Clément Roumégoux

Biotechnology for Sustainability

Toward a role for LOTUS domain containing proteins during sexual reproduction in *Arabidopsis thaliana*

Dissertation for the obtention of the master's degree in

Biotechnology for Sustainability

Supervisor: Jörg Becker, principal investigator, ITQB

Oeiras, 2022

Clément Roumégoux

Biotechnology for Sustainability

Toward a role for LOTUS domain containing proteins during sexual reproduction in *Arabidopsis thaliana*

Dissertation for the obtention of the master's degree in

Biotechnology for Sustainability

Supervisor: Jörg Becker, principal investigator, ITQB

Jury: Isabel Abreu

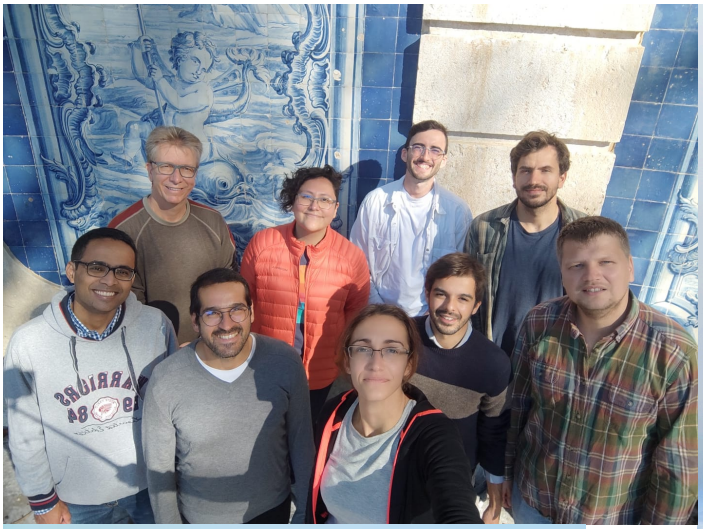
Arguer: Maria Fernanda Niño-González

Vowels: Jörg Becker

Location for dissertation defense: Instituto de Tecnologia Química e Biológica António Xavier (ITQB NOVA), Oeiras, 30th of November 2022

Acknowledgments

No words can express how grateful I am, but pictures can.



Abstract

Plant reproduction study is key to increase crop yield and resilience to climate change. The LOTUS domain comprises ~100 amino acid and structurally form a winged helix-turn-helix. In animal, LOTUS domain containing proteins (LDCPs) have been associated with oogenesis and spermatogenesis but their role in plant remains elusive. The goal of this project is to understand the role of LDCPs in the reproductive process of *Arabidopsis thaliana*. Data acquired from gene expression suggested that gene AT2G05160 (*LDCP1*) and AT3G52980 (*LDCP2*) are specifically and highly expressed in reproductive tissues. Previously published results indicate *LDCP2* to be important for pollen development. The role of *LDCP1* and *LDCP2* was investigated through reverse genetic approaches using T-DNA lines and CRISPR-Cas9 technology to obtain mutants, including a double mutant. Characterization of the lines included the comparison of seed set, germination rate and morphology of pollen against those of wild-type plants. The results obtained showed that LDCPs may have an important role in pollen formation. None of the single mutants showed a significant disturbance in gametogenesis. However, the double heterozygous (*ldcp1/+ ldcp2/+*) plants revealed defects in pollen development, leading to a ~50% reduction in the number of seeds per silique. This result is contrasted by the full recovery of normal seed set in the double homozygous mutant (*ldcp1 ldcp2*). Furthermore, the peculiar reverse transcription PCR pattern and the pollen defect in a heterozygous line but not in a homozygous hint at a possible technical problem related to T-DNA lines. In parallel, we developed CRISPR-Cas9 mutants targeting both genes at the same time. We were able to introduce the construct into the plants but could not identify mutant. As a conclusion, it is worth considering using CRISPR technology to clarify whether the gametogenesis defect is linked to a technical issue or an actual phenotype.

Key words: Arabidopsis, LOTUS Domain, T-DNA mutant, CRISPR-Cas9

Resumo

Estudos reprodutivos em plantas são fundamentais para aumentar o rendimento de culturas e a resiliência às mudanças climáticas. O domínio LOTUS é composto por ~100 aminoácidos e forma estruturalmente uma hélice-volta-hélice alada. Em animais, proteínas que contêem o domínio LOTUS (LDCPs) têm sido associadas à oogênese e espermatogênese, mas o seu papel em plantas permanece indefinido. O objetivo deste projeto é compreender o papel dos LDCPs no processo reprodutivo de *Arabidopsis thaliana*. Os dados adquiridos da expressão genética sugeriram que os genes AT2G05160 (*LDCP1*) e AT3G52980 (*LDCP2*) são especifica e altamente expressos em tecidos reprodutivos. Resultados previamente publicados indicam que o gene LDCP2 é importante para o desenvolvimento do pólen. O papel dos genes *LDCP1* e *LDCP2* foi investigado por uma abordagem genética reversa usando linhas de T-DNA e tecnologia CRISPR-Cas9 para obter mutantes, incluindo um mutante duplo. A caracterização das linhagens obtidas incluiu a comparação do conjunto de sementes, taxa de germinação e morfologia do pólen com plantas wild-type. Os resultados obtidos mostraram que os LDCPs podem ter um papel importante na formação do pólen. Nenhum dos mutantes únicos apresentou diferências significativas na gametogênese. No entanto, as plantas duplamente heterozigoticas (*ldcp1/+ ldcp2/+*) revelaram defeitos no desenvolvimento do pólen, levando a uma redução de ~50% no número de sementes por síliqua. Este resultado é contrastado pela recuperação completa do conjunto de sementes normais no mutante homozigotico duplo (*ldcp1 ldcp2*). Além disso, o padrão peculiar de PCR de transcrição reversa e o defeito do pólen na linha heterozigotica, mas não na linha homozigotica, sugerem um possível problema técnico relacionado às linhas de T-DNA. Em paralelo, desenvolvemos mutantes CRISPR-Cas9 visando ambos os genes ao mesmo tempo. Conseguimos introduzir o vector nas plantas, mas até agora não conseguimos identificar mutantes. Como conclusão, é importante considerar o uso da tecnologia CRISPR para esclarecer se o defeito na gametogênese está relacionado com um problema técnico ou a um fenótipo real

Palavras-chave: Arabidopsis, Domínio LOTUS, T-DNA mutante, CRISPR-Cas9

Table of Contents

List of figures and tables

INTRODUCTION

MATERIALS AND METHODS

RESULTS

SUPPLEMENTAL MATERIAL

List of abbreviations

Cas9 – CRISPR-associated protein 9

Col-0 – Columbia 0 variety

CRISPR – clustered regularly interspaced short palindromic repeats

DNA – Deoxyribonucleic acid

gDNA – Genomic deoxyribonucleic acid

LDCP – LOTUS domain containing protein

LOTUS – Limkain, Oskar and Tudor domain-containing proteins

MT – Mutant

O/N - Overnight

PCR – Polymerase chain reaction

RNA – Ribonucleic acid

RT-PCR – reverse-transcriptase polymerase chain reaction

sgRNA – single guide ribonucleic acid

T-DNA – Transfer deoxyribonucleic acid

TE – Transposable element

WT – Wild type

Introduction

Contextualization

Any change in the planet's climate, whether permanent or temporary, is referred to as climate change. It is the sum of two related sources of climate change: anthropogenic and natural climate change. However, the most violent climatic event has occurred over the last two centuries and was, with a 95 percent certainty, of anthropogenic source (IPCC, 2013). Since pre-industrial age, the average global temperature has increased by nearly 1°C and if this trend persists, global warming will eventually reach, with a high confidence, 1.5°C in the next two decades. (IPCC, 2022).

Globally, agriculture and related land uses accounted for 17% of global emissions in 2018 (FAO, 2020). The likelihood of experiencing more frequent extreme weather events, drought, river flooding, and chronic sea level rise along with global temperature increases are expected to increase. This places agriculture as a direct victim of climate change as a 17% reduction in global yield by 2050 is expected (Nelson et al, 2014). Moreover, according to projections, geographical north–south inequalities might arise with regions in southern Europe being affected more often by global warming than those in the north. Elevated temperatures, limited precipitation and risk of wildfire in southern Europe are expected to reduce soil water availability, among other factors, ultimately impacting yields (EPRS, 2020).

To efficiently respond to these challenges, measures around the world are being implemented. Among them, gene editing has appeared as an approach to limit the effects of climate change on agriculture, through better understanding of biological systems and implementation of genetically improved organisms. Environmental domestication is not new. Starting 12,000 years ago, humans have been selectively breeding for crop traits such as reduced growing seasons, enhanced resistance to diseases and pests, increased seeds and fruits size and nutritional content (Purugganan, 2019). As our understanding of plant biology improved with an exponentially growing population, we switched our approach from classical breeding to the less time-consuming random mutagenesis and finally to precise genome editing (Wieczorek & Wright, 2012). Seeds improvement and adaptation to climate change are of particular interest. The foundation of our societies is based on seed production. Historically, the common point among emerging states (around 3000-year BC) is the production of large quantities of cereals. As discussed by James C. Scott, only seeds are adapted to the concentration of wealth, taxation, long-term storage and rationing (Scott, 2018). Nowadays, plants grown from seeds and cereals have kept their critical importance from a cultural, nutritional, and economical point of view (food, feed, fiber and fuel). Among the steps of seed production, plant sexual reproduction is of critical importance, especially considering the sensitivity reported for climate change. So far, multiple undesired effects have been linked to higher temperature such as a reduction in pollen viability, pollen germination rates, and pollen tube growth rate and seed set (Hedhly et al, 2009). Plant reproduction studies have the potential to help increase crop yields and provide food security, as well as guide effective conservation strategies in the face of unparalleled population growth and biodiversity loss.

