# Molecular and evolutionary study of chromosomal gene clusters active during male reproductive development in *Arabidopsis thaliana*

Karolina Borowiec

Independent project • 30 credits
Swedish University of Agricultural Sciences, SLU
Department of Plant Biology
Master's Programme in Genetic and Molecular Plant Science
Examensarbete / Institutionen för växtbiologi, 208
Uppsala, 2024

# Molecular and evolutionary study of chromosomal gene clusters active during male reproductive development in *Arabidopsis thaliana*

*Molekylär och evolutionär studie av kromosomala genkluster aktiva under manlig reproduktiv utveckling i Arabidopsis thaliana*

Karolina Borowiec

| | |
|---|---|
| **Supervisor:** | **Jens Sundström, SLU, Department of Plant Biology** |
| **Assistant supervisor:** | Sonam Yadav, SLU, Department of Plant Biology |
| **Examiner:** | Adrien Sicard, SLU, Department of Plant Biology |

| | |
|---|---|
| **Credits:** | 30 credits |
| **Level:** | Second cycle, A2E |
| **Course title:** | Master thesis in Biology |
| **Course code:** | EX0895 |
| **Programme/education:** | Master's Programme in Genetic and Molecular Plant Science |
| **Course coordinating dept:** | Department of Plant Biology |
| **Place of publication:** | Uppsala |
| **Year of publication:** | 2024 |

| **Title of series:** | Examensarbete / Institutionen för växtbiologi |
| **Part number:** | 208 |

| | |
|---|---|
| **Keywords:** | Pollen development, Gene clusters, Multiplex CRISPR/Cas9 |

**Swedish University of Agricultural Sciences**
Faculty of Natural Resources and Agricultural Sciences
Department of Plant Biology

## Abstract

This project aimed to develop a gene-editing system to knock out (KO) genes in Cluster 8 of *Arabidopsis thaliana*, which are involved in pollen development. Building on the findings of Reimegård et al. (2017) and Xie et al. (2015), the project generated a synthetic dicistronic sgRNA gene (*SDRG*) and CRISPR/Cas9 system to create targeted KOs of genes with high linkage disequilibrium. The Final Vector containing this system was successfully assembled and can be used for follow-up functional analyses. Phylogenetic analysis indicated significant genetic redundancy within the PME gene family, complicating functional studies. Removing gene redundancy is expected to reveal more pronounced phenotypes. Initial T1 generation screening showed promising results, with 15% of genotyped individuals scoring positive for the inserted transgene. The technology developed herein can help further elucidate plants' reproductive development and holds the potential for developing alternative methods to induce male sterility, contributing to hybrid vigour in crops. Overall, this project underscored the complexity of multifactorial genetic experiments and provided a foundation for future investigations into the functional roles of Cluster 8 genes in pollen development.



*Keywords:* Pollen development, Gene clusters, Multiplex CRISPR/Cas9

## Sammanfattning

Detta projekt syftade till att utveckla ett genredigeringssystem för att slå ut (KO) gener i Kluster 8 hos Arabidopsis thaliana, vilka är involverade i pollenutveckling. Med utgångspunkt i resultaten från Reimegård et al. (2017) och Xie et al. (2015) skapades en syntetisk dikistronisk sgRNA-gen (*SDRG*) och ett CRISPR/Cas9-system för att specifikt slå ut gener med hög genetisk koppling. Den slutliga plasmiden som innehåller detta system sattes samman framgångsrikt och kan nu användas för vidare funktionella analyser. Fylogenetiska analyser visade på betydande genetisk överlappning inom PME-genfamiljen, vilket försvårar funktionella studier. Genom att eliminera denna genetiska överlappning förväntas mer uttalade fenotyper kunna observeras. Den initiala screeningen av T1-generationen visade lovande resultat, där 15 % av de analyserade individerna var positiva för den insatta transgenen. Den teknik som utvecklades kan bidra till en djupare förståelse av växters reproduktiva utveckling och har potential att utveckla alternativa metoder för att inducera manlig sterilitet, vilket kan öka heterosiseffekten i grödor. Sammanfattningsvis betonade projektet komplexiteten i multifaktoriella genetiska experiment och lade en grund för framtida studier av de funktionella rollerna hos gener i Kluster 8 i pollenutvecklingen.



*Keywords:* Pollenutveckling, Genkluster, Multiplex CRISPR/Cas9

# Table of contents

# List of Figures

# Abbreviations

| | |
|---|---|
| 1/2MS | Half strength Murashige and Skoog medium |
| aa | Amino acid |
| ABA | Abscisic Acid |
| *BAR* | BASTA resistance gene |
| BASTA | Phosphinotricin, an analogue of Glufosinate Ammonium |
| BMs | Basic Modules |
| CARK | Cytosolic ABA receptor kinase |
| CDS | Coding DNA sequence |
| CMS | Cytoplasmic male sterility |
| Col-0 | Columbia-0 |
| CRISPR/Cas | Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) - associated (CRISPR-Cas) endonucleases |
| crRNA | CRISPR RNA |
| CW | Cell wall |
| DSB | Double-strand break |
| dsDNA | Double-stranded DNA |
| ECM | Extracellular matrix |
| F1 | Filial generation |
| gDNA | Genomic DNA |
| GG | Golden Gate |
| GMOs | Genetically modified organisms |
| GO | Gene Ontology |
| GOIs | Genes of interest |
| HG | Homogalacturonan |
| indel | Insertion-deletion |
| KO, KOs, KOd | Knock out/s, Knocked out |
| LB | Lysogeny broth |
| *MS1* | *MALE STERILITY 1* |
| NHEJ | Nonhomologous end-joining |

| | |
|---|---|
| *NOS* | Nopaline synthase |
| P0 | Parental generation |
| PAM | protospacer-adjacent motif |
| PCD | Programmed cell death |
| PCP | Pollen coat proteins |
| PME | Pectin methylesterase |
| *PME4/VGDH1* | *PECTIN METHYLESTERASE 4 / VANGUARD HOMOLOG 1* |
| *PME5/VGD1* | *PECTIN METHYLESTERASE 5 / VANGUARD 1* |
| *PMEIL* | *PME INHIBITOR-LIKE* |
| PMEIs | Pectin methylesterase inhibitors |
| Pol II and Pol III | RNA polymerase II and III |
| RFLP | Restriction fragment length polymorphism |
| RS | Recognition site |
| RT | Room temperature |
| *SDRG* | Synthetic Dicistronic sgRNA Gene |
| SDs | Segmental duplications |
| sgRNA | Single guide RNA |
| SM | Supplementary Material |
| SNPs | Single nucleotide polymorphisms |
| SpCas9 or Cas9 | *Streptococcus pyogenes* Cas9 |
| T1, T2… | Transformant generation 1, 2… |
| TALENs | Transcription Activator Like Effector Nucleases |
| TDS | Transcribed DNA sequence |
| TEs | Transposing elements |
| TF | Transcription factor |
| TI | Transcriptional interference |
| TII REs | Type II restriction enzymes |
| tracrRNA | Trans-activating CRISPR RNA |
| *U3p* | *U3 snoRNA* promoter |
| *U6-26p* | *U6-26* promoter |
| *UBIp* | Ubiquitin promoter |
| UTRs | Untranslated regions |
| WT | *Wild-type* |
| ZFPs | Zinc-Finger Proteins |

# 1. Introduction

## 1.1 Flowering and Pollen

According to the latest reports, 288,735 angiosperm species, or flowering plants, are currently recognised (Jiang et al., 2023). Given that there are about 342,953 accepted vascular plant species (Govaerts et al., 2021), this highlights the flowering apparatus's critical role in these species' long-term survival. Additionally, the vast majority of globally cultivated crops, both staple and otherwise, are flowering plants (Leff et al., 2004). This underscores the importance of understanding the processes that guide the healthy development of these vital plant organs.

*Arabidopsis thaliana* has been established as the model organism for plant studies (Koornneef & Meinke, 2010). The flowering process in this ubiquitous weed is still to be fully understood, relying on complex genetic pathways (Quiroz et al., 2021). Its flowers are composed of distinct organs, including the central carpel that houses the female gamete cells and the anther that produces the male gametes found within the pollen (Scott et al., 2004). The pollen grain is a vegetative cell containing two sperm cells (Ischebeck, 2016; Yao et al., 2022), Figure 1a.

The anther contains four locules, each housing pollen-producing mother cells. These cells undergo meiosis to form tetrads, which then separate into microspores eventually maturing into pollen. The locules are lined with a tapetum cell layer that provides nutrition, enzymes to dissolve pectin and callose for microspore release, and precursors for the pollen wall and coat constituents. Failure to develop a tapetum layer results in male sterility (Crang et al., 2018; reviewed in Yao et al., 2022).

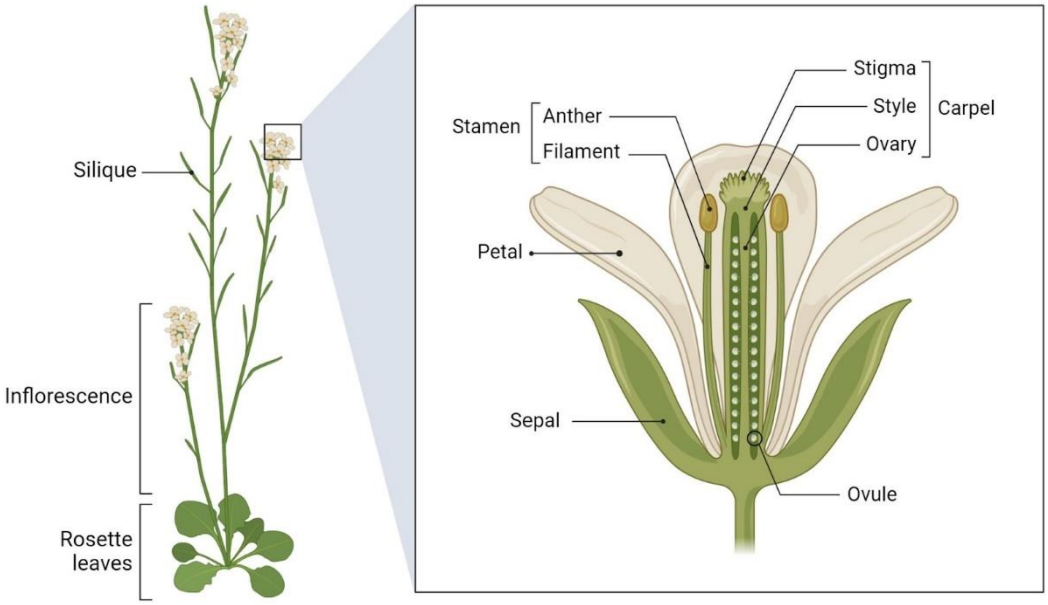
The pollen is encased in a multilayered wall and a coat called tryphine, each produced through different mechanisms. The intine layer is gametophyte-derived, while the nexine and sexine layers are derived from the tapetum or sporophyte. The tryphine is formed from an extracellular matrix (ECM) produced by tapetal cells after programmed cell death (PCD), as well as contributions from sporophytic anther cell layers and gametophytes. The tryphine serves multiple functions, including protection from environmental factors, facilitating pollen-stigma interactions, and aiding in grain hydration (Lu et al., 2020). As the pollen matures,

the tapetum degrades, leaving the pollen protected only by the epidermis and the endothecium layers until the anther opens (Pacini & Dolferus, 2016), Figure 1b.

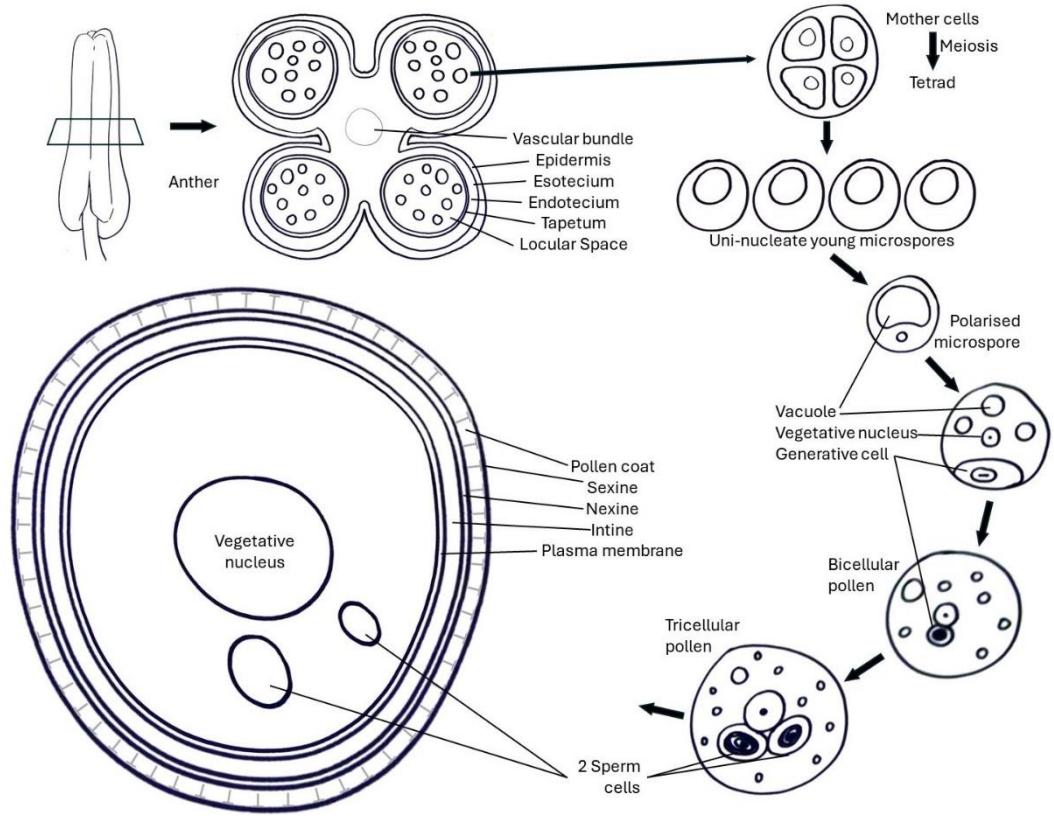
When plants reproduce sexually, fertilisation occurs through the successful delivery of sperm to egg cells. The process begins with pollen grains being deposited onto the stigma. If the interaction is compatible, the pollen germinates and forms a pollen tube, a highly polarised process characterised by tip growth (Jiang et al., 2005; Wang et al., 2008). Various gametophytic pollen coat proteins (PCPs) have been reported to be involved in pollen-stigma recognition (Lu et al., 2020). The pollen tube invades the pistil and extends down to the egg cells in the ovule, enabling the double fertilisation characteristic of angiosperms (Jiang et al., 2005; Wang et al., 2008). In species such as *A. thaliana*, the style is closed, and the intercellular space of the tract is filled with a nutrient-rich ECM that supports pollen tube growth (Jiang et al., 2005; Wang et al., 2008), Figure 1c.

*A. thaliana*, in contrast to its outcrossing relative *Arabidopsis lyrata*, predominantly reproduces via self-pollination, or selfing. In fact, it displays a strong selfing syndrome which includes small and unscented flowers, and little separation between the female and male parts, which aids in selfing (Luo & Widmer, 2013).

a.



b.




c.

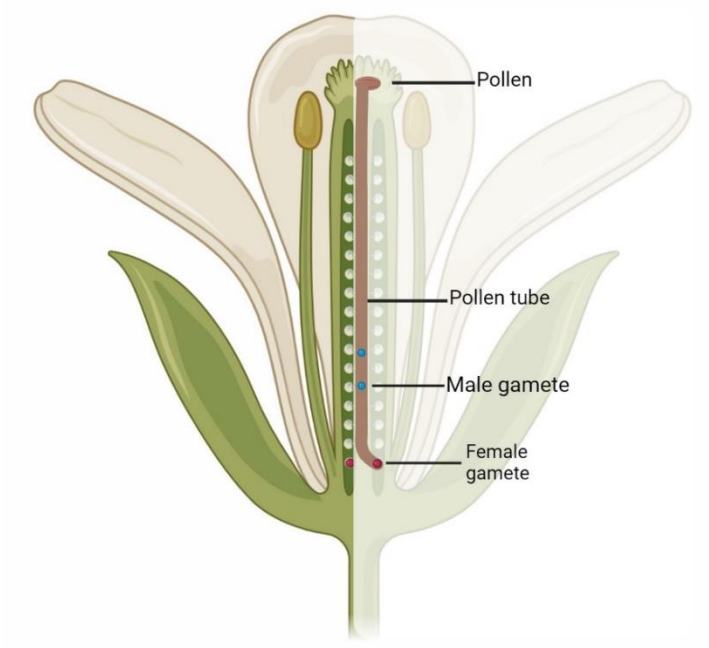



*Figure 1: Labelled Arabidopsis thaliana flower components. a. The stamen encompasses the male portion of the flower and the carpel encompasses the female parts. The siliques are pods filled with fertilised seeds. Figure generated on Biorender.com; b. Schematic drawings of a section of an anther, male gametophyte development in angiosperms, and pollen's wall layers and coat. Adapted from Pacini & Dolferus, 2016 and Yao et al., 2022; c. Labelled double fertilisation process. Figure generated on Biorender.com*

## 1.2 Genetic Cluster Formation

Genetic clustering refers to the grouping of functionally related genes. In eukaryotes, these clusters differ from prokaryotic operons as they are not transcribed as a single transcriptional unit. Instead, each gene within a cluster operates as an independent unit that can be co-activated and co-transcribed (Field & Osbourn, 2012; Reimegård et al., 2017). Clustered genetic architecture is thought to confer a selective advantage by increasing the likelihood of co-inheritance of genes necessary to maintain a functional pathway. This organisation can also simplify the co-regulation of the components by streamlining their regulatory pathways (Field & Osbourn, 2012).

Traditionally, genetic clustering was recognised to exist in eukaryotes but was considered rare in plants. However, a decade-old paper suggests this phenomenon is more common than previously assumed. In recent years, multiple examples of clusters involved in secondary metabolite synthesis have been identified (Field & Osbourn, 2012). Some examples include clusters involved in the synthesis of defence-related secondary metabolites in *Zea mays* (Frey et al., 1997), *Avena strigosa* (Qi et al., 2004; Qi et al., 2006; Mugford et al., 2009), *Oryza sativa* (Wilderman et al., 2004; Shimura et al., 2007; Swaminathan, et al., 2009; Wang et al., 2011), *Sorghum bicolor*, *Manihot esculenta* and *Lotus japonicus* (Takos et al., 2011) and last but not least, in *A. thaliana*, with the synthesis of thalianol and marenal triterpenes (Field & Osbourn, 2012). Development-related Clusters have also been identified in *A. thaliana* (Reimegård et al., 2017).

The two examples of genetic clusters already described in *A. thaliana* are the thalianol and marenal triterpene clusters. They can provide insight into how such structures develop. They are thought to have first appeared in the last common ancestor of *A. thaliana* and *Arabidopsis lyrata* and originated from an ancestral duplication event of a gene pair, followed by subsequent specialisation and the recruitment of additional genes. Their shared origin suggests they may possess features that facilitated their clustering. Interestingly, phylogenetic analyses indicate that both thalianol and marenal triterpene clusters were located very close to the centromere in their ancestor. Similarly, the aforementioned *Z. mays*, *S. bicolor* and *A. strigosa* clusters have been identified in sub-telomeric regions. Both chromosomal regions are known to be poor in genes and rich in transposing elements (TEs). For example, sub-telomeres are hotspots for chromosomal recombination and segmental duplications (SDs). The two *Arabidopsis* species diverged approximately 13 million years ago, during which time the two clusters underwent relocation within the *A. thaliana* karyotype, moving away from the centromere. Still, they have been shown to accept and retain SDs, particularly in the form of helitrons, which can trans-duplicate genes. This suggests that these regions remain highly dynamic (Field & Osbourn, 2012).

It has been proposed that TEs may have played a role in cluster assembly by promoting ectopic recombination, indicating that enhanced recombinational activity may be crucial for cluster formation. The unusually high turnover of SDs may increase the chances of favourable genetic combinations arising. Furthermore, eukaryotes typically suppress meiotic recombination at centromeres and many telomeres. This feature would protect against the breakup of functional clusters, effectively linking the genes strongly together and fixing the clusters in genetically heterogeneous populations. In other words, genetic clusters may arise in chromosomal regions that are highly dynamic whilst also being protected from meiotic recombination, thus increasing the cluster's internal genetic linkage once it is formed (Field & Osbourn, 2012).

As the *A. thaliana* is a predominantly self-pollinating species, the outcrossing events are rare, meaning that genetic combinations are maintained through generations by selfing (Brock et al., 2007). Highly selfing plants are almost entirely homozygous, meaning that meiotic recombination events are unlikely to "break" clusters, independently of their loci. This could explain why the thalianol and marenal triterpene clusters were able to be relocated within the karyotype into areas subject to recombination events and still be retained after millions of years. This implies that, within selfing plant species, selective advantage plays a lesser role than co-regulation in maintaining the clusters within the genome.

## 1.3 Cluster 8 is Involved in the Pollen Development via Cell Wall Pectin Fine-Tuning

### 1.3.1 Previous Research

Previous research by Reimegård et al. (2017) on *A. thaliana* found that stamen development is linked to the co-expression of co-located and functionally related genes. They identified seventeen clusters spread across multiple chromosomes. Interestingly, most of these genes shared neither homology nor significant promoter element similarity, suggesting that their transcriptional activation is associated with local chromatin decondensation rather than common regulatory elements (Reimegård et al., 2017). This finding is consistent with observations in *Avena strigosa*, where the avenacin cluster also appears to be linked to chromatin decondensation (Wegel et al., 2009), as well as with the marenal and thalianol cluster genes, which are regulated by chromatin remodelling factors (Field & Osbourn, 2012).

These seventeen clusters act downstream of *MALE STERILITY 1* (*MS1*). The *MS1* gene encodes a PHD-finger transcription factor (TF) in the tapetum and mediates the expression of sporophytic PCPs and the tapetum's PCD. Its mutants

have abnormal exine bacula, lack vesicles associated with intine formation, have no coat and are male sterile. Meaning that *MS1* is likely involved in pollen wall and coat formation, and pollen viability. MS1 acts downstream of MYB DOMAIN PROTEIN 80 (MYB80)/MYB103/MALE STERILE 188 (MS188). The *MYB80/MYB103/MS188* gene is a key regulator of sexine formation; its mutants have no sexine skeleton or pollen coat (Lu et al., 2020).

Among Reimegård's (2017) clusters, Cluster 8 was an agreeable candidate for a functional study as it has a minimal size and a satisfactory amount of information is available about its individual genes. It contains three genes of interest (GOIs): AT2G47030, AT2G47040, and AT2G47050. The first gene, AT2G47030, is known as *PECTIN METHYLESTERASE 4* or *VANGUARD HOMOLOG 1* (*PME4/VGDH1*); the second, AT2G47040, as *PME5* or *VANGUARD1* (*PME5/VGD1*); and the third, AT2G47050, as *PME INHIBITOR-LIKE* (*PMEIL*), Figure 2. Notably, PME4/VGDH1 and PME5/VGD1 are paralogous, sharing 85% protein identity and 92% similarity. Nearly two decades ago, Jiang et al. (2005) too noted the co-positioning of these functionally related genes, all of which have PME domains.


*Figure 2: Graphical representation of the three genes of interest within Cluster 8, as Reimegård et al. (2017) identified via bioinformatical analysis. The chromosomal region of the cluster is indicated by position number times thousand (k = x1000). The genes represented by faint grey arrows are the flanking genes, those in blue are paralogous to one another and the orange one has no paralogues within this cluster. Adapted from Reimegård et al. (2017), supplementary material.*

### 1.3.2 Background to the Functions of the Genes of Interest

The plant cell wall (CW) provides structural integrity and modulates growth. It is composed of cellulose, hemicellulose, proteins and pectins (homogalacturonan (HG) and rhamnogalacturonan). Pectins are synthesised in the Golgi apparatus, with HGs being methyl-esterified by pectin methyltransferases and subsequently modified by PMEs in the CW. This demethylesterification creates carboxyl groups that complex with $Ca^{2+}$ ions, increasing CW firmness. PMEs also prepare pectins for degradation by polygalacturonases, contributing to both the firming and softening of the CW. By regulating the mechanical and chemical properties of the CW, PMEs indirectly contribute to the plant's physiological and developmental processes. Higher plants possess different isoforms of these enzymes, encoded by multiple gene families. Acidic PMEs are thought to perform random demethylesterification, while basic PMEs carry out linear demethylesterification (Jiang et al., 2005; Louvet et al., 2006).

