of the *PoGT61_7* transgene. The outer mucilage disperses into the RR solution forming a cloudy layer while the inner layer is like a capsule surrounding the seed (Figure 5C). The outer layer is water soluble and easily extracted, so it cannot be observed

after agitation (Figure 6). The inner layer is in-soluble in water (Naran et al., 2008) and still adheres to the seed after shaking. Macquet et al. (2007) found that even though the main component of these two layers are the same, (RGI), the structures are different, being unbranched in the outer layer. In contrast, RGI molecules in the inner layer are branched and attached to the seed coat by covalent bonds making it difficult to extract (Macquet et al., 2007). Voiniciuc et al. (2015) demonstrated that *MUCI21*, which is an *Arabidopsis GT61* in clade B (Figure 1), may be necessary for pectin attachment to the seed coat by decorating xylan backbone with xylose. Knocking out this gene caused a seed mucilage defect. If *PoGT61_7* (clade A) has a similar function like *MUCI21* facilitating attachment of RGI to the seed coat, it is possible that heterologous expression of *P. ovata GT61_7* can affect the *Arabidopsis* mucilage inner layer.

Although RR is an easy and quick method to differentiate between wild-type and mutant seed mucilage phenotypes, this method is not comprehensive enough for accurate analyses. To better examine changes in xylan structure and composition, immunolabeling and monosaccharide analysis are needed. This requires many seeds but T_2 stocks from T_1 plants in this experiment were limited and needed for obtaining T_3 generation. Therefore, immunolabeling and monosaccharide analysis could be carried out on the T_3 seeds.

AT1G71690 is the only seed specific *DUF579*

AT1G71690 is a member of clade B together with *GXM1*, *GXM2*, and *GXM3* (Figure 2). *GXMs* are glucuronoxylan methyltransferases since gxm mutants show a significant decrease in the relative amount of methylglucuronic acid (MeGlcA) compared to glucuronic acid (GlcA) and MeGlcA (Lee et al., 2012). But unlike other clade B members, AT1G71690 is only expressed in generative phase (Figure 3) hence Lee et al. (2012) did not include it in their functional analysis even though it is a *GXM* homolog. They focused on genes that are expressed in *Arabidopsis* stems (Lee et al., 2012). This study focuses on xylan biosynthesis in seed mucilage; the target gene should be expressed during mucilage production. In *Arabidopsis*

mucilage is produced during seed development and deposited in the outer layer of the seed coat (Windsor et al., 2000) peaking at approximately five days post anthesis (Esfandiari et al., 2013). Figure 3 shows the only gene highly expressed during this period is AT1G71690. Therefore, knocking out this gene may affect xylan content and structure in seed mucilage.

The clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR associated 9 (Cas 9) system is powerful new technology that could knockout AT1G71690 (Liu et al., 2015; Stemmer et al., 2015). Two important components are the single guide RNA (sgRNA) for target recognition and the Cas9 endonuclease that can introduce a double-stranded DNA cleavage at the specific site determined by the sgRNA (Liu et al., 2015; Stemmer et al., 2015). Since the sgRNA is only 20 nts long, off target mismatches can occur so correct design is important to avoid lengthy and costly incorrect transformation (Liu et al., 2015; Stemmer et al., 2015). Therefore, designing sgRNAs using several online tools (Figure 8), *in silico* analysis (Figure 9) and *in vitro* testing in protoplasts (Figure 10) were performed to test the CRISPR/Cas9 system.

Low-efficiency protoplast transfection could be the reason for unsuccessful genome editing

Although delivery of DNA was successful into the protoplast cells (Figure 10), genome editing was not detected (Figure 11), possibly due to the design of the sgRNAs or the protoplast transfection rate. The best sgRNA candidates were selected from those designed using multiple online sources for which the probability to create edits is relatively high (Figure 8). Simulations run using Geneious 8.1 software show all three sgRNAs could create protein frameshifts and terminations (Figure 9). However, the amount of isolated protoplasts was relatively low (3.5 x 10^5 from 40 leaves) compared to the amount recommended in the TEAMP protocol. Yoo et al. (2007) pointed out that low yields of protoplasts may be due to accession, poor plant growth and leaf condition, and inefficient enzyme digestion.

Low transfection efficiency (26%) could be another reason for unsuccessful genome editing. Several possible issues include the PEG solution, low quality DNA, the ratio between protoplasts and DNA, and poor protoplast quality (Yoo et al., 2007). Of these, the ratio could certainly be a problem since balancing the number of cells and the DNA concentration is critical but is technically challenging. The protoplast quality could contribute to this low efficiency because the broken cells cannot be transfected. The transfection efficiency reported by Yoo et al. (2007) was about 60-90%, with genome editing analysis considered to be reliable if transfection efficiency was higher than 50%. Therefore, optimization of the protoplast isolation and transfection protocols is likely to improve the chance of successfully being able to test CRISPR/Cas9 constructs prior to embarking on plant transformation experiments. For the future experiments, isolating protoplasts using the Tape-Arabidopsis Sandwich method could be an option (Wu et al., 2009). Wu et al. (2009) reported that using two kinds of tape, Time tape (Time Med, Burr Ridge, IL) attached to the upper leaf epidermis and 3M Magic tape to the lower epidermis are more efficient and less tissue damaging in isolating protoplast compared to the cutting method of Yoo et al. (2007).

CONCLUSION

Two methods have been described for testing two candidate genes from different gene families for their involvement in heteroxylan substitution. Heterologous expression of *GT61_7* in transgenic *Arabidopsis* lines produced a phenotype in the inner adherent seed mucilage layer whilst design and protoplast infiltration of three sgRNAs against a selected *DUF579* gene was explored for genome editing potential.

MATERIAL AND METHODS

Plant material and growth conditions

Arabidopsis thaliana Columbia-0 (Col-0) ecotype plants were grown for floral dip transformation and protoplast transfection following Bent (2006). Seeds were stratified for three days at 4°C on prepared soil mix and transferred to a growth room in The Plant Accelerator, University of Adelaide, at 22°C under long days (16h light).

Phylogenetic analyses of *GT61* and *DUF579* genes

A bioinformatics approach was used to identify homologs of *P. ovata GT61* and *DUF579* genes in *Arabidopsis*. Using the PFAM PF04577 for *GT61* (Anders et al. 2012) and PF04669 for *DUF579* (Lee et al., 2012) as filters, sequences in FASTA format from several species were downloaded from EnsemblPlants (http://plants.ensembl.org/biomart/martview). Sequences were also downloaded from NCBI databases. *Arabidopsis* genes were previously identified by Lee et al. (2012) and 7 *P. ovata* contigs were selected from the ARC Centre of Excellence in Plant Cell Walls unpublished data.

All cDNA sequences were imported in FASTA format from the online databases. Sequences were curated to remove duplicates, short sequences, and those not containing an open reading frame (ORF). All sequences were aligned using the translation alignment option (MUSCLE alignment) in the Geneious 8.0 software and alignment editing was conducted to remove sequences that did not contain a homologous site using BMGE (Block Mapping and Gathering with Entropy) online software (https://galaxy.pasteur.fr/forms::BMGE). Finally, two phylogenetic trees were built using the FastTree plugin, then visualised using the FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

RNA sequence data for *Arabidopsis* DUF579 genes were obtained from public data (http://trace.ddbj.nig.ac.jp/DRASearch/), project number SRP075604 and submission number SRA428850. The raw data were processed in CLC Genomics Workbench 9.5.2.

Constructing an expression vector to overexpress *Plantago GT61_7* in *A. thaliana*

The construction of the expression vector followed the protocol of Curtis and Grossniklaus (2003). The *PoGT61_7* cDNA was isolated by Dr. Wei Zeng (University of Melbourne), ligated into entry vector pCR®8 and transformed into *Escherichia coli* One Shot ® TOP10 competent cells using a chemical transformation procedure (heat shock) following Invitrogen's instructions. Two restriction enzymes were used for verifying the insert, Pvull-HF for checking the presence of insert, and HincII for confirming the correct orientation. The insert (*PoGT61_7*) was transferred to the destination vector by LR reaction (Gateway system) containing the *Arabidopsis* seed coat promotor (AT4G11180) and BAR resistance gene and kindly supplied by Kum Foeng Ang (University of Adelaide). The insert was verified by sequencing at the Australian Genome Research Facility Ltd (AGRF), Adelaide.

***Agrobacterium* and Plant transformation**

The expression vector containing *PoGT61_7* driven by the seed-coat promoter (Figure S3) was transformed into *Agrobacterium tumefaciens* via a freeze/thaw method as per Wise et al. (2006). Transformed *Agrobacterium* colonies were verified by PCR and cultures were used on flowering *Arabidopsis* plants using the floral dip method followed by a floral spray one week later (Bent, 2006).