Arabidopsis thaliana as a model to understand genetic function

Arabidopsis thaliana is a relatively small flowering plant that is frequently used in plant biology as a model organism. Arabidopsis belongs to the Brassicaceae family, which includes cultivated plants like cabbage and radish. Arabidopsis is not relevant for agriculture, but it has significant benefits for fundamental research in genetic and molecular biology. Arabidopsis offers a number of advantages as a research model. Firstly, its minimal maintenance and modest space needs allow the production of vast populations of plants in quite basic facilities. Secondly, Arabidopsis only possesses five chromosomes, which is unusual for a flowering plant. Thirdly, and most crucially, because of its short generation period and high number of offspring, each specimen can generate a vast number of seeds in a relatively short time (Leonelli, 2007). Arabidopsis was adopted as the plant model organism in the 1980s and extensively studied since then. A search on Scopus (the largest abstract and citation database of peer-reviewed literature) for the term “Arabidopsis” results in over 100,000 mentions in either the title, abstract or keywords of research papers, giving an idea of the importance of this small plant in the research world (Scopus, 2022). Arabidopsis is the model organism to identify crop gene function. Crop genetic characterization is extremely expensive and laborious, where key components such as whole genome sequencing and other molecular tools are still lagging behind. Over the years, Arabidopsis gene identification has confirmed itself as a proof of concept from where crop improvement can be facilitated (Zhang et al, 2004).

The use of mutants created from T-DNA insertion, such as SALK-lines, is important to define the functional significance of selected genes. *Agrobacterium* species have the unique and natural capacity to perform horizontal gene transfer across organisms from different phylogenetic groups. *Agrobacterium tumefaciens* is well studied for its pathogenic ability to transfer DNA (T-DNA) plasmid into the plant genome, resulting in the over-expression of hormones, such as auxins, inducing the formation of tumors, also called crown galls (Mashiguchi et al, 2019). The first *A. tumefaciens* transformation for biotechnological purposes was achieved in 1986. The original pathogenic plasmid is replaced by a customized cassette containing the necessary gene for the replication, conservation and selection of transformed organisms (Lee & Gelvin, 2008), giving the opportunity to introduce novel genes into plants, as well as the deactivation of genes via insertion mutagenesis. T-DNA mutational library collections have proven to be the most effective way to perform reverse genetic analysis and to characterise large parts of the coding genome. This widespread usage of the Col-0 T-DNA collection is attributable, in part, to the fact that the *Arabidopsis thaliana* Col-0 accession has been the dominant model in plant studies for decades, coupled with a near-whole genome assembly and the uncomplicated access to seeds (Provart et al, 2016). Another reason for the success of T-DNA insertion lines is *A. tumefaciens’s* natural ability to randomly insert plasmid DNA (more than 5-kbp fragment) coupled with effective protocols to disrupt the host’s DNA (O’Malley et al, 2015). The production of a large library of mutant lines was made possible by tracking a known sequence inserted in the left border of the plasmid, which enables the identification of mutants in a high-throughput fashion (Session et al, 2002). This international effort resulted in over 700,000 mutant lines with gene-affecting insertions produced to date, representing possible disruption mutants for the majority of the Arabidopsis genes (Jupe et al, 2019). Nevertheless,

evidence of limitations, such as genomic alterations and epigenetic stability in transgenic plants are starting to appear in the literature. It was estimated that 17% of the mutants produced by T-DNA insertion exhibited chromosomal rearrangements (Castle et al, 1993) and recently epigenetic analysis, performed in T-DNA insertion lines, revealed a variety of consequences ranging from the silencing of non-native integrated DNA to changes in chromatin marks and hence chromatin arrangement and activity (Jupe et al, 2019). Another limitation of using T-DNA insertion lines is the potential insertion of several T-DNA sequences per plant, with recent studies showing that multiple insertions of T-DNAs can trigger chromosome translocation (Pucker et al, 2021; Clark & Krysan, 2010). As our understanding of gene functions advanced, the need for flexible, quick and precise gene editing techniques had to be met.

CRISPR, standing for clustered regularly interspaced short-palindromic repeat, is a revolutionary gene editing technique involving the enzyme CAS9. This genome engineering system has been widely adopted due to its robustness and precision. The identification of these clustered regularly interspaced short-palindromic repeats in *Escherichia coli*, dates back to 1987 (Ishino et al). The term was later coined by Francisco Mojica from the University of Alicante, Spain, who was able to link the repeated short-palindromic repeats found in bacteria to sequences found in their natural predators, bacteriophages (Mojica et al., 2005). It was experimentally proven to be a defence mechanism in 2007 by Barrangou et al. Finally, researchers could reveal the function of each of the components involved in the nuclease process and to engineer the complex to target a specific sequence of their choice (Gasiunas et al., 2012 ; Jinek et al., 2012). CRISPR-Cas systems are now categorized into six categories based on their distinct effector proteins, which are then divided into two major groups. Class 1 systems use a number of complex proteins to accomplish interference, whereas class 2 systems use a single nuclease effector such as Cas9, Cas12, or Cas13 (Zhang, 2019). Type II CRISPR systems are receiving the most attention due to their low number of components involved and their simplicity. The process requires 3 components; the protein CAS9, the trans-activating crRNA (tracrRNA) and the specific guide RNA sequence (sgRNA) (see figure 1a). Once the Cas9 complex is formed, it can target the DNA sequence complementary to the sgRNA and cleaves it. This triggers the natural DNA repair mechanism enzyme to correct the broken segment by insertion or deletion of a few nucleic acid base pairs (see figure 1b), potentially creating a frame shift (translating into a non-functioning protein) or a premature stop codon. This process requires the DNA sequence to be next to a protospacer adjacent motif (PAM). This molecular biotechnological tool has now spread to nearly every life science field and is considered one of the most powerful and versatile gene editing method for crop improvement (Ricroch, 2019). Some examples of major crops improvement include Maize transformation for the production of higher content of amylopectin and phytic acid synthesis (Liang et al, 2014). Knock out mutant for the gene CLAVATA3 in rapeseed triggered the production of significantly higher number of seeds per silique and higher seed weight (Yang et al, 2018). The CRISPR-Cas technology was also shown to be efficient in genome editing of cotton, increasing lateral root system, and leading to higher biomass (Wang et al, 2017). However, the CRISPR technology is quite recent, and a number of limitations have emerged.

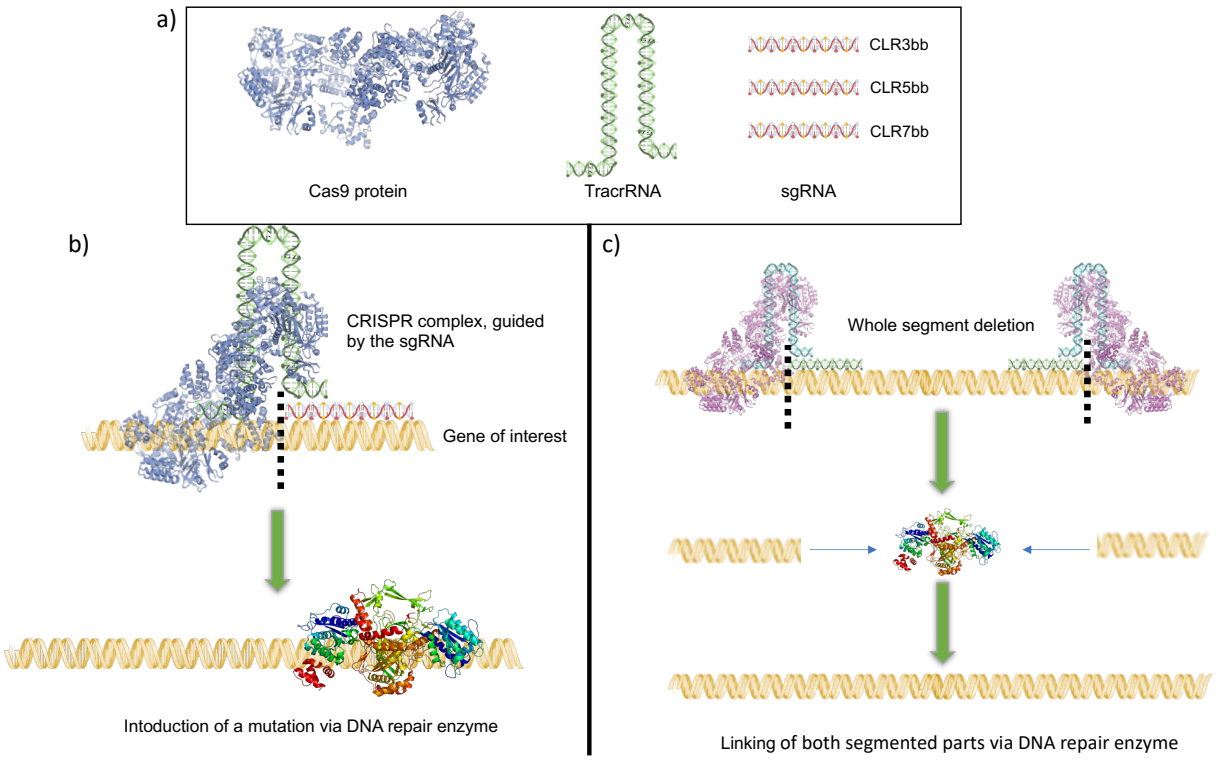


Figure 1 - Schematic representation of CRISPR-Cas9 components and mechanism. a) Single component forming the CRISPR-Cas9 complex, including the ribonuclease Cas9 protein, the TracrRNA acting as a molecular scaffold and the three candidate protospacers (CLR3bb, CLR5bb and CLR7bb) allowing recognition of the specific regions. b) The CRISPR complex, guided by the RNA protospacers, targets the specific gene of interest and cleaves it. The natural DNA repair mechanism process this damage by either inserting or deleting a few nucleotides. c) The dual sgRNA CRISPR technology targets two regions at the same time, allowing to either disrupt distinct regions in one event or to delete a large part of genome.