Male fertility is strongly affected by the pollen grain and tube's CW composition. The pollen CW is made up of two layers, an inner sheet containing callose and cellulose, and an outer layer made up primarily out of pectins. The most

abundant form of pectin is HG, which is highly methyl-esterified and can be acted upon by PMEs. This process is tightly regulated by PME inhibitors (PMEIs), making this a highly dynamic process. Their balance is crucial for the adequate fine-tuning of the rigidity and plasticity of the pollen's CW (Wang et al., 2020).

### 1.3.3 PME5/VGD1 and PME4/VGDH1

The *PME5/VGD1* gene, AT2G47040, encodes a 595 amino acid (aa) protein homologous to PMEs. It contains three distinct domains: a secretion-related transmembrane domain, a PMEI-homologous domain, and a pectinesterase-homologous domain (Jiang et al., 2005). *PME5/VGD1* was found to be stably expressed in desiccated mature pollen grains, hydrated pollen grains and pollen tubes (Wang et al., 2008). Functional studies have shown that disrupting *PME5/VGD1* significantly reduces PME activity in these areas, leading to abnormal pollen tube extension. Pollen lacking *PME5/VGD1* can germinate on the stigma but exhibits delayed extension into the style and transmitting tract. In vitro germination results in burst pollen tubes, severely affecting male fertility (Jiang et al., 2005). A different study found that *PME5/VGD1* is also expressed throughout the silique developmental process, though at lower levels. This suggests that its role may extend beyond pollen tube growth, potentially affecting silique and seed development (Louvet et al., 2006).

Several mechanisms of action have been proposed for PME5/VGD1. First, the basic isoelectric point of PME5/VGD1 suggests it facilitates linear demethylesterification of HGs, promoting the formation of a $Ca^{2+}$ pectate gel that stiffens the pollen tube wall and aids in cell attachment. Loss of *PME5/VGD1* would hinder gel formation, weakening the pollen tube wall and potentially causing bursting during in vitro growth. However, interaction with the stigma may have stabilising effects on the pollen tube, explaining why bursting was not observed in on-plant germination. Second, PME5/VGD1 may contribute to the degradation of papillary, stigmatic, and stylar CWs, further facilitating pollen tube invasion. This protein might be secreted by the tube into female floral tissues, which have environments favouring random demethylesterification. Indeed, this enzyme has the ability to act randomly or linearly depending on the surrounding conditions. Random demethylesterification of HGs leads to the release of protons, promoting the action of endopolygalacturonases, likely contributing to the degradation of the stylar cell walls. A mutant lacking *PME5/VGD1* would exhibit weaker PME activity, delaying the invasion of female reproductive tissues. Finally, the delay in extension could be due to ineffective interactions of the tubes with the ECM in the style and transmitting tract. Linear demethylesterification by PME, under suitable conditions within the transmitting tract, could produce Ca2+ pectate gel that supports tube-ECM interactions necessary for adequate extension. Loss of

*PME5/VGD1* function would hinder this interaction, delaying pollen tube extension (Jiang et al., 2005).

To summarise, *PME5/VGD1* is expected to regulate the mechanical strength and rigidity of the pollen tube wall, help degrade the stylar CWs, and promote the interaction between the pollen tube and the stylar ECM. Overall, it is crucial for efficient sperm nuclei delivery to egg cells by enabling pollen tube extension and interaction with female reproductive tissues (Jiang et al., 2005). Additionally, *PME5/VGD1* seems to be involved in the development of siliques (Louvet et al., 2006).

As mentioned, *PME5/VGD1* and *PME4/VGDH1* (AT2G47030) are paralogous; not only do they have 85% aa sequence identity, but they also exhibit the same expression pattern and function, as observed from a complementation assay (Jiang et al., 2005). Still, as the *PME5/VGD1*'s expression is 21 times higher than that of the *PME4/VGDH1*, and knock out (KO) of *PME5/VGD1* alone does not completely prevent pollen tube extension due to functional redundancy provided by *PME4/VGDH1*, the former is expected to be the main contributor to normal pollen tube extension beyond the stigma (Jiang et al., 2005). On a different note, PME4/VGDH1 was identified to be secreted by two-week-old etiolated seedlings germinated and grown in liquid culture (Charmont et al., 2005), suggesting a stress response-related function.

### 1.3.4 PMEIL

The last gene product, PMEIL, is distinct from the others in Cluster 8. It is only 216 aa long, shares just 27% identity with the other gene products, and contains domain, PMEI, suggesting a unique function (Jiang et al., 2005), later confirmed (Wang et al., 2020).

*PMEIL* was found to be stably expressed in the desiccated mature pollen grains, hydrated pollen grains and pollen tubes (Wang et al., 2008; Wang et al., 2020). KO studies revealed it regulates pollen viability and tube growth, with significant implications for male fertility. Female fertility was mostly, but not entirely, unaffected, suggesting it may have roles in floral differentiation and ovary formation (Wang et al., 2020).

A study investigating the effects of Abscisic Acid (ABA) treatment on flowering *A. thaliana* found that *PMEIL* directly affects pollen fertility. ABA is a phytohormone crucial for preparing plants to respond to stress by altering global gene expression patterns, leading to significant phenotypic changes. One of its effects, when applied to flowering plants, is a strong reduction in pollen fertility. This decrease was traced to a member of the cytosolic ABA receptor kinase (CARK) gene family, *CARK3/PTI1-4*, an ABA-responsive gene. *PMEIL* is located 244 bp downstream of this gene, both are *in cis* on the antisense DNA strand. This genetic organisation has made it so that *CARK3/PTI1-4* was shown to function as

*PMEIL*'s promoter. ABA treatment of flowering plants leads to the transcription of *CARK3/PTI1-4*, causing transcriptional interference (TI) of *PMEIL*. The TI phenomenon can occur when genes are arranged in tandem, and it involves the RNA polymerase II (Pol II) complex at the upstream gene intruding onto the promoter of the downstream gene, suppressing its transcription by interrupting interactions with TFs. A reliable method to study the impact of TI is the use of T-DNA insertion mutants. These types of mutations act as insulators between genes and block read-throughs, restoring the expression of the downstream genes. In summary, the transcription of *CARK3/PTI1-4* upon ABA treatment leads to pollen sterility. The study revealed a connection between ABA-mediated stress responses and decreased pollen fertility as a strategy to cope with stresses during flowering (Wang et al., 2020).

### 1.3.5 Other PMEs the Genes of Interest have Homology to

Homologues of PME5/VGD1 and PME4/VGDH1 involved in pollen-based male fertility in *A. thaliana* include PME48 (AT5G07410.1), PPME1 (AT1G69940.1), and PMEI2 (AT3G17220.1). These PMEs regulate the level of homogalacturonan (HG) methylesterification during pollen germination and pollen tube growth (Wang et al., 2020), suggesting they may provide redundancy in the Cluster 8 triple KO mutant. Notably, the absence of *PME48* results in delayed pollen tube germination (TAIR, 2024), a phenotype similar to that associated with the*PME5/VGD1* KO (Jiang et al., 2005).

AT3G62170, also known as *PMEI-PME37* or *VANGUARD 1 HOMOLOG 2* (*VGDH2, PMEI-PME37/VGDH2*), has the highest homology to these two genes. Its protein sequence scored 70% identity to *PME5/VGD1* and 68% identity to *PME4/VGDH1*'s gene products. *PMEI-PME37/VGDH2* is specifically expressed in the pollen grain and later in the pollen tube. However, attempts to complement mutations did not restore the WT phenotype, highlighting that amino acid sequence homology and expression patterns do not necessarily predict functional interchangeability (Jiang et al., 2005).

*PMEIL* (AT2G47050) appears to have only one homologue, AT3G62180. It is expressed during the embryo stage, flowering, and pollen development and primarily enables enzyme inhibitor activity (TAIR, 2024).

## 1.4 Cluster 3

Cluster 3, as identified by Reimegård et al. (2017), was also briefly looked at during this project. Querying the Tair and Phytozome online databases (TAIR, 2024 and Phytozome, 2024), results in limited outputs. It appears that these genes in general are less researched than others. The few keywords, functions and expression patterns are listed in Table 1.

*Table 1: Keywords associated with the genes of interest within Cluster 3 as identified by Reimegård et al. (2017) after querying the Tair and Phytozome online databases.*

| Gene | Protein Superfamily, Main motif, Description | InterPro DOMAIN/s | Expressed in: | Functions in: |
|---|---|---|---|---|
| AT1G20120 | GDSL-like Lipase / Acylhydrolase<br><br>GDSL-motif esterase / acyltransferase / lipase<br><br>Enzyme group with broad substrate specificity that may catalyse acyltransfer or hydrolase reactions with lipid and non-lipid substrates. | Lipase, GDSL, active site (IPR008265), Lipase, GDSL (IPR001087) | flowering stage, 4 anthesis, petal differentiation and expansion stage carpel, collective leaf structure guard cell, pollen, sepal, stamen, stem | lipid metabolic process, lipase activity, hydrolase activity, acting on ester bonds, carboxylesterase activity |
| AT1G20130 | | Pistil-specific extensin-like protein (IPR003882), Lipase, GDSL, active site (IPR008265), Lipase, GDSL (IPR001087) | | lipid metabolic process, structural constituent of cell wall, lipase activity, hydrolase activity, acting on ester bonds |
| AT1G20132 | | Lipase, GDSL, active site (IPR008265), Lipase, GDSL (IPR001087) | | lipid metabolic process, hydrolase activity, acting on ester bonds, lipase activity |
| AT1G20135 | | | flowering stage, 4 anthesis, leaf whorl, petal differentiation and expansion stage, carpel, collective leaf structure, flower, petal, stamen | lipid metabolic process, lipase activity, hydrolase activity, acting on ester bonds |
| AT1G20150 | Subtilisin-like serine endopeptidase | IProtease-associated PA (IPR003137), Proteinase inhibitor, propeptide (IPR009020), Peptidase S8/S53, subtilisin/kexin/sedolisin (IPR000209), Peptidase S8/S53, subtilisin, active site (IPR022398), Proteinase inhibitor I9, subtilisin propeptide (IPR010259), Peptidase S8, subtilisin-related (IPR015500) | F mature plant embryo stage, leaf whorl, petal differentiation expansion stage, plant embryo cotyledonary stage collective leaf structure, flower, root, seed, sepal, E expanded cotyledon stage | proteolysis, negative regulation of catalytic activity, identical protein binding, serine-type endopeptidase activity |

## 1.5 CRISPR/Cas9

Many biotechnological methods can be used to edit genomes or to generate genetically modified organisms (GMOs). Some of the more popular tools are Zinc-Finger Proteins (ZFPs), Transcription Activator Like Effector Nucleases (TALENs), and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) - associated (CRISPR-Cas) endonucleases. These tools share programmable DNA-binding domains, which grant them specificity for the chromosomal locus of interest. However, ZFPs and TALENs can be quite inaccessible, as they require the researchers to engineer a brand new target-specific endonuclease for each application. In contrast, the CRISPR-Cas system is unique because it only requires an RNA molecule's design. This makes the process more accessible and less resource-intensive and provides scientists with greater flexibility (Shamshirgaran et al., 2022). For this reasons, CRISPR/Cas9 was chosen as the gene editing tool in this project.

The CRISPR-Cas system evolved as a form of adaptive immunity by allowing the detection and degradation of invasive DNA from bacteriophages and plasmids. In short, when a virus attacks a bacterium, a short sequence of the invading DNA is integrated as a unique "spacer" into the bacterium's genome within the CRISPR locus. Upon a subsequent encounter with a similar virus, this spacer is transcribed and processed into a small CRISPR RNA (crRNA), which binds to a trans-activating CRISPR RNA (tracrRNA). Together, they bind the Cas protein and guide it towards the target DNA. Once the target is identified, the Cas endonuclease cleaves and invalidates the invading genetic material. This defence system, in its various forms, is common in prokaryotes. The classification of the CRISPR-Cas system is based on the properties of the Cas protein. Class-1 Cas complexes are made up of multiple small Cas genes, each with its distinct function; Class-2 Cas proteins are, on the other hand, large multifunctional Cas proteins (Jinek et al., 2012; Shamshirgaran et al., 2022).

Among the most commonly used endonucleases for genome editing is *Streptococcus pyogenes* Cas9 (SpCas9 or Cas9). It is a Class-2 endonuclease, particularly suitable for use in artificial genetic editing. For instance, it can interact with the protospacer-adjacent motif (PAM), bind and unwind DNA, pair with the crRNA, target the right sequence and still fulfil its endonuclease activity (Jinek et al., 2012; Shamshirgaran et al., 2022). However, these characteristics also impose certain limitations. Specifically, Cas9 can only bind DNA sequences with the 5'-N20-NGG-3' structure (where "N" stands for any base, "N20" indicates the gRNA spacer sequence, and "NGG" refers to the PAM). Although DNA is generally rich in appropriate PAM sequences, the necessity for the PAM to be close to the target sequence can limit the number of optimal targeting sites (Xie et al., 2015).

As the CRISPR-Cas9 tool continues to grow in popularity, it also undergoes continuous improvements. One significant advancement involves the optimisation

of the gRNA component. GuideRNAs found in bacteria are made up of two RNA molecules, a target-specific spacer (crRNA) and a conserved 76 nts Cas9-binding scaffold sequence with a specific stem-and-loop structure (Trans-activating crRNA, tracrRNA). These two molecules were later artificially merged into a single gRNA (sgRNA) to streamline their use and simplify the system. In this configuration, the expressed Cas9 is associated with the sgRNA and scans the DNA for the appropriate PAM sequence. Upon finding a match, Cas9 compares the spacer sequence within the sgRNA with the upstream sequence of PAM. When the target sequence is recognised, Cas9 introduces a double-strand break (DSB) ~3 bp upstream of the PAM, effectively knocking out the gene (Jinek et al., 2012; Shamshirgaran et al., 2022).

The gene KO process exploits endogenous DNA-repair mechanisms. The DSB is repaired either via the error-free homologous recombination mechanism or the error-prone nonhomologous end-joining (NHEJ) pathway. The latter is more prevalent but often results in an insertion-deletion (indel) or substitutions at the repair site, disrupting the target gene's coding sequence and effectively knocking it out. If the initial repair is accurate and does not introduce an indel, the Cas9 endonuclease will continue to target and cleave the sequence until an indel is introduced, ensuring the gene is eventually KOd (Jinek et al., 2012; Shamshirgaran et al., 2022). The introduction of long indels and chromosomal deletions renders the mutants easily detectable via simple PCR assays (Hui et al., 2019). Such generated mutations are genetically stable, meaning they are transmitted to the next generation (Hui et al., 2019).

## 1.6 Need for a Multiplex CRISPR/Cas9 system

The field of functional genomics aims to identify the roles of specific genes by observing the changes in an organism’s phenotype when those genes are disrupted. The most common method to nullify the effects of genes on phenotypes is by knocking them out (Bouché & Bouchez, 2001). However, this process can become more complicated in complex organisms that possess genetic redundancies. These redundancies are hypothesised to be an evolutionary safeguard against the negative outcomes derived from deleterious mutations (Zhang, 2012). In such cases, should a gene be lost, compensatory mechanisms can take over via functionally redundant genes. Consequently, to an external observer, the DNA mutation introduced may appear to have no effect on the phenotype (Peng, 2019). This can make it challenging to ascertain the function of the specific Gene Of Interest (GOI, Xie et al., 2015).

When studying the effects of multiple mutations, the standard practice is to introduce single mutations in individual organisms and then cross them to stack them in one individual (Jones, 2015). However, this method is unlikely to work

when the loci of interest are strongly genetically linked, a condition known as linkage disequilibrium, which is often the case with clustered genes. Such technical constraints have long hindered the understanding of the mechanisms regulating genetic clusters (Xie et al., 2015). Many methods have been developed to overcome this issue. One of the simplest approaches is to combine the Cas endonuclease with multiple sgRNAs. However, both the endonuclease and the multiple sgRNA-expressing cassettes require their own functional promoters and terminators, making the construct size difficult to manage. With each sgRNA cassette averaging between 400-500 base pairs, fitting multiple such cassettes into a single plasmid vector quickly becomes limited by the plasmid's capacity (Xie et al., 2015).

## 1.7 Polycistronic sgRNA Genes

Being able to compact multiple functional sgRNAs into a single synthetic gene to save space is a sought-after feature. One promising approach, developed by Xie et al. (2015), was inspired by a plant-specific dicistronic gene that simultaneously codes for a $tRNA^{Gly}$ and a small nucleolar RNA (snoRNA). Both RNAs are transcribed as a single unit in the form of a primary tRNA-snoRNA (tsnoRNA) precursor. The plant's endogenous post-transcriptional processing mechanisms then liberate both molecules, activating them (Kruszka et al., 2003). The ribonucleases responsible for this processing are RNase P, which cleaves off the six superfluous nts on the 5' side of the pre-tRNA, and RNase Z, which cleaves the link between the two RNAs. Both RNases recognise the secondary structure of the pre-tRNA, rather than the tsnoRNA itself (Kruszka et al., 2003; Xie et al., 2015). This implies that they could be taken advantage of for the targeting of other sequences containing pre-tRNA secondary structures.

Xie et al. (2015) designed a gene editing method that hijacks this endogenous system. Their construct contains a Cas9 cassette followed by a single polycistronic gene coding for multiple sgRNAs, each preceded by a pre-tRNA. This repetition is the foundation upon which the multiplexity of this method stands. They proposed that these tandem repeat conformations would mimic the tsnoRNA precursors and recruit the same RNases. Upon excision of mature tRNA molecules, several functional sgRNA molecules would be released as a side effect, ready to couple with the Cas9 protein, Figure 3. The research group found success in using paired sgRNAs to target single genes, leading to the deletion of short chromosomal fragments between two cleavage sites. They found optimal outcomes with distances ranging from 350 to 750 base pairs, with higher efficiency when the sites are closer (Xie et al., 2015). Woo et al. (2015) was able to introduce a 223 bp deletion between target sites that were originally separated by 201 bps.

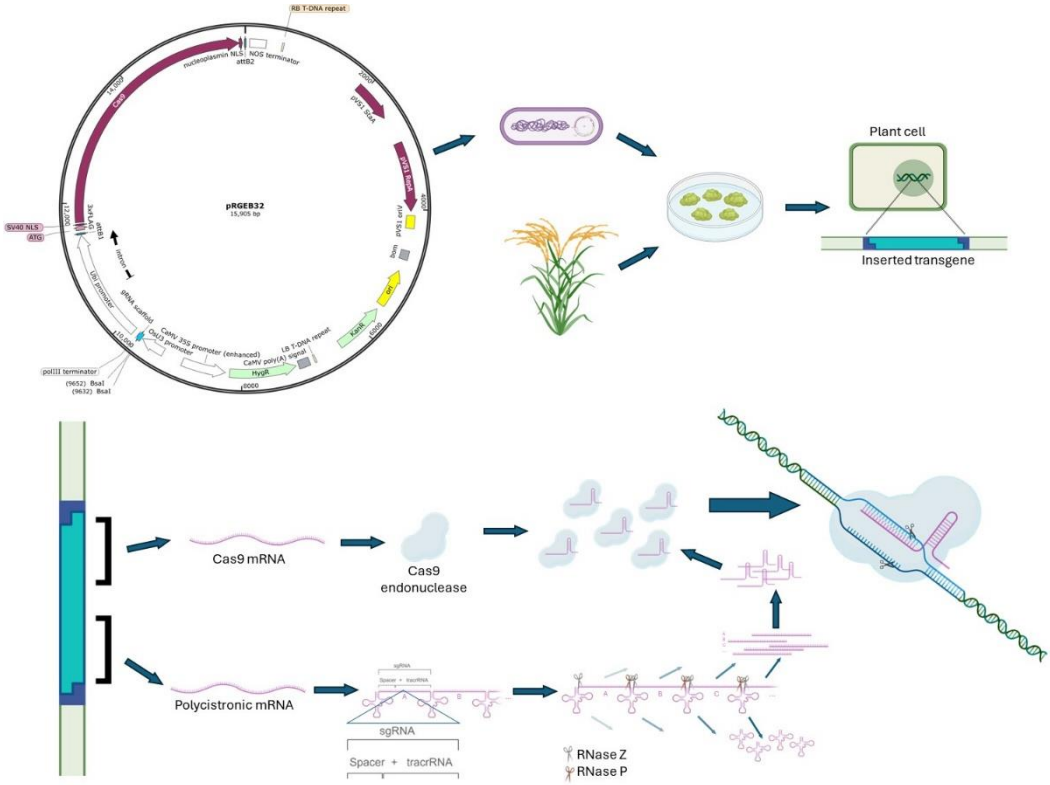



*Figure 3: Simplified schematic representation of the mode of action of the gene editing system designed by Xie et al. (2015). Transgenic rice plants were generated via Agrobacterium-mediated transformation of mature seed-derived calli. The destination vector they used was pRGEB32 (by SnapGene). The delivered genetic material (Cas9 and polycistronic sgRNA genes) integrates randomly into the plant's genome. The synthetic polycistronic gene is made up of tandemly arrayed pre-tRNA-gRNA units, with each gRNA containing a target-specific spacer (identified with plum-coloured capital letters) and a conserved gRNA scaffold. Upon transcription, the Cas9 mRNA is translated into a functional protein, whereas the polycistronic sgRNA gene's pre-tRNA portions assume the expected clover-like secondary structures. These secondary structures are readily recognised by the RNases P and Z (represented by scissors), which then cleave the tRNAs off and liberate the various sgRNAs. The scaffold portion of the sgRNAs then folds accordingly and is recognised by the Cas9 protein. Upon their coupling, the complexes are ready to scan the DNA for the introduction of appropriate double-strand breaks.*

This novel method has several interesting features. Firstly, this post-transcriptional processing is a plant endogenous system that is nearly universally conserved, making gene editing methods based on this process feasibly applicable to other plant species. Secondly, multiplexing the sgRNAs overcomes the need for some polymerases to start transcription at a specific nt, which would otherwise limit the targetable loci on the DNA. Thirdly, fitting multiple sgRNAs into one cassette, as opposed to each sgRNA having its own, makes it so that the sgRNAs are expressed simultaneously and at the same level, which would be otherwise hard to achieve (Xie et al., 2015). Another advantageous feature is the rather unusual presence of internal transcription-promoting boxes, box A and box B, within

tRNAs. These elements, known as type II promoters, are recognised by the Pol III complex, boosting transcription (Dieci et al., 2007).

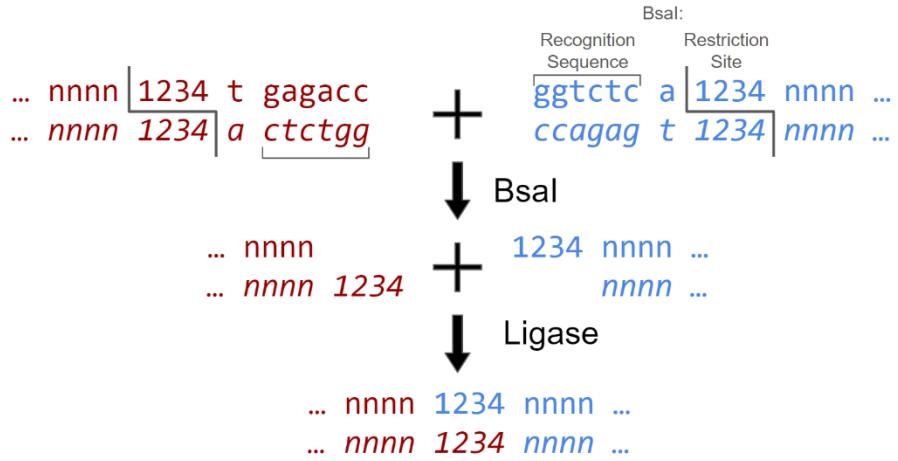
This method was shown to produce the desired sgRNA molecules at greater numbers than through conventional means. The spacer sequences had no additional nts except for the last sgRNA in the row, which could have a tail of one to seven additional uracil nts on the 3' side of the gRNA scaffold. These tails were reported to make no difference in the efficiency of locus targeting. Multiplex genome editing was successful, introducing indels via NHEJ and even chromosomal fragment deletions. This system was designed to be flexible, scalable, and compatible with the Golden Gate assembly of individual PCR components (Xie et al., 2015).