BASTA selection and transgene detection

Seeds from T_0 plants were harvested approximately 4 weeks after the last transformation step, sown directly onto soil and sprayed with a 60 mg/L herbicide solution (Basta, or glufosinate-ammonium). Genomic DNA from plants surviving Basta selection (T_1 plants) was extracted using Edward buffer (Edwards et al., 1991) and genotyped by PCR using the GT61 primers listed in Table S1 to detect the *GT61* transgene. Genomic PCR fragments were sequenced at AGRF.

RNA extraction, cDNA synthesis, and real time qPCR analysis

RNA extraction, cDNA synthesis and real time qPCR analysis were carried out as per Burton et al. (2008). Total RNA was extracted from frozen T_2 developing seed in intact siliques. Post-prep DNAse (Ambion Turbo DNA-free) was applied to extracted RNA which was run on a 1% agarose gel at 100 V for 20 minutes to evaluate RNA quality. Between 2-4 µl of RNA was used for synthesizing first strand cDNA using oligo $(dT)_{18}$ and Superscript III Reverse Transcriptase kit (Invitrogen, USA) according to the manufacture's protocol. qPCR analysis was performed using three biological replicates of Columbia WT and 6 putative transgenic plants. The transgene expression levels were normalised using *Arabidopsis* Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) (Table S1).

Ruthenium red whole seed staining

Whole *Arabidopsis* T_2 seeds were stained using 0.01% (w/v) ruthenium red (RR) (ProSciTech (PST) Pty Ltd, C075) staining with shaking following McFarlane et al. (2014) and without shaking following Voiniciuc et al. (2015) with modification excluding pre-hydration in water for 5 min. For observing adherent and non-adherent layers, three biological replicates for every plant type were used. Individual seed was observed under 50X magnification at three different times, directly after adding RR solution, 30 minutes and one hour later. For observing the adherent inner layer only, 20 seeds for each sample were treated with 50 mM EDTA for 2 hours with shaking (400 rpm) followed by 1 hour in RR solution with shaking, then the seeds were observed under 32X magnification. All images were captured on a Zeiss Stemi 2000 microscope with an attached AxioCam ERc 5s camera. The thickness of mucilage inner layers and the seeds was measured using Zeiss® Zen. Significant differences (Tukey's HSD) were analysed using SPSS 24.

Selection of sgRNA targets on an *Arabidopsis DUF579* gene (AT1G71690)

Single guide RNAs were designed using six online tools.

1. CCTop http://crispr.cos.uni-heidelberg.de (Stemmer et al., 2015);

2. CRISPR-P http://crispr.hzau.edu.cn/cgi-bin/CRISPR2/CRISPR (Lei et al., 2014);
3. Cistrome http://cistrome.org/SSC/ (Xu et al., 2017);
4. ATUM https://www.atum.bio/products/crispr;
5. MIT http://crispr.mit.edu:8079/;
6. CRISPRSCAN http://www.crisprscan.org (Moreno-Mateos et al., 2015)

The first exon sequence of AT1G71690 was used as the target where 20 nucleotide (nts) upstream of the protospacer adjacent motif (PAM) (Liu et al., 2015) were selected on each website. Only sgRNA that had high score evaluations and low or no off-targets were selected. Sequences of selected guide RNAs were mapped onto the targeted genomic region. Three single guide RNAs were chosen based on the frequency of consensus candidate sgRNAs from all six websites.

Construction of AT1G71690 sgRNA and Cas9 expression vectors

CRISPR/Cas9 expression vectors were generated as per Schiml et al. (2016). Double stranded DNA fragments were generated by annealing the forward and reverse protospacers (sgRNA1, sgRNA2, and sgRNA3 separately). Details of the three sgRNAs are provided in Table S1. The PAM sequences (NGG) were not included in the oligo. The dsDNA and the entry vector with flanking attR sites for Gateway cloning (pEn-Chimera) were digested by BbsI in separate reactions then ligated using T4 ligase. The sgRNAs from pEn-Chimera were transferred into the destination vector (pDe-Cas9) using LR Clonase II according to the manufacturer's instructions. All vectors used were a gift from Johannes Daniel Scharwies (University of Adelaide). The final construct contained an sgRNA driven by the *Arabidopsis* U6-26 promoter and Cas9 driven by the constitutive Ubi4-2 promotor from parsley (Figure S4, http://www.botanik.kit.edu/molbio/990.php). The pDe-Cas9 is registered as number 6531115762 in TAIR database.

Transient expression of the CRISPR/Cas9 construct in *Arabidopsis* protoplasts

Protoplast isolation and DNA transfection were carried out according to Yoo et al. (2007) using a protocol called transient expression in *Arabidopsis* mesophyll protoplast (TEAMP). Leaves from four-week-old plants before flowering were harvested, diced with a razor blade and digested in enzyme solution containing Cellulose R10 and Macerozyme R10 (Yakult Pharmaceutical Ind. Co. Ltd, Japan) and protoplasts released counted using a hemocytometer under the light microscope. Highly concentrated DNA (2 μg/μl) was used to transfect the protoplasts. The YFP:35S construct (MT466) from Dr. Matthew Tucker (University of Adelaide) was used to evaluate the success rate of protoplast transfection. After transfection, protoplast cultures in six-well plates were incubated up to 32 hours in the dark at room temperature (23-25°C) (Li et al., 2014).

Mutagenesis detection

Genomic editing generated by CRISPR/Cas9 constructs was evaluated using the Tracking of Indels by Decomposition (TIDE) online software (Brinkman et al., 2014). Genomic DNA from each of the three replicates for control and transfected protoplasts (sgRNA1, sgRNA2, sgRNA3) was extracted using DNeasy Plant Mini Kit (Qiagen, ID: 69104). Then 700 bp of the targeted DUF579 gene was PCR amplified using the DUF579_14F and DUFR579_753R primers (Table S1). After PCR purification samples were sent for sequencing at AGRF, Adelaide. The TIDE online software was used to compare the 20nt guide sequence upstream PAM (5'-3') against the sequencing chromatogram generated for the PCR product from the control and each sgRNA sample.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Dr. Neil Shirley for providing *Arabidopsis* RNA sequence data and performing real time qPCR, Dr. Julian Schwerdt for assistance with phylogenetic analysis, Dr. Wei Zeng for *PoGT61_7* and Kum Foeng (Maple) Ang for the destination vector and assistance with bacterial transformation, Dr. Kelly Houston for sharing the knowledge in designing sgRNA, Johannes Daniel Scharwies for CRISPR/Cas9 vectors and help with vector construction, Taj Arndell for assistance with protoplast experiments, Dr. Matthew Tucker for providing YFP construct, and Shi Fang (Sandy) Khor, Dr. Ali Hasan, and James Cowley for technical advice.

LITERATURE CITED

Anders N, Wilkinson MD, Lovegrove A, Freeman J, Tryfona T, Pellny TK, Weimar T, Mortimer JC, Stott K, Baker JM, Defoin-Platel M, Shewry PR, Dupree P, Mitchell RAC (2012) Glycosyl transferases in family 61 mediate arabinofuranosyl transfer onto xylan in grasses. Proc. Natl. Acad. Sci. U.S.A **109:** 989-993

Arsovski AA, Haughn GW, Western TL (2010) Seed coat mucilage cells of *Arabidopsis thaliana* as a model for plant cell wall research. Plant Signal Behav **5:** 796-801

Bent A (2006) *Arabidopsis thaliana* floral dip transformation method. *In* K Wang, ed, Agrobacterium protocols, Ed 2nd Vol 1. Humana Press, New Jersey, pp 87-104

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

Brown D, Wightman R, Zhang Z, Gomez LD, Atanassov I, Bukowski JP, Tryfona T, McQueen-Mason SJ, Dupree P, Turner S (2011) *Arabidopsis* genes *IRREGULAR XYLEM* (*IRX15*) and *IRX15L* encode DUF579-containing proteins that are essential for normal xylan deposition in the secondary cell wall. Plant J. **66:** 401-413

Burton R, Jobling S, Harvey A, Shirley N, Mather D, Bacic A, Fincher G (2008) The genetics and transcriptional profiles of the cellulose synthase-like *HvCslF* gene family in barley. Plant Physiol **146:** 1821-1833

Chiniquy D, Sharma V, Schultink A, Baidoo EE, Rautengarten C, Cheng K, Carroll A, Ulvskov P, Harholt J, Keasling JD, Pauly M, Scheller HV, Ronald PC (2012) XAX1 from glycosyltransferase family 61 mediates xylosyltransfer to rice xylan. Proc. Natl. Acad. Sci. U.S.A **109:** 17117-17122

Curtis MD, Grossniklaus U (2003) A gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant Physiol **133:** 462-469

Ebringerová A (2005) Structural diversity and application potential of hemicelluloses. Macromol. Symp. **232:** 1-12

Edwards K, Johnstone C, Thompson C (1991) A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Res. **19:** 1349