Issues regarding off-target gene processing are particularly highlighted. Solution are being designed to address these issues. In silico tools are used to provide confidence levels on expected off-targets or can be used to assess the specificity of the sgRNA (Alkan et al, 2018). Whole genome sequencing is also an option but is rarely used (Sturne et al, 2022). Engineering of the existing Cas protein can also improve its specificity and functionality (Vora et al, 2020). Optimization of the process also enables the targeting of several DNA segment using dual sgRNA techniques, making it a feasible option to delete large segment of the DNA and increase the chances of a loss-of-function (see figure 1c) (Pauwels et al, 2018)

Hegemony of flowering plants and pollen development

Flowering plants are one of the most successful land plant group, accounting for over 90% of all species and colonising nearly every single terrestrial habitat. Their complex reproductive biology is of particular importance to understand, partly, the hegemony of flowering plants. Evolutionary, gametes in the Plantae kingdom, can be either motile with similar size and form (isogamy), motile with a discrepancy in size (anisogamy) or can reach the oogamy stage where the female gamete is large and static with a small sperm (Umen and Coelho 2019). This oogamy stage, common to flowering plants, is associated with elevated genome size and complexity (Fang et al. 2017), as well as an evolutionary advantage of enclosing the static female gamete, providing a desirable condition against environmental threats necessary for the maintenance of the female egg and the development of the future embryo (Gasser and Skinner 2019). The pollen grain is also of particular importance to the success of flowering plants. Evolutionary, the plant kingdom exhibits a large diversity of sperm motility. Conjugating green algae are missing sperm motility, whereas liverworts, mosses, and hornworts referred to as bryophytes, being an even closer relative to flowering plants and representative model to understand the colonisation of land, produce motile sperm. In the vascular plants, a few members from the gymnosperms and fern reproduce using motile flagellate sperm, whereas angiosperms all developed non-motile gametes, which are instead transported using pollen tubes (Higo et al, 2018).

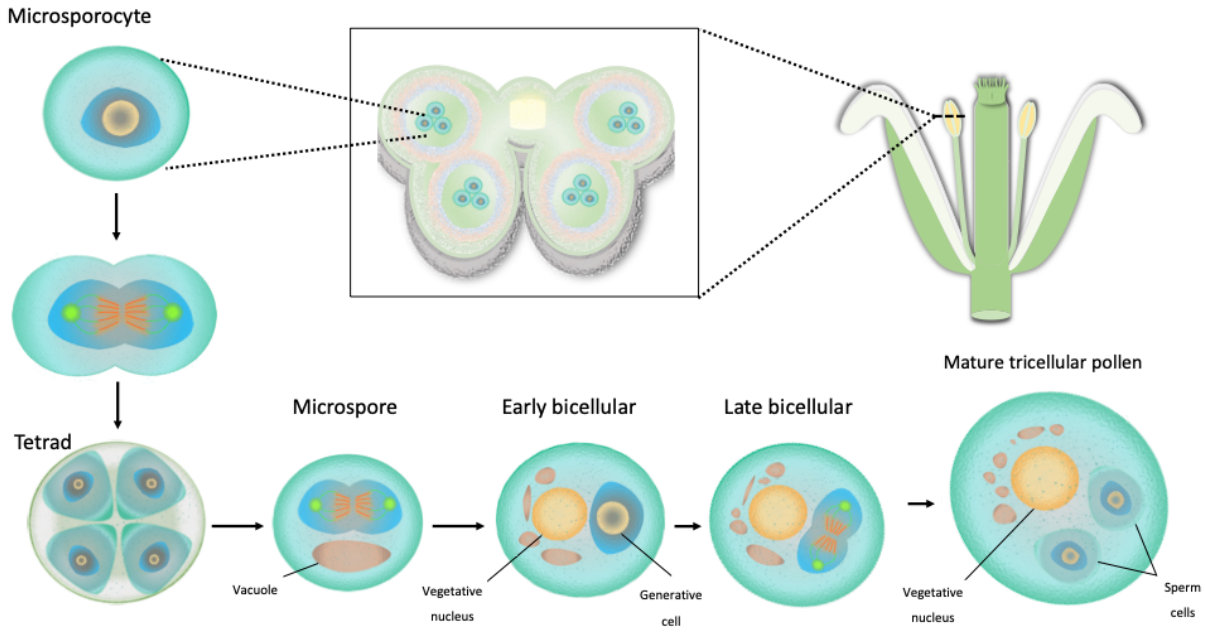


Figure 2 - Schematic representation of male gametogenesis process. The whole flower is represented in the right top corner. The section in the center is a cross-sectional representation of the anther containing the microsporocyte enclosed in the tapetum. Following the two meiosis events, a tetrad cell (lower left corner) is generated containing four microspore cells. Once detached, each indivivdual microspore cell contain a large vacuole and a nucleus. From the nucleus, mitosis I generates a vegetative nucleus and a generative cell. The generative cell divides into 2 sperm cells during mitosis II, producing the mature pollen grain.

The pollen grain's features have expanded the wide variety of habitats in which fertilisation can be achieved successfully. Sperm cells being enclosed in the protective pollen grain allows them to reach the female organ intact. Pollen grains have also developed a number of dispersal techniques to ensure fertilisation, whether biotic (animals, insects) or abiotic (wind and water). We can find this same characteristic in seeds as well, increasing the likelihood of survival by colonising different environments. The seed coat and endosperm also allow for extended periods of dormancy until conditions are desirable for development. This unique set of reproductive characteristics has resulted in the evolution of a wide variety of breeding and pollination mechanisms, which we are only starting to comprehend. Pollen development is of particular interest to the scientific community. Firstly, the amount of information regarding pollen grain development is larger when compared to eggs cells. The main reason for this discrepancy is that pollen is readily available and accessible compared to egg cells (Gehring, 2019). Secondly, the egg cells seem to have a less complex epigenetic system. CG and CHH methylation remain quite stable during female gametogenesis (Ingouff et al, 2017) and methylation of genes and transposable elements in general is lower compared to sperm cells (Park et al., 2016). For these reasons, pollen grains and male gametogenesis are of particular interest to us.

Microsporogenesis and microgametogenesis are the two main steps in the establishment of a pollen grain (see figure 2). Microsporogenesis, also known as male meiosis, is the first step in pollen development. Microsporogenesis is characterised by the meiosis division of a diploid microsporocyte cell into four haploid microspore cells. This forms a tetrad, composed of four microspores separated from each other by a thick wall, called callose. Each microspore detaches itself from the tetrad and the central vacuole starts to grow, pushing the nucleus to the pollen wall. This event, where the nucleus loses its central position in the cytoplasm, triggers the shift from a microspore to an uninucleate pollen grain. The microgametogenesis stage starts with the first steps being pollen mitosis I, resulting in the formation of a relatively larger vegetative cell compared to the generative cell. This generative cell is enclosed into a callose wall, which will disappear quickly after DNA synthesis, providing the appropriate isolation required for gametophytic genome expression to occur without interaction from neighbours (Shivanna, 2003 via Ünal et al, 2013). The generative cell is then separated from the pollen wall and eventually located in the cytoplasm. The generative cell undergoes pollen mitosis II, giving rise to two sperm cells. This second mitosis event can either occur in the anthers or later during pollen tube growth (Edlund et al. 2004).

LOTUS Domain Containing Proteins

The LOTUS (**L**imkain, **O**skar and **Tu**dor) domain comprises ~100 amino acid and is structurally characterised by its winged helix-turn-helix. This domain is evolutionary conserved among bacteria, fungi, animal, and plant. The original study identifying the LOTUS domain was based on the proteins Oskar and TDRD5/TDRD7 and highlighted its potential as RNAse, possibly playing an important role in post-transcriptional RNA-protein complexes (Anantharaman and Zang. 2010). Two forms of LOTUS domain exist, namely the extended LOTUS (eLOTUS) and minimal LOTUS (mLOTUS) domains, the latter lacking the C-terminal (Jeske et al. 2017). Depending on the type, the LOTUS domain can re-

arrange either in monomers, in dimers or with a protein complex such as Vasa (Jeske et al, 2017). See table 1 for a summary of literature review containing some of the important LDCPs observations reported.

Studies investigating LDCPs' exact function in plants are still elusive, although the literature points at their involvement in male gamete development. Recently, a paper discussed the intron structure and the cellular localisation of the *LDCP2*. Their results pointed out that the transcript of the gene *LDCP2* contained only one exon from the two predicted. They also found that during heat exposure, the protein encoded by *LDCP2* can co-localize with markers of processing bodies and stress granules, forming cytoplasmic foci. They also showed that the transcript *LDCP2* was highly and specifically expressed in pollen grains (Xu et al, 2020a). Another study investigated the function of LDCPs in Chinese cabbage (*Brassica campestris*). The genes investigated were also highly and specifically expressed in pollen grain throughout micro gametogenesis and during pollen germination. Following the construction of a mutant for both genes, the team saw a partial abortion in pollen grains and a reduction in pollen germination in the seemingly healthy pollen population (Xu et al, 2020b). Furthermore, the construction of an over-expressing mutant for the genes BCMF30A and BCMF30C was achieved. It revealed that the over-expression mutant led to abnormal pollen development and male sterility, similar to knock out mutants. Based on their previous study, they hypothesized that the gene's over expression triggers an increase in cytoplasmic foci, interfering with pollen development (Xu et al, 2020c). The mechanism by which the LDCPs are involved in pollen development is still unknown, but an *in vitro* experiment has shed light on the affinity between LDCPs and specific nucleic acid structure. The paper investigated the evolutionary conserved function among species of LDCPs to bind to specific molecules. AT2G15560 and AT3G52980 (*LDCP2*) were among the gene being investigated. They showed that indeed, both genes had a particular affinity to RNA G-quadruplex structure (Ding et al, 2020). Those G-quadruplex RNA structures have a regulatory role in plant development (Yang et al, 2020). Moreover, DNA G-quadruplex structures have been found inside plant retrotransposons (Lexa et al, 2014). Another pair of *LDCPs* in *Arabidopsis thaliana* have been characterised, namely AT5G64710 and AT5G09840. These *LDCPs* have not been linked to gametogenesis, but rather to RNA processing. They have been shown to be highly expressed in the mitochondria. The null mutant for both genes triggered a clear reduction in mature RNA and further experiments demonstrated the involvement of the two LDCPs with mitochondrial RNA 5' processing (Stoll and Binder, 2016).