## 1.8 Golden Gate Cloning

Despite their utility in manipulating DNA, DNA recombination systems have some limitations due to the requirement for special sequences that must accompany both the GOI and the final destination vector. These special sequences persist after the recombination event, potentially interfering with the desired construct and limiting the number of compatible plasmid vector options. Consequently, recombination-based cloning techniques are less flexible compared to classical cloning technologies based on restriction enzymes and DNA ligases (Kotera & Nagai, 2008). Still, the traditional creation of a DNA library via cloning requires many successive steps (Engler et al., 2009).

In contrast, Golden Gate (GG) cloning offers a streamlined approach that reduces the process to two main steps: preparing intermediate constructs and assembling the final vector construct (Engler et al., 2009). GG was first developed in 1996 when it was shown that it is possible to introduce multiple inserts into a vector backbone (Lee et al., 1996; Padgett & Sorge, 1996). GG cloning leverages the unique properties of type II restriction enzymes (TII REs), which cut double-stranded DNA (dsDNA) outside of their recognition site (RS). This feature is exploited by designing DNA sequences flanked by the RS of the preferred TII RE, allowing the digestion of fragments to remove the site and generate unique short overhangs. The digested DNA fragments can then align and be ligated by a T4 DNA ligase. A benefit of this strategy is that only the desired products will accumulate, as unwanted ligation products will be immediately re-digested. Consequently, when two DNA sequences are designed in a manner to be digested into having complementary sticky ends, they can be ligated in a scar-free manner. A common type of TII RE used is the BsaI which recognises the non-palindromic sequence GGTCTC(N) and asymmetrically restricts both strands of the DNA. More specifically, it cuts the sense strand right after the N nt and four nts downstream on the antisense strand, Figure 4. The resulting four nt long sticky ends can have 256 possible combinations (Engler et al., 2008; Engler et al., 2009).

a.




b.

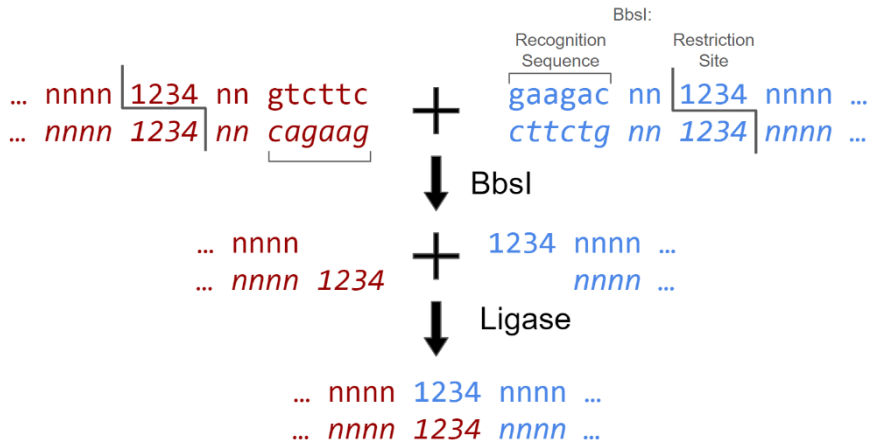



*Figure 4: Schematic representation of the concepts behind Golden Gate Assembly. a. Two DNA ends terminated by the same four nucleotides (here represented by numbers 1234, complementary nucleotides noted in italics) flanked by a BsaI recognition sequence, form two complementary DNA overhangs after digestion by the BsaI restriction enzyme. The two DNA fragments can then align and be securely ligated by the T4 DNA Ligase. Adapted from Engler & Marillonnet, (2011); b. The same principle stands for the BbsI restriction enzyme.*

An example of successfully applied GG cloning was described by Engler et al. (2009). Their goal was to linearly assemble nine distinct DNA fragments and insert this product into an acceptor vector, in only one step and only one tube. The protocol proved to be highly efficient, with over 90% of recombinant clones containing the desired construct. Their high success rate led them to suggest that even more fragments could be ligated just as easily. They suggested the observed high efficiency to have been multifactorial. First, by avoiding a PCR amplification step, they reduced the chances of single nucleotide polymorphisms (SNPs) arising, which can take place during PCR. Second, the one-step GG strategy minimises DNA handling, such as DNA purification, thus preventing some DNA damage. Lastly, they recognised the need for a higher degree of complexity in the design of

overhangs when using the one-step one-tube protocol to prevent them from creating unwanted pairs. When using the BsaI RE, the overhang conformations that should be avoided are the possible 16 palindromes and partial cross complementarities between overhangs that are not intended to match. Specifically, four nt long overhangs which accidentally share three compatible nts could bind aberrantly and introduce SNPs. Still, a large variety of compatible overhangs can be easily designed (Engler et al., 2009).

High efficiency was also reported by Kotera and Nagai (2008) for the cloning of multiple DNA fragments directly from crude PCR solution. By introducing TII REs and DNA ligases into the solution, they could recombine fragments in a single step. Efficiency was further increased by inhibiting DNA polymerase with aphidicolin and selectively digesting hemi-methylated template DNA with DpnI. This protocol was described to be simple and efficient, taking about 15 minutes to complete (Kotera & Nagai, 2008).

Overall, GG cloning facilitates the efficient and precise assembly of multiple DNA fragments. The method's high efficiency can be attributed to careful design considerations, minimal DNA handling, and leveraging the unique properties of TII REs. This makes GG cloning a powerful tool for generating recombinant gene libraries and other complex genetic constructs in a streamlined and reliable manner.

## 1.9 Aims

This project builds on the work of Xie et al. (2015). It aims to design a Multiplex CRISPR/Cas9 system for knocking out Cluster 8 to explore its function and role in pollen development. This study involved creating a synthetic dicistronic sgRNA gene (*SDRG*) that could simultaneously produce two unique sgRNAs, following the principles of Golden Gate Assembly.

This cluster was an appropriate candidate to study as it is small, with only three genes, the two paralogous genes have a high aa sequence identity score, meaning they could be targeted by the same sgRNA, and the KO phenotypes of all three genes are well established. As a follow-up to this project, mutant *A. thaliana* plants can be generated, and pollen-related phenotypes may be observed after careful breeding.

Overall the technology developed in herein has the potential to be used to better understand the sexual development of flowering plants. This increased knowledge could help the plant science and crop breeding communities to develop new crop varieties that are better adapted to meet agricultural and economic demands.

# 2. Materials and Methods

## 2.1 Plant Material and BASTA Selection

The seeds from *wild-type* (WT) *Arabidopsis thaliana* ecotype Columbia-0 (Col-0) and transgenic plants in Col-0 background (T1 and T2) were incubated at 37°C for a few days, then at -20°C for another few days and finally kept at room temperature (RT). The WTs were germinated directly in soil and perlite mixture (3:1, v/v) in a pot. At the full rosette stage, they were exposed to long-day conditions (16h light at 22°C and 8h darkness at 20°C). Post-flowering, they were ultimately used for floral dip transformation with *Agrobacterium tumefaciens*.

The transformed T1 seeds (0.5 ml), and the following T2 seeds (0.3 ml), on the other hand, were first sterilised with 25% bleach (+0.02% Triton X-100) and 70% ethanol and rinsed four times with autoclaved water. They were then poured onto 12x12 cm Petri dishes with half-strength Murashige and Skoog medium (1/2MS) containing a basal salt mix, no vitamins, 1% (w/v) sucrose and BASTA (Phosphinotricin, an analogue of Glufosinate Ammonium, 20 mg/l, by Duchefa Biochemie, Haarlem, Netherlands, Cat #P0159) for selection. The dishes were sealed with microporous tape and exposed to two days of vernalisation at 4°C in darkness. Next, they were transferred into a growth chamber with the following light conditions: 22°C/20°C, 16h light/8h dark daylight cycle, 110 uE. Over two-hundred survivors of the preliminary BASTA selection were transferred into a soil-perlite mixture in pots after two weeks, placed in a different growth chamber with long-day conditions, and grown in high humidity. After the seedlings were well established in the soil, they were sprayed with the BASTA herbicide (20 mg/l) about three times over a few weeks to eliminate false positives; which halved the number of plants.

## 2.2 Gene and Protein Sequence Analysis

### 2.2.1 Paralogous Genes Identity Scoring

The genes AT2G47030 and AT2G47040 were found to be paralogous by Reimegård et al. (2017). This factor was taken advantage of to construct one sgRNA to target both genes. Their identity was assessed using two softwares, BioEdit and GeneDoc. First, their genomic DNA (gDNA) sequences, including 5' and 3' UTRs and introns, were analysed and then compared with the equivalent coding DNA sequence (CDS). As CDSs are expected to be under stronger conservation pressures, compared to gDNA portions that do not end up being transcribed, the CDS-based identity is more relevant for the design on sgRNA spacers.

To begin with, both pairs of sequences were pairwise aligned and then analysed for identity scores using the BLOSUM 62 matrix score table. In BioEdit this was done using the tool option "optimal GLOBAL alignment" which also directly outputs the identity score. In GeneDoc, pairwise alignment was used with the default settings. The statistics report tool was then used to obtain the identity score.

### 2.2.2 Homologous Gene Scouting and Phylogenetic Tree Construction

The amino acid (aa) sequences of the three genes within Cluster 8 were obtained from the Phytozome online database (Phytozome, 2024). These sequences were used as queries within the BLASTP program, which this website offers, to identify other protein sequences with high identity scores. BLASTP scanned the *A. thaliana* proteome using the "compositional matrix adjust" method and the BLOSUM62 comparison matrix. As a rule of thumb, hits with the e-value closest to "0" were chosen as they represent sequences with the highest identity to the query sequence.

The identity and similarity values were further strengthened by checking them manually on Geneious. The aa sequences of the two paralogues were first pairwise aligned using the tool "Global alignment with free end gaps" with the Blosum62 cost matrix which had the gap open penalty set to 10 and the gap extension penalty set to 3. The aligned sequences were imported into GeneDoc, where the "Statistics Report" tool was used.

A small-scale protein family analysis was performed by building phylogenetic trees centred on the two paralogous genes from Cluster 8. Two distinct trees were constructed. The first was based on querying the PhyloGenes (2024) online database, limited to the organism *A. thaliana*, for gene *PME5/VGD1* (AT2G47040) and taking note of other genes in its most immediate branch, for a total of eighteen genes. This was not possible to do for *PMEIL* (AT2G47050) due to limited data. The second tree was constructed by assembling all the other homologues listed in this report, which were compiled together by researching literature.

The aa sequences of all the genes pictured in the trees were obtained from the Phytozome (2024) online database. The trees were constructed using Geneious software by first aligning the sequences of the given tree. The tool used to do so was “Global alignment” with the Blosum62 cost matrix which had the gap open penalty set to 10, the gap extension penalty set to 3, and the “build guide tree via alignment” option checked. The whole process was set to ten refinement iterations. The actual tree was built using the “Geneious Tree Builder” tool set to use the “Jukes-Cantor” genetic distance model, the “Neighbor-Joining” as the tree-building method, the “Bootstrap” as the resampling method and a random seed. The “Create Consensus Tree” option was checked and the support threshold was set to 50%. The number of replicates was set to a hundred. Outgroups were chosen to match the needs of the tree.

## 2.3 Construction of the Synthetic Dicistronic sgRNA Gene

### 2.3.1 Construction of the Synthetic Dicistronic sgRNA Gene Transcribed DNA Sequence

The synthetic dicistronic sgRNA gene (*SDRG*) was designed to contain two sgRNAs, named 30/40 and 50, each preceded by a pre-tRNA sequence. They were designed to target and KO genes AT2G47030, AT2G47040 and AT2G47050, located within Cluster 8 (Reimegård et al., 2017). The names of the sgRNA spacers and primers were based on the last two digits of the gene loci. Only one sgRNA was designed for genes AT2G47030 and AT2G47040 because they are paralogous, with their genomic DNA sequence having 89% identity to one another, therefore both are interchangeably susceptible to being targeted by the 30/40 sgRNA. Both also had identical PAMs located in the right spot in relation to the targeted sequence; Table 2.

*Table 2: Genomic DNA sequences of the three genes of interest and their respective sgRNA target sites. Each gene is titled with the organism of origin, the gene's locus, its chromosomal coordinates and its position relative to the conventional DNA reading directionality. The different areas are colour-coded, in order of appearance: 5'UTR, coding DNA sequence (CDS), **sgRNA target site**, relative PAM, intron, 3'UTR. Details obtained from Phytozome (2024).*

>A.thaliana TAIR10|AT2G47030|Chr2:19324232..19326397 reverse

ACTGGCCTAAGTCATTCAACAATTTGGCTTTAATCTAATAAAGCCTCTCCCTCTCCGATTAACCTTAATAAATCCGGAGGG
ATCGGAGAGAAAAAAGAACCAAAAAAAAAAAAAAAACAGAAAAAAAAAAATGATTGGAAAAGTTGTGGTCTCGGTGGCC
TCCATCCTCTTAATAGTCGGAGTTGCCATAGGAGTCGTTGCCTTCATAAACAAAAATGGCGATGCTAATCTGTCTCCCCAA
ATGAAAGCGGTTCAAGGAATTTGCCAGTCGACATCCGACAAAGCCTCATGTGTCAAAACTCTCGAGCCGGTCAAGAGCGA
GGACCCAAACAAGCTGATCAAGGCTTTCATGCTCGCTACAAAAGACGAATTAACCAAATCGTCTAACTTCACGGGTCAAA
CCGAAGTAAACATGGGCTCGAGCATCTCGCCAAACAACAAAGCCGTTCTTGATTACTGCAAGAGAGTTTTCATGTACGCT
CTTGAGGATCTCGCTACCATTATTGAAGAAATGGGTGAAGATCTTAGCCAGATCGGGAGCAAAATCGACCAGCTCAAACA
ATGGTTAATCGGTGTTTACAATTACCAAACTGATTGTCTTGACGATATCGAAGAAGATGATTTAAGAAAGGCGATTGGTG
AGGGCATTGCAAACTCCAAGATTCTCACTACCAACGCTATTGACATCTTCCACACTGTCGTTAGCGCCATGGCCAAGATTA
ACAACAAGGTTGACGACTTGAAGAACATGACAGGCGGAATCCCCACTCCCGGAGCCCCTCCTGTTGTTGATGAATCTCCC
GTGGCCGACCCAGATGGTCCTGCTCGTCGTCTTCTTGAAGACATTGACGAGACC**GGAATCCCAACATGGGTTTC**AGGTG
CTGACAGGAAGCTCATGGCTAAGGCTGGACGTGGCCGTAGAGGCGGCCGCGGAGGCGGTGCTAGGGTCAGAACGAACTT
TGTGGTGGCTAAGGATGGAAGCGGACAGTTTAAGACGGTTCAACAAGCCGTTGATGCTTGTCCTGAGAATAACCGAGGCC
GATGCATCATCTACATCAAGGCCGGTCTCTACAGAGAGCAAGTAATCATCCCTAAGAAGAAGAACAACATCTTCATGTTT
GGAGATGGTGCAAGAAAGACCGTCATTTCTTACAACAGAAGTGTTGCTCTCAGCCGTGGAACCACCACATCCCTCAGTGC
CACAGTTCGTAAGTCTCATCAAACCCTAATTCAGTAGTTAGGTTACACATATTTTTCTCTACATAATTCATGTTTTCGATAT
ATATATTGTGCAGAGGTTGAATCGGAAGGATTCATGGCGAAATGGATGGGATTCAAGAACACCGCCGGTCCAATGGGAC
ACCAAGCCGCGGCTATCCGAGTGAACGGAGACCGTGCCGTGATCTTCAACTGCAGATTCGACGGTTATCAAGACACATTG
TACGTTAACAACGGTCGTCAGTTCTACAGAAACTGTGTGGTCTCGGGAACAGTAGACTTCATCTTCGGCAAATCCGCAAC
CGTGATCCAAAACACACTAATTGTTGTCCGTAAAGGAAGCAAGGGACAATACAACACGGTCACAGCCGATGGAAACGAA
TTGGGCTTAGGGATGAAAATTGGTATCGTCCTCCAAAACTGCCGCATCGTGCCCGACAGGAAACTAACGCCAGAGAGACT
TACGGTGGCGACATACTTGGGAAGGCCGTGGAAGAAATTCTCGACCACCGTGATAATGAGCACCGAGATGGGAGATTTG
ATTAGACCAGAAGGGTGGAAGATTTGGGATGGGGAAAGTTTCCACAAGTCATGTAGGTACGTTGAGTACAACAACCGTG
GACCAGGAGCATTCGCTAACAGAAGAGTTAACTGGGCTAAGGTTGCCAGGTCTGCAGCTGAGGTCAATGGCTTCACAGCT
GCTAACTGGTTAGGTCCTATTAACTGGATTCAAGAAGCCAATGTTCCTGTCACAATTGGATTATAATTTATAAAGCGTATA
TGAAAATTATTTTAGTTTTTAAGTATATGAAAATTGGTGTATGAAATTGGTTTATGAAAACTTTTTAAGAAAGTCTATAATGT
TTTTCCTCTGTTAATTTGCTTCATCTAATTAAGGTCATATTGCAATGCAGACAAGTACTCATGTTGCTTTATTATAAGTTCC
TT

>A.thaliana TAIR10|AT2G47040|Chr2:19327895..19330197 reverse

ACTGGCCCAAGTCATTCAACAACTTGGCTTTTATCTAATAAAGCCCCTCCCTCTCCGATAAACCTTAAATAAAACCGGAG
GGAGAGCAAAAAATAAAAACAAAAATATACAAAAACAAAAACAAAAAATCATAAAAAATGATTGGAAAAGTTGTGGTCTC
GGTGGCCTCCATCCTCTTAATAGTTGGAGTTGCCATAGGAGTCGTTGCCTACATAAACAAAAATGGTGATGCTAATCTGT
CTCCCCAAATGAAAGCGGTCCGAGGAATTTGCGAGGCGACTTCCGACAAAGCCTCATGTGTCAAAACTCTCGAGCCGG
TCAAGAGCGATGACCCAAACAAGCTGATCAAGGCTTTCATGCTCGCTACCAGAGATGCTATAACCCAATCATCAAACTT
CACGGGTAAAACCGAAGGAAACTTGGGTTCGGGCATCTCGCCAAACAACAAAGCCGTTCTTGATTACTGCAAGAAAGTT
TTCATGTACGCGCTTGAGGATCTCAGCACCATTGTTGAGGAAATGGGTGAAGATCTTAACCAGATCGGGAGCAAAATCG
ACCAGCTTAAACAATGGTTAACCGGTGTTTACAATTACCAAACTGATTGTCTTGACGATATCGAAGAAGATGATTTAAGA
AAGACTATTGGAGAGGGCATTGCAAGCTCCAAGATTCTCACTAGCAACGCTATTGACATCTTCCACACTGTCGTTAGCG
CCATGGCCAAGCTTAACTTGAAGGTCGAGGACTTCAAGAACATGACAGGCGGAATCTTCGCTCCTTCAGACAAAGGAG
CAGCTCCCGTCAACAAAGGAACCCCTCCTGTTGCTGATGATTCTCCCGTGGCCGATCCAGATGGTCCTGCTCGCCGTC
TTCTTGAAGACATTGACGAGACC**GGAATCCCAACATGGGTTTC**AGGTGCTGACAGGAAGCTCATGACTAAGGCTGGAC
GTGGCAGTAATGACGGTGGTGCTAGGATCAGAGCAACCTTTGTGGTGGCTAAGGACGGAAGTGGACAGTTTAAGACTG
TTCAACAAGCCGTTAATGCTTGTCCTGAGAAAAACCCAGGCCGTTGCATCATCCACATCAAGGCTGGTATCTACAGAGA
GCAAGTGATCATCCCTAAGAAGAAGAACAACATCTTCATGTTCGGAGATGGTGCAAGAAAGACCGTCATTTCTTACAAC
AGAAGTGTTAAACTCAGCCCTGGAACCACCACTTCCCTTAGTGGCACAGTCCGTAAGTCTCACCAAACCCTAATTCACT
AGTTAGGTTAACAAATATTTTTCTCTACATATATAGCTCATGTTATATATTATTGCGCAGAGGTTGAGTCCGAAGGATTCA
TGGCAAAATGGATCGGATTCAAGAACACCGCCGGTCCAATGGGACACCAAGCTGTGGCTATCAGAGTGAACGGAGACC
GTGCCGTGATCTTCAACTGCAGGTTCGACGGTTACCAAGACACCTTGTACGTCAACAATGGTCGTCAGTTCTACAGAAA
CATTGTCGTGTCAGGAACAGTCGACTTCATCTTCGGCAAATCCGCAACCGTGATCCAAAACTCACTCATTGTTGTCCGT
AAAGGAAACAAGGGACAATTCAACACGGTCACAGCCGATGGAAACGAAAAGGGTCTAGCGATGAAAATCGGTATCGTC
CTCCAAAACTGTCGCATCGTTCCCGACAAGAAACTAGCGGCAGAGAGATTAATCGTGGAGTCATACTTGGGAAGGCCA
TGGAAGAAATTCTCTACCACCGTGATTATCAACTCCGAGATAGGAGATGTGATTAGACCAGAAGGTTGGAAGATTTGGG
ATGGGGAAAGTTTCCACAAGTCATGTAGGTACGTTGAGTACAACAACCGTGGACCAGGAGCTATCACTAACCGAAGAGT
TAACTGGGTTAAGATCGCCAGGTCTGCAGCTGAGGTCAATGACTTCACTGTTGCTAACTGGTTAGGTCCTATTAACTGG
ATTCAAGAAGCCAATGTTCCCGTCACGCTTGGATTATAAGGCGTATAAATCAAATCAAATCAAATCAAAAAAAAAAAAAA
GAGTAGTACGTGCTCATGTAAAAGGTGAAAATACAACGTCGTTTTCGGTAGGGTAATATGCGCGGGGAGATGAAGTTGT
AAGATTTGCTTTTTGTTTGTCTCACGTATGAAAGATGAGGAAAAAAAAAATATTGAAAATTTATGTTGTTTGATTTTTGTTT
CTATGTTTTGGGATATAATAATTTATTCGCTGTGTTGTACCTATGTTTAATAACCTTTTTAAAATATTATAATTTAACTAATT
TATTT

>A.thaliana TAIR10|AT2G47050|Chr2:19331684..19332646 reverse

AATCCATTCCTTTCCAATCCACGTCCAAAAAACAAAACCAAAAAAAAAAAAAAAAAAAAGACAAAAAAAAATGAACACTCCA
ATAAAACTCGCCTTTCTCATTCTCTGCATTGCCCTAACCGCAACCGCATTTATAGTCCCAGCAAAACGTGACGCCGTTTC
ACCACACCACCAAAAAGCCGTCGATGGAATCTGCAGCGTTGTGCAAGACAAACGTCTATGCAGCATTACCTTAAGAAAC
GTCCCAAGCGATGATCCCGCCGTTTTGGTTCGTTACTTAGCCACGGCAGCAGAAGCGTCCGTTAAAAAGGGTTTGAAG
TTCCTCTCCGGAATCAAACCAAAATACAAAGGAAACGCCTTCGCCACAACATGCATCACCGGCTGCGAGAAACAGCTAC
ACAACGCCTTGGACGACTTTGCAGATTTCTGGAAAGCCGCAGGAAAAAATCTAACGAGCATGGCTGAGAATTACTTCAC
ATGTAAGAAGAAGATGACTTCCATCTTCACTTACCATTCCACTTGTCTCGATGACATTTACGACAAGACGTTGCACAAAG
TGGTCGAAGGAGGGATTGGCCTTGGGAAAAGAATGAGTGGTGAGTCT**GTGGATGTGTTCGCTGGAAT**GGGCAAAGTC
TTTAACACTTTCAACATTAAGACCAAACTTAACCAGAAAGATACCGACACGTTGCTCCCACCACCTTTGTCCTTTTACTAC
TATTGATTTTATTTATCAATATATATTATATATCATAAAACACAACAAAATGTACTGCTTCGGTCGATGGTGATGAATAATT
TTCATCGATTGGGTTTTCAGATATTTTGGGGATCTTGTAACCAAAGGTTCCTTTTGATGTATGTTTGTATTTAATATGTTAT
GGATTATAATCATGTATGGACCTCGTGATTGGAAAGGATATGGAATAATGTGATTATCTAATAATAAAAACGTAATTAATA
TCG

The spacer sequences were designed on benchling.com and optimised for SpCas9, with the NGG PAM on the 3’ side. As this tool outputs a long list of possible spacers, only the ones on the sense strand, with as few repetitions as possible, and beginning with either a G or a C nt, were chosen. Target specificity was confirmed by BLAST of the *A. thaliana* genome (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

The transcribed DNA sequence (TDS) of the *SDRG* was assembled from three BsaI-digestible blunt-ended dsDNA sequences called Basic Modules (BMs), using the Golden Gate Assembly strategy. These BMs were designed to have four nt recombination sites, corresponding to the middle nts of the spacer sequence. The full annotated TDS can be seen in Supplementary Material (SM) 1 (SM1). Importantly, it ends with a poly-t terminator and is flanked by adapters containing BsaI RS, for the restriction-ligation step necessary to ligate it with the *A. thaliana U6-26* promoter.