Esfandiari E, Jin Z, Abdeen A, Griffiths JS, Western TL, Haughn GW (2013) Identification and analysis of an outer-seed-coat-specific promoter from *Arabidopsis thaliana*. Plant Mol. Biol. **81:** 93-104

Fischer MH, Yu N, Gray GR, Ralph J, Anderson L, Marlett JA (2004) The gel-forming polysaccharide of psyllium husk (*Plantago ovata* Forsk). Carbohydr. Res. **339:** 2009-2017

Francoz E, Ranocha P, Burlat V, Dunand C (2015) *Arabidopsis* seed mucilage secretory cells: regulation and dynamics. Trends Plant Sci. **20:** 515-524

Guo Q, Cui SW, Wang Q, Young JC (2008) Fractionation and physicochemical characterization of psyllium gum. Carbohydr Polym **73:** 35-43

Haughn GW, Western TL (2012) *Arabidopsis* seed coat mucilage is a specialized cell wall that can be used as a model for genetic analysis of plant cell wall structure and function. Front Plant Sci. **3:** 64

Jensen JK, Johnson N, Wilkerson CG (2013) Discovery of diversity in xylan biosynthetic genes by transcriptional profiling of a heteroxylan containing mucilaginous tissue. Front Plant Sci. **4:** 183

Jensen JK, Johnson NR, Wilkerson CG (2014) *Arabidopsis thaliana* IRX10 and two related proteins from psyllium and *Physcomitrella patens* are xylan xylosyltransferases. Plant J. **80:** 207-215

Jensen JK, Kim H, Cocuron JC, Orler R, Ralph J, Wilkerson CG (2011) The DUF579 domain containing proteins IRX15 and IRX15-L affect xylan synthesis in *Arabidopsis*. Plant J. **66:** 387-400

Khaliq R, Tita O, Antofie MM, Sava C (2015) Industrial application of psyllium: An overview. ACTA Universitatis Cibiniensis **67:** 210-214

Lee C, Teng Q, Zhong R, Yuan Y, Haghighat M, Ye Z-H (2012) Three *Arabidopsis* DUF579 domain-containing GXM proteins are methyltransferases catalyzing 4-O-methylation of glucuronic acid on xylan. Plant Cell Physiol. **53:** 1934-1949

Lei Y, Lu L, Liu H-Y, Li S, Xing F, Chen L-L (2014) CRISPR-P: A Web Tool for Synthetic Single-Guide RNA Design of CRISPR-System in Plants. Mol Plant **7:** 1494-1496

Li J-F, Zhang D, Sheen J (2014) Cas9-based genome editing in *Arabidopsis* and tobacco. *In* Methods in Enzymology, Vol 546, pp 459-472

Liu W, Zhu X, Lei M, Xia Q, Botella JR, Zhu J-K, Mao Y (2015) A detailed procedure for CRISPR/Cas9-mediated gene editing in *Arabidopsis thaliana*. Sci. Bull. **60:** 1332-1347

Macquet A, Ralet M-C, Kronenberger J, Marion-Poll A, North HM (2007) In situ, chemical and macromolecular study of the composition of *Arabidopsis thaliana* seed coat mucilage. Plant Cell Physiol. **48:** 984-999

McFarlane HE, Gendre D, Western TL (2014) Seed coat ruthenium red staining assay. Bio-protocol **4**

Moreno-Mateos MA, Vejnar CE, Beaudoin J-D, Fernandez JP, Mis EK, Khokha MK, Giraldez AJ (2015) CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. Nat. Methods **12:** 982

Naran R, Chen G, Carpita NC (2008) Novel rhamnogalacturonan I and arabinoxylan polysaccharides of flax seed mucilage. Plant Physiol **148:** 132-141

Naran R, Chen G, Carpita NC (2008) Novel rhamnogalacturonan I and arabinoxylan polysaccharides of flax seed mucilage. Plant Physiology **148:** 132-141

North HM, Berger A, Saez-Aguayo S, Ralet M-C (2014) Understanding polysaccharide production and properties using seed coat mutants: future perspectives for the exploitation of natural variants. Ann. Bot. **114:** 1251-1263

Phan JL, Tucker MR, Khor SF, Shirley N, Lahnstein J, Beahan C, Bacic A, Burton RA (2016) Differences in glycosyltransferase family 61 accompany variation in seed coat mucilage composition in *Plantago* spp. J. Exp. Bot. **67:** 6481-6495

Ralet M-C, Crépeau M-J, Vigouroux J, Tran J, Berger A, Sallé C, Granier F, Botran L, North HM (2016) Xylans provide the structural driving force for mucilage adhesion to the *Arabidopsis* seed coat. Plant Physiol **171:** 165-178

Rennie EA, Scheller HV (2014) Xylan biosynthesis. Curr. Opin. Biotechnol. **26:** 100-107

Schiml S, Fauser F, Puchta H (2016) CRISPR/Cas-Mediated site-specific mutagenesis in *Arabidopsis thaliana* using Cas9 nucleases and paired nickases. *In* M Murata, ed, Chromosome and Genomic Engineering in Plants: Methods and Protocols, Vol 1469. Humana Press, New York, NY, pp 111-122

Stemmer M, Thumberger T, del Sol Keyer M, Wittbrodt J, Mateo JL (2015) CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. PLoS ONE **10:** e0124633

Voiniciuc C, Günl M, Schmidt MH-W, Usadel B (2015) Highly branched xylan made by IRREGULAR XYLEM14 and MUCILAGE-RELATED21 links mucilage to *Arabidopsis* seeds. Plant Physiol **169:** 2481-2495

Western TL (2012) The sticky tale of seed coat mucilages: production, genetics, and role in seed germination and dispersal. Seed Sci. Res. **22:** 1-25

Western TL, Burn J, Tan WL, Skinner DJ, Martin-McCaffrey L, Moffatt BA, Haughn GW (2001) Isolation and characterization of mutants defective in seed coat mucilage secretory cell development in *Arabidopsis*. Plant Physiol **127:** 998-1011

Western TL, Skinner DJ, Haughn GW (2000) Differentiation of mucilage secretory cells of the *Arabidopsis* seed coat. Plant Physiol **122:** 345-356

Windsor JB, Symonds VV, Mendenhall J, Lloyd AM (2000) *Arabidopsis* seed coat development: morphological differentiation of the outer integument. Plant J. **22:** 483-493

Wise AA, Liu Z, Binns AN (2006) Three methods for the introduction of foreign DNA into *Agrobacterium*. *In* K Wang, ed, Agrobacterium protocols, Ed 2nd Vol 1. Humana Press, New Jersey, pp 43-54

Wu F-H, Shen S-C, Lee L-Y, Lee S-H, Chan M-T, Lin C-S (2009) Tape-Arabidopsis Sandwich-a simpler Arabidopsis protoplast isolation method. Plant Methods **5:** 16

Xu H, Xiao T, Chen C-H, Li W, Meyer C, Di Wu LC, Zhang F, Liu JS, Brown M, Liu S (2017) Sequence determinants of improved CRISPR sgRNA design. Submitted. . *In*, Vol 2017

Yoo S-D, Cho Y-H, Sheen J (2007) *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. Nat. Protoc. **2:** 1565-1572