More information regarding LDCPs is available in other model organisms. Similar to plants, LDCPs have been investigated for their involvement in germline development. In mice, the LOTUS domain containing protein MARF1 has been identified as a regulator of female gamete formation (cytoplasmic maturation, embryogenesis, meiosis and the silencing of genome integrity threats, such as transposon elements). The mutant lacking MARF1 showed a clear meiotic arrest and, following the micro-injection of mRNA coding for oocyte maturation-promoting factor, over half of the oocytes were able to complete meiosis. (You-Qiang Su et al, 2012). Furthermore, Zhu et al, 2018 determined that MARF1 was exclusively expressed in ovaries, specifically in late-stage oocytes (stages 12-14). The egg number produced by MARF1-null mutants were similar to the control, but none of them hatched. Stage 14 of oocyte

Table 1 - Summary of literature review containing some of the important LDCPs observations reported.

Authors / Year	Species	Main point
Callebaut and Mornon, 2010	Human	LOTUS domain is a conserved feature of the protein complexes present in the piwi containing germline-specific granules in different organisms.
Jeske et al, 2015	Drosophila	In drosophilid LOTUS domain forms dimers .No RNA binding activity from drosophila Oskar LOTUS domain.
Stoll and Binder, 2016	Arabidopsis	MNU1/2 function might be a mitochondrial endonucleases active in 5' processing of mRNA as part of a processing complex or have a role in scaffold for larger protein assemblies.
Jeske et al, 2017	Drosophila / Silk moth Bombyx	LOTUS interacts with Tejas and Tapas protein. LOTUS domain stimulate Vasa which is required for proper embryonic development. Interaction is conserved in insects silk moth.
Zhu et al, 2018	Drosophila	LOTUS domains as a post-transcriptional effector domain to recruit the CCR4-NOT deadenylase complex to shorten target mRNA poly-A tails and suppress translation of the mRNAs.
Ding et al, 2020	Mouse / Human / Drosophila/ Arabidopsis	Lotus domain selectively binds with high affinity G-rich RNA. LOTUS recognise and binds G4 RNA but not DNA. LOTUS domain interractions are evolutionarly conserved.
Xu et al, 2020	Arabidopsis	Perfect colocalization of AT2G05160 with marker of stress granule and processing bodies. AT2G05160 was expressed exclusively in pollen.
Xu et al, 2022	Arabidopsis	Knock-out mutant of AT2G05160 displayed no phenotype. The over-expression mutant displayed shrunken pollen morphology leading to male sterility which was always associated with the formation of high number of granules.

development was undoubtedly involved in female sterility, as no meiotic spindles or polar bodies were observed in MARF1 null mutants. On the other hand, male-related sterility is also a common phenotype in *LDCPs* mutants. Drosophila mutants for *LDCPs* (tej and tap) resulted in the loss of germline cells (Patil et al, 2014). Using a TDRD5 knock out in mice, the team observed a clear male sterility with defects in the assembly of intermitochodrial cements and the chromatoid bodies, both being ribonucleic protein granules taking part in RNA processing for male germline development. (Yabuta et al., 2011).

As in plants, it seems that LDCPs have a particular affinity with RNA. *In vitro* and *in vivo* results showed a specific and strong affinity between the LOTUS domain of TDRD5 from human and RNA G-quadruplex. They could also demonstrate that the TDRD5 LOTUS domain did bind preferentially to G-quadruplex single-stranded RNA rather than double-stranded RNA or to DNA G-quadruplex. Using a diversity of representative species (human, mice, Drosophila, plant and bacteria) containing LDCPs, evidence of evolutionary conserved function such as binding G4 RNA was provided (Ding et al, 2020). Another paper confirmed that LOTUS shows a sturdy and specific interaction with a well-known regulator of mRNAs translation, the Vasa protein. (Jeske et al, 2015).

So far, no reports have shown that plant's LDCPs have been linked with transposable elements, but studies using different model organisms have shown such a relationship. A study has linked TDRD5 to transposon silencing through the class of non-coding PIWI-interacting RNAs (piRNA), involved in silencing transposons and assuring germline genome integrity. They observed a strong decrease in total piRNA levels. They also found that the total number of miRNA was down 70%, compare to wild type, clearly indicating the essential role of TDRD5 in the production of a large subset of pachytene piRNA (Ding et al, 2018). Another set of LDCPs (Tej and Tap) were linked with piRNA. Tap and Tej mutants showed a decrease in piRNAs in the germline and a mis-localisation of piRNA related proteins from the nuage. De-Repression of TEs was also observed (Patil et al. 2014). Another study links the LDCP TDRD5 to retrotransposon silencing involved in male germline development of mice. A 20-fold increase in expression of open reading frame 2 of the LINE-1 retrotransposon was observed in TDRD5 mutants, being possibly driven by disruption of the re-methylation of the LINE-1 promoter (Yabuta et al, 2011). Using a *marf1* mutant, the team could observe an increase in Line1 and Lap retrotransposon RNA, correlating with a rise of nuclear DNA double-strand breaks, suggesting that MARF1 might be involved in meiosis and retrotransposon of female of germline (Su et al, 2012).

Epigenetic and transposons in pollen development

Epigenetics is defined as gene expression changes, without direct alteration to DNA sequence. It is mostly triggered by environmental factors and has an element of stability over generations (Lacal and Ventura, 2018). In plants, epigenetic changes are largely controlled by a family of dynamic DNA fragments called Transposable Elements (TE) (Ramakrishnan et al, 2021). Originally thought to be "parasitic" or "Junk" DNA (Orgel & Crick, 1980) due to their DNA independent replication cycle and gene disruption ability (Fedoroff, 2012). TEs have been re-categorized as one the main contributors of evolution and diversity (Oliver et al. 2013) and have drawn attention due to their omnipresence among species and their abundance (Galindo-González et al, 2017). Their transfer to the next generations is guaranteed by their large copy number, their prevalence in many chromosomes or by inserting into a genetically active part (Bennetzen, 2000). In plants, epigenetic changes are driven by DNA methylation, usually triggered by environmental change. RNA-directed DNA methylation (RdDM) is a complex multi-step pathway involving at least 40 components. Recent model establishes that, upon recognition of an unmethylated region, NUCLEAR RNA POLYMERASE 4 initiates the transcription of a single stranded RNA from a double stranded DNA. A strand is then added by RNA-DEPENDENT RNA POLYMERASE 2 proteins and the double stranded RNA is processed by DICER-like proteins (Singh et al, 2019). The siRNAs are introduced into Argonaute proteins, which lead them to the complimentary sequence. Here, the NUCLEAR RNA POLYMERASE V scaffold is recruited and the DNA methyltransferase enzyme methylates a cytosine base (Sigman et al. 2021). This methylation process results in the transcriptional silencing of loci from transposable elements and other repetitive DNA. This silencing mechanism is of particular importance during sexual reproduction to avoid transposable elements-related damage to the genome. The pollen grain is a particular cell, as it is composed of two distinct cell type, the vegetative nucleus and two sperm cells. It has been established that the sperm cells and the vegetative cells have different roles during the sexual reproduction. In the mature pollen grain, the vegetative nucleus

chromatin is completely de-condensed, allowing molecular activity to occur. In contrast, the chromatin of the sperm cell remains tightly packed in a effort to remain intact. It was also shown that some TEs are activated in pollen grains and their origin is found in the vegetative nucleus. The accumulation of siRNAs in sperm cells was also demonstrated, confirming the communication between the vegetative nucleus and the sperm cells (Slotkin et al, 2009).

Aim of the study

Our starting hypothesis was that LOTUS Domain Containing Proteins are involved in the development of pollen grains in Arabidopsis. We think LDCPs are involved in the sub-cellular communication between the vegetative nucleus and the sperm cells. One of the key stages during pollen development is the silencing of the TEs through small interfering RNA. We believe that through its zinc-finger domain, RNA recognition motif and LOTUS domain, LDCPs could have the ability to recognize and process RNA, possibly playing a key part in pollen development.

In this study, we aim to understand the role of two *LDCPs*, namely AT2G05160 (*LDCP1*) and AT3G52980 (*LDCP2*) in the development of pollen grains. This will be performed through reverse genetic approaches, using T-DNA lines and CRISPR-Cas9 technology.

Materials and methods

Plant materials

All plants used in this work originated from the *Arabidopsis thaliana* ecotype Columbia (Col-0). WT, *ldcp1* (SALK_022373.32.75.x) and *ldcp2* (SALK_115366.24.15.x) seeds were obtained from SALK library from the NASC. Seeds were disinfected using ethanol (70% absolute: 30% distilled water) for 15 mins, in a rotating bath. Seeds were plated on *Arabidopsis* growing media (AGM) (1% (m/v) sucrose, 0.43% (m/v) Murashige and Skoog salts, 0.05% (m/v) MES buffer and 0.8% (m/v) plant agar. Stratification (2 days at 4°C) was used to break the dormancy of the seeds, followed by growth under short day conditions (8 hours light, 16 hours dark). Two weeks after germination, the seedlings were transferred to soil under long day conditions (16 hours of light, 8 hours of dark). The chamber condition was kept stable at 23°C and approximately 60% humidity.