*sgRNA Primer Design*

The sgRNA spacers were used to generate sgRNA primers with extender capacity. The forward primers included a few flanking bases to support cleavage by BsaI, the BsaI recognition site (RS), the 12 nucleotides on the 3’ side of the spacer (with the first four forming the unique overhang sequence), and an 18-nucleotide sequence compatible with the sgRNA scaffold. The reverse primers also included the same BsaI RS, the 12 nucleotides on the 5’ side of the spacer (reverse complemented), and a 17-nucleotide sequence compatible with the pre-tRNA, Figure 5. These primers were designed to take advantage of the BsaI RE's typology. By positioning its recognition site upstream of the spacer's middle four nucleotides on both sides, digestion with BsaI results in the generation of the desired unique overhangs, facilitating targeted ligation of the Basic Modules (BMs) further ahead. This method is essentially based on reverse-engineering the mode of action of the TII REs. Adapter primers were also designed to include the BsaI adapter sequence, followed by a portion complementary to either the pre-tRNA sequence or the sgRNA scaffold sequence. All primers were ordered from TAG Copenhagen A/S and are listed in SM2.

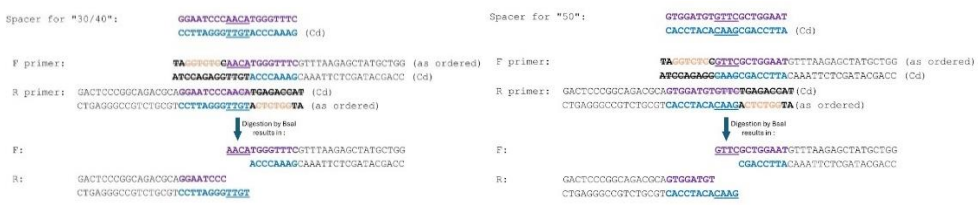



*Figure 5: Schematic representation of the sgRNA design process. It starts with identifying the gene-relative sgRNA spacer sequence (sense strand in plum and complementary strand in aqua), splitting it into two halves, each with a four-nucleotide long complementary overhang (underlined),*

*introducing BsaI recognition sequences (in orange) upstream of it in a way that the restriction of dsDNA sequences in the following GG step would result in the creation of these overhangs.*

*Overlap PCR and Basic Module Assembly*

The BMs were generated by first carrying out an overlap PCR to create a 164 nt dsDNA fragment made of the sgRNA scaffold sequence followed by the pre-tRNA sequence (both provided as oligonucleotides). The overlap PCR reaction was carried out in a volume of 50 µl and contained 0.5 µl (1U) of Phusion™ High-Fidelity DNA Polymerase (by Thermo Scientific, Cat# F530S and F530L), plus the other components and 1 µl each of the sgRNA scaffold and pre-tRNA oligonucleotides (10 µM working stock), both with overlapping 30 nt long sequences (listed in SM2). The PCR program was as follows: 98°C for 30 s, 10 cycles of 98°C for 10 s, 60°C for 10 s, and 72°C for 15 s, and a final extension at 72°C for 2 min.

The newly assembled two-component fragments were then PCR amplified. A total of 200 µl (4x 50 µl) of reaction was made up using the same Phusion Polymerase as above, 0.5 µl each of "F gRNA→tRNA" and "R gRNA←tRNA" primers (10 µM working stock, listed in SM2) and 1 µl of overlap PCR product. The PCR program was the same as above but consisted of 30 cycles. The PCR products were separated by electrophoresis on 1.5% agarose gel in 1xTAE buffer. The 164 bp gel bands were excised with a scalpel and eluted using the Zymoclean Gel DNA Recovery Kit (Cat #D4002, by Zymo Research), following the manufacturer's protocol.

The BMs were assembled by using the new 164 bp sgRNA scaffold - pre-tRNA DNA fragment as the template, and the sgRNA primers to extend and amplify it. These primers contain 17-18 nts long stretches complementing the template, and 21 nts long stretches containing parts of the spacer sequence and BsaI RS, as template extenders. In other words, the BMs were amplified by PCR with primers designed to add flanking BsaI sites on each side of the modules. Most importantly, the BsaI cleavage sites overlapped with the recombination sites. The first and third BMs additionally used adapter primers, one each. The PCR followed the same PCR amplification protocol as above.

BM1 was constructed by using primers AdF and 30/40R and for this reason, it was also called the "AdF-30/40R" fragment. The intermediate BM2 was amplified with the primer pair corresponding to the unique primer-coded fragment, "30/40F-50R". The last fragment, BM3, was amplified with the 50F-AdR primer pair. The detailed sequences of all BMs are listed in SM1. The amplicons were separated by electrophoresis and eluted.

A graphical representation of the workflow that was followed to assemble the BM2 fragment is depicted in Figure 6. The basic structures are as follows: BM 1: (5′) 2 restriction supporting nts::BsaI RS::1 restriction supporting nt::pre-tRNA::12 nts on the 5′ side of the 30a spacer::1 restriction supporting nt::BsaI RS::2

restriction supporting nts (3′). The structure of the intermediary BMs followed the same principle. As an example, the structure of BM 2: (5′) 2 restriction supporting nts::BsaI RS::1 restriction supporting nt::12 nts of the 3’ side of the 30a spacer::sgRNA scaffold::pre-tRNA::12 nts on the 5′ side of the 40a spacer::1 restriction supporting nt::BsaI RS::2 restriction supporting nts (3′). The only part that changes for the following intermediate fragments is the spacer portion. The last BM fragment structure: (5′) 2 restriction supporting nts::BsaI RS::1 restriction supporting nt::12 nts of the 3’ side of the 50a sgRNA scaffold::poly-t terminator::1 restriction supporting nt::BsaI RS::2 restriction supporting nts (3′).

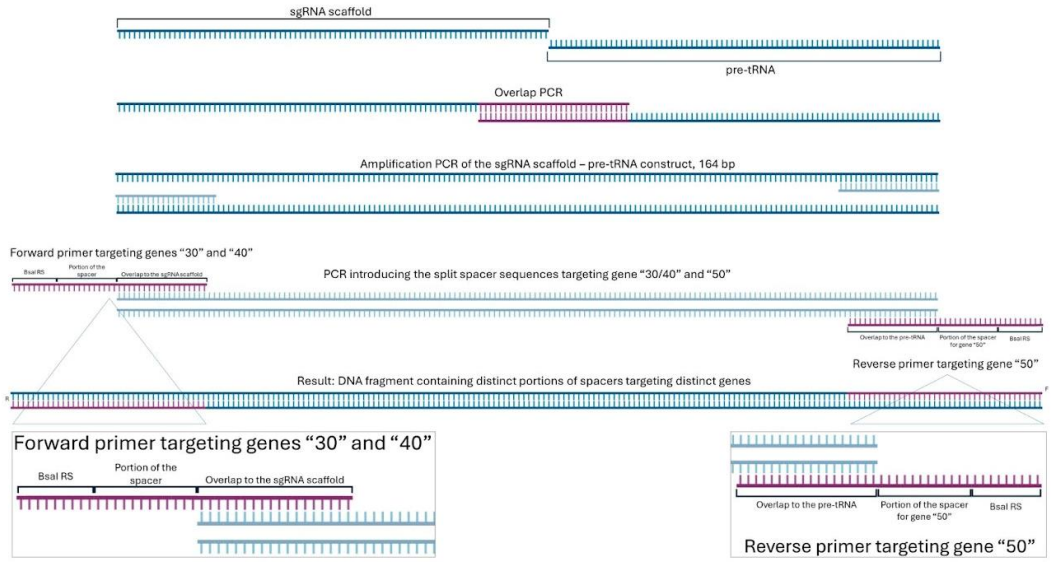



*Figure 6: Basic module assembly workflow: Graphical representation of the steps taken to assemble the middle basic module. The oligonucleotides are to scale relative to one another. The process begins with a simple overlap PCR to create a two-component 164 bp fragment consisting of the sgRNA scaffold and the pre-tRNA sequence. The constructed fragments were then PCR amplified using the sgRNA primers, thereby extending the sequence. Result: DNA fragments containing distinct portions of spacers targeting distinct genes. Figure generated on Biorender.com*

The propensity of the primers and BMs to create secondary structures, such as hairpins, or inappropriate annealing to self and others was assessed through the OligoCalc tool (http://biotools.nubic.northwestern.edu/OligoCalc.html). Due to the inevitable propensity of these sequences rich in repetitions, concessions had to be made towards accepting less-than-ideal sequences.

### *Golden Gate Assembly of the Synthetic Dicistronic sgRNA Gene Transcribed DNA Sequence*

The construction of the *SDRG* transcribed DNA sequence (TDS) commenced with the restriction-ligation of the three BMs following the GG assembly method. In short, the BMs were digested by the BsaI RE, which left behind 4 nt long

overhangs designed to be compatible; all carried out in only one tube. At their annealing, T7 DNA ligase permanently seals the ligations, Figure 7.

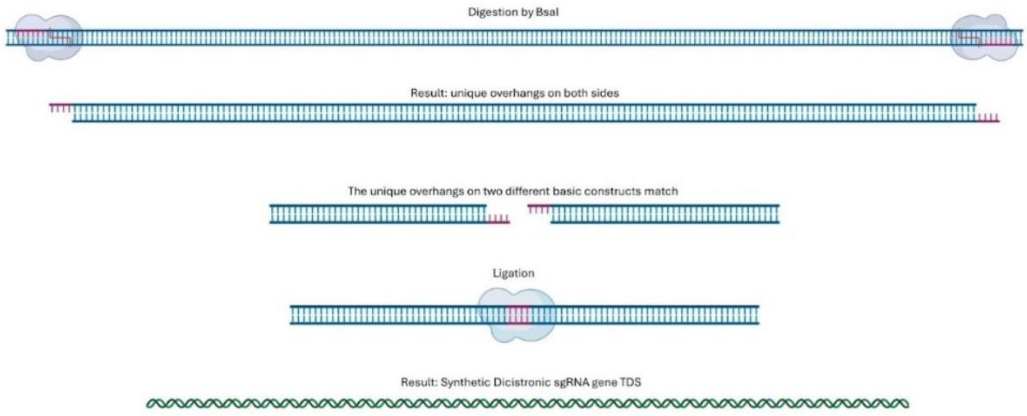



*Figure 7: Schematic diagram of the Golden Gate assembly. The basic modules are digested by BsaI outside of its recognition site, highlighted. It results in a sequence with unique overhangs, highlighted. The unique overhangs on two different basic modules were designed to match. Upon their alignment, they are seamlessly ligated by the T4 DNA Ligase. Result: Synthetic Dicistronic sgRNA gene TDS.*

The reaction was performed in a 20 µl mixture including 16 ng of each DNA fragment, 10 µl of 2× StickTogether™ DNA Ligase Buffer (by New England Biolabs (NEB), Ipswich, Massachusetts, United States, Cat# B0535 AVIAL), 2 µl of 1 mg/ml Bovine Serum Albumin (BSA) (by Sigma-Aldrich, Merck, Cat# A7030), 0.5 µl of T7 DNA Ligase (by NEB, Cat# M0318S), and 0.5 µl of BsaI-HFv2 (by NEB, Cat# R3733S). The temperature cycling was as follows: 35 cycles of 37°C for 5 min and 20°C for 10 min, followed by a final incubation at 20°C for 60 min. As a note, the BsaI RE is active between temperatures of 37 to 50°C.

The 20 µl ligation mixture was diluted with 40 µl of Nanopure water (MilliQ, MQ) and used for a PCR step to specifically amplify the assembled construct and re-introduce the adapter sequences on both sides of the fragments. A total of 200 µl (4x 50 µl) of reaction was made up using the same Phusion Polymerase as above, 0.5 µl of AdF and AdR primers (10 µM working stock, listed in SM2) and 1 µl/tube of the diluted ligation mix as template. The PCR program was the same as above but consisted of 30 cycles. Here too, the PCR products were separated by gel electrophoresis and the bands of the appropriate size were eluted using the same Zymoclean Gel DNA Recovery Kit.

### 2.3.2 Ligation of the Synthetic Dicistronic sgRNA Gene TDS to the Promoter

The *SDRG* expression cassette was designed to be driven by the *U6-26* promoter (*U6-26p*), which was cloned from the pICSL90002 plasmid (by AddGene, Cat# 68261). This promoter is naturally found upstream of the gene *U6 SMALL NUCLEOLAR RNA26* (*U6-26p*, AT3G13855) and was chosen to avoid post-transcriptional modifications (e.g. capping and polyadenylation) as it is recognised by the RNA polymerase III (Pol III; Castel et al., 2019). Xie et al. (2015) also used a promoter recognised by this polymerase.

In a single GG assembly step, the *SDRG* for Cluster 3 was ligated to *U6-26p* and the whole was cloned into the pGGAselect plasmid, which provided a backbone for the assembly. This plasmid's size is 2200 bp, but restriction by BsaI releases a short 87 bp fragment and linearises the plasmid, resulting in a 2,133 bp long sequence. This step was performed using the NEBridge Golden Gate Assembly Kit (BsaI-HFv2) (by NEB, Cat# E1601S), which contains the BsaI-HFv2 RE, T4 DNA Ligase and the pGGAselect plasmid.

The assembly was as follows. The ratio of pGGAselect to inserts was 1:2. The size of the pICSL90002 plasmid is 2453 bp. However, under this specific instance, only the size of the *U6-26p* is taken into consideration, which is just 206 bp. To do so, the Neb Converter (2024) tool was used. Knowing the length of the DNA fragments, in bp, and using the molarity best suited for the ligation, the mass of the fragments can be calculated. The calculations are illustrated in Table 3. Since small volumes are hard to handle, concentrated DNA can also be diluted to bulk up the pipetting volume.

The reaction solution received the relative volumes of DNA fragments, 10 µl of the buffer and 1 µl of the GG Mix solution provided in the kit. The volume was made up to 20 µl with MQ.

*Table 3: Example of the calculations of the optimal DNA fragment load for a successful kit-based Golden Gate assembly attempt.*

| DNA Fragment | Size, in bp | Molarity advised. The values add up to 0.20 pmol | Required mass, in ng | Concentration, in ng/µl | Volume to use, in µl |
|---|---|---|---|---|---|
| pGGAselect | 2,133 | 0.04 | 52.5 | 75 | 0.7 |
| *SDRG* TDS | 409 | 0.08 | 10.2 | 55 | 0.18 |
| *U6-26p* | 206 | 0.08 | 20.1 | 89 | 0.22 |

The assembled vector was used as a PCR template for amplifying the *U6-26p* using the primers U6 F and 40R, SM2. This allowed to produce DNA fragments made up of the full promoter and a short portion compatible with the Cluster 8 *SDRG* and a BsaI RS. At the same time, the full *SDRG* TDS was used to create DNA fragments essentially equivalent to BM2+BM3, Figure 9 Level 0 or SM1, by using the primers 40F and AdR, SM2.

The two fragments were assembled just like in the "Golden Gate Assembly of the Synthetic Dicistronic sgRNA Gene Transcribed DNA Sequence" section above. Two reactions were carried out, one with 16 ng of both DNA samples and the other one with 35 ng of each. The reactions were used for a colony PCR and the obtained vector was sent for Sanger sequencing.

## 2.4 Final Vector Construction

The final vector was constructed following the same GG modular assembly principles as described in the "Golden Gate Assembly of the Synthetic Dicistronic sgRNA Gene Transcribed DNA Sequence" section above. The only difference was in the use of a different RE, namely instead of continuing using the BsaI RE, this step was performed with the BbsI (BpiI) RE (by NEB, Cat# R0539S). This TII RE recognises the GAAGAC sequence and cuts the dsDNA upstream of it and introduces four base 5'-overhangs overhangs upon cutting, Figure 4b.

The final vector was assembled from four different components: the newly assembled *U6-26p::SDRG* expression cassette DNA fragment, a BASTA resistance gene (*BAR*) carried in pICSL11017 (pICH47732::*NOSp-BAR-NOST*; by AddGene, Cat# 51145), the Cas9 expression cassette carried in the plasmid BCJJ345 (by AddGene, Cat# 117504) and the plasmid backbone pICSL4723-P1 (by AddGene, Cat# 86173) which contains a kanamycin resistance gene for positive colony selection. The reactions were used for a colony PCR and the vector that was obtained was sent for Sanger sequencing.

## 2.5 Quality Control Sequencing: Bacterial Transformation, Colony PCR and Plasmid Extraction

The various plasmids carrying sequences of interest were transformed into the appropriate competent bacteria to be assessed by Sanger sequencing.

### 2.5.1 Bacterial Transformation

*For the SDRG TDS*

When the *SDRG* TDS was successfully assembled, as suggested by the gel band size, it was eluted from agarose gel and cloned into a pCR™4Blunt-TOPO™ vector using the Zero Blunt™ TOPO™ PCR Cloning Kit for Sequencing (by Invitrogen™, Cat# 450159). These plasmids were then transformed into One Shot™ TOP10 Chemically Competent Cells (by Invitrogen™, Cat# C404004, now discontinued), following the manufacturer's instructions. In short, 4 µl of purified PCR product, and 1 µl each of the salt solution and vector, were incubated at room temperature RT for 10 min and added to the competent cells. These were gently flicked and incubated for 20 min on ice, followed by a 30 sec heat shock in a 42°C water bath, and immediately placed back onto ice for 2 min. Under sterile conditions, each tube received 250 µl of S.O.C. medium (by Invitrogen, Cat# 15544-034) and incubated on a moving platform at 37°C for 1 h. The transformed cells were plated on LB agar plates containing Kanamycin (50 µg/ml) and were grown at 37°C overnight.

*For the U6-26p::SDRG*

As the process of the assembly of *U6-26p::SDRG* followed a different protocol, a different method was used. The GG process ligated the *U6-26p*, the *SDRG* TDS and the pGGAselect plasmid. The modified plasmid is therefore ready to be transformed into 10-beta Competent *E. coli* (High Efficiency, by NEB, Cat# C3019H). In short, 4 µl of the ligation reaction was added to the cells, which were then incubated for 20 min on ice, followed by a 30 s heat shock at 42°C in a water bath, and again 2 min on ice. Under sterile conditions, each tube received 950 µl of the outgrowth medium that comes with the cells. After a 1 h incubation at 37°C, the medium was spun down for 5 min at 4000 rpm, to concentrate the cells and allow to remove 700 µl of supernatant. The cells were resuspended and 100 µl was used for plating on LB agar enriched with Chloramphenicol (33 µg/ml). The plates were incubated overnight at 37°C.

*For the Final Vector*

The Final Vector was assembled in two GG reactions, called "0.02 pmol" and "20 ng". These names are based on the amount of the relative DNA mass each reaction solution received. This was done to test which protocol has better efficacy. The total volume of each reaction was 20 µl, of these 10 µl were used to transform the same One Shot™ TOP10 Chemically Competent Cells as above using the same protocol. The transformed cells were plated on LB agar plates containing Kanamycin (50 µg/ml) and were grown at 37°C overnight.

### 2.5.2 Colony PCR

Colony PCR was performed to identify colonies born from a successful transformation event. Tens of colonies were screened each time. The goal of this process is to identify colonies carrying the desired plasmid, which can be identified by observing the resulting gel. Colonies scored as positive are then used for Sanger sequencing to make sure no SNPs are present in the sequence of interest.

The genetic material was collected by resuspending a portion of the colony in 10 µl of MQ and using only 1 µl of this suspension as DNA template. The DNA polymerase used was DreamTaq™ Green PCR Master Mix (2X) (by Thermo Scientific, Cat# K1081), in a total volume of 15 µl. The primers used depended on the vector insert being assessed. For the *SDRG* TDS, the insert-specific AdF_only and AdR_only primers were used, for the *U6-26p::SDRG* the vector-specific CW F and CCW R were used and for the Final Vector, the NOS-BAR primer pair was used. All primer sequences can be seen in SM2. The thermocycler was set to 95°C for 3 min, 34 cycles of 95°C for 30 s, 60°C for 30 s, and 72°C for 30 s, and a final extension at 72°C for 5 min. Gel electrophoresis of the amplicons allowed the identification of which colonies carried inserts of the correct size by comparing the gel band sizes to the relative DNA ladder. The chosen *SDRG* TDS and Final Vector colonies were then inoculated into 5 ml of LB broth with Kan (50 µg/ml), and incubated on a rotating platform at 37°C overnight.

### 2.5.3 Plasmid Extraction and Sequencing

The cloned plasmids were extracted using the GeneJET Plasmid Miniprep Kit (by Thermo Scientific, Cat# K0503), following the manufacturer's protocol. In principle, the bacteria are resuspended and subjected to alkaline lysis to liberate the plasmid DNA. After neutralisation and purification steps, the DNA that adsorbed onto the silica membrane was eluted.

The *SDRG* TDS and *U6-26p::SDRG* inserts were sequenced unidirectionally by Sanger Sequencing (by Eurofins Genomics, Copenhagen), using T7 sequencing primers. Vectors containing the right sequences, with no SNPs or other indels, were selected for further work. The Final Vector was sequenced in the same manner but using insert the specific primer pair, NOS F2 FW and YAO:Cas9:E9-F2 RV, SM2.

## 2.6 Plant Transformation via Floral Dip

The pICSL4723-P1 Final Vector plasmid carrying the desired Cluster 3 SDRG/Cas9 were transformed to *Agrobacterium tumefaciens* strain GV3101. A liquid 50 ml culture in high salt LB broth (Lysogeny broth), enriched with 5 µl gentamicin (30 mg/ml), 1 µl rifampicin (50 mg/ml), 2.5 µl tetracycline (10 mg/ml) and 5 µl kanamycin (50 mg/ml), was established from a transformed single colony

and grown overnight at 28°C on a shaker. The following day, four 50 ml liquid LB broth flasks, antibiotic-free, were each inoculated with 50 µl of bacteria-rich broth. The flasks were again incubated overnight at 28°C on a shaker. On the transformation day, the flasks containing *A. tumefaciens*-rich broth were centrifuged for 10 min at 5000 rpm at RT and re-suspended in 5% (w/v) sucrose and 0.02% Silwet L-77 solution in autoclaved water.

The WT *A. thaliana* plants, parental generation (P0), were induced to flower by long-day conditions. The majority of the mature flowers and seed pods were removed before the transformation procedure. The transformation was performed by dipping flower buds in bacterial suspension and gently stirring for about 15 sec; protocol adapted from Clough & Bent (1998). Thirty-two and twenty-one pots, each made up of four individual plants, were created.

The plants were grown under controlled long-day conditions until the plants produced T1 seeds. When the seed pods matured, the plants were left to dry out and were ultimately bagged in cellophane and allowed to desiccate fully. The seeds were stored at 37°C for three days and transferred to a -20°C freezer for a week.