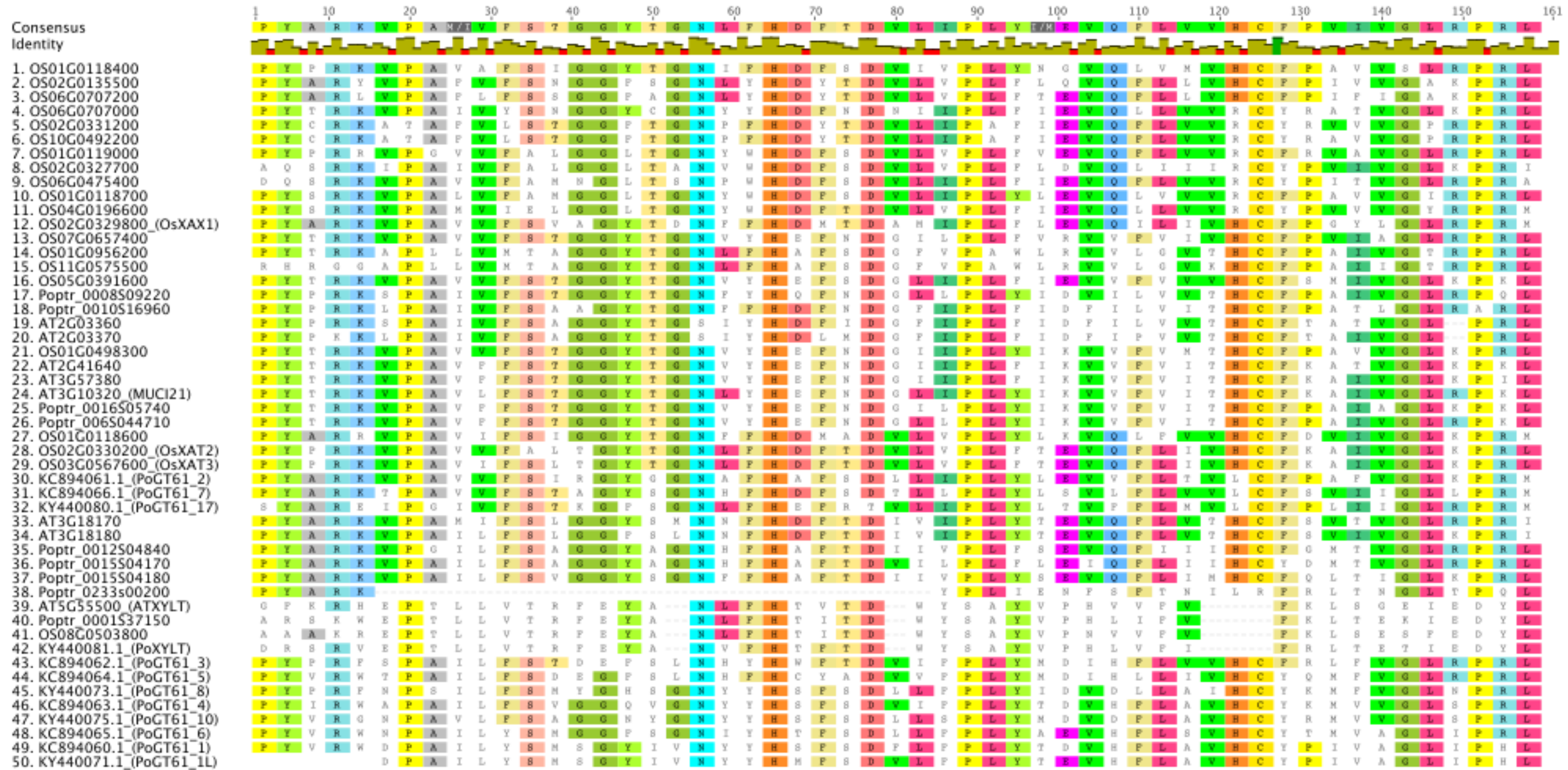


Figure S1. An alignment of GT61 protein from *P. ovata*, *A. thaliana*, *O. sativa,* and *P. trichocarpa*

Consensus
Identity
1. AT4G24910
2. Migut.I00536.1
3. contig_27741_CDS
4. AT2G15440
5. Migut.N01174.1
6. Migut.B00370.1
7. AT5G67210_AtIRX15L
8. AT3G50220_AtIRX15
9. contig_23673_CDS_IRX15
10. contig_33361_CDS
11. Migut.L00176
12. AT1G67330
13. AT1G27930
14. Migut.E00015.1
15. contig_18354_CDS
16. AT1G71690
17. Migut.C01349.1
18. AT4G09990_GXM2
19. AT1G33800_GXM1
20. contig_18060_CDS
21. Migut.C01347.1
22. contig_14698_CDS
23. Migut.N00852.1
24. contig_26654_CDS
25. AT1G09610_GXM3

Figure S2. An alignment of DUF579 protein from *A. thaliana*, *P. ovata*, *M. guttatus*