Cross-pollination method

Plants no younger than 40 days old were used for cross-pollination. The mature siliques and the open flowers of the mother plant were removed using fine scissors or forceps. The small buds were removed and 3 to 5 large unopened buds per branched were kept. The flower bud was opened by inserting the tip of tweezers between petals and sepals. The sepals were removed, and the tip of the tweezers were slide in between the petal and anthers to access the interior of the buds. All petals and anthers were removed carefully. The stigma was covered using a small paper bag and left for 24 hours. For the father plant, an open flower was removed by the pedicel and tapped onto the exposed stigma of the mother plant, ensuring pollen was released on the stigma. The pollinated inflorescence was covered with a small paper bag to avoid unwanted cross-pollination. 15 to 25 days after the pollination (or when the silique turned yellow), the siliques were harvested and placed into a bag. The seeds were separated from the silique by shaking the bag and left to dry in a dissector until they reached a 30% humidity level. Figure 3 summarizes the generation of the mutant line.

Phenotypic characterisation

The phenotypic characterisation included germination assay and sexual and non-sexual macro-observation. The germination assay was carried out to investigate deficiencies in germination between mutants and wildtype Col-0. The seeds were disinfected as previously described and plated on AGM media. Following a stratification step (2 days at 4°C), the plates containing the seeds were placed under short-day conditions for 10 days. Pictures were acquired and the germination rate was calculated. It was established that the seed germinated following the appearance of the first true leaves. The non-sexual macro-observation included the measurements of the height of the fully grown plants and the number of leaves on the rosette. Both measurements were acquired on plants no younger than 60 days old. Characterisation of reproductive phenotype was performed by measuring the silique length of plants no younger than 60 days old. The siliques were picked on the main branch, the first five siliques were

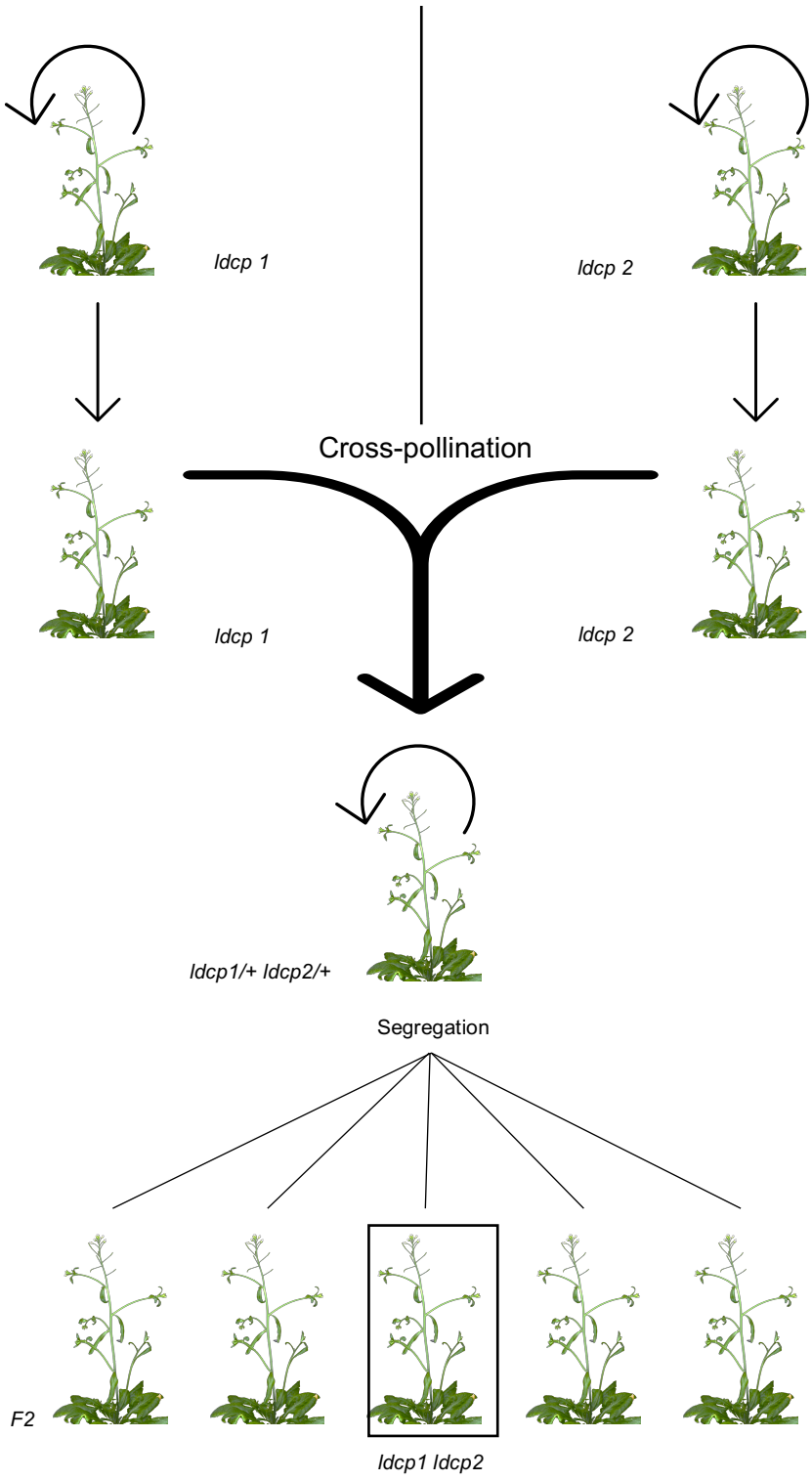


Figure 3. Schematic diagram of the generation of double homozygous mutant (*ldcp1 ldcp2*) by crossing two single mutant plants containing a SALK T-DNA insertion. The two single mutants used were the line *ldcp1* (SALK_022373.32.75.x) and line *ldcp2* (SALK_115366.24.15.x) targeting the genes AT2G05160 and AT3G52980, respectively. Following the cross-pollination process, the double heterozygous mutant (*ldcp1/+ ldcp2/+*) was self-pollinated. In the segregation progeny, we identified the double homozygous individual (*ldcp1 ldcp2*) by genotyping.

measured. The mean value of each mutant was compared to the wildtype Col-0, using a t-test to investigate the possible statistical difference.

Fixation and clearing of siliques

Siliques were preserved using the technique described by Flores-Tornero et al, 2021. Briefly, mature siliques (~15mm long) were collected upside down in a 1.5mL microcentrifuge tube filled with 1.2mL of the fixing solution (99.8% ethanol absolute, acetic acid (3:1) (v/v)). The tube was incubated overnight at room temperature. The following day, the fixing solution was replaced with 1.2 mL of 70% ethanol and the tubes were stored away from direct sunlight at room temperature. To acquire pictures, a small petri dish was filled with 70% ethanol and placed under a SMZ800 stereoscope (NIKON). The silique was collected from the centrifuge tube and placed in the petri dish. A 3D printed holder was kindly provided by the 3D printing facility at Instituto Gulbenkian de Ciência. This apparatus enabled us to fit a smart phone to the stereoscope and take pictures.

Pollen grain staining

A microscope slide was placed under a stereomicroscope. Freshly opened flower was collected by grabbing its pedicel with tweezers. The flower was placed upside down and dab against the microscope slide. The release of pollen grains on the slide was verified using a stereomicroscope. 7 µL of DAPI dye (1 µg/ml), staining DNA, was added on the pollen grains. A cover slip was carefully placed on the drop, ensuring minimum air bubble formation. The slide was observed using DM6 B Upright Microscope (Leica) under brightfield and UV settings (405nm). The Fiji software (https://imagej.net/software/fiji/downloads) was used to visualize and colorize the picture.

Genomic DNA extraction

The protocol used to extract gDNA is derived from the technique reported by Edwards et al, 1991. Well grown rosette leaves (no younger than 20 days old) were collected and preserved in eppendorf tubes, containing glass beads. The Eppendorf tubes were then dipped into liquid nitrogen and ground using a Tissuelyser II (30s at a frequency of 30 Hz). 400µL of DNA extraction buffer (Edwards Solution) was immediately added and quickly vortexed. 150 µL of 3M sodium acetate (pH 5.2) was added, and the solution was vortexed for 10s. The tubes were incubated at -20°C for 10 mins and centrifuged for 5 mins at 10 000g. The supernatant was collected into a new Eppendorf tube. 400µL of isopropanol was added, and the solution was mixed and incubated for 10 min at 4°C. Following a centrifugation step of 5 min, at 12 000g, the supernatant was discarded. The tube was then washed with 250 µL of Ethanol 70% and vortexed until the DNA sediment was detached from the wall of the tube. The tube was centrifuged for 3 mins at 10 000g and the supernatant was discarded. The DNA sediment was dried on a 65°C hot plate for 5 mins. It was resuspended using 50 µL of deionized water and stored at 4ºC.

PCR reaction

Wild type and T-DNA mutants were genotyped by Polymerase Chain Reaction (PCR) to confirm homozygosity. DreamTaq DNA Polymerase from Thermo Scientific was used to perform the PCR (See supplemental table 1 for the list of reagents). Two sets of reactions were designed to determine, for each gene, whether the line was a homo- or heterozygous mutants. The T-DNA insertion contains a sequence known to be located on the left border (LB 1.3), which was used to identify the mutant. The identification of the wild type used a pair of primers spanning the region of the predicted T-DNA insertion. Differentiation is based on the inability of the polymerase, under normal PCR condition, to amplify the entire region containing the T-DNA insertion. Table 2 contains the list of primers, their sequences and conditions used to perform the PCR.

Table 2 - List of primers used for PCR amplification.