## 2.7 Plant DNA Extraction and Detection of Transformants via Genotyping

A total of 102 BASTA selection survivors from the T1 generation plated in February were sampled. Total DNA was extracted from leaf disks made by punching the leaf with an Eppendorf tube cap. These disks were dropped into screw-cap tubes containing 200 µl of DNA extraction buffer (250 mM of NaCl, 25 mM of EDTA with pH 8.0 and 200 mM of Tris adjusted to pH 7.5 HCl) and a few mini glass spheres. The tissues were disrupted with the Precellys 24 Touch Homogeniser (by Bertin technologies) with a setting of two 30 sec rounds at 5500 rpm, with a 45 sec break in between. Subsequently, 10 µl of 10% SDS solution was added to each tube, vortexed and incubated at 55°C for 10 min in a water bath. The tubes were centrifuged at max rpm for 3 min and 100 µl of supernatant was transferred to new Eppendorf tubes containing 100 µl of 2-propanol, followed by a 10 min incubation at 4°C. The DNA was pelleted by centrifugation at max speed for 5 min and washed with 50 µl of 70% ethanol, and re-pelleted again. Finally, the pellet was left to dry and resuspended in 30-40 µl of MQ.

The quality of the extracted DNA was checked by performing a positive control PCR with primers for *ACTIN* (provided by a colleague), an *A. thaliana* endogenous gene. All samples were positive. The concentration and biological contamination levels were checked by NanoDrop; both parameters were highly variable between the different DNA samples.

Having assessed the quality of the DNA extracted from the T1 generation, the presence of the Final Vector for Cluster 3, containing the BASTA resistance gene (*BAR*) was confirmed by PCR. The primer pair was designed to amplify a segment of 435 bp spanning from the *NOS* promoter and covering part of the *BAR* gene; the primer pair sequences, NOS-BAR F and NOS-BAR R, SM2.

For each reaction, 2-3 µl of DNA suspension (depending on the DNA's concentration) was used in a 28 µl volume containing the DreamTaq™ Green PCR Master Mix (2X) (by Thermo Scientific, Cat# K1081), MQ and primers. The amplification was carried out as follows: 95°C for 3 min, 34 cycles of 95°C for 30 s, 60°C for 30 s, and 72°C for 45 s, and a final extension at 72°C for 5 min. The amplification products were analysed by electrophoresis in 1.5% agarose gels.

# 3. Results

## 3.1 Gene and Protein Sequence Analysis

### 3.1.1 Genomic DNA Identity Analysis of the Two Paralogues

The genes AT2G47030 and AT2G47040 were found to be paralogous by Reimegård et al. (2017). Their DNA sequence identity was assessed using two different software programs. When considering the full genomic sequences, including the 5' and 3' untranslated regions (UTRs) and introns, BioEdit scored the sequences as having 83.7% identity, while GeneDoc scored them as 82% identical. However, CDSs had much higher identity scores, with BioEdit assigning a 90% identity score and GeneDoc an 89% score. This discrepancy highlights the need for the CDS to remain more conserved than the non-transcribable regions nearby.

### 3.1.2 Amino Acid Sequence Identity Scores within Cluster 8 and to Proto-Cluster 8

Using PME4/VGDH1's (AT2G47030.1) 589 amino acid (aa) sequence as the query for the BLASTP tool available in the Phytozome (2024) online database resulted in two homologous genes being identified, both with an e-value of 0. PME5/VGD1 (AT2G47040.1) scored an 85% identity and only 2% gapping. As expected, PMEI-PME37/VGDH2 (AT3G62170.1) also appeared on the list; it scored 68% identity and only 4% gapping. Similarly, using PME5/VGD1's (AT2G47040.1) 596 aa sequence, resulted in coinciding outputs, both with an e-value of 0. PME4/VGDH1 (AT2G47030.1) scored an identity of 85% and a gapping of 2%; whereas PMEI-PME37/VGDH2 (AT3G62170.1) scored an identity of 70% and a gapping of 7%. These values are further corroborated by GeneDoc, were they were shown to score 85% identity and 92% similarity.

Things were different in the case of PMEIL (AT2G47050.1) 217 aa sequence. It did not align with either of the two members of Cluster 8, displaying low or absent identity scores to them. It did, however, align with AT3G62180.1, Plant invertase/pectin methylesterase inhibitor superfamily protein. Whilst the e-value was greater than 0, it scored a 54% identity and a gapping of only 2%.

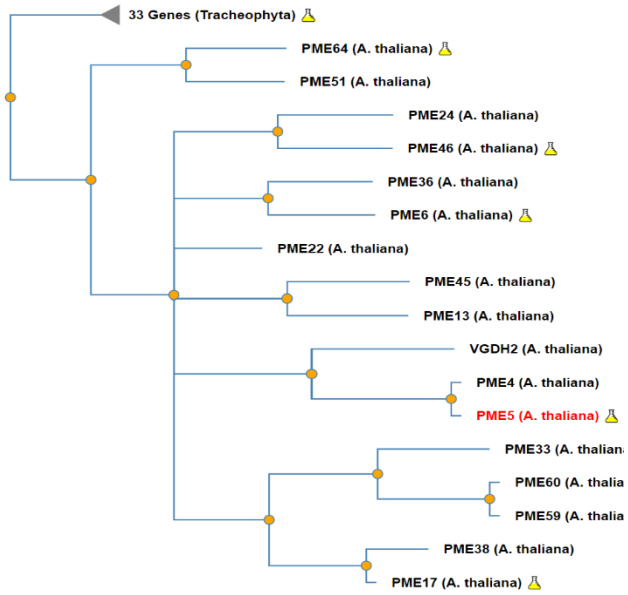
### 3.1.3 Phylogenetic Tree Construction

Two phylogenetic trees were constructed. The first was based on querying the PhyloGenes (2024) online database for any genes in the *A. thaliana* genome that may have homology to the gene products of *PME5/VGD1* (AT2G47040) or *PME4/VGDH1* (AT2G47030). Doing so gave fifty hits of genes belonging to the PME gene family. This tree was made up of two branches, one with thirty-three hits and one with seventeen, Figure 8a. The GOIs were located in the smaller branch so the tree was pruned to it alone. All of the protein sequences from this pruned tree were compiled together and used to build a phylogenetic tree manually, Figure 8b. As a means of comparison, a second tree was built following the same methods, but using only the genes mentioned in this report's introduction, which were compiled together by literature research rather than aa sequence homology, Figure 8c.

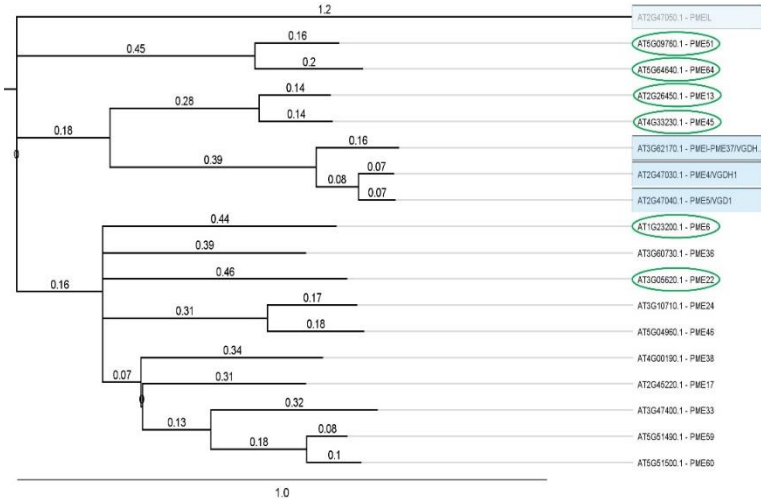
Comparing the two trees that were built manually (Figure b and c), the only protein sequences that are in common between them are identified with the light blue boxes. The PMEIL did not belong to the PhyloGenes-based tree, it was manually inserted to act as an outgroup and put the relative homologies into perspective. As the overlap is thus minimal, the TAIR (2024) online database was queried for the unique genes listed in the PhyloGenes based trees, Table 4. This was done to infer if the protein homology-based genes may provide a source of redundancy for the genes of Cluster 8 after they are KOd. From this surface-level scan, it can be noticed that they all have PME and/or PMEI domains and that six are expressed in pollen, Table 4. Notably, few of these genes do not seem to have much information immediately available.

As a side note, the PhyloGenes (2024) database could not be queried for the *PMEIL* gene, AT2G47050. It remains unclear why that should be so. Additionally it did not show up in the list of fifty genes compiled by PhyloGenes, Figure 8a.

*a.*




*b.*




*c.*

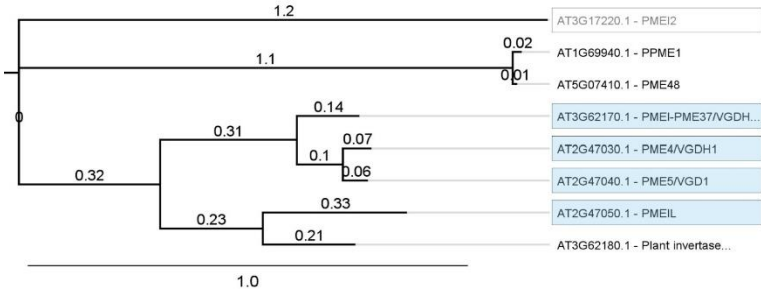



*Figure 8: Phylogenetic trees. a. Snapshot of a phylogenetic tree compiled by PhyloGenes online database around the two paralogous genes from Cluster 8. The less relevant branch is collapsed but all of the gene names listed included PME or PMEI keywords within them; b. Tree built manually from sequences from PhyloGenes Database. The circled genes have been found to be expressed in pollen (TAIR, 2014). PMEIL inserted manually to serve as the outgroup; c. Phylogenetic tree built manually from genes listed in this report alone. The outgroup, still mentioned in this report, was chosen for its lowest score and is identified by the pale colour; b and c. The numbers next to the nodes represent substitutions per site. The names provided on PhyloGenes are not always the same as on TAIR (2024), refer to Table 4. The blue boxes identify the genes in common between the two trees and the pale boxes identify the outgroup sequences.*

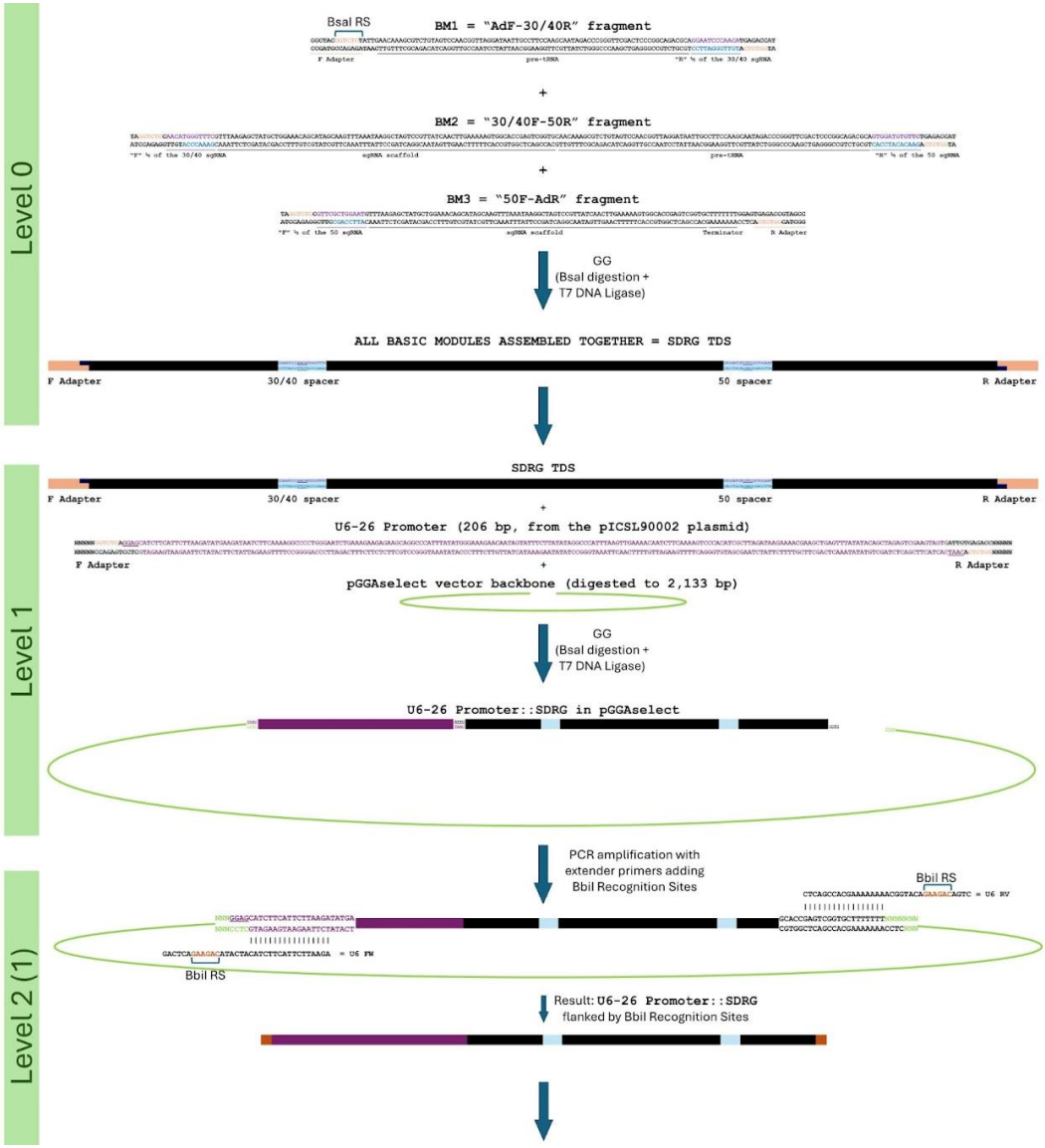
*Table 4: Surface-level screening of the TAIR (2024) online database for key words associated with the genes with the highest homologies to the two paralogous genes within Cluster 8, AT2G47030 and AT2G47040. The choice of genes to investigate was based on the phylogenetic analysis results depicted in Figure 8b. The genes encapsulated in blue boxes were excluded from the screening. Key words associated with pollen or plant sperm are identified by the bold font.*

| Locus | Gene Name | Details |
|---|---|---|
| AT5G09760 | PMEI-PME51 | Pectin methylesterase inhibitor; INVI-PMEI protein isoform which clustered with a PME.<br>Expressed during: LP.02 two leaves visible stage, LP.04 four leaves visible stage, LP.06 six leaves visible stage, LP.08 eight leaves visible stage, LP.10 ten leaves visible stage, LP.12 twelve leaves visible stage, flowering stage, mature plant embryo stage, petal differentiation and expansion stage, plant embryo bilateral stage, plant embryo cotyledonary stage, plant embryo globular stage, vascular leaf senescent stage<br>Expressed in: carpel, cauline leaf, collective leaf structure, cotyledon, flower, flower pedicel, guard cell, hypocotyl, inflorescence meristem, leaf apex, leaf lamina base, petal, petiole, plant embryo, plant sperm cell, **pollen**, root, seed, sepal, shoot apex, shoot system, stamen, stem, vascular leaf |
| AT5G64640 | PME64 or PMEI-PME94 | Pectin methylesterase inhibitor; INVI-PMEI protein isoform which clustered with a PME.<br>Expressed during: **L mature pollen stage**, LP.02 two leaves visible stage, LP.04 four leaves visible stage, LP.06 six leaves visible stage, LP.08 eight leaves visible stage, LP.10 ten leaves visible stage, LP.12 twelve leaves visible stage, **M germinated pollen stage**, flowering stage, mature plant embryo stage, petal differentiation and expansion stage, plant embryo bilateral stage, plant embryo cotyledonary stage, plant embryo globular stage, vascular leaf senescent stage<br>Expressed in: carpel, cauline leaf, collective leaf structure, cotyledon, flower, flower pedicel, guard cell, hypocotyl, inflorescence meristem, leaf apex, leaf lamina base, petal, petiole, plant embryo, plant sperm cell, **pollen, pollen tube cell**, root, seed, sepal, shoot apex, shoot system, stamen, stem, vascular leaf<br>GO (extra): defence response by callose deposition |
| AT2G26450 | PME13 or PMEI-PME66 | Pectin methylesterase inhibitor; INVI-PMEI protein isoform, clusters with a PME.<br>Expressed during: **L mature pollen stage, M germinated pollen stage**, flowering stage, petal differentiation and expansion stage, plant embryo globular stage<br>Expressed in: carpel, collective leaf structure, flower, petal, plant embryo, **pollen, pollen tube cell,** sepal, stamen |
| AT4G33230 | PME45 or PMEI-PME45 | Pectin methylesterase inhibitor; INVI-PMEI protein isoform which clustered with a PME.<br>Expressed during: **L mature pollen stage, M germinated pollen stage**, flowering stage<br>Expressed in: flower, plant sperm cell, **pollen, pollen tube cell**, sepal |
| AT1G23200 | PME6 or HMS | ProPME pectin methyl esterase involved in embryo development, stomatal function, and mucilage release.<br>Expressed during: the flowering stage, petal differentiation and expansion stage |

| | | |
|---|---|---|
| | | Expressed in: carpel, collective leaf structure, flower, guard cell, petal, **pollen**, seed, sepal, stamen, stem<br>Located in: embryo plant cell, seed coat |
| AT3G60730 | PME36 or PMEI-PME65 | Pectin methylesterase inhibitor; INVI-PMEI protein isoform which clustered with a PME.<br>Expressed in: flower |
| AT3G05620 | PME22 or PMEI-PME22 | Pectin methylesterase inhibitor; INVI-PMEI protein isoform which clustered with a PME.<br>Expressed during: petal differentiation and expansion stage<br>Expressed in: flower, hypocotyl, leaf apex, **pollen tube cell**, shoot system |
| AT3G10710 | PME24 or PMEI-PME24, RHS12 | Pectin methylesterase inhibitor; INVI-PMEI protein isoform which clustered with a PME; involved in root hair development.<br>Expressed during: flower, root hair cell, trichoblast<br>Expressed in: root hair cell |
| AT5G04960 | PME46 or PMEI-PME46 | Encodes a protein that modulates the activity of pectin methylesterase within the cell wall; pectin methylesterase inhibitor.<br>Expressed in: flower, root, root hair cell, trichoblast |
| AT4G00190 | PME38 | Pectin methylesterase inhibitor; probable pseudogene.<br>Expressed in: flower, synergid |
| AT2G45220 | PME17 | Pectin methylesterase involved in pectin remodelling. Regulated by its PRO region that triggers PME activity in the resistance to Botrytis cinerea.<br>Expressed during: flowering stage, petal differentiation and expansion stage, plant embryo globular stage<br>Expressed in: flower, fruit, inflorescence meristem, non-hair root epidermal cell, plant embryo, root, sepal, stamen, stem, vascular leaf |
| AT3G47400 | PME33 or PMEI-PME33 | Pectin methylesterase inhibitor; INVI-PMEI protein isoform which clustered with a PME.<br>Expressed during: petal differentiation and expansion stage<br>Expressed in: carpel, flower, guard cell, hypocotyl, root |
| AT5G51490 | PME59 or PMEI-PME59 | Pectin methylesterase inhibitor; INVI-PMEI protein isoform which clustered with a PME.<br>Expressed in: flower |
| AT5G51500 | PME60 or PMEI-PME60 | Pectin methylesterase inhibitor; INVI-PMEI protein isoform which clustered with a PME.<br>Expressed during: flowering stage<br>Expressed in: flower, plant sperm cell, root |

## 3.2 Construction of the Cluster 8-SDRG/Cas9 system

The genes in Cluster 8, as identified by Reimegård et al. (2017) were chosen to be knocked out in order to carry out functional analysis of these genes and of the cluster as a whole. To do so, a Cluster 8-SDRG/Cas9 system containing two sgRNAs targeting the three genes of Cluster 8 were constructed for this project. Their target sequences can be seen in Table 2. A workflow of the whole process can be seen in Figure 9.



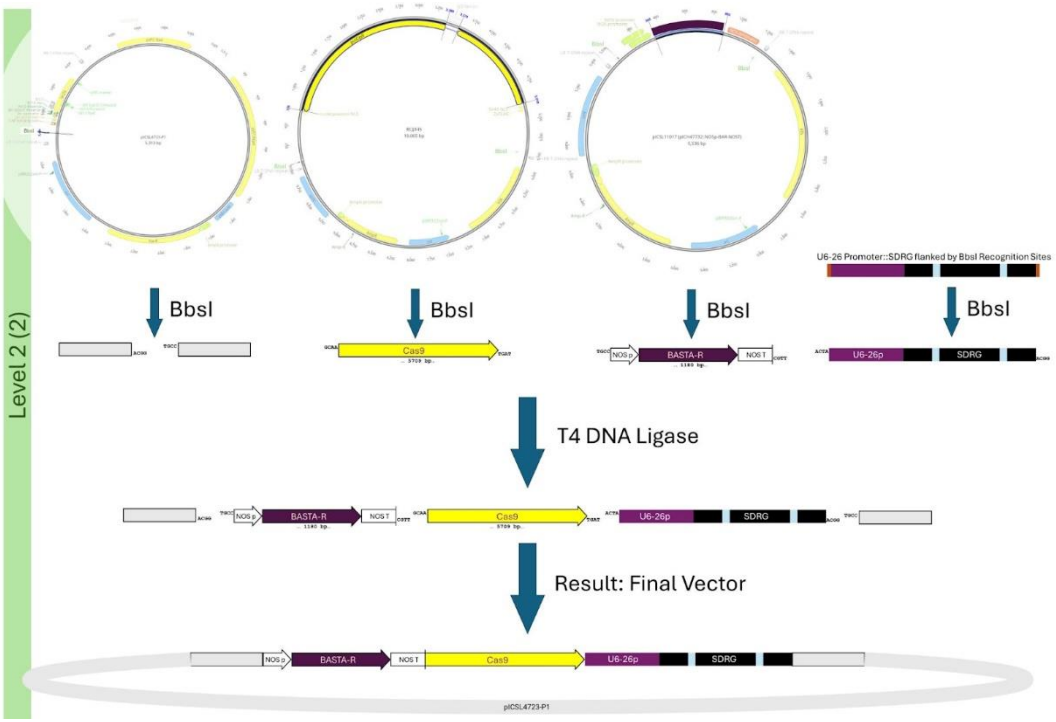



*Figure 9: Workflow of the Cluster 8-SDRG/Cas9 system construction procedure; Level 0: The synthetic dicistronic sgRNA gene (SDRG) TDS was first produced by PCR amplifying three Basic Modules with the common sgRNA scaffold - pre-tRNA fragments as template. These Basic Modules were assembled by Golden Gate cloning, using BsaI as the restriction enzyme and ligated in order; Level 1: The SDRG TDS and the U6-26p were ligated together and cloned into the pGGAselect acceptor vector (in a single reaction); Level 2: Three distinct expression units were assembled in a final destination vector. Plasmids containing the BASTA resistance gene, pICSL11017 (pICH47732::NOSp-BAR-NOST), the complete Cas9 expression cassette, BCJJ345, and the U6-26p::SDRG were cloned into an empty destination vector, pICSL4723-P1. This GG reaction was performed the same way as the prior two ones with the exception of a different restriction enzyme being used, the BbsI instead of the BsaI.*

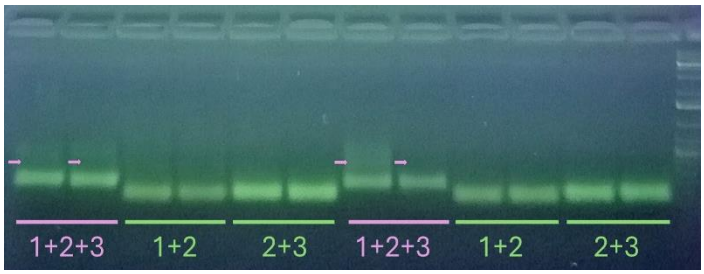
### 3.2.1 Constructing the Synthetic Dicistronic sgRNA Gene TDS

The successful construction of the synthetic dicistronic sgRNA gene (*SDRG*) TDS was accomplished via GG cloning, Figure 9. To ensure accuracy, assembly was performed in three ways: combining BMs in pairs (BM1+BM2 and BM2+BM3), and assembling the full TDS from BM1+BM2+BM3. All GG reactions were used as templates for PCR amplification, with their relative specific primers. Analysis of the agarose gel, Figure 10a, indicated that all three GG reactions were successful. Since the full construct was successfully assembled on the first attempt, the paired BM assemblies, BM1+BM2 and BM2+BM3, were given lesser priority. Before moving forward, it is worth mentioning that the gel photograph in Figure 10a has some very bright bands, but these had the "wrong" size. The faint bands right above them were of the right size, and whilst they were hard to capture on film they were much more vibrant live. The "multiple band" phenomenon was due to the primers covering sequence portions that were

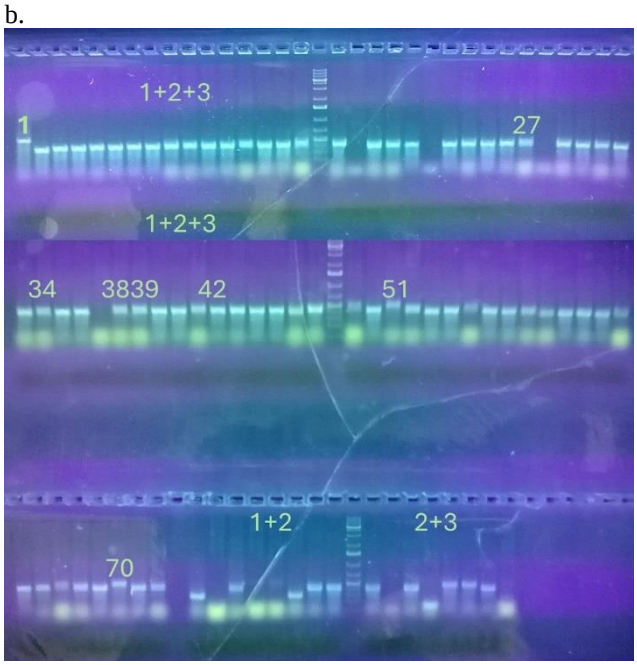
inherently repetitive within any sequence made up of more than one single Basic Module (BM).