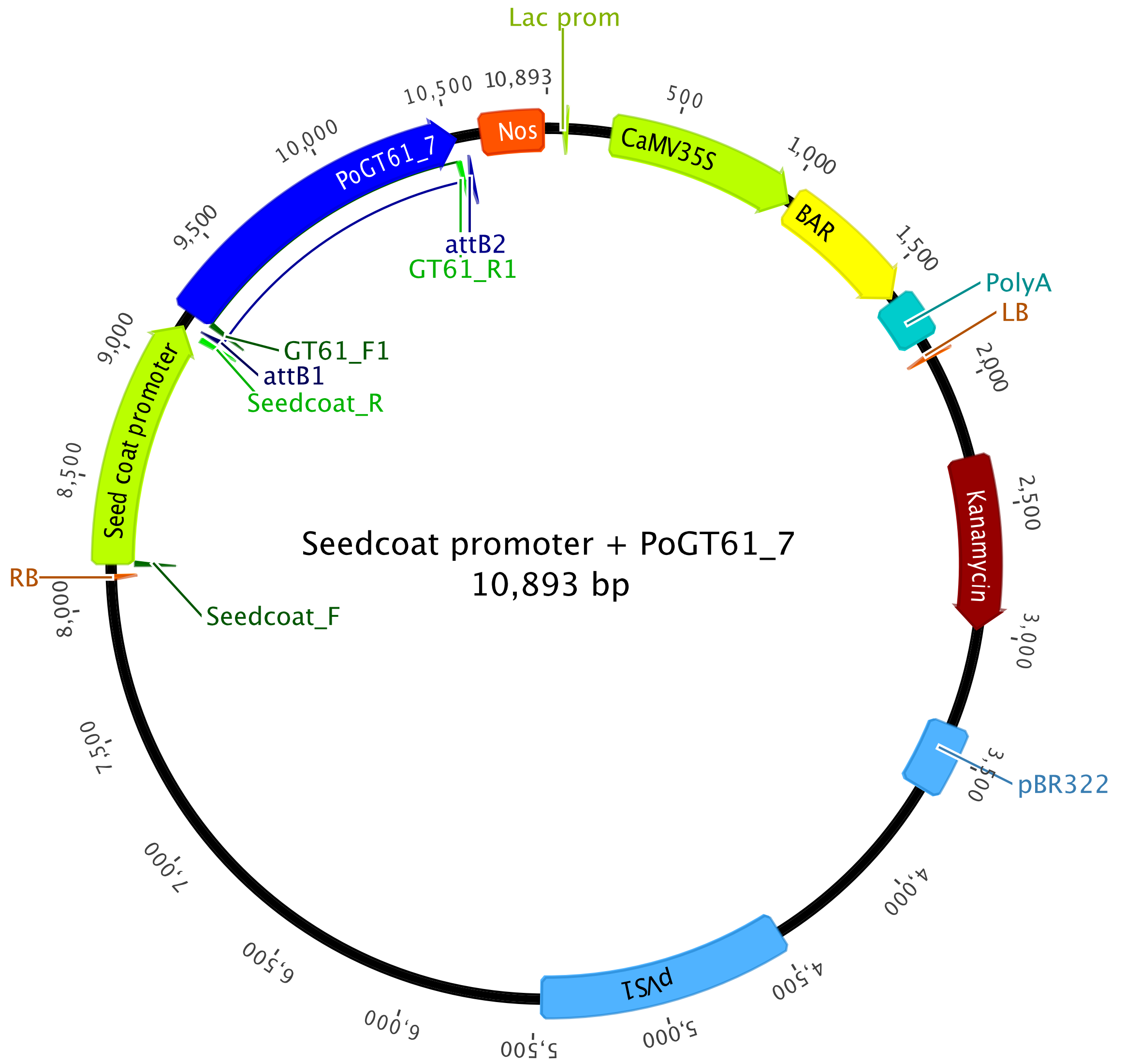


Figure S3. *PoGT61_7* with seed-coat promoter in an expression vector

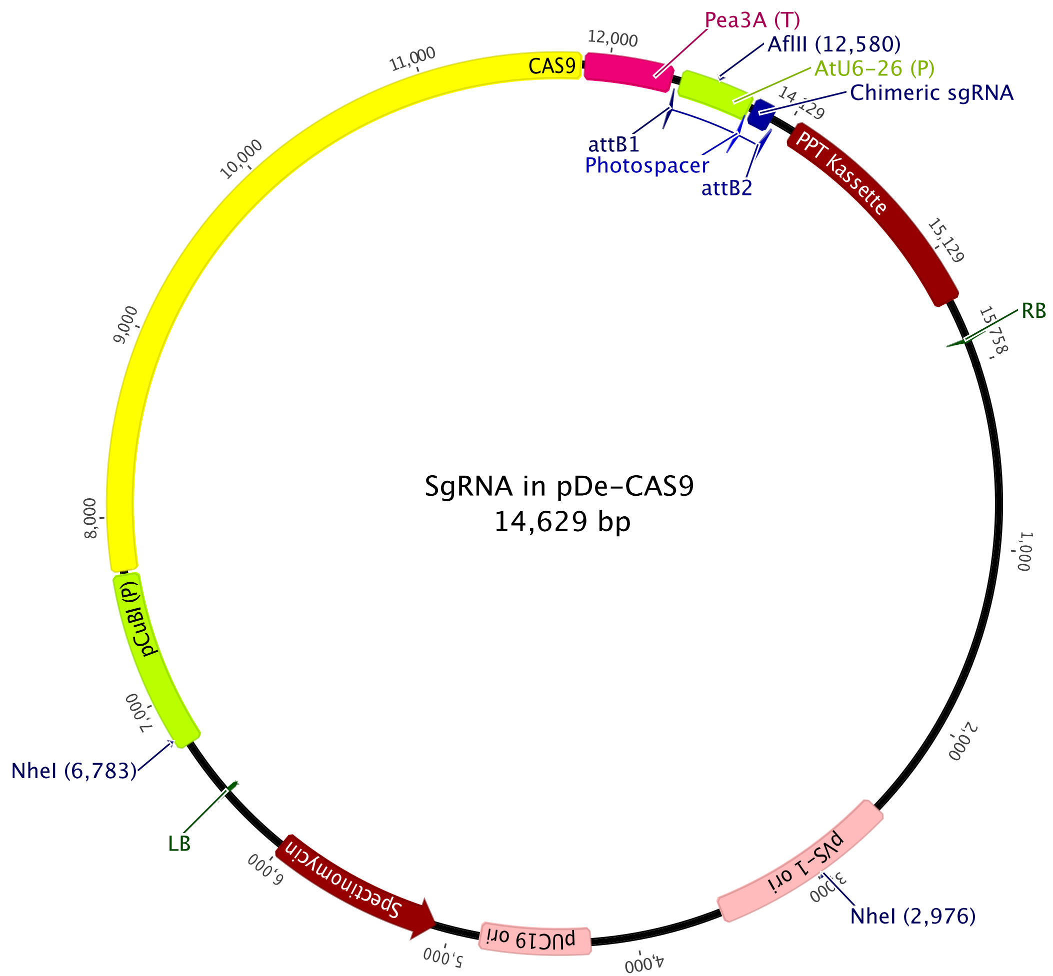


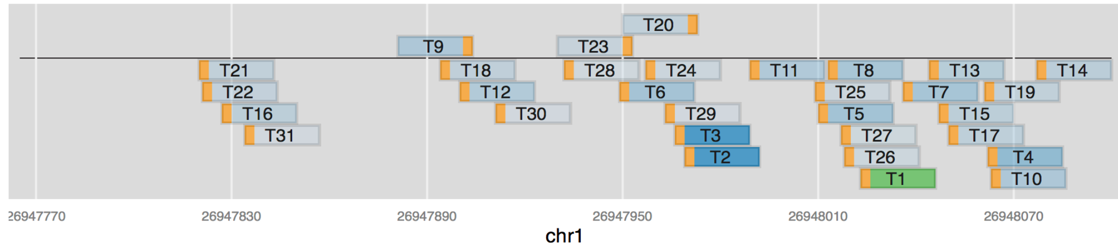
Figure S4. An example of expression vector (pDe_Cas9) with sgRNA

Results for DUF579

Download full results file here.
Download sgRNAs target sites as fasta file here.

Detailed results

Species: Arabidopsis (Arabidopsis thaliana TAIR10)	PAM:	NGG
Input: GTAACGTCTCTCACTTAAGCACAATCTCTCTGCATCTCTCAATGACGACTTATCTCCCAT	Target site length:	20
TTCCGATGACCACGAGACTCATACTCTCTAGCTTCGTCTCTCTCGTCGTACTCACTCTCT	Target site 5' limitation:	NN
TCTTCATCACCAGAACCGGTTTCTCACCGTCATCTTCATTTCATCAGCCTCTGAACAACA	Target site 3' limitation:	NN
CTCTCCGGATCTCCACTTCCTCCACCGGCAGTAAATTACAGAGTCCAAGAAGTTGTAATA	Core length:	12
AGATCCCTCCTTCCCTCGCCGATGCTCTCGTCCACTATGCCTCCTCCAATGTCACTCCCC	Core MM:	2
AGCAAACTCTCTCCGAAATCTCCGTCACCAAGAAG	Total MM:	4



Legend for off-target site positon: E = exonic; I = intronic; - = intergenic

T1 out of 31
<Previous Next>
Sequence: GGCATAGTGGACGAGAGCATCGG
Oligo pair fwd: ATTGGGCATAGTGGACGAGAGCAT rev: AAACATGCTCTCGTCCACTATGCC

Coordinates	strand	MM	target_seq	PAM	distance		gene name	gene id
chr1:26948023-26948045	-	0	GGCATAGT [GGACGAGAGCAT]	CGG	0	E		AT1G71690

T2 out of 31
<Previous Next>
Sequence: TGGACTCTGTAATTTACTGCCGG
Oligo pair fwd: ATTGTGGACTCTGTAATTTACTGC rev: AAACGCAGTAAATTACAGAGTCCA

Coordinates	strand	MM	target_seq	PAM	distance		gene name	gene id
chr1:26947969-26947991	-	0	TGGACTCT [GTAATTTACTGC]	CGG	0	E		AT1G71690

T3 out of 31
<Previous Next>
Sequence: ACTCTGTAATTTACTGCCGGTGG
Oligo pair fwd: ATTGACTCTGTAATTTACTGCCGG rev: AAACCCGGCAGTAAATTACAGAGT

Coordinates	strand	MM	target_seq	PAM	distance		gene name	gene id
chr1:26947966-26947988	-	0	ACTCTGTA [ATTTACTGCCGG]	TGG	0	E		AT1G71690

T4 out of 31
<Previous Next>
Sequence: GATTTCGGAGAGAGTTTGCTGGG
Oligo pair fwd: ATTGGATTTCGGAGAGAGTTTGCT rev: AAACAGCAAACTCTCTCCGAAATC

Coordinates	strand	MM	target_seq	PAM	distance		gene name	gene id
chr1:26948062-26948084	-	0	GATTTCGG [AGAGAGTTTGCT]	GGG	0	E		AT1G71690
chr5:11558991-11559013	-	4	GAGATCGA [AGAGAGATTGCT]	AGG	190	-		AT5G30450
chr3:3757428-3757450	+	4	GATTAGGG [AGAGAGTTTGAG]	AGG	0	E		AT3G11890
chr5:25629710-25629732	-	4	GATTTTTG [AGAGAGTTTGTG]	TGG	95	-	PSAN	AT5G64040

T5 out of 31
<Previous Next>
Sequence: AGAGCATCGGCGAGGGAAGGAGG
Oligo pair fwd: ATTGAGAGCATCGGCGAGGGAAGG rev: AAACCCTTCCCTCGCCGATGCTCT

Coordinates	strand	MM	target_seq	PAM	distance		gene name	gene id
chr1:26948010-26948032	-	0	AGAGCATC [GGCGAGGGAAGG]	AGG	0	E		AT1G71690
chr5:20195715-20195737	+	3	AAAGCACC [GGCGAGGGAAGT]	GGG	0	E		AT5G49710

Figure S5. An example of CRISPR/Cas9 target from CCTop with the off-target prediction. Target gene is AT1G71690. After T3, sgRNA targets other genes. (https://crispr.cos.uni-heidelberg.de/)

Table S1. Sequences of all primers and sgRNAs

Oligo name	Sequence 5'-3'	Orientation
M13	CACAGGAAACAGCTATGAC	Reverse
SS42	TCCCAGGATTAGAATGATTAGG	Forward
SS43	CGACTAAGGGTTTCTTATATGC	Reverse
sgRNA1	ATTGGGCATAGTGGACGAGAGCAT	Forward
	AAACATGCTCTCGTCCACTATGCC	Reverse
sgRNA2	ATTGACTCTGTAATTTACTGCCGG	Forward
	AAACCCGGCAGTAAATTACAGAGT	Reverse
sgRNA3	ATTGCTGTAATTTACTGCCGGTGG	Forward
	AAACCCACCGGCAGTAAATTACAG	Reverse
Seed-coat promoter	CTTTTCTGGGAAGCTCGTTG	Forward
	CGCACAAAATTGGTTTGTTG	Reverse
GT61_P1	ATGGATAACGAGGAGCAAATGT	Forward
	CTACGGCTCAAGTAGCTCTT	Reverse
GT61_P2	GGGAAGTACCCAATGGACAGT	Forward
	TGGCTTTCAGAATGGTTTCTC	Reverse
DUF579_14F	AACGTCTCTCACTTAAGCACA	Forward
DUF579_753R	TACCAAGAAGTTGCACGGAG	Reverse
GAPDH	TGGTTGATCTCGTTGTGCAGGTCTC	Forward
	GTCAGCCAAGTCAACAACTCTCTG	Reverse