Pair of primer	Sequence (5'-3')	Annealing temperature (˚C)	Extension time (sec)	Expected product size (bp)	Purpose
CLR1	GGGTTCATGAGTGAGTATCACTC	61	40	950	Identification of WT *LDCP1*
CLR2	CTTGGTGAGACTATAGCCAGC				
LBb1.3	ATTTTGCCGATTTCGGAAC	57	15	500	Identification of MT *LDCP1*
CLR2	CTTGGTGAGACTATAGCCAGC				
CLR45	GCAAGATCAAAACGACCGTG	57	70	1140	Identification of WT *LDCP2*
CLR46	AGCAAGAATGCGTTTGACG				
LBb1.3	ATTTTGCCGATTTCGGAAC	57	20	650	Identification of MT *LDCP2*
CLR46	AGCAAGAATGCGTTTGACG				

Gel Electrophoresis

DNA fragment size observations, regarding T-DNA line genotyping, were all performed in 1% (w/v) agarose gel electrophoresis in 1X TAE (40 mM Tris-HCl, 20 mM, 1 mM EDTA) and 0.05 µL/mL RedSafe DNA Stain was added to the agarose solution. 7 µL of 1x Purple loading dye (NEB) was added to each sample before gel loading. GeneRuler DNA Ladder Mix (Thermo Scientific) was used as molecular weight marker. The electrophoretic process was performed at 120 V for 45 min. Visualization under UV light revealed the bands and pictures were acquired using a Gel Doc XR+ (Biorad).

RNA Extraction and RT-PCR

Unopened inflorescences from the mutant lines and wild type plants were collected for RNA extraction. RNA extraction was carried out using Spectrum Plant Total RNA Kit (Sigma) following the manual provided. Quality and quantity check of all the samples were performed using NanoDrop 2000/2000c spectrophotometers. Samples were then delivered to Instituto Gulbenkian de Ciência for further quality verification and reverse transcription step using the SMART-SEQ2 protocols. DreamTaq DNA

Table 3 - List of primers used for RT-PCR amplification.

Pair of primer	Sequence (5'-3')	Annealing temperature (˚C)	Expected gDNA product size (bp)	Expected cDNA product size (bp)	Purpose
AK133	ATGTGAGGATTCCTTGCCAGC	67	560	400	RT-PCR identification of *LDCP1*
Ak135	AGCTCAACCGTTCTGCTTGTTC				
AK130	TTAGAGATGCAGGCATCAAGAGCGC	67	400	320	RT-PCR identification of *UBC21*
Ak132	CACACAGACTGAAGCGTCC				
CLR60	TTCTCTGAAGTTGGACCAGTGG	57	460	370	RT-PCR identification of *LDCP2*
CLR61	GCTGCTTAAATCAAGAAGAGGTGC				

Polymerase (Thermo Scientific) was used to perform the PCR from the cDNA. The primers were designed to span introns (primer information can be found in table 3). WT cDNA and gDNA were used as control and the gene Ubiquitin 21 (UBC21) was used as reference gene.

Plasmid reception and preparation

The protocol as well as the plasmids were kindly provided by Dr. Corinna Thurow from the Albrecht-von-Haller-Institute for Plant Sciences, Georg-August-University Göttingen, Germany. Upon receival, the plasmid pBCsGFPEE (GEE) was introduced into *E.coli DH5α*. The heat shock protocol is as follows: 0.5 µL of GEE plasmid (concentration around 2 µg /µL) were mixed with 50 µL of competent cells. The solution was then heated up to 37℃ for 40 secs and then placed on ice. 950µL of LB containing 100µg/mL of spectinomycin were added to the mix and left to incubate for 30 mins at 37℃, 300rpm using the Thermomixer Compact 5350 Mixer (Eppendorf). 250 µL of the solution were plated onto LB plate (spectinomycin 100 µg/mL) and incubated for 24 hours at 37℃. Colony PCR was performed by picking a small amount of the colony with a tip and diluting it in 50 µL H2O (nuclease-free). The solution was incubated at 95℃ for 10 min and quickly centrifuged (1 min at 10 000g). 2 µL of the supernatant were used as template for the PCR. The positive colonies were stored in glycerol (200µL of sterile glycerol for 800µL of LB) in -80℃. The competent cells containing the plasmid were then grown in liquid LB containing 100µg/mL spectinomycin (24 hours at 37℃). The plasmid was extracted using the Plasmid Miniprep Kit (Zymo Research) following the manufacturer's instructions. Following the integrity check via sequencing (data not available) the plasmid was then digested (see supplemental table 2). The entire volume of the digested plasmid was run on a 0.5% TAE agarose gel (60 volt/ 2hours) and based on the size, the plasmid was extracted from the gel, purified using PCR & Gel Band Purification Kit (GRISP) and stored at -20℃ until future use.

Table 4 – List of sgRNA candidates

Oligos with BsaI overhangs sequences		
sgRNA candidate	Oligos ID	sequence (5'-3')
CLR3bb	CLR3 (oligo UP)	GATTG TCACAAAGACATGGCAAAGC
	CLR4 (oligo BOTTOM)	AAACGCTTTGCCATGTCTTTGTGA C
CLR5bb	CLR5 (oligo UP)	GATTGTCAGTGAAACTACTCTCTGC
	CLR6 (oligo BOTTOM)	AAACGCAGAGAGTAGTTTCACTGA C
CLR7bb	CLR7 (oligo UP)	GATTGATCTGCCATTACTTCAACAA
	CLR8 (oligo BOTTOM)	AAACTTGTTGAAGTAATGGCAGAT C
CLR9bb	CLR9 (oligo UP)	GATTGATAACTGAATGTTGTCCATG
	CLR10 (oligo BOTTOM)	AAACCATGGACAACATTCAGTTAT C
CLR13bb	CLR13 (oligo UP)	GATTGATAACTGAATGTTGTCCATG
	CLR14 (oligo BOTTOM)	AAACCATGGACAACATTCAGTTAT C
CLR39bb	CLR39 (oligo UP)	GATTGCATTCTCTGGTTCAAGTTGA
	CLR40(oligo BOTTOM)	AAACTCAACTTGAACCAGAGAATG C

Protospacer design and preparation

The protospacer design was performed by Anton Kermanov. Open reading frames for both *LDCP1* and *LDCP2* were provided as input to CRISPys tool (http://multicrispr.tau.ac.il/) considering Cas9 PAM sequence. The output provided the potential sgRNA sequences. The best matches were selected for further verification using CRISPOR tool (http://crispor.tefor.net/), for the targeting of undesired regions. Based on the score of each predicted sgRNA, 6 candidates were chosen for implementation (CLR3, 5, 7, 9, 13, 39). The oligo sequences were ordered from IDT (https://eu.idtdna.com/pages). The sequence of each candidate is available in table 4. Upon receival, the single stranded oligo (UP and BOTTOM) were hybridized to each other. Briefly, 5uL of oligo UP and 5uL of oligo BOTTOM were incubated at 98℃ for 3 min and then brought back to 4℃ on a 0.1 ℃/s ramp and stored at 4°C using the T100 Thermal Cycler (Bio-Rad).

Ligation and transformation

Briefly, the GEE plasmid digested with BsaI was ligated to the double stranded oligos containing the BsaI overhangs using T4 Ligase (ThermoScientific). The ligation reaction was performed according to supplemental table 3. For transformation, 10µL of GEE plasmid (containing the sgRNA) was added to *E.coli DH5α* competent cells on ice. Following a heat shock reaction identical to the one explained above, the competent cells were plated on LB containing 100µg/ml of spectinomycin and incubated o/n at room temperature. Colony PCR was performed using the primer (oligo BOTTOM) of each candidate and the primer L236 (GAGCCGGAAGCATAAAGTGTAAAGC) targeting the Lac promoter (see supplemental table 4). Positive candidates were sent for sequencing using the same set of primers (oligo BOTTOM and L236).

sgRNA generation by *in vitro* transcription

The protocol for the Invitro CRISPR-Cas9 assessment is derived from the technique developed by Bente et al., 2020 and was redesigned by Anton Kermanov. Briefly, GEE plasmid was used as a template to amplify the sgRNA using a T7 promoter. A modified sgRNA primer as described by Bente et al, 2020 had to be designed including a promoter for T7 RNA polymerase. Using DreamTaq DNA Polymerase from Thermo Scientific, a 130 base pair fragment was amplified (see supplemental table 5). This sequence was used as a template for *in vitro* RNA transcription (see supplemental table 6). DNAse treatment was performed by adding 0.5 µL of DNAse I (Roche) and 0.3 µL of calcium (1M) to the solution. It was then incubated at room temperature for 30 mins. 1 µL of EDTA (50mM) was added to the solution. The quality and yield were estimated by running a 2% TAE gel electrophoresis (120 volt / 15 mins). The transcription reaction was then purified with RNA clean and concentrator kit (Zymo).

Generation of templates for *in vitro* cleavage assay

Primers were selected to amplify the region of interest (see supplemental table 8) in an asymmetric way, where each cleaved fragment should be distinguishable based on size. The amplification was done using DreamTaq DNA Polymerase (Thermo Scientific) on the T100 Thermal Cycler (Bio-Rad). The DNA template was extracted from Col-0 plants. The fragments were then purified using the PCR & Gel Band Purification Kit (GRISP).

Cas9 cleavage of PCR products and selection of best sgRNA candidates

The *in vitro* CRISPR-Cas9 reaction was set up as described in the supplemental table 7. Briefly, commercially available Cas9 protein (NEB) were introduced to the sgRNA candidates, forming a complex. This complex was then exposed to DNA templates of our gene of interest. The solution was then incubated at 22℃ for one hour and heat-inactivated for 15 mins at 65℃. 20µl of the reaction was run in a 1% TAE agarose gel for 60 mins at 90V. The gel was observed under UV light. We based our final selection on whether, the candidates could cleave both genes of interest.

Agrobacterium transformation

Based on the *in vitro* CRISPR-Cas9 reaction, three sgRNA candidates were retained for Agrobacterium transformation (CLR3bb, CLR5bb and CLR7bb). The *Agrobacterium tumefaciens* strain GV3101::pMP90 was used for transformation. Briefly, an aliquot of *Agrobacterium* competent cells was thawed completely on ice, then 30 µl were mixed with 3 µl of purified pGEE plasmid (~300-500 ng) and incubated on ice for 30 min. The mix was placed in liquid nitrogen for 1 min , then transferred to a 37℃ water bath for one minute and finally incubated on ice for 5 min. 1ml of LB was added to the mix and incubated in a shaker (Eppendorf Thermomixer Compact 5350 Mixer) at 28ºC, at 180 RPM for 2 hours. The tube was then centrifuged for 1 min at 10 000g and the excess LB removed. The pellet was re-suspended in the left-over LB and plated on LB containing rifampicin (50 µg/l), gentamycin (25 µg/l), and spectinomycin (100 µg/l). The plates were left to incubate for 2 days at 28℃. The presence of the plasmid insertion was verified by colony PCR (see supplemental table 5). The positive colonies were transferred to 1 ml of LB media containing rifampicin (50 µg/l), gentamycin (25 µg/l), and spectinomycin (100 µg/l) and incubated overnight at 28-30ºC, 180 RPM (Eppendorf Thermomixer Compact 5350 Mixer). The candidates were stored in glycerol (200µl of sterile glycerol for 800µl of LB) at -80℃.