Colony PCR was performed, Figure 10b, and Sanger sequencing of eight colonies identified colony number 1 as carrying the complete *SDRG* TDS from one adapter to the other, Figure 10c. This sequence contained two SNPs: one within each adapter portion. In the forward adapter, an additional guanine (G) was found at position seven (relative to the start of the adapter), where an extra G was inserted alongside two existing Gs. The other SNP, located at the opposite end of the sequence, involved a missing thymine (T) in the terminator portion. These SNPs might not be present, having been erroneously identified by the sequencing process. Alternatively, they could be true SNPs introduced by the polymerase during amplification. Regardless of the cause, PCR amplification of this DNA template using SNP-free primers was able to "fix" the problem.

a.




b.



c.

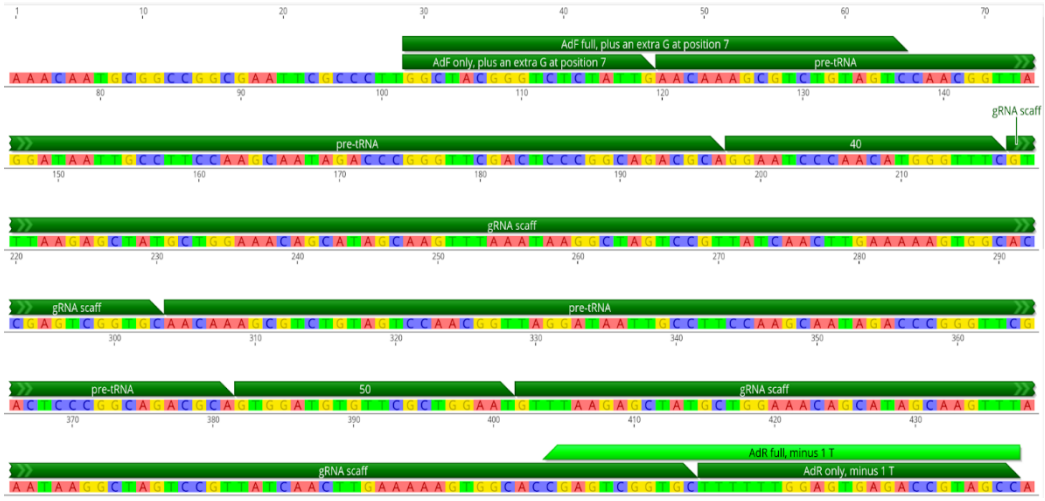



*Figure 10: Synthetic Dicistronic sgRNA Gene TDS. a. Agarose gel: the bands represent the size of the PCR amplicons of three combinations of Golden Gate reactions: full synthetic dicistronic sgRNA gene TDS assembly (1+2+3), and partial TDS assembly, where the basic modules were paired (1+2 and 2+3). Each well represents a distinct GG reaction. The first half of the gel contains amplicons amplified at 60.5°C and the other at 59.6°C. The small arrows indicate the band that was eluted (hard to see in the photograph, was clearer live). Ladder: 1 Kb Plus DNA Ladder; b. Agarose gel: Colony PCR results of the screening of competent bacteria transformed with vectors containing each of the combinations of fragments assembled. Ladder: 1 Kb Plus DNA Ladder; c. Sanger sequencing results of the synthetic dicistronic sgRNA gene (SDRG) TDS amplified within colony number 1. The sequence is annotated with the identity of each sequence portion; the numbers 40 and 50 indicate the spacer portions of sgRNAs targeting the two relative genes. Two single nucleotide polymorphisms were found at either end of the sequence: one extra G in the G trio in the AdF portion and one missing T from the terminator portion.*

### *Shortcomings and Challenges*

Many attempts were undertaken to construct the *SDRG* TDS via GG assembly, but the success rate was initially very low. The original plan was to create constructs containing a total of six sgRNAs, two for each of the three GOIs within Cluster 8, as successfully demonstrated by Xie et al. (2015). This approach was considered optimal because the simultaneous action of two sgRNAs on the same gene could introduce chromosomal deletions, simplifying genotyping (Xie et al., 2015).

The next approach, which ultimately proved successful, took secondary structures into deeper consideration. New sgRNA primers were ordered, designed to minimise the formation of problematic secondary structures. Although the formation of some secondary structures was inevitable due to the high nucleotide repetition levels within the BMs, the new BMs were predicted to be less problematic, as shown in the SM under "secondary structures (2ry structures, new tab)."

However, due to the low success rate of GG assembly, the approach had to be reconsidered. The first compromise was to abandon the idea of paired sgRNAs, which meant giving up the easy genotyping associated with chromosomal deletions.

The chances of such deletions being introduced were low, as Xie et al. (2015) noted that even with paired sgRNAs, NHEJ DNA repair mechanisms were still prioritised.

Whilst it is not clear what was behind the initial low success rate, it was proposed that the propensity of the BMs and primers to form aberrant secondary structures might be responsible. Many of the oligonucleotides used had a tendency to form secondary structures such as hairpins or to inappropriately anneal to one another due to repeating nucleotide patterns. A surface-level scan using the Multiple Primer Analyzer (Thermo Fisher Scientific, 2024) revealed that secondary structure formation was a valid concern, SM3 (old 2ry structures tab).

The next approach, which ultimately proved successful, took secondary structures into deeper consideration. New sgRNA primers were ordered, designed to minimize the formation of problematic secondary structures. Although the formation of some secondary structures was inevitable due to the high nucleotide repetition levels within the BMs, the new BMs were predicted to be less problematic, SM3 (2ry structures, new tab).

### 3.2.2 Ligation of the Synthetic Dicistronic sgRNA Gene TDS to the Promoter

The ligation of the *U6-26p* to the *SDRG* TDS was done via GG assembly, Figure 9 Level 1. The functional pGGAselect constructed for Cluster 3 was used as a template to amplify the promoter and a short portion of the following gene. This amplicon was eluted from the gel and used for the following GG reaction.

Two different concentrations were considered; in the first reaction 16 ng of each DNA fragment was used, and in the second 35 ng of each. Using the lower concentration proved to be successful, Figure 11a. Again, in the agarose gel figure, two bands of different sizes are seen because the primers can bind to different regions of the original template as it contain a few repeating areas. The band indicated by the arrow was the correct one. Colony PCR was performed and of the six colonies that developed only one, number 1, carried the desired construct, Figure 11b. This colony was sent for Sanger sequencing, and the sequencing results are nearly ideal, Figure 11c. Four nts are missing right upstream of the *U6-26p* and there seems to be an insertion of four nts between the promoter and the *SDRG*. Fortunately these small indels are localised in a manner as to not be expected to interfere with appropriate functioning.

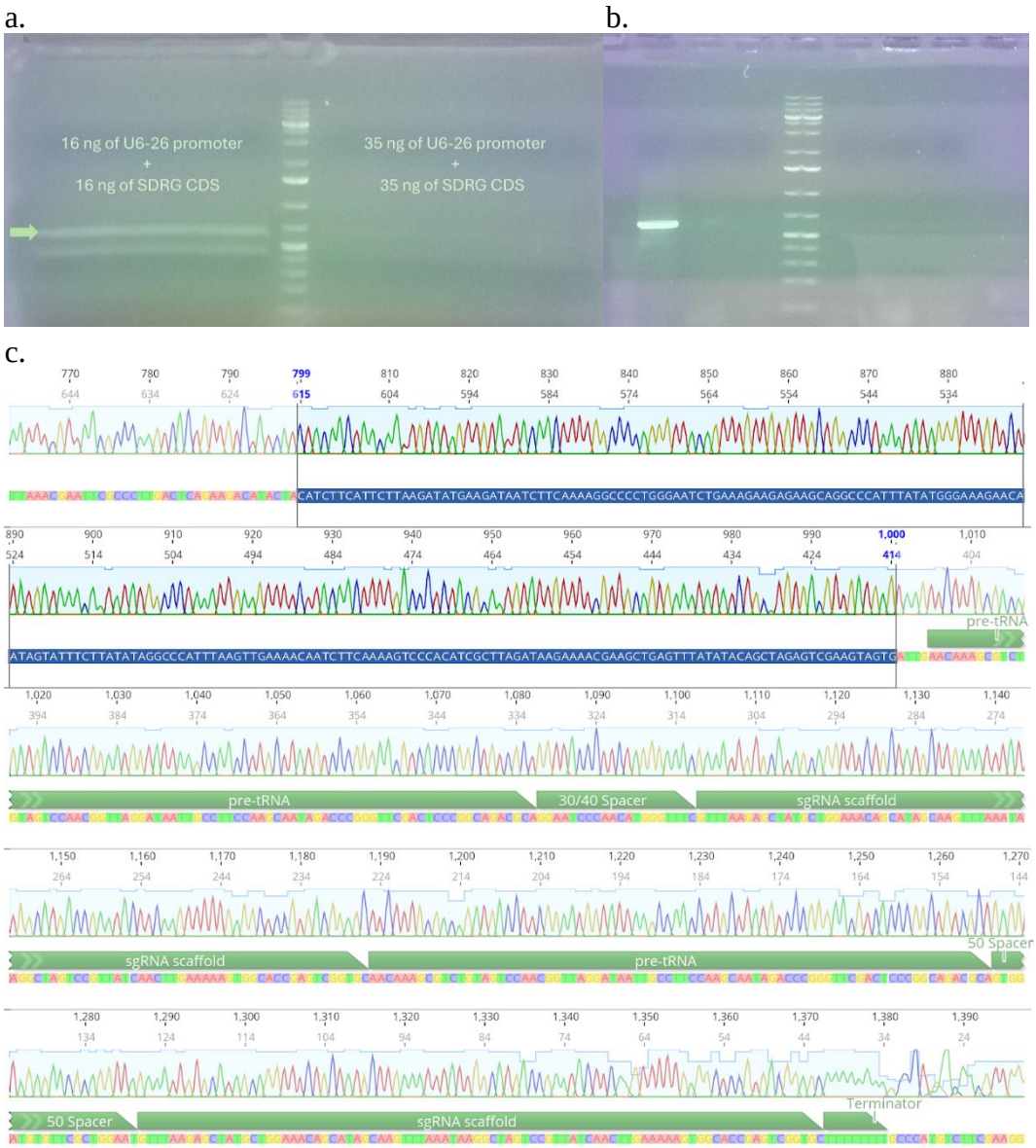



*Figure 11: Ligation of the Synthetic Dicistronic sgRNA Gene TDS to the Promoter a. Agarose gel: the bands represent the size of the PCR amplicons resulting from the Golden Gate reaction that ligated the U6-26 promoter to the full synthetic dicistronic sgRNA gene TDS. The arrow indicates the band that was eluted, >600 bp. Ladder: 1 Kb Plus DNA Ladder; b. Colony PCR agarose gel. Only one out of six colonies scored as positive; the band size matched the expected size of ~600 bp. Ladder: 1 Kb Plus DNA Ladder; c. Sanger sequencing results from U6-26p::synthetic dicistronic sgRNA gene amplified within colony number 1, identified by colony PCR. The sequence is annotated with the identity of each sequence portion: the portion highlighted in blue represents the U6-26p that matches 100% to the original reference sequence, the first four nts do not match to the reference (not highlighted); the numbers 30/40 and 50 indicate the spacer portions of sgRNAs targeting the two relative genes.*

*Shortcomings and Challenges*

Due to a misalignment between the *SDRG* TDS and the pGGAselect plasmid that remained unnoticed till the very end, whereby their digestion by BsaI did not produce compatible overhangs, Figure 9 Level 1, a shortcut was taken.

In short, as the complete pGGAselect vector for Cluster 3 was fully functional, it was used for a PCR reaction where the promoter and a short sequence of Cluster 3 sgRNA gene was amplified using the non-paired primers U6 F and 40R, SM2. Using the 40R primer substituted the spacer portion from Cluster 3 to that from Cluster 8. Doing so created amplicons made up of the full *U6-26p* followed by the pre-tRNA sequence, the 30/40R half of the sgRNA and the BsaI RS; or in other words, *U6-26p*+BM1.

In parallel, the fully assembled Cluster 8 *SDRG* was amplified with the primer pair 40F and AdR_only, in essence turning the BM1+BM2+BM3 construct into BM2+BM3 only, Figure 9 Level 0. Since the BM1 and BM2 portions of their relative constructs have compatible overhangs, carrying out a GG reaction ligates them together and generates the construct *U6-26p*+BM1+BM2+BM3, or *U6-26p::SDRG*.

### 3.2.3 Final Vector Construction

The final vector was constructed via GG assembly of four different components, of which three were located within plasmids, Figure 9 Level 2. For this step a different RE was used but every other factor was kept exactly the same. The Final Vector thus assembled was used to transform competent cells directly,

As mentioned, the reaction was carried out with two different DNA mass loads. The relative reactions were called “0.02 pmol” and “20 ng”. The goal was to further optimise the process but both reactions performed interchangeably well as shown by the 100% success rate displayed by the colony PCR, Figure 12a.

The colony PCR step was actually expected to face a challenge related to the Final Vector backbone, pICSL4723-P1. As this plasmid is only targeted by the BbsI RE once, as opposed to twice, two compatible overhangs are created, Figure 9 Level 2. These could be expected to be prioritised over other compatible overhangs during the ligation reaction. However, the colony PCR assay, Figure 12a, showed that incorporation of the desired inserts, as opposed to empty plasmid re-circularisation, was successful in all colonies tested. The quality of the assembly remains to be assessed by Sanger sequencing.

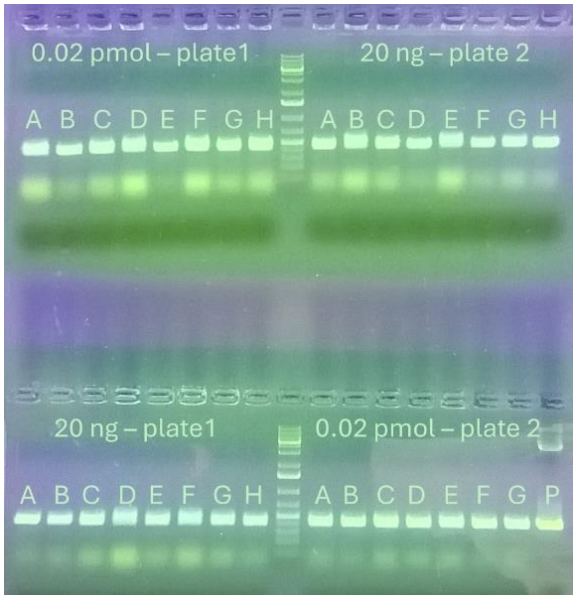



*Figure 12: Final Vector Construction. Colony PCR agarose gel. Out of thirty-one colonies screened, spread over four different plates and labelled alphabetically, all thirty-one tested positive for genetic material corresponding to the latter part of the NOS promoter and the first part of the BASTA resistance gene. The original BASTA resistance plasmid, pICSL11017, was used as a positive control and is identified by the letter "P". The band size matched the expected size of 435 bp. Ladder: 1 Kb Plus DNA Ladder.*

## 3.3 BASTA Selection and Genotyping in Search of Transformants

### 3.3.1 BASTA Selection

The parental generation (P0) was transformed with the Final Vector containing the Multiplex-CRISPR/Cas9 system targeting portion of the Cluster 3. The T1 generation was plated on BASTA-enriched media on two occasions, February 1$^{st}$ and June 11$^{th}$. The individuals from the T1 planted first that scored positive for carrying a portion of the transgene were selected for the collection of T2 seeds. The T1 generation plated in June has not yet been transferred to soil.

The T1 seeds, plated in February, were generated by thirty-two transformed P0 lines, each line spread on a distinct plate. Due to issues with contamination many plates had to be thrown away; only eight lines ultimately remained for further work. Over two-hundred surviving seedlings were transferred into soil, SM4. They were then treated with BASTA again during the juvenile stage, which removed half of the false positives. Over a hundred plants were allowed to develop into adulthood. Figure 13b depicts four genotypes that later failed to be genotyped as carrying BASTA resistance-related transgenic DNA, alongside genotype number 45, which did score positively, Figure 15. No striking phenotype could be observed at the time of taking this picture.

The T1 seeds plated in June were obtained from twenty-one lines of flower-dipped P0. After seventeen days, out of twenty-one plates, ten were fully dead (all germinated seedlings were either pale or brownish, e.g. #15), six were mostly dead (few individuals could be spotted starting to develop true leaves) and only five

plates had areas rich in greenery (e.g. #16 and 11), Figure 13a. Twenty-one days post plating, only four plates had any surviving seedlings (plates #7, 9, 11, 16). These are still awaiting being transferred into soil. The plates that are thriving had many individual seedlings developing true rosettes.

a.

b.

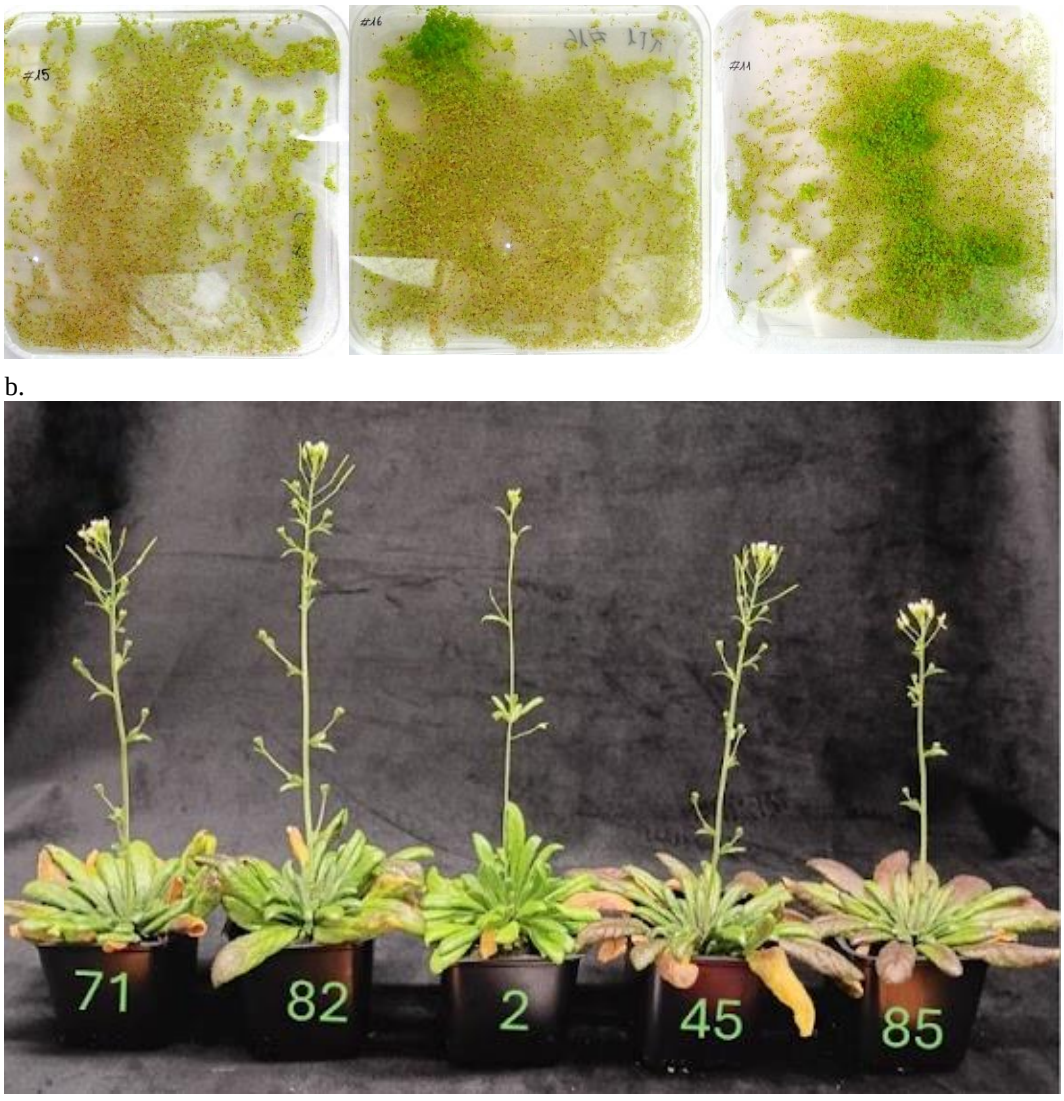



*Figure 13: BASTA Selection. a. BASTA-enriched plates used for the germination of the T1 mutant generation plated in June. Only three out of twenty-one plates are displayed. Each line generated by floral dipping of the parental line was germinated on its own distinct plate. Pictures taken seventeen days post-plating; b. Mature representatives of the T1 generation plated in February. Each individual genotype was identified by an assigned number. These pictures were taken before successful genotyping. Later findings showed that only number 45 scored as carrying BASTA resistance-related transgenic DNA.*

The seeds of the T2 generation were produced by allowing all members of the T1 generation plated in February to self and inter pollinate. These seeds were collected but their amount was much lower than that of the previous generation. Attempts were made to germinate a small portion of the seeds, ~0.3 ml of an Eppendorf tube. This volume corresponded to ~20% of the available seed stock for each line. This first attempt did not result in any survivors; another attempt will be carried out.

### 3.3.2 Positive Control and Genotyping PCRs

The DNA was extracted from the *A. thaliana* plants grown from T1 seeds plated in February, generated by floral dipping the parental generation (P0) in transformed *Agrobacterium* solution. The quality of the extracted DNA was first checked by performing a control PCR for an endogenous housekeeping gene, in this case, *ACTIN* (primers provided by a colleague), Figure 14. The assay was conducted for all individuals but the rest of the pictures were either lost or not taken.