APPENDIX 2 – PLANT PHYSIOLOGY INSTRUCTIONS FOR AUTHORS

Plant Physiology®

INSTRUCTIONS FOR AUTHORS

Plant Physiology, 2017 Last Edited August 14, 2017

- **Scope**
  - Plant Physiology is an international journal that publishes on the broadest aspects of plant biology. The journal welcomes original submissions that offer new and fundamental insights into the origins, development, and function of plants from the molecular to the whole organism and its interactions within the biotic and abiotic environment. Plant Physiology encourages submissions that span a range of technologies, including those of structural, molecular, and cellular biology, biochemistry, biophysics, bioenergetics, genetics, physiology, and field-based approaches as well as those making use of synthetic, bioinformatics, and -omics tools. Manuscripts submitted to Plant Physiology must not be under simultaneous consideration or have been published elsewhere, either in part or in whole. Prior publication on preprint servers, as a poster abstract or an oral presentation is not considered previous publication of the research.
  - A Plant Physiology article should not:
  - • be purely descriptive in content
  - • present confirmatory, preliminary, or incomplete research, including research based primarily on analysis of a single allele or transgenic line.
  - • report on well-known processes in another species
  - Permissions policy
  - Ethics Policy
- **Quick guide: Submission requirements**
  - **Please see Full submission guidelines for more detail** Before you submit, please have ready: **Cover letter**: A decision on whether, and how, to handle a manuscript is based initially on this letter Clearly state the significance of the work to understanding in plant biology in the broadest terms Explain how the work will advance knowledge of plant biology and/or enable research in the future Do not simply duplicate what is stated in the manuscript abstract Justify any disqualification of potential reviewers Indicate if the journal is to consider more than one corresponding author in publication - sound, material justification for multiple corresponding authorship is needed or the journal will not consider requests **Manuscript text file Author Contributions Please include on the first page of your manuscript after the titles and author list a breakdown of the authors' contributions for a first-page footnote (new style). The format should be similar to the following: A.A. conceived the original screening and research plans; B.B. and C.D. supervised the experiments; Z.Y. performed most of the**

experiments; Y.X. provided technical assistance to Z.Y.; A.B. designed the experiments and analyzed the data; B.B. conceived the project and wrote the article with contributions of all the authors; Z.Z. supervised and complemented the writing. One-sentence summary Figure files: 1 figure per file; include the legend as part of the file [**Help With Digital Art**]() **Supplemental Materials in one pdf**

- **Quick guide: Submitting a revision**
  -
  -
  - Before you submit, please have ready: **Response to reviewers with annotations to changes in manuscript Revised text file with all changes highlighted Final, revised one-sentence summary Figure files (1 figure per file with legends included) vTOC icon Funding information Next-gen sequence or microarray data (if applicable)**
- **Research Areas and Editor Expertise**
  -
  - These categories are not rigid and you may select an editor from any area Editor-in-Chief: Mike Blatt (membrane transport physiology and biophysics, stomata, membrane trafficking, quantitative systems modelling)
  - **Biochemistry and Metabolism**
  - **Breakthrough Technologies**
  - **Cell Biology**
  - **Ecophysiology and Sustainability**
  - **Genes, Development, and Evolution**
  - **Membranes, Transport, and Biogenetics**
  - **Signaling and Response**
  - **Systems and Synthetic Biology**
  - **Reviews**
- **Article types**
  - **Research Articles** Research Articles present original findings with new and fundamental insights into the biological processes of plants and/or set out novel approaches, tools, or resources that will enable scientific progress. **Research Reports** Research Reports are a forum for new and original findings and/or methods that are highly focused, and challenge current approaches or thinking in the plant sciences. Like full-length Research Articles, a Research Report must address hypotheses about the origins, development, and/or functions of plants or add substantially to approaches, tools, and resources that enable scientific progress.are Ideally, a Report should be 6 pages, including tables, figures, and references (4 or 5 figures or tables and ~25,000 characters). **Breakthrough Technologies** Breakthrough technologies present advances of exceptional significance, broad applicability, and interest that include new analytic or biochemical methods, bioinformatic tools, datasets, or algorithms as well as genome sequences or databases relevant to new model species. **Letters to the Editor** Letters to the Editor are usually solicited. Uninvited Editorials and Letters to the Editor on topics of interest or controversy will be considered for publication. Authors should contact the Editor-in-Chief (Mike Blatt; eic-plantphys@glasgow.ac.uk) before submission. **Commentaries** Commentaries appear occasionally to accompany articles of special interest and are solicited by the Editor-in-Chief or Associate Editor for Reviews. Commentaries offer insights into the background and significance of the article for the lay reader. **Scientific Correspondence** Scientific

Correspondence provides a forum for new scientific ideas and findings, typically based on an analysis of the existing literature and/or limited experimentation. They are strictly limited to three journal pages including tables, figures, and references. Authors proposing to submit Scientific Correspondence should contact the Editor-in-Chief in advance. Publication costs are waived for approved Scientific Correspondence articles. Scientific Correspondence will be considered subject to the extent to which thinking is challenged on topics of interest to the Journal. Scientific Correspondence should not include an abstract and a summary of the work should be provided within the first four sentences. The Correspondence format also should not include separate sections for methods, results or discussion, although these components of the work may be interwoven within the text and figure legends. **Updates** Updates are solicited reviews of recent progress and meant to be resources for research and advanced teaching tools. Updates should not exceed 5000 words in the main body of the text, and may include figures and tables. In addition, Updates may include Boxes with figures, tables, and/or text that highlights particular topics and/or concepts or introduces outstanding questions and/or challenges in the field. Updates are primarily associated with Focus Issues. Authors wishing to contribute an Update to a regular issue must first correspond with the Editor-in-Chief or Associate Editor for Reviews. Items for Inclusion in Updates (this text is not counted towards the main body:

-
- Advances Box (900 characters, including spaces, required)
-
- • The Advances box is a short collection of bullet point statements (3-5) that concisely convey to the reader the recent advances in the area, including emerging concepts and/or distinctions, that constitute a main motivation for the discussion developed in the article.
- • As the box aims to focus on recent developments, conclusions and future directions should be discussed in the Concluding Remarks section and/or the Outstanding Questions box.
- • The text in the Advances Box is not called out in the text.
- • The Advances Box does not count towards the total number of allowed display elements in the manuscript.
-
- Outstanding Questions Box (900 characters, including spaces, required)
-
- • Important questions for future research should be summarized in a box (not included in box count or element limit). This is an excellent opportunity to offer input and guidance on new directions for the field.
- • Please write succinct questions in list format, with bullet points to indicate the start of a new concept.
- • The Outstanding Questions Box should not include references.
- • The box should be called out in an appropriate section in the text, generally the Concluding Remarks section, as 'see Outstanding Questions'. This element will be placed as the last box in the paper, although it should not be numbered with the other boxes.
-
- Text Boxes
- • Ideal for providing explanations of basic concepts or theories, giving detailed mechanisms, or discussing case studies.
- • Please cite text boxes in the main text as: Box 1.
- • Boxes should have a single sentence title (no more than 8 words).

- ◦ • Text boxes can occasionally contain a small figure or table. Please cite the element in the box text.
- ◦ • 400 words max per box.
- ◦ • References for citations in the box should be included in the main reference list and must also be cited in the main text.
- ◦ • No more than three text boxes per article.
- ◦ **Topical Reviews** Topical Reviews are usually invited and are intended to provide experts and non-experts alike with the conceptual and technical background behind the most important areas of research at the forefront of plant biology. Topical Reviews are normally limited to 8,000 words in the main body of the text, and may include figures and tables. In addition, Topical Reviews may include Boxes with figures, tables, and/or text that highlights particular topics and/or concepts or introduces outstanding questions and/or challenges in the field. Items for Inclusion in Topical Reviews (this text is not counted towards the main body): Advances Box (900 characters, including spaces, required) • The Advances box is a short collection of bullet point statements (3-5) that concisely convey to the reader the recent advances in the area, including emerging concepts and/or distinctions, that constitute a main motivation for the discussion developed in the article. • As the box aims to focus on recent developments, conclusions and future directions should be discussed in the Concluding Remarks section and/or the Outstanding Questions box. • The text in the Advances Box is not called out in the text. • The Advances Box does not count towards the total number of allowed display elements in the manuscript. Outstanding Questions Box (900 characters, including spaces, required) • Important questions for future research should be summarized in a box (not included in box count or element limit). This is an excellent opportunity to offer input and guidance on new directions for the field. • Please write succinct questions in list format, with bullet points to indicate the start of a new concept. • The Outstanding Questions Box should not include references. • The box should be called out in an appropriate section in the text, generally the Concluding Remarks section, as 'see Outstanding Questions'. This element will be placed as the last box in the paper, although it should not be numbered with the other boxes. Text Boxes • Ideal for providing explanations of basic concepts or theories, giving detailed mechanisms, or discussing case studies. • Please cite text boxes in the main text as: Box 1. • Boxes should have a single sentence title (no more than 8 words). • Text boxes can occasionally contain a small figure or table. Please cite the element in the box text. • 400 words max per box. • References for citations in the box should be included in the main reference list and must also be cited in the main text. • No more than three text boxes per article. **Founders' Reviews** Founders' Reviews are by invitation only. They highlight the work of preeminent scientists in plant biology and provide in-depth perspectives on their fields of research. Like Topical Reviews, they are intended to provide experts and non-experts alike with the conceptual and technical background of research of exceptional importance in plant biology. Founders' Reviews should not exceed 14,000 words with references (but excluding supplemental material), and may include up to 6 figures and tables.