Arabidopsis thaliana transformation with Agrobacterium

A. thaliana plants used for this transformation were 40 days old with inflorescence full of young and closed buds and few formed or mature siliques. The mature siliques were cut at the time of plant transformation to avoid a strong background from untransformed seeds. *A. thaliana* plants were transformed using the floral dipping protocol designed by Clough and Bent, 1988. Isolated colonies from *A. tumefaciens* GV3101::pMP90 containing the transformed plasmids (pBCsGFPEE + CLR3bb,

pBCsGFPEE + CLR5bb and pBCsGFPEE + CLR7bb) were transferred to 5 ml of LB media containing rifampicin (50 µg/l), gentamycin (25 µg/l), and spectinomycin (100 µg/l) and incubated overnight at 28-30ºC, 180 RPM (Innova 44 from New Brunswick). 1ml of the overnight 5ml culture was transferred to 200ml of LB, containing the previously used antibiotic, and incubated overnight at 28ºC, 180 RPM. The culture was centrifuged using Allegra X-12R (Beckman Coultier) for 15 minutes, at 3 000g, at room temperature. The supernatant was discarded, and the cell pellets were resuspended in 150 mL H2O with 5% (m/v) sucrose, 0.43% (m/v) Murashige and Skoog and 0.05% (m/v) MES buffer. Each of the three 150ml solution were divided into 50ml flask. Each solution was reciprocally mixed together to form 3 transformation events (CLR3bb+CLR5bb, CLR3bb+CLR7bb and CLR5bb+CLR7bb) with a total volume of 100ml per transformation solutions. Before dipping the plants, 0.03% (v/v) Silwett-77 was added to each of the 3 transformation solutions. A total of eight *A. thaliana* plants were used for each transformation event. The branches of the plant were dipped into the solution, 2 times for 2 minutes, while in the solution the pot was slowly spinned. The branches were untangled before each dip. The plants were left to dry horizontally, on absorbing paper protected from direct light for 24 hours. The plants were then moved back to long day and allowed to grow. Seeds were collected, dried, sterilized, and plated on AGM media containing BASTA (10 µL/mL). The BASTA resistant plants were transferred to soil. Genotyping of each of the plants was performed to identify plants which received both insertions (see supplemental table 5)

Results

Identification of 2 *LDCP* genes with a potential role in male gametogenesis in Arabidopsis

The *Arabidopsis thaliana* genome encodes 6 *LDCPs*. To identify *LDCPs* possibly involved in gametogenesis, we used previously acquired RNA expression data from the database EVOREPRO (https://evorepro.sbs.ntu.edu.sg). Figure 4a shows the expression data of the 6 *LDCPs* in Arabidopsis. *LDCP1* and *LDCP2* were chosen based on their specific expression patterns. Both genes are highly and specifically expressed during key stages of pollen development and their expression is low in other organs, except for *LDCP1*, which shows relatively high expression in egg cells. The visualization clearly highlights that *LDCP1* is highly expressed in microspores, bicellular-pollen and in late bicellular pollen. *LDCP2* is also of interest, as it shows a relatively high expression during pollen development stages. Moreover, the genes *LDCP1* and *LDCP2* have a high structural similarity, considering that *LDCP1* is relatively larger than *LDCP2*, the similarity reaches 82% (see supplementary figure 1) at the gene level and 55% at the protein level (see figure 4b). The protein structure is very similar with both of them containing a Zinc-Finger domain, LOTUS domain and RNA recognition motif in the same order (see Figure 4c).

Single mutant *ldcp1* and *ldcp2* do not display gametophytic defects

The main goal of using mutants generated by T-DNA insertion, such as SALK-lines, is to define the functional significance of selected genes. The phenotypic observation was carried comparing the SALK lines *ldcp1* (SALK_022373.32.75.x) and *ldcp2* (SALK_115366.24.15.x) targeting the genes AT2G05160 and AT3G52980, respectively, to the wild type control *Arabidopsis thaliana* ecotype Columbia-0 (Col-0). Genotyping, via PCR targeting the T-DNA insertions, was carried to investigate the homozygosity of the SALK lines received. The stability of the T-DNA insertion was evaluated over two generations of each SALK line and homozygosity was found to be stable (see figure 5a). No developmental difference was observed between *ldcp1, ldcp2* and Col-0 (see figure 6a). We investigated further possible male reproductive phenotype defects based on two grounds. Firstly, previous studies about the functional role of LDCPs in several models (Yabuta et al, 2011; Patil et al, 2014; Jeske et al, 2017; Zhu et al, 2018) clearly show that LDCPs play a key role in reproduction. Secondly, the transcriptional information available from the EVOREPRO database revealed that *LDCP1* and *LDCP2* are highly and specifically transcribed in reproductive organ, mainly during the pollen developmental stages and in the female egg cell for *LDCP1*. Reproductive phenotypic characterization included flower structure observation, epifluorescence microscopic recording of pollen grain and morphology observation, and the number of seeds per silique. When comparing *ldcp1*, *ldcp2* and Col-0, a significantly higher number of seeds per silique for *ldcp2* was observed (see figure 7c).

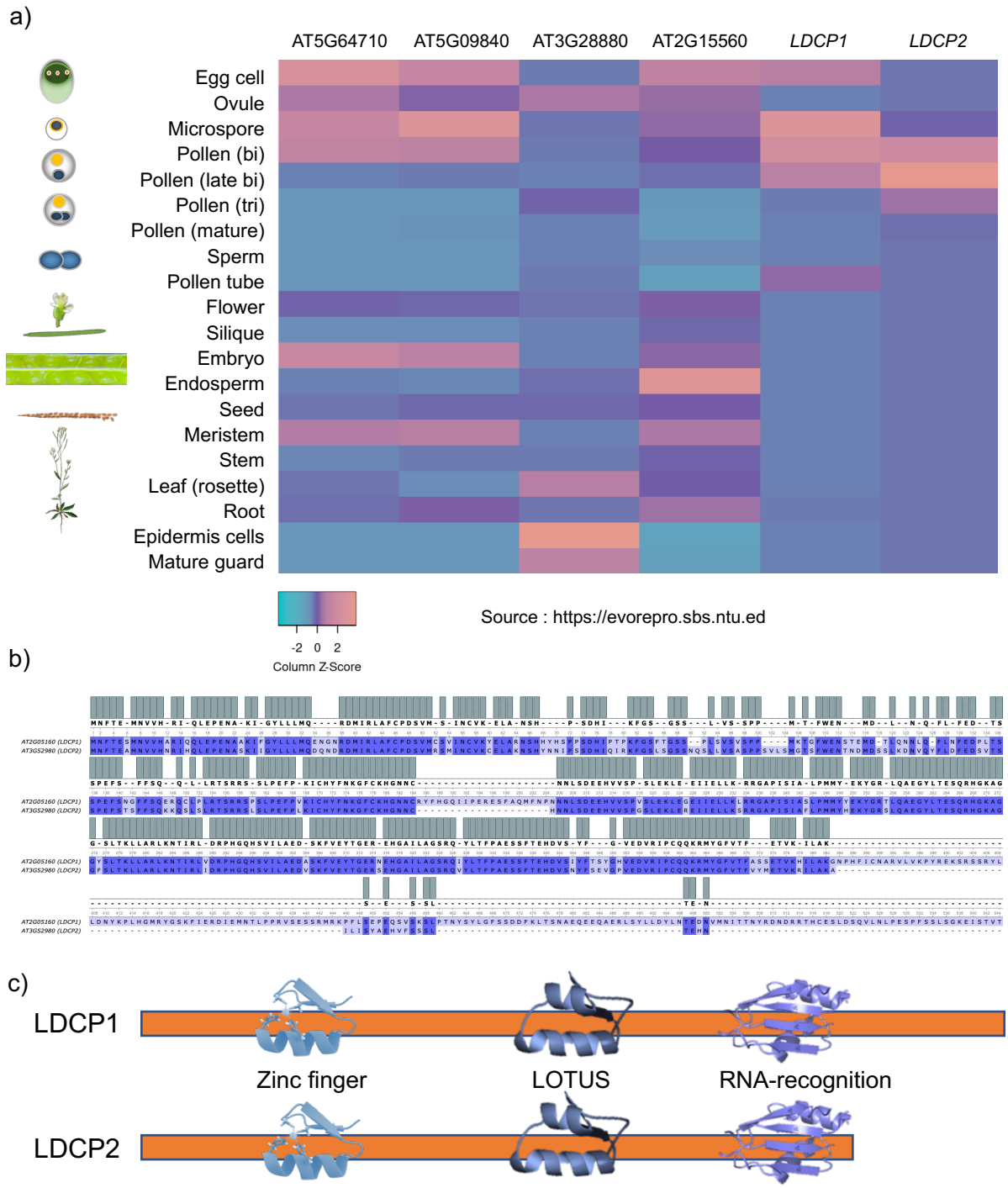


Figure 4 – RNA expression data, amino acid sequence and domain structure similarities reported between *LDCP1* and *LDCP2*. a) The data were acquired from the database EVOREPRO (https://evorepro.sbs.ntu.edu.sg). The heatmap was designed using the software HeatMapper (https://www.heatmapper.ca). It is be read by column where each column gives the RNA expression level of the particular gene in every organ. Orange indicates a high expression of the gene, purple a relatively lower and turquoise, a low expression. b) Amino acid alignment of LDCP1 and LDCP2 with the consensus agreement. This alignment was provided by MUSCLE through the software Unipro UGENE. c) Schematic representation of the domain distribution of LDCP1 and LDCP2.