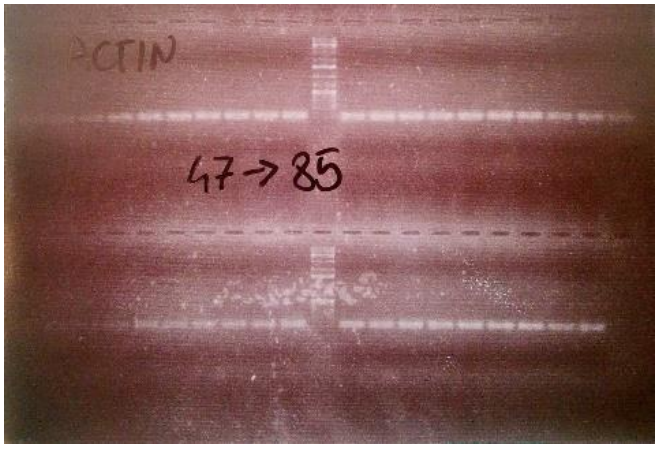
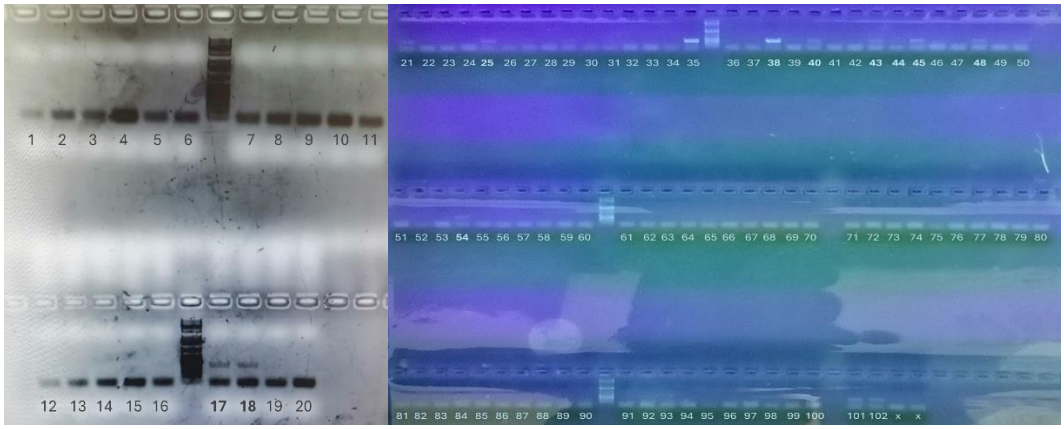



*Figure 14: Positive control PCR: Picture of the printout of an agarose gel with a positive control PCR targeting ACTIN. Only the genotypes 47 to 85 are pictured. Ladder: 1 kb Plus DNA Ladder.*

These same plants were genotyped by PCR using the NOS-BAR primers, SM2, targeting the newly chromosome-inserted BASTA resistance gene. Of the 102 plants screened, Figure 15a, only fifteen scored positive, Figure 15b. The genotyping carried out so far was limited to only one portion of the transgenes. Genotyping or sequencing of the actual target genes has not yet been performed but is advised.

a.




b.

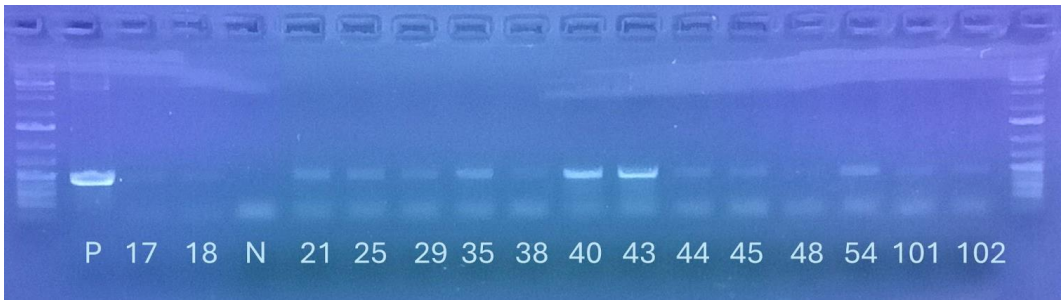



*Figure 15: Genotyping for presence of transgenes: Agarose gel pictures. The band of interest was 435 bp. The numbers identify the individual plant that was genotyped; P = pICSL11017, (pICH47732::NOSp-BAR-NOST plasmid for positive control; N = genotype number 1 as negative control. Ladder: 1 Kb Plus DNA Ladder; a. Two distinct gels used to genotype all of the 102 T1 plants. The genotypes that scored positive are identified by the bold font; b. Final agarose gel with all the genotypes of A. thaliana T1 generation that scored positive for the BASTA resistance gene.*

# 4. Discussion

## 4.1 Multiplex CRISPR/Cas9 as Means to Study Clustered Genes

This project was designed to follow up on a portion of the findings about gene clustering in relation to stamen development described by Reimegård et al. (2017). It focused on developing a gene-editing system to KO the three genes in Cluster 8. This novel method was developed because it can be challenging to stack mutations of genes in high linkage disequilibrium, which is the characteristic most defining of clustered genes. A similar Multiplex CRISPR/Cas9 system was developed by multiple research groups (e.g. Xie et al., 2015; Hui et al., 2019). In particular, Xie et al. (2015) developed a method to produce multiple sgRNAs from a single polycistronic gene by hijacking the plant's endogenous tRNA-processing system, Figure 3. Technology developed in this project was used to generate a T1 generation able to withstand BASTA herbicide selection, Figure 13, and has scored positive for genetic material corresponding to said gene, Figure 15. The introduction of indels at the tarted loci remains to be ascertained.

The three Cluster 8 GOIs are crucial for normal pollen function. Previous studies disrupting each of the two paralogues individually found them essential for normal pollen tube extension (Jiang et al., 2005). Technical limitations in the past prevented the creation of KO mutants deficient in two or more genes with high linkage disequilibrium (Xie et al., 2015), and since the two paralogous genes provide redundancy for one another (Jiang et al., 2005), this can potentially mask the true KO phenotype. The observed pollen tube elongation delay in single KO mutants might be significantly stronger if the confounding factor of redundancy was removed.

The role of *PMEIL* within Cluster 8 differs slightly from the other two. Whilst all three genes have PMEI domains, adequate production of PMEIL enzymes is necessary to ensure pollen viability and pollen tube growth. When its transcription is disrupted, such as by ABA-mediated transcriptional interference, male fertility is significantly affected (Wang et al., 2020). Knocking out the full Cluster 8 has the potential to create mutants whose vegetative growth appears normal, all the while male fertility is nearly fully, or fully, decimated.

## 4.2 Phylogeny Analysis Revealed Possible Sources of Redundancy

Phylogenetic analysis revealed that the two paralogues within Cluster 8 belong to a large gene family. According to the PhytoGenes (2024) online database, PME4/VGDH1 (AT2G47030.1) and PME5/VGD1 (AT2G47040.1) are most closely related to each other but also align with forty-eight other gene products, Figure 8a. *A. thaliana* is believed to have a 66-member gene family encoding putative PME proteins, of which ten could not be shown to be expressed at any developmental stage, making them probable pseudogenes (Louvet et al., 2006). The family of fifty splits into two branches, Figure 8a, and the branch containing the GOIs was considered further. A surface-level scan of keywords associated with the non-GOIs indicated that six have functions within the pollen, Table 4. While a more thorough literature review is necessary before drawing definitive conclusions, it is reasonable to suspect that these genes may provide some level of redundancy.

Redundancy provided by other genes may impede a thorough functional analysis. Previous studies found that knocking out *PME5/VGD1* (AT2G47040) leads to an 18% decrease in total PME activity within pollen grains (Jiang et al., 2005), implying that over 80% of this function is upheld by other PME family members. *PME4/VGDH1* (AT2G47030) is likely to be highly involved despite its expression level being 21 times lower than that of its paralogue (Jiang et al., 2005). Overall, it is likely that the workload is distributed among multiple PME family members. Additionally, pollen grains are complex structures, Figure 1b, meaning that the 80% of the PME function carried out other family members may take place in distinct areas of the grain.

Whilst the possibility of redundancy is an obstacle to functional studies, the KOs or TI of each of the Cluster 8 GOIs do result in visible phenotypes (Jiang et al., 2005; Wang et al., 2020). This suggests that PME family member-derived redundancy may be rather minimal. The only exception pertains *PME48*, whose KO phenotype mimics that of the two paralogues (TAIR, 2024). Indeed, it is worth noting that the tree constructed from aa sequence homology alone, Figure 8b, differs from the one compiled from genes mentioned in the literature during the background research, Figure 8c. Notably, *PME48* (AT5G07410) was not included in the pruned aa sequence similarity-based phylogenetic tree, but has similar KO phenotypes (TAIR, 2024). This contrast underscores the complexity of the PME gene family. Future functions studies may need consider knocking out both the Cluster 8 paralogues and *PME48*.

On a separate note, as the two paralogues also have PMEI domains, many genes on the phylogenetic tree did too, Table 4. Seventy-eight PMEI proteins, part of a large multigene family, have been identified in *A. thaliana* (Müller et al., 2013). PMEIL (AT2G47050.1) also contains this domain but was not listed in the full

phylogenetic tree compiled by PhyloGenes (2024), indicating it was not part of the list of fifty genes with protein sequence homology. This absence could be due to PMEIL sharing only 27% identity with the two paralogues or possibly because it is not included in the PhyloGenes database. Indeed, attempting to construct a phylogenetic tree in PhyloGenes (2024) for PMEIL specifically was not possible.

Using gene editing methods known to have off-target effects could, unintuitively, provide an advantage in the functional analysis of genes belonging to large gene families. As the CRISPR/Cas9 system is known for its potential to act on off-target loci (Guo et al., 2023), and it has been established that the Cluster 8 GOIs belong to PME and PMEI gene families, which are highly conserved, it is reasonable to propose that the gene-editing system designed here may act on any number of these genes. Although the chances of this occurring are very low, identifying individuals with both KOs in Cluster 8 and other PMEs suspected to provide redundancy could offer a unique perspective on pollen development. To assess the likelihood of such off-target effects taking place, the CDSs of all homologues should be compiled and loci compatible with the sgRNA spacers (allowing for two to three nucleotide mismatches) and PAMs should be identified.

### 4.2.1 Putative Proto-Cluster 8 Discovery and its Redundancy Potential

Phylogeny analysis revealed that the two paralogues most closely group with PMEI-PME37/VGDH2 (AT3G62170.1), with identity scores of 70% and 68%. PMEIL has only one homologue, AT3G62180.1 (Plant invertase/pectin methylesterase inhibitor superfamily protein), with a 54% identity score. These homologues were observed to be neighbours on the chromosome and share a similar relative positional arrangement to the GOIs in Cluster 8. Indeed, Cluster 8 is found on chromosome 2, at locus “99%” of the way. Whist many more genes are found between in it and the telomere, its position is still quite extreme in the karyotype. The same is true for the Proto-Cluster 8, found on chromosome 3, also at locus “99%” of the way. This coincidence suggests that they may have indeed arisen from a chromosomal duplication event. Additionally, their modest proximity to telomeres may indeed have benefitted their arising as this area is known for being highly dynamic (Field & Osbourn, 2012). Given that the *A. thaliana* genome underwent duplication in three distinct events (De Smet et al., 2017), it is plausible that Cluster 8 may have formed during one of these events by being duplicated from the Proto-Cluster 8 and further differentiating under distinct evolutionary pressures. A similar origin was described for the thalianol and marenal triterpenes (Field & Osbourn, 2012).

The ability of *PMEI-PME37/VGDH2* (AT3G62170) to compensate for *PME5/VGD1* (AT2G47040) in homozygous KO mutants of was tested by overexpressing it under the *PME5/VGD1* promoter in pollen and pollen tubes. None

of the over forty T1 transformants showed a restored phenotype, indicating that PMEI-PME37/VGDH2 functions differently from PME5/VGD1. Conversely, overexpressing *PME4/VGDH1* (AT2G47030) under the same promoter restored the phenotype to various degrees in 93% of T1 transformants (Jiang et al., 2005).

Unfortunately, similar analyses cannot yet be made for *PMEIL* and AT3G62180 due to insufficient information. However, given their lower identity percentages compared to the two paralogues, their potential to provide redundancy is likely even lower.

Despite the seemingly low redundancy provision potential of the Proto-Cluster 8 genes, assumptions cannot be made without empirical testing. Following the generation of homozygous Cluster 8 KO mutants, a follow-up study could be conducted whereby a T-DNA insertional mutant for either of the Proto-Cluster 8 genes could be obtained and further genetically edited to KO the other gene. Since Cluster 8 and Proto-Cluster 8 are on distinct chromosomes, they exhibit no linkage disequilibrium, making the chances of the F1 generation carrying both KOs 50%, if using a Proto-Cluster 8 heterozygous T1 individual. Removing this potential source of redundancy could reveal clearer phenotypes associated with Cluster 8 KO, but also introduce other phenotypes that are unrelated to Cluster 8.

## 4.3 The Multiplex CRISPR/Cas9 System Designed Herein has the Potential to Generate Successful Transformants

The preliminary BASTA selection, specifically the germination step, was highly efficient in starving out entire lines of non-transformed seedlings. In fact, seventeen out of twenty-one plates were completely exterminated. Only four plates persisted, showing green patches surrounded by wilted seedlings. While these patches seem to be developing well on a BASTA-rich medium at seventeen days post sowing, Figure 13a, it is likely that many are false positives. The plates were expected to host thriving seedlings surrounded by many wilted ones. Instead, clumps of greenery could be seen. Sowing seeds too close together can result in each seed receiving a smaller dose of herbicide, allowing them to develop despite having the WT genotype. Additionally, uneven suspension of the herbicide in the medium before pouring could have led to these green islands, although this possibility is low but technically feasible.

The low survival rate of seedlings may have been tied to the transformation method, as it can significantly impact its efficacy. *Arabidopsis spp* are most often transformed via floral dip, a method that, while more common, is less efficient compared to tissue regeneration methods (Hui et al., 2019). Another *Agrobacterium*-related factor to consider is that the length of the genetic material it

may integrate into the recipient's genome can vary. It is possible that the of the three transgenes only part was integrated in any of the infected individuals. As BASTA selection was carried out, the genotypes that tested positive for the *BAR* should carry at least this gene. It, however, remains unclear at this moment if any other Final Vector transgenes were integrated alongside it. It can only be verified by conducting another genotyping PCR whereby other portions of the transgenes are targeted.

The T1 adult generation did not display any obvious divergent phenotypes, Figure 13b. However, as the stamens were not specifically inspected, any divergence may have been overlooked. Currently, all that can be said with confidence is that at least fifteen T1 individuals tested positive for a portion of the inserted transgenes. As no genotyping or sequencing has yet been done for the Cluster 3 target loci, it is impossible to know if any indels have been introduced at all.

Currently not much is known about the functions of the genes found in Cluster 3, making it hard to predict what their KOs would entail. The majority of the functions related to these genes mention lipid metabolism, Table 1. Pollen is rich in the same lipids found in other parts of the plant, but their composition differs. These include: triacylglycerol (essential for pollen development and pollen tube growth), extraplastidial galactolipids (synthesized during pollen tube growth) and sphingolipids and sterols (very unique composition in pollen). The production of the pollen tube requires high levels of fatty acid and membrane lipids. The unique lipid composition of the pollen allows that to happen. Knocking out genes involved in lipid metabolism often results in the organisms being incompatible with life (Ischebeck, 2016). The Cluster 3 genes are mostly expressed in the flowering stage but AT1G20150 was also expressed in the embryo stage, Table 4. The KO phenotypes may be dependent on which of the clustered genes are KOd. Overall, the mutants may display anything from non-functional pollen grains, to misshapen flowers, all the way to seedlings dying soon after germination.

Nevertheless, the T1 generation was not expected to show significant phenotypic changes, as they still carry a paternal WT allele that can compensate for any loss of function in the other allele. This is because whilst the *BAR* was driven by the *NOS* (nopaline synthase) promoter, which is active in a wide range of plant cell types and regulated by various development and environment related factors (Kimet al., 1994); the Cas9's expression was driven by the *YAO* promoter, which is active in egg cells specifically (Hui et al., 2019). This implies that the Cas9 endonuclease can only be expressed prior to the egg being fertilised, leaving the paternal allele unaffected.

The best chance of observing any phenotype is tied to the genotype of the plant, which requires both alleles to be non-functional. Hui et al. (2019) self-crossed the T1 generation and found that 60% of the T2 offspring were homozygous,

heterozygous, or bi-allelic for the targeted mutations. Meaning that functional analyses could be carried out in as few as three generations. Ideally, double allele mutants could have been achieved in this project if the breeding of the fifteen transformant candidates had been deliberately managed. Whilst the *A. thaliana* is notorious for its self-pollination rather than outcrossing reproductive style, the latter still takes place. Reports suggest that natural populations outcross at a rate from 0.3 to over 10%, promoted by this plant's flower morphology (Luo & Widmer, 2013). As all the plants used in this project were kept crowded, the ratio of selfing to outcrossing is likely to be different from that in natural populations. With that in mind, all 102 individuals of the mature T1 generation were kept together and allowed to breed freely. Despite their selfing preference, this approach may have lowered the chances of desirable genotypes arising within the T2 generation. Indeed, a preliminary small-scale screen involving germination on BASTA-enriched medium did not yield any survivors. However, since only about one-fifth of the T2 seeds was tested, the screening can be repeated and scaled up to increase the chances of identifying the desired genotypes.

The chances of obtaining double allele mutants in the T2 generation can be increased by genotyping the candidates before flower appearance. Restricting the T1 breeding pool to individuals that tested positive for the transgenes would increase the likelihood of the T2 generation inheriting two mutated alleles to at least 50%. While inheritance calculations can become more complex when multiple genes are targeted, the use of polycistronic sgRNA genes simplifies this process. These genes generate equal amounts of each sgRNA they code for (Xie et al., 2015), so if one gene is successfully KOd, the chances are high that the other targeted genes will also be KOd.

The T1 generation has the potential to carry the targeted KO mutations. Hui et al. (2019), who used a similar approach to the one described in this report, found that about a quarter of the T1 generation were chimeric and had PCR-detectable deletions in the GOIs, with most editing occurring within somatic cells. They used the Ubiquitin promoter (*UBIp*) to drive the expression of Cas9 and the *U3 snoRNA* promoter (*U3p*) to drive the expression of the sgRNA gene. They later noted that editing efficiency in T1 plants increased when using egg or meristem-specific promoters, such as those of the genes *YAO* and *EGG CELL 1.2* instead of the constitutively expressing promoters like the 35S or *UBIp* (Hui et al., 2019). And indeed, to construct the gene editing system for this project, the *YAO* promoter was chosen to drive the expression of Cas9.

The chimerality they observed could be explained by *Agrobacterium* successfully transforming the egg cells, but as the promoters chosen were not optimal for egg cells, the transgenes may have only begun expressing after the egg cell divided into a multi-celled embryo. Chimerality poses a significant obstacle to the vertical transmission of GOI KOs. If the meristem tissue does not arise from the

cell line that was transformed, it has no transgene or targeted mutation to transmit. Using egg cell-specific promoters could lead to Cas9 being expressed directly in the egg cell, preventing the formation of chimaeras. This would result in the mature T1 plant being fully heterozygous for the inserted transgenes and the targeted KO.

It is worth noting that of the 102 individuals genotyped, only fifteen scored positive for *BAR*. If the *Agrobacterium* transformed an embryo rather than an egg cell, chimerality could become a confounding factor. The DNA was collected from a small portion of a leaf from the rosette. If any of the candidates tested were indeed chimeric for the inserted transgenes, and if the cell line that developed into the leaf was not transformed, there would be no insertion to genotype. Individual plants could score positive while only carrying the transgenes in leaves and not in the flower meristem, or vice versa. Collecting DNA material from a flower instead could provide a more accurate assessment of the transgene presence and expression in reproductive tissues.

## 4.4 The Multiplex CRISPR/Cas9 System can Introduce a Variety of Mutation Types

Whilst it is still unclear what kind of mutations within the Cluster 8 targets have the potential to be introduced, if any at all, they are most likely NHEJ-derived SNPs and short indels (Xie et al., 2015; Hui et al., 2019). However, as these genes are very close together, the individual sgRNAs may mimic the mode of action of paired sgRNAs as described by Xie et al. (2015). This may then introduce any of the combinations possible of chromosomal deletions, Figure 16. And whilst genotyping chromosomal deletions is easier as simple methods like restriction fragment length polymorphism (RFLP) analysis can be employed (Woo et al., 2015), "small" mutations can be identified by techniques such as qPCR (Falabella, et al., 2017) or T7E1 and Surveyor Mismatch Cleavage Assays (Vouillot et al., 2015).

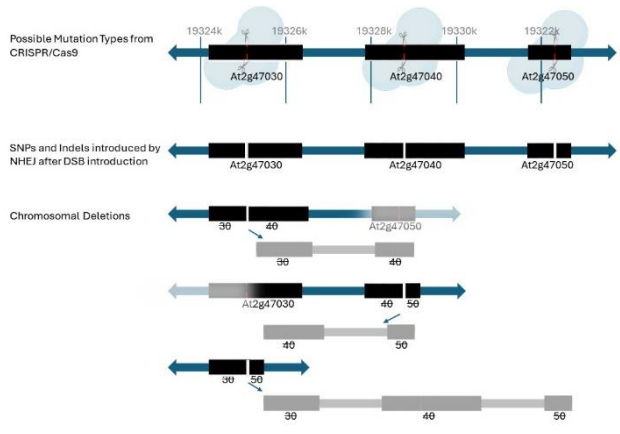



*Figure 16: Graphical representation of the possible outcomes of Cluster 8-SDRG/Cas9 system introducing double strand breaks at the target loci. The individual genes are represented by the blue boxes and the target spacer sequences are represented by the orange accent. Mutations introduced via DNA repair mechanisms are identified by a white strip. The section about chromosomal deletions shows their possible combinations in the event that the single sgRNAs mimic paired sgRNAs.*

## 4.5 Future Perspective

Plant breeding is a discipline focused on improving crop production. Success in this field is influenced by many factors, including the inherently limited genetic diversity within a crop's gene pool. However, genetic diversity can be artificially increased through methods ranging from deliberate crossings, to complex biotechnologies. Scientific advances are continuously finding new ways to introduce novel traits into staple crops. Not only have classic crop breeding methods become more efficient and systematised, but new biotechnologies are continuously being developed to introduce beneficial traits (Sharma et al., 2013; Tiwari et al., 2023).

With the imminent uncertainties brought forth by climate change, more efficient plant breeding approaches are essential for improving crop yield and quality. This need is particularly pressing as the expanding global population is expected to drastically increase demand (Steinwand & Ronald, 2020). These approaches can also reduce the environmental impact by lessening the reliance on pesticides, herbicides, and water, which in turn can help decrease greenhouse gas emissions (Brookes, 2021; Das et al., 2023).

The vast majority of globally cultivated crops are flowering plants. Leff et al. (2004) list about 150 crop types, noting varying levels of diversity among them (Leff et al., 2004). In contrast, it can be difficult to think of crops of non-angiosperm origin, beyond mushrooms, pine nuts and some select algae. The importance of flowering plants to agriculture and global ecosystems is evident. Thus, understanding the processes involved in the regulation of flowering is crucial for science's contribution to the current food supply chain.

This project is highly influenced by the findings of Xie et al. (2015). Their methods were used to design a DNA construct containing the Cas9 endonuclease followed by two unique sgRNAs, preceded by pre-tRNA sequences. This synthetic dicistronic sgRNA gene (*SDRG*) was designed to simultaneously produce two sgRNAs with the aim of targeting and knocking out a cluster of genes involved in pollen development. This Multiplex CRISPR/Cas9 system was assembled per the principles of the Golden Gate Assembly. Successfully knocking out all targeted genes will allow exploration of the functions of this gene cluster and clarify the significance of their co-regulation. Currently, little is known about the need for their co-expression or the underlying regulatory mechanisms.

To our current understanding, whilst single gene KOs have been carried out (Jiang et al., 2005), a full Cluster 8 KO has not yet been attempted. The results described here have the potential to be used for the introduction of the triple KO and further phenotypic observations. Accomplishing so could provide novel insights into the function of this genetic locus. Plants generated in this study are also suitable for further research, particularly involving targeted crossings for the generation of homozygous or bi-allelic individuals.

This project may have a chance to contribute practically to the breeder's "male sterility" toolbox. Male sterility is a well-established tool for producing field crops that exhibit hybrid vigour, or heterosis. Hybrid vigour is a phenomenon where the filial generation (F1) demonstrates increased yield, stress tolerance, and adaptability compared to the parental generation (P0). By inducing male sterility in hermaphroditic plants, self-pollination is prevented, allowing for controlled pollination. This control is crucial for selecting pollen donors that maximise the heterosis effects. Traditionally, male sterility was achieved through the manual removal of anthers. However, today the most commonly used method is cytoplasmic male sterility (CMS). CMS works by rendering the plant's male gametophytes, or pollen, non-functional (Bohra et al., 2016). Cluster 8 KO has the potential to provide an alternative route to the achievement of male sterility in crops. Harnessing hybrid vigour is essential for meeting the increasing demand for higher yield and improved stress tolerance in field crops. This approach helps sustain the enhanced yield gains necessary to support global agricultural productivity.