- **Full submission guidelines**
- ◦ **Research Article organization**
- ▪ **Short title not to exceed 50 characters and spaces Corresponding**

author(s) details Article title All author names and affiliations

-
- **One sentence summary**
- One single, short sentence (200 characters max) State the **most important concept** to come from your research Focus on the biology, not the methods, unless a it a Breakthrough Technologies paper Write it for the non-expert Use the active voice Avoid abbreviations Do not use 'we' and phrasing such as 'this paper shows' Examples of well-written summaries: ***An acyltransferase reduces cross linking in grass cell walls, yielding grass leaves and stems that can be more easily broken down to make biofuels.*** [link to article] ***An epiphytic fungus induces plant resistance against pathogens.*** [link to article] ***Microbial elicitors and the plant defense hormone jasmonic acid differentially modulates the plant's innate immune response.*** [link to article] ***An inactive invertase may indirectly stimulate the activity of active cell wall invertases.*** [link to article]
- **Footnotes in the following order:**
- **List of author contributions** Please include in your manuscript a breakdown of the authors' contributions for a first-page footnote (new style). The format should be similar to the following: A.A. conceived the original screening and research plans; B.B. and C.D. supervised the experiments; Z.Y. performed most of the experiments; Y.X. provided technical assistance to Z.Y.; A.B. designed the experiments and analyzed the data; B.B. conceived the project and wrote the article with contributions of all the authors; Z.Z. supervised and complemented the writing. Funding information (if any) Present addresses (if any) Corresponding author email
- **Abstract (250 word max) Introduction, Results, Discussion, and Conclusions Materials and Methods** This section should reference all standard procedures but must be complete enough so that results can be verified by other laboratories. **Accession numbers** List at the end of the Material and Methods any data or materials available in a public repository. Novel DNA sequences must be deposited in GenBank (http://www.ncbi.nlm.nih.gov) and accession numbers listed in text. If a number has not yet been assigned at the time of submission, use xxx as a place holder to be updated later. **Large datasets** Datasets too large for inclusion in the manuscript must be uploaded as supplementary material. **Acknowledgments**
- **Tables**
- Provide a heading for each column Use Word's create table feature—not tabs! Color, shading, and graphics are not supported; make it a figure
- **Figure legends**
- **Literature Cited**
- Use plain text (no bold, italics, etc.) Remove Word fields Extract from EndNote Reformatting is not necessary in many cases!!! X No unpublished data X No abstracts X No

posters X No personal communications The accuracy in the Literature Cited section is the responsibility of the authors.

- **TOC icons**
  - The journal uses a graphic icon to represent each paper on the Table of Contents, which you should submit with the final production version of the manuscript:
  -
  - 96 x 96 pixels
  - JPEG format (in RGB color mode)
  - Please note that animated GIFS are no longer permitted
  - If using type, ensure it is legible in the final icon.
  - **Making an icon from a figure**
  - Remove unessential text and graphic elements
  - Reduce the figure to its strongest, most important graphic elements
  - Useful tools for creating and editing icons are Adobe Illustrator, Adobe Photoshop, and GraphicConverter. If an original figure is an EPS file, Illustrator can be used to edit graphic elements and adjust the font size and thickness.
- **Abbreviations**
  - A list of abbreviations that can be used without defining is at: https://pphys.msubmit.net/html/Abbreviations.pdf.) Introduce all other abbreviations parenthetically after the first use of the term both in the abstract and the text. Abbreviations must be used three times in the text (this includes table and figure legends) or the term must be spelled out. Abbreviations in **gene and protein names** are acceptable when used 3 or more times. Numbers should be closed up in gene names that are spelled out (i.e., *APETALA2*). Avoid using an abbreviation first, unless spelling out the name of a gene would make the sentence indecipherable. If possible, define genes and proteins and their abbreviations in functional context when first used. Taxonomic prefixes (At, Os, Nt, etc.) should be closed up to the gene/protein names and match the style (italic or normal type) of the name.
- **Figure Preparation**
  - · One file per figure; include the figure number and legend · Number figures consecutively according to the order in which they are called out in the text · Format your figures according to our "Help With Digital Art" document · Provide enough information so that the reader can understand them without significant input from the text · Designate panels with capital letters, no parentheses, and no periods after letters, in the upper left-hand corner · Whenever possible, position panels vertically for one-column reproduction in the Journal · Format the width of sequence data in the paper to one column **Figure manipulation** *Plant Physiology* does not allow enhancements or manipulations of micrographs, gels, or other digital images using Photoshop or any other software. Refer to J Cell Biol 158: 1151 (http://www.jcb.org/cgi/content/full/158/7/1151) and the joint Editorial by Cathie Martin and Mike Blatt (http://www.plantphysiol.org/content/163/1/3.full) for guidance on acceptable and unacceptable digital image manipulation. *Plant Physiology* reserves the right to ask authors to provide supporting data on which figures were based. A few pointers: * separate multiple

images when collected into a single figure * apply contrast, brightness, or color adjustments uniformly to an entire image (or image set if combined) * targeted alteration is prohibited

- **Image data and micrography** must provide information about the acquisition and processing of the images, including: · Make and model of microscope · Type, magnification, and numerical apertures of the objective lenses · Imaging medium · Fluorochromes and related material · Make and model of any camera systems · Acquisition software and version · Software used for image processing subsequent to data acquisition. Please include details and types of operations involved (e.g., type of deconvolution, 3D reconstitutions, surface or volume rendering, gamma adjustments, etc.). **All micrographs** must include a bar to indicate the scale. In general, images representing experimental data must be supported by statistical analyses based on multiple datasets. **NOTE ON PREPARING FIGURES FOR COLOR VISION-DEFICIENT READERS** Many readers of the Journal (1 in 12, on average) have some form of color-deficient vision; therefore, when preparing your figures, please observe the following guidelines to ensure that all readers will be able to comprehend your data.
- In fluorescent double-staining micrographs and DNA chips, do not use the combination of red and green; use magenta and green instead.
- For micrographs with triple or more channels, additionally show either grayscale image of each channel or the combination of the two most important channels in magenta and green.
- For graphs and line drawings, label elements on the graph itself rather than making a separate color-coded key. Do not try to convey information in color only, but use BOTH color and shape (solid and dotted lines, different symbols, various hatchings, etc.).
- For more information, see the following web site: http://jfly.iam.u-tokyo.ac.jp/color/

- **Cover figure submissions.**
- **Statistical analysis**

- **Large-scale datasets and other supplemental data**

- **Supplemental data** All supplemental data must be included at submission and will be peer reviewed. Where possible, combine all supplemental materials and legends into one PDF. Supplemental materials are subject to the same copyright restrictions as published manuscripts and cannot be presented elsewhere without proper citation. All materials included as Supplemental data must be cited in the manuscript with direct reference to each item, such as Figure S1, Table S1, etc. Include a list of all Supplemental data at the end of the Materials and Methods section under the heading "Supplemental Material" along with a brief descriptive title. Supplemental materials that cannot be submitted via eJPress are occasionally hosted by authors, or for authors, on a third-party hosting site. For supplemental data of this nature, it is the responsibility of the authors to both provide instruction for access to this material within the manuscript under the heading "Supplemental Material" and remove or disable access to this material upon completion of the peer-review process. Authors providing this sort of supplemental data are responsible for the security of the data and accept that third-party access to the confidential account may be a risk. We encourage reviewers to use an anonymous web-browsing site, such as

http://anonymouse.org, to further promote confidentiality. **Large-scale datasets** Manuscripts that contain gene expression profiling data are required to describe the experiments according to MIAME guidelines (Brazma et al. 2001 Nature Genetics 29: 365-371; http://www.mged.org). Unreplicated gene expression profiling experiments will not be accepted for publication. Transcript profiling must include the complete set of genome sequences analyzed, ESTs identified, and genes queried in transcript profiling. Data supporting transcript profiling experiments must include complete sequence information. Independent validation is expected for electrophoresis, proteomic and MS-based experiments, where biologically important differences in protein (gene) expression are reported. Include the methods, search and statistical parameters, and details of any software used in mass spectrometry to create peak lists. For MS/MS, include the number of peptides used to identify the protein as the sequence and charge state of each peptide. For peptide mass fingerprinting, the number of peptides that match the sequence and the total percent of sequence coverage should be quoted. Reports of post-translational modifications should include the methods used to discover the modification(s). When mapped to amino acid(s) by fragmentation analysis, data should be reported as ambiguous if mapping to a single amino acid is not possible. Refer to "Minimum Information About a Proteomics Experiment" for further information (http://www.psidev.info/MIAPE). At the time of publication, supplemental data must be placed in a permanent public repository if one is available, or if none is available, in *Plant Physiology*. Examples of accepted public gene expression repositories are SRA (http://www.ncbi.nlm.nih.gov/sra,) GEO (http://www.ncbi.nlm.nih.gov/geo), and ArrayExpress (http://www.ebi.ac.uk/arrayexpress). **Spreadsheets** Data should be presented in a single spreadsheet with multiple tabs. **Next-generation sequence data** As with microarray data, complete data sets from next-generation sequencing must be deposited in an accepted format in a permanent public repository with free access (e.g., GEO or NCBI's Short Read Archive sequence database). This includes data from small RNA, mRNA, specialized RNA libraries, ChIP-seq, whole-genome re-sequencing or genotyping, whole-genome bisulfite sequencing, etc. Links to web sites other than a permanent public repository are not an acceptable alternative because they are not permanent archives.

- **The process: what to expect**
  -
  - **Peer review**
  - **Proof**
- **Policies**
  - **ORCID IDs** *Plant Physiology now requires all Corresponding Authors of Accepted articles to have an ORCID ID prior to production of the article. ORCID provides a persistent digital identifier that distinguishes an author from other researchers and, through integration in key research workflows such as manuscript and grant submissions, supports automated linkages between you and your professional activities, ensuring that your work is recognized. For more information about ORCID, please visit http://orcid.org/about/what-is-orcid/mission.* ***Corrections and retractions*** *If necessary, corrections of significant errors in published articles will be published in a later issue of the journal. Authors are requested to bring any errors to the attention of the Production Manager as soon as possible. Corrections published in the journal will also be linked to the original article. Articles may be retracted by their authors, academic or institutional sponsor, editor, or publisher because of*

pervasive error or unsubstantiated or irreproducible data. A correction or retraction, so labeled, will appear in a prominent section of the journal, listed in the contents page, and include the title of the original article. The text of a retraction will explain why the article is being retracted and include a bibliographic reference to it. ***Preprint policy*** *Plant Physiology will review manuscripts that include data posted on an author's website or posted on preprint servers such as the bioRxiv. Final published manuscripts will reside on the Journal site. Once the published article has appeared online at the Journal site, a toll-free link providing barrier-free access to the HTML and PDF versions of the article will be e-mailed to all authors. These links may be deposited in open access institutional repositories.* ***NIH Public Access Policy*** *Plant Physiology deposits the final version of published articles in PubMed Central for release 12 months after the date of publication. If a paper was funded in whole or in part by the National Institutes of Health, this fulfills the requirements of the NIH Public Access Policy. However, authors of NIH-funded papers, if they choose, may deposit the peer-reviewed, accepted version of their manuscript in PubMed Central, provided the release date is set for 12 months after the date of final publication in the Journal.* ***OPEN articles*** *The journals of the American Society of Plant Biologists (ASPB) make all articles free 12 months after publication. Authors can purchase immediate free access in one of the following ways: 1. Corresponding authors publishing in Plant Physiology who are members of ASPB have their articles made freely available upon publication as a member benefit. Nonmembers who work at a subscribing institution pay $950. Nonmembers working at an institution without a subscription pay $1500. ASPB retains copyright. 2. For authors whose funders require that the work be published under a Creative Commons or similar license, ASPB offers a CC-BY license for $2000. The CC-BY license allows noncommercial and commercial reuse of all or part of a paper without restriction. On behalf of all authors, the ASPB journals deposit final published articles in PubMed Central for release 12 months after the date of publication (unless a free-access option applies). Permission to reuse part or all of a copyrighted work published in an ASPB journal is granted without fee for personal or educational use, provided that copies are not made or distributed for profit or commercial advantage and that copies bear the full citation, the journal URL (www.plantphysiol.org; www.plantcell.org), and the following notice: "Copyright American Society of Plant Biologists." Please request permission in writing from the Copyright Clearance Center (www.copyright.com) if the use is commercial or if you wish to make multiple copies other than for educational purposes. Please note that the ASPB journals allow authors to reuse their content for any purpose (noncommercial or commercial) without written permission from ASPB, so long as the original source and a copyright notice are cited.* ***Authorship*** *Authorship credit should be based only on substantial contributions to (a) conception and design, or analysis and interpretation of data; (b) on drafting the article or revising it critically for important intellectual content; and (c) on final approval of the version to be published. Conditions a, b, and c must all be met. All authors must have participated sufficiently in the work to take public responsibility for the content, consistent with the authorship policy adopted by the International Committee of Medical Journal Editors and published in the Uniform Requirements for Manuscripts Submitted to Biomedical Journals, 1994. Manuscripts must include Author Contributions following any Acknolwedgments. The Author Contributions must list the contributions of each*

author contributed to the research and/or the writing of the manuscript. Note which of the following tasks each author performed: designed the research; performed research; contributed new analytic/computational/etc. tools; analyzed data; or wrote the paper. After submission, any change in authorship will require the agreement of all coauthors. ***Ethics*** *Please click on the following link for information regarding ASPB policies and procedures for handling allegations of author misconduct: http://www.aspb.org/publications/aspb-journals/policies-procedures/.*

Conflicts of Interest *Please click on the following link for ASPB's statement on potential conflicts of interest for editors and reviewers: http://www.aspb.org/publications/coi.cfm.* ***Distribution of materials*** *Plant Physiology requires that all materials described in a manuscript be made freely available to academic investigators for non-commercial research purposes. Upon request, all novel materials described in the publication must be made available in a timely manner for non-commercial research purposes, subject to requisite permissions from any third-party owners of all or parts of the material. Obtaining permission from third parties is the responsibility of the requestor. No restrictions or conditions may be placed on the use of any materials described in a paper published in Plant Physiology that would limit their use for non-commercial research purposes. The provision of novel enzymes, antibodies, and material products described in the manuscript can be reasonably limited in case that substantial effort is required for their isolation or preparation. All manuscripts submitted to Plant Physiology must include the following statement as an unnumbered footnote: "The author(s) responsible for distribution of materials integral to the findings presented in this article in accordance with the Journal policy described in the Instructions for Authors (http://www.plantphysiol.org) is (are): J.D. Author (author@college.edu)." Contact Information for the author(s) responsible for distribution of materials must be provided. To facilitate sharing of materials, authors may choose to use one of the following services. Addgene (http://www.addgene.org) Arabidopsis Biological Resource Center (http://www.biosci.ohio-state.edu/pcmb/Facilities/abrc/a/a>) Maize Genetics Cooperation Stock Center (http://maizecoop.cropsci.uiuc.edu) ASPB provides these links as a service to our author community. The Society does not take responsibility for or endorse the services these companies offer and cannot attest to the quality of their work. If you have questions, or if you do use one of these services and would like to send ASPB feedback on your experience, please contact Nancy Winchester, ASPB Director of Publications, at nancyw@aspb.org.*

- ***Gene nomenclature***
  - *Plant Physiology adheres to community standards for gene and protein nomenclature. The two-letter prefix indicating species (e.g., At, Zm, Os) is not part of the gene symbol and is used to prevent possible confusion between species. For Arabidopsis mutant genes, authors must follow the guidelines for naming genes outlined by Meinke and Koornneef ([1997] Community standards for Arabidopsis genetics. Plant J 12: 247-253), and register new mutant gene symbols at http://www.Arabidopsis.org/doc/submit/functional_annotation/123. This URL also provides other useful information and links on plant gene and protein classification. Other resources include: http://www.Arabidopsis.org/portals/nomenclature/guidelines.jsp (Arabidopsis) http://www.maizegdb.org/maize_nomenclature.php (Maize) http://wheat.pw.usda.gov/ggpages/wgc/98/ (Wheat) http://tgrc.ucdavis.edu*

[(Tomato)]() [http://www.chlamy.org/nomenclature.html (Chlamydomonas)]() VandenBosch, A., and Frugoli, J. 2001. Guidelines for genetic nomenclature and community governance for the model legume Medicago truncatulata. MPMI 14, 1364-1367. [http://www.brassica.info/info/reference/gene-nomenclature.php (Brassica)]()

- ***Fees and charges***
    - ***ARTICLE PUBLICATION FEES Research Articles (Non-ASPB Members): $1900 Research Articles (ASPB Members): $1600 Research Reports (Non-ASPB Members): $900 Research Reports (ASPB Members): $800 OPEN ACCESS FEES OPEN: $1500 if CA institution does not subscribe $950 if CA's institution subscribes to Plant Physiology FREE if corresponding author (CA) is a member of ASPB. For authors whose funders require that the work be published under a Creative Commons or similar license, ASPB offers a CC-BY license for $2000. The CC-BY license allows noncommercial and commercial reuse of all or part of a paper without restriction. Join ASPB: http://www.aspb.org/membership/***

American Society of Plant Biologists