Double heterozygous mutants *ldcp1/+ ldcp2/+* shows gametophytic defects

By crossing the single homozygous mutants, *ldcp1* and *ldcp2*, double heterozygous plants, *ldcp1/+ ldcp2/+*, were generated. Heterozygosity of the lines created was confirmed by genotyping (see figure 5a). Phenotypic observation was performed comparing the heterozygous *ldcp1/+ ldcp2/+*line to the control Col-0. No phenotypic difference was observed between *ldcp1/+ ldcp2/+* plants and Col-0 in terms of non-sexual development, such as plant height, number of rosette leaves or stem structure. The seeds obtained from the cross-pollination showed a similar germination rate to seeds from Col-0. Nevertheless, a detailed observation of the flowers, pollen grains and siliques of *ldcp1/+ ldcp2/+* was performed to identify any potential phenotypic difference. Flowers of the line *ldcp1/+ ldcp2/+* displayed shorter filament than Col-0 (see figure 6b). This result can only be considered preliminary as further characterization with a higher number of flowers from various lines is required to draw significant conclusions. Then, a representative population of pollen grains were observed by light microscopy and over 40% of the pollen grains from the heterozygous double mutant, *ldcp1/+ ldcp2/+* (see figure 7a), were shrunken whereas the rest displayed WT morphological shape (see figure 6c). Next, to check the presence of sperm cells and vegetative nucleus inside the pollen grains, we used Hoechst as nucleic acid staining method and observed the sample by epifluorescence microscopy. This observation revealed that the shrunken pollen grains were also lacking both sperm cells and vegetative nucleus (see figure 6c).

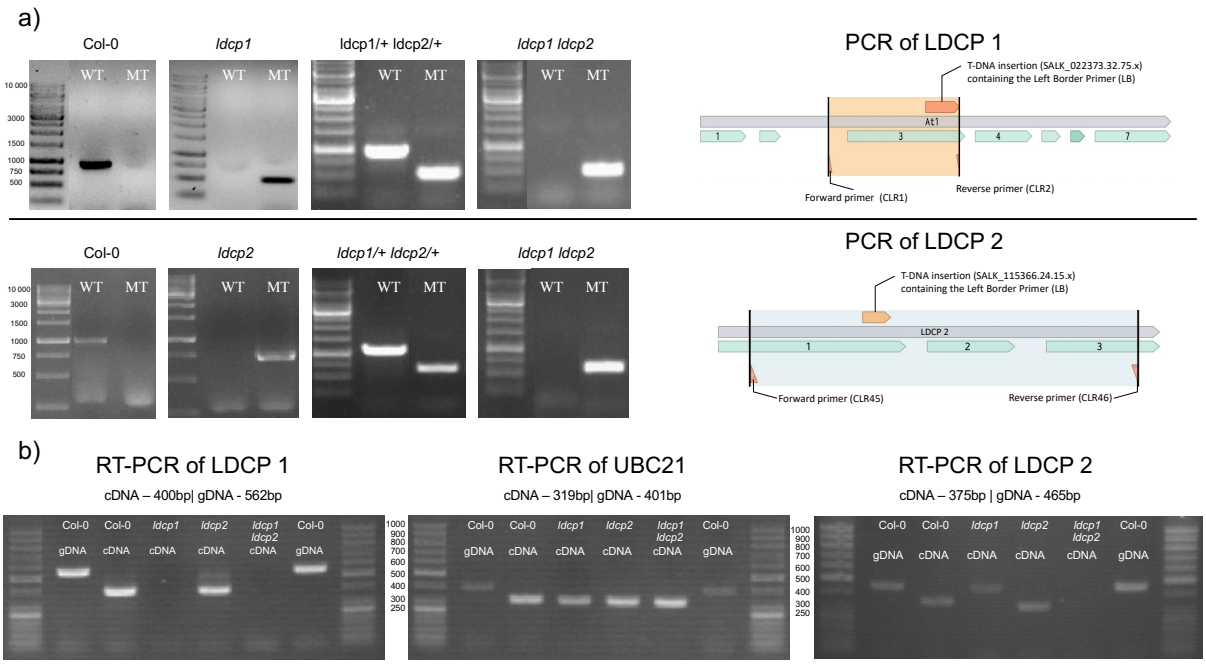


Figure 5 – Genotyping verification of T-DNA mutant lines. a) Gel electrophoresis of PCR from *ldcp1*, *ldcp2* homozygous single mutant and the *ldcp1 ldcp2* double homozygous mutant. The right panel shows the schematic representation of the gene of interest with the localization of the forward and reverse primer, including the predicted insertion of the T-DNA. The upper panel corresponds to the PCR of gene LDCP1 (Expected size for Col-0 wild type band is 980bp and for *ldcp1* band is 450bp) and the lower panel corresponds to the gene LDCP2 (Expected size for Col-0 wild type band is 1150bp and for *ldcp2* band is 700bp). b) Gel electrophoresis of Reverse transcriptase-PCR to detect the presence of transcript. Extracted RNA from unopened flower buds was send for SMART-SEQ2 protocol to the IGC Genomics Unit to obtain cDNA and then a PCR was performed. Ubiquitin 21 (UBC21) was used as control to verify the quality of the cDNA sample.

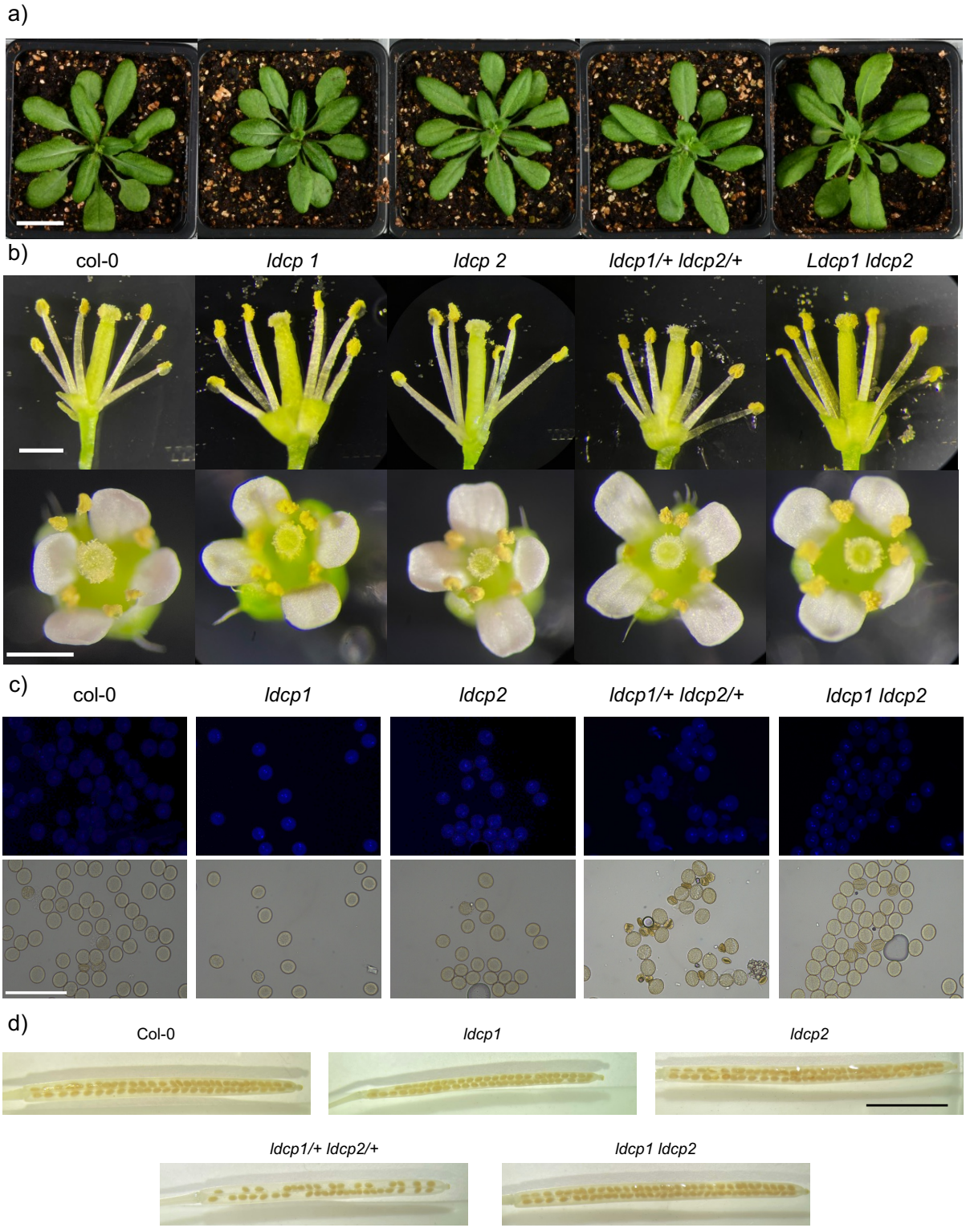


Figure 6 – Phenotypic characterisation of T-DNA line mutants a) Picture of *Arabidopsis thaliana* rosette leaves at day 18 of growth in wild type and mutants. Scale bar = 2.5 cm b) Stereoscope observation of stamen and filaments (top panel, scale bar = 1.0 mm) and flower (bottom panel, scale bar = 2.0 mm) of *A. thaliana* wild type and mutants. c) Microscope observation pollen grains using Höchst staining (top panel) and bright field (bottom panel). Scale bar = 100 μm. d) Stereoscope observation of mature siliques cleared with 99% Ethanol and acetic acid (3:1) solution. Scale bar = 5.0 mm