## 4.6 Conclusion

In conclusion, this project focused on developing a Cluster 8-SDRG/Cas9 system to KO clustered genes related to pollen development. A follow-up project, cantered on generating and phenotyping the resulting mutants, could verify the accuracy of the initial predictions. It is anticipated that the homozygous or bi-allelic Cluster 8 *A. thaliana* mutants is most likely exhibit mostly normal vegetative growth but significant disruption in pollen development. Indeed, whilst *PME5/VGD1* and *PME4/VGDH1* facilitate normal pollen tube extension (Jiang et al., 2005), *PMEIL* is responsible for the pollen's viability (Wang et al., 2020).

Phylogeny analyses revealed multiple sources of genetic redundancy, complicating future functional studies. Future research may benefit from knocking out Cluster 8 and Proto-Cluster 8 and knocking out the two paralogous genes in conjunction with *PME48*, to quantify their influence upon pollen viability and pollen tube extension. Knowledge gained from these follow-up studies could help enrich the current understanding of the flower developmental process and have practical applications in developing alternative methods to induce male sterility, thereby providing a new tool for harnessing hybrid vigour.

# References


Bohra, A., Jha, U.C., Adhimoolam, P., Bisht, D. and Singh, N.P., (2016). Cytoplasmic male sterility (CMS) in hybrid breeding in field crops. *Plant Cell Reports*, *35*(5), pp.967-993. https://doi.org/10.1007/s00299-016-1949-3

Bouché, N. and Bouchez, D., (2001). Arabidopsis gene knockout: phenotypes wanted. *Current opinion in plant biology*, *4*(2), pp.111-117. https://doi.org/10.1016/S1369-5266(00)00145-X

Brock, M.T., Tiffin, P. and Weinig, C., 2007. Sequence diversity and haplotype associations with phenotypic responses to crowding: GIGANTEA affects fruit set in *Arabidopsis thaliana*. *Molecular Ecology*, *16*(14), pp.3050-3062. https://doi.org/10.1111/j.1365-294X.2007.03298.x

Brookes, G., (2021). Environmental Impacts of Genetically Modified (GM) Crop Use: Impacts on Pesticide Use and Carbon Emissions. In: Ricroch, A., Chopra, S., Kuntz, M. (eds) *Plant Biotechnology: Experience and Future Prospects*. Springer, Cham. pp.87-101. https://doi.org/10.1007/978-3-030-68345-0_7

Castel, B., Tomlinson, L., Locci, F., Yang, Y. and Jones, J.D., (2019). Optimization of T-DNA architecture for Cas9-mediated mutagenesis in *Arabidopsis*. *PloS one*, *14*(1), pp.1-20. https://doi.org/10.1371/journal.pone.0204778

Charmont, S., Jamet, E., Pont-Lezica, R. and Canut, H., (2005). Proteomic analysis of secreted proteins from *Arabidopsis thaliana* seedlings: improved recovery following removal of phenolic compounds. *Phytochemistry*, *66*(4), pp.453-461. https://doi.org/10.1016/j.phytochem.2004.12.013

Clough, S.J. and Bent, A.F., 1998. Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *The plant journal*, *16*(6), pp.735-743. doi: 10.1046/j.1365-313x.1998.00343.x.

Crang, R., Lyons-Sobaski, S., Wise, R., (2018). Flowers and Male Reproductive Structures. In: Crang, R., Lyons-Sobaski, S. and Wise, R. *Plant Anatomy: A Concept-Based Approach to the Structure of Seed Plants*. Springer Cham. pp.579-613. https://doi.org/10.1007/978-3-319-77315-5_20

Das, A., Mahanta, M., Pramanik, B. and Purkayastha, S., (2023). Genetically Modified Crops and Crop Species Adapted to Global Warming in Dry Regions. In: Anandkumar Naorem, Deepesh Machiwal (eds) *Enhancing Resilience of Dryland Agriculture Under Changing Climate: Interdisciplinary and Convergence Approaches*. Springer Nature Singapore. pp.385-409. https://doi.org/10.1007/978-981-19-9159-2_19

Dieci, G., Fiorino, G., Castelnuovo, M., Teichmann, M. and Pagano, A., (2007). The expanding RNA polymerase III transcriptome. *TRENDS in Genetics*, *23*(12), pp.614-622. https://doi.org/10.1016/j.tig.2007.09.001

Engler, C., Gruetzner, R., Kandzia, R. and Marillonnet, S., (2009). Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PloS one*, *4*(5), pp.e5553. https://doi.org/10.1371/journal.pone.0005553

Engler, C., Kandzia, R. and Marillonnet, S., (2008). A one pot, one step, precision cloning method with high throughput capability. *PloS one*, *3*(11), p.e3647. https://doi.org/10.1371/journal.pone.0003647

Engler, C. and Marillonnet, S., (2011). Generation of Families of Construct Variants Using Golden Gate Shuffling. In: Lu, C., Browse, J., Wallis, J. (eds) *cDNA Libraries: Methods and Applications.* Methods in Molecular Biology, vol 729. Humana Press. pp.167-181. https://doi.org/10.1007/978-1-61779-065-2_11

Falabella, M., Sun, L., Barr, J., Pena, A.Z., Kershaw, E.E., Gingras, S., Goncharova, E.A. and Kaufman, B.A., (2017). Single-step qPCR and dPCR detection of diverse CRISPR-Cas9 gene editing events in vivo. *G3: Genes, Genomes, Genetics*, *7*(10), pp.3533-3542. https://doi.org/10.1534/g3.117.300123

Field, B. and Osbourn, A., (2012). Order in the playground: Formation of plant gene clusters in dynamic chromosomal regions. *Mobile genetic elements*, *2*(1), pp.46-50. https://doi.org/10.4161/mge.19348

Frey, M., Chomet, P., Glawischnig, E., Stettner, C., Grun, S., Winklmair, A., Eisenreich, W., Bacher, A., Meeley, R.B., Briggs, S.P. and Simcox, K., (1997). Analysis of a chemical plant defence mechanism in grasses. *Science*, *277*(5326), pp.696-699. https://www.science.org/doi/10.1126/science.277.5326.696

Govaerts, R., Nic Lughadha, E., Black, N., Turner, R. and Paton, A., (2021). The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. *Scientific Data*, *8*(215), pp.1-10. https://doi.org/10.1038/s41597-021-00997-6

Guo, C., Ma, X., Gao, F. and Guo, Y., (2023). Off-target effects in CRISPR/Cas9 gene editing. *Frontiers in bioengineering and biotechnology*, *11*. doi: 10.3389/fbioe.2023.1143157

Hui, L., Zhao, M., He, J., Hu, Y., Huo, Y., Hao, H., Hao, Y., Zhu, W., Wang, Y., Xu, M. and Fu, A., (2019). A simple and reliable method for creating PCR-detectable mutants in Arabidopsis with the polycistronic tRNA–gRNA CRISPR/Cas9 system. *Acta Physiologiae Plantarum*, *41*(170), pp.1-14. https://doi.org/10.1007/s11738-019-2961-3

Ischebeck, T., 2016. Lipids in pollen—They are different. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids*, *1861*(9), pp.1315-1328. https://doi.org/10.1016/j.bbalip.2016.03.023

Jiang, K., Wang, Q., Dimitrov, D., Luo, A., Xu, X., Su, X., Liu, Y., Li, Y., Li, Y. and Wang, Z., (2023). Evolutionary history and global angiosperm species richness–climate relationships. *Global Ecology and Biogeography*, *32*(7), pp.1059-1072. https://doi.org/10.1111/geb.13687

Jiang, L., Yang, S.L., Xie, L.F., Puah, C.S., Zhang, X.Q., Yang, W.C., Sundaresan, V. and Ye, D., (2005). VANGUARD1 encodes a pectin methylesterase that enhances pollen tube growth in the Arabidopsis style and transmitting tract. *The Plant Cell*, *17*(2), pp.584-596. https://doi.org/10.1105/tpc.104.027631

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E., (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*(6096), pp.816-821. https://www.science.org/doi/10.1126/science.1225829

Jones, H.D., (2015). GM foods: is there a way forward?. *Proceedings of the Nutrition Society*, *74*(3), pp.198-201. https://doi.org/10.1017/S0029665115002098

Kim, Y., Buckley, K., Costa, M.A. and An, G., 1994. A 20 nucleotide upstream element is essential for the nopaline synthase (nos) promoter activity. *Plant molecular biology*, *24*, pp.105-117. https://doi.org/10.1007/BF00040578

Koornneef, M. and Meinke, D., (2010). The development of Arabidopsis as a model plant. *The Plant Journal*, *61*(6), pp.909-921. https://doi.org/10.1111/j.1365-313X.2009.04086.x

Kotera, I. and Nagai, T., (2008). A high-throughput and single-tube recombination of crude PCR products using a DNA polymerase inhibitor and type IIS restriction enzyme. *Journal of Biotechnology*, *137*(1-4), pp.1-7. https://doi.org/10.1016/j.jbiotec.2008.07.1816

Kruszka, K., Barneche, F., Guyot, R., Ailhas, J., Meneau, I., Schiffer, S., Marchfelder, A. and Echeverría, M., (2003). Plant dicistronic tRNA–snoRNA genes: a new mode of expression of the small nucleolar RNAs processed by RNase Z. *The EMBO journal*, *22*(3), pp.621-632. https://doi.org/10.1093/emboj/cdg040

Lee, J.H., Skowron, P.M., Rutkowska, S.M., Hong, S.S. and Kim, S.C., (1996). Sequential amplification of cloned DNA as tandem multimers using class-IIS restriction enzymes. *Genetic analysis: biomolecular engineering*, *13*(6), pp.139-145. https://doi.org/10.1016/S1050-3862(96)00164-7

Leff, B., Ramankutty, N. and Foley, J.A., (2004). Geographic distribution of major crops across the world. *Global biogeochemical cycles*, *18*(1), pp.1-27. https://doi.org/10.1029/2003GB002108

Louvet, R., Cavel, E., Gutierrez, L., Guénin, S., Roger, D., Gillet, F., Guerineau, F. and Pelloux, J., (2006). Comprehensive expression profiling of the pectin methylesterase gene family during silique development in *Arabidopsis thaliana*. *Planta*, *224*(4), pp.782-791. DOI 10.1007/s00425-006-0261-9 https://www.jstor.org/stable/23389479

Lu, J.Y., Xiong, S.X., Yin, W., Teng, X.D., Lou, Y., Zhu, J., Zhang, C., Gu, J.N., Wilson, Z.A. and Yang, Z.N., (2020). MS1, a direct target of MS188, regulates the expression of key sporophytic pollen coat protein genes in *Arabidopsis*. *Journal of Experimental Botany*, *71*(16), pp.4877-4889. https://doi.org/10.1093/jxb/eraa219

Luo, Y. and Widmer, A., (2013). Herkogamy and its effects on mating patterns in *Arabidopsis thaliana*. *PLoS One*, *8*(2), p.e57902. doi:10.1371/journal.pone.0057902

Mugford, S.T., Qi, X., Bakht, S., Hill, L., Wegel, E., Hughes, R.K., Papadopoulou, K., Melton, R., Philo, M., Sainsbury, F. and Lomonossoff, G.P., (2009). A serine carboxypeptidase-like acyltransferase is required for synthesis of antimicrobial compounds and disease resistance in oats. *The Plant Cell*, *21*(8), pp.2473-2484. https://doi.org/10.1105/tpc.109.065870

Müller, K., Levesque-Tremblay, G., Fernandes, A., Wormit, A., Bartels, S., Usadel, B. and Kermode, A., (2013). Overexpression of a pectin methylesterase inhibitor in *Arabidopsis thaliana* leads to altered growth morphology of the stem and defective organ separation. *Plant signaling & behavior*, *8*(12), p.e26464. https://doi.org/10.4161/psb.26464

Pacini, E. and Dolferus, R., (2016). The trials and tribulations of the plant male gametophyte—Understanding reproductive stage stress tolerance. In: Arun K. Shanker and Chitra Shanker (eds) *Abiotic and Biotic Stress in Plants - Recent Advances and Future Perspectives.* IntechOpen. pp. 703-754. http://dx.doi.org/10.5772/61671

Padgett, K.A. and Sorge, J.A., (1996). Creating seamless junctions independent of restriction sites in PCR cloning. *Gene*, *168*(1), pp.31-35. https://doi.org/10.1016/0378-1119(95)00731-8

Peng, J., (2019). Gene redundancy and gene compensation: An updated view. *Journal of Genetics and Genomics*, *46*(7), pp.329-333. https://doi.org/10.1016/j.jgg.2019.07.001

Qi, X., Bakht, S., Leggett, M., Maxwell, C., Melton, R. and Osbourn, A., (2004). A gene cluster for secondary metabolism in oat: implications for the evolution of metabolic diversity in plants. *Proceedings of the National Academy of Sciences*, *101*(21), pp.8233-8238. https://doi.org/10.1073/pnas.0401301101

Qi, X., Bakht, S., Qin, B., Leggett, M., Hemmings, A., Mellon, F., Eagles, J., Werck-Reichhart, D., Schaller, H., Lesot, A. and Melton, R., (2006). A different function for a member of an ancient and highly conserved cytochrome P450 family: from essential sterols to plant defence. *Proceedings of the National Academy of Sciences*, *103*(49), pp.18848-18853. https://doi.org/10.1073/pnas.0607849103

Quiroz, S., Yustis, J.C., Chávez-Hernández, E.C., Martínez, T., Sanchez, M.D.L.P., Garay-Arroyo, A., Álvarez-Buylla, E.R. and García-Ponce, B., (2021). Beyond the genetic pathways, flowering regulation complexity in *Arabidopsis thaliana*. *International Journal of Molecular Sciences*, *22*(11), pp.1-28 or pp.5716. https://doi.org/10.3390/ijms22115716

Reimegård, J., Kundu, S., Pendle, A., Irish, V.F., Shaw, P., Nakayama, N., Sundström, J.F. and Emanuelsson, O., (2017). Genome-wide identification of physically clustered genes suggests chromatin-level co-regulation in male reproductive development in *Arabidopsis thaliana*. *Nucleic Acids Research*, *45*(6), pp.3253-3265. https://doi.org/10.1093/nar/gkx087

Scott, R.J., Spielman, M. and Dickinson, H., (2004). Stamen structure and function. *The Plant Cell*, *16*(suppl_1), pp.S46-S60. https://doi.org/10.1105/tpc.017012

Shamshirgaran, Y., Liu, J., Sumer, H., Verma, P.J. and Taheri-Ghahfarokhi, A., (2022). Tools for efficient genome editing; ZFN, TALEN, and CRISPR. In: Verma, P.J., Sumer, H., Liu, J. (eds) *Applications of genome modulation and editing*. Methods in Molecular Biology, vol 2495. Humana, New York, NY. pp.29-46. https://doi.org/10.1007/978-1-0716-2301-5_2

Sharma, S., Upadhyaya, H.D., Varshney, R.K. and Gowda, C.L.L., (2013). Pre-breeding for diversification of primary gene pool and genetic enhancement of grain legumes. *Frontiers in plant science*, *4*(Article 309), pp.1-14. https://doi.org/10.3389/fpls.2013.00309

Shimura, K., Okada, A., Okada, K., Jikumaru, Y., Ko, K.W., Toyomasu, T., Sassa, T., Hasegawa, M., Kodama, O., Shibuya, N. and Koga, J., (2007). Identification of a biosynthetic gene cluster in rice for momilactones. *Journal of Biological Chemistry*, *282*(47), pp.34013-34018. https://doi.org/10.1074/jbc.M703344200

De Smet, R., Sabaghian, E., Li, Z., Saeys, Y. and Van de Peer, Y., (2017). Coordinated functional divergence of genes after genome duplication in *Arabidopsis thaliana*. *The plant cell*, *29*(11), pp.2786-2800. https://doi.org/10.1105/tpc.17.00531

Steinwand, M.A. and Ronald, P.C., (2020). Crop biotechnology and the future of food. *Nature Food*, *1*, pp.273-283. https://doi.org/10.1038/s43016-020-0072-3

Swaminathan, S., Morrone, D., Wang, Q., Fulton, D.B. and Peters, R.J., (2009). CYP76M7 is an ent-cassadiene C11α-hydroxylase defining a second multifunctional diterpenoid biosynthetic gene cluster in rice. *The Plant Cell*, *21*(10), pp.3315-3325. https://doi.org/10.1105/tpc.108.063677

Takos, A.M., Knudsen, C., Lai, D., Kannangara, R., Mikkelsen, L., Motawia, M.S., Olsen, C.E., Sato, S., Tabata, S., Jørgensen, K. and Møller, B.L., (2011). Genomic clustering of cyanogenic glucoside biosynthetic genes aids their identification in Lotus japonicus and suggests the repeated evolution of this chemical defence pathway. *The Plant Journal*, *68*(2), pp.273-286. https://doi.org/10.1111/j.1365-313X.2011.04685.x

Tiwari, A., Tikoo, S.K., Angadi, S.P., Kadaru, S.B., Ajanahalli, S.R. and Vasudeva Rao, M.J., (2023). Use of Molecular Technologies in Plant Breeding. In: (no eds) *Market-Driven Plant Breeding for Practicing Breeders.* Singapore: Springer Nature Singapore. pp.157-203. https://doi.org/10.1007/978-981-19-5434-4_5

Vouillot, L., Thélie, A. and Pollet, N., (2015). Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. *G3: Genes, Genomes, Genetics*, *5*, pp.407-415. doi: 10.1534/g3.114.015834

Wang, H.H., Qiu, Y., Yu, Q., Zhang, Q., Li, X., Wang, J., Li, X., Zhang, Y. and Yang, Y., (2020). Close arrangement of CARK3 and PMEIL affects ABA-mediated pollen sterility in *Arabidopsis thaliana*. *Plant, Cell & Environment*, *43*(11), pp.2699-2711. https://doi.org/10.1111/pce.13871

Wang, Q., Hillwig, M.L. and Peters, R.J., (2011). CYP99A3: functional identification of a diterpene oxidase from the momilactone biosynthetic gene cluster in rice. *The Plant Journal*, *65*(1), pp.87-95. https://doi.org/10.1111/j.1365-313X.2010.04408.x

Wang, Y., Zhang, W.Z., Song, L.F., Zou, J.J., Su, Z. and Wu, W.H., (2008). Transcriptome analyses show changes in gene expression to accompany pollen germination and tube growth in *Arabidopsis*. *Plant Physiology*, *148*(3), pp.1201-1211. https://doi.org/10.1104/pp.108.126375

Wegel, E., Koumproglou, R., Shaw, P. and Osbourn, A., (2009). Cell type–specific chromatin decondensation of a metabolic gene cluster in oats. *The plant cell*, *21*(12), pp.3926-3936. https://doi.org/10.1105/tpc.109.072124

Wilderman, P.R., Xu, M., Jin, Y., Coates, R.M. and Peters, R.J., (2004). Identification of syn-pimara-7, 15-diene synthase reveals functional clustering of terpene synthases involved in rice phytoalexin/allelochemical biosynthesis. *Plant Physiology*, *135*(4), pp.2098-2105. https://doi.org/10.1104/pp.104.045971

Woo, J.W., Kim, J., Kwon, S.I., Corvalán, C., Cho, S.W., Kim, H., Kim, S.G., Kim, S.T., Choe, S. and Kim, J.S., (2015). DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nature biotechnology*, *33*(11), pp.1162-1164. https://doi.org/10.1038/nbt.3389

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

Yao, X., Hu, W. and Yang, Z.N., (2022). The contributions of sporophytic tapetum to pollen formation. *Seed Biology*, *1*(Article 5), pp.1-13. https://doi.org/10.48130/SeedBio2022-0005

Zhang, J., (2012). Genetic Redundancies and Their Evolutionary Maintenance. In: Soyer, O. (eds) *Evolutionary Systems Biology.* Advances in Experimental Medicine and Biology, vol 751. Springer, New York, NY. pp.279-300. https://doi.org/10.1007/978-1-4614-3567-9_13

# Database and Software Resources

Jbrowse, 2024. https://jbrowse.arabidopsis.org/?data=Araport11
Multiple Primer Analyzer (Thermo Fisher Scientific, 2024).
https://www.thermofisher.com/se/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/multiple-primer-analyzer.html
Neb Converter, 2024 (Mass from Moles Converter).
https://nebiocalculator.neb.com/#!/dsdnaamt tab “Moles → Mass”
PhyloGenes, 2024. https://phylogenes.arabidopsis.org/tree
Phytozome, 2024. https://phytozome-next.jgi.doe.gov/
TAIR, 2024. https://www.arabidopsis.org

BioEdit: https://bioedit.software.informer.com/
GeneDoc: https://nrbsc.org/gfx/genedoc
Geneious: https://www.geneious.com/
MEGA11: https://www.megasoftware.net/

# Unlocking the Secrets of Plant Fertility: A Gene Editing Adventure

Plant breeding has always aimed to create more productive, tastier, and resilient crops. Achieving these goals isn't just about mixing and matching different varieties; it often requires understanding the specific genes responsible for various traits. One of the most exciting tools available for this purpose is CRISPR/Cas9, a powerful gene-editing technology that lets scientists precisely tweak plant DNA.

In our project, we explored a specific group of genes in Cluster 8, identified by previous research, which are vital for pollen development in the model plant *Arabidopsis thaliana*. Pollen is crucial for plant reproduction, so understanding these genes could have significant implications for crop fertility and productivity.

Previous studies suggested that Cluster 8 genes might compensate for each other if one becomes non-functional, making it hard to see their true effects. This genetic "backup system" is known as redundancy. To overcome this challenge, we designed a Multiplex CRISPR/Cas9 system to disrupt the function of all three Cluster 8 genes simultaneously.

Our phylogenetic analysis revealed that two Cluster 8 genes are part of a larger family of related genes, many of which might also support pollen development. This redundancy means that disrupting Cluster 8 alone might not show much effect because other genes step in to help. Interestingly, we found that genes closely related to Cluster 8, which we called Proto-Cluster 8, might have a historical connection due to genome duplication events. These genes also seemed to play a role in pollen development, adding another layer of complexity.

To fully understand the role of Cluster 8 and its relatives, future studies will need to create plants that lack both Cluster 8 and Proto-Cluster 8 genes. This could help reveal hidden effects and improve our understanding of pollen development. The knowledge gained from this research could have real-world applications, such as developing crops with male sterility. This trait is useful for producing hybrids that combine the best qualities of different plants, leading to better yields and more resilient crops.

In conclusion, while our project faced challenges, it highlighted the intricate web of plant genetic interactions. By continuing to unravel these mysteries, we can pave the way for innovations in agriculture that will help feed a growing global population.

# Acknowledgements

I would like to express my sincere gratitude to several individuals whose support and guidance were crucial to the success of this project.

Firstly, my deepest thanks to my supervisor, Jens Sundström, for his unwavering support and insightful feedback throughout this research.

I am also profoundly grateful to my assistant supervisor, Sonam Yadav, for her involved supervision and continuous mentorship.

Additionally, I appreciate my colleagues and the entire research department for their assistance. With special thanks to Laxmi Mishra and Luboš Říha.

This work would not have been possible without the collective effort and support of everyone involved.

Thank you.

# Supplementary Material List

SM1: Alignment
SM2: Primers, Oligonucleotides and Other Sequences
SM2: Secondary structures
SM4: T1 generation plated February 1st

## Publishing and archiving

Approved students' theses at SLU are published electronically. As a student, you have the copyright to your own work and need to approve the electronic publishing. If you check the box for **YES**, the full text (pdf file) and metadata will be visible and searchable online. If you check the box for **NO**, only the metadata and the abstract will be visible and searchable online. Nevertheless, when the document is uploaded it will still be archived as a digital file. If you are more than one author, the checked box will be applied to all authors. You will find a link to SLU's publishing agreement here:

- https://libanswers.slu.se/en/faq/228318.

☒ YES, I/we hereby give permission to publish the present thesis in accordance with the SLU agreement regarding the transfer of the right to publish a work.

☐ NO, I/we do not give permission to publish the present work. The work will still be archived and its metadata and abstract will be visible and searchable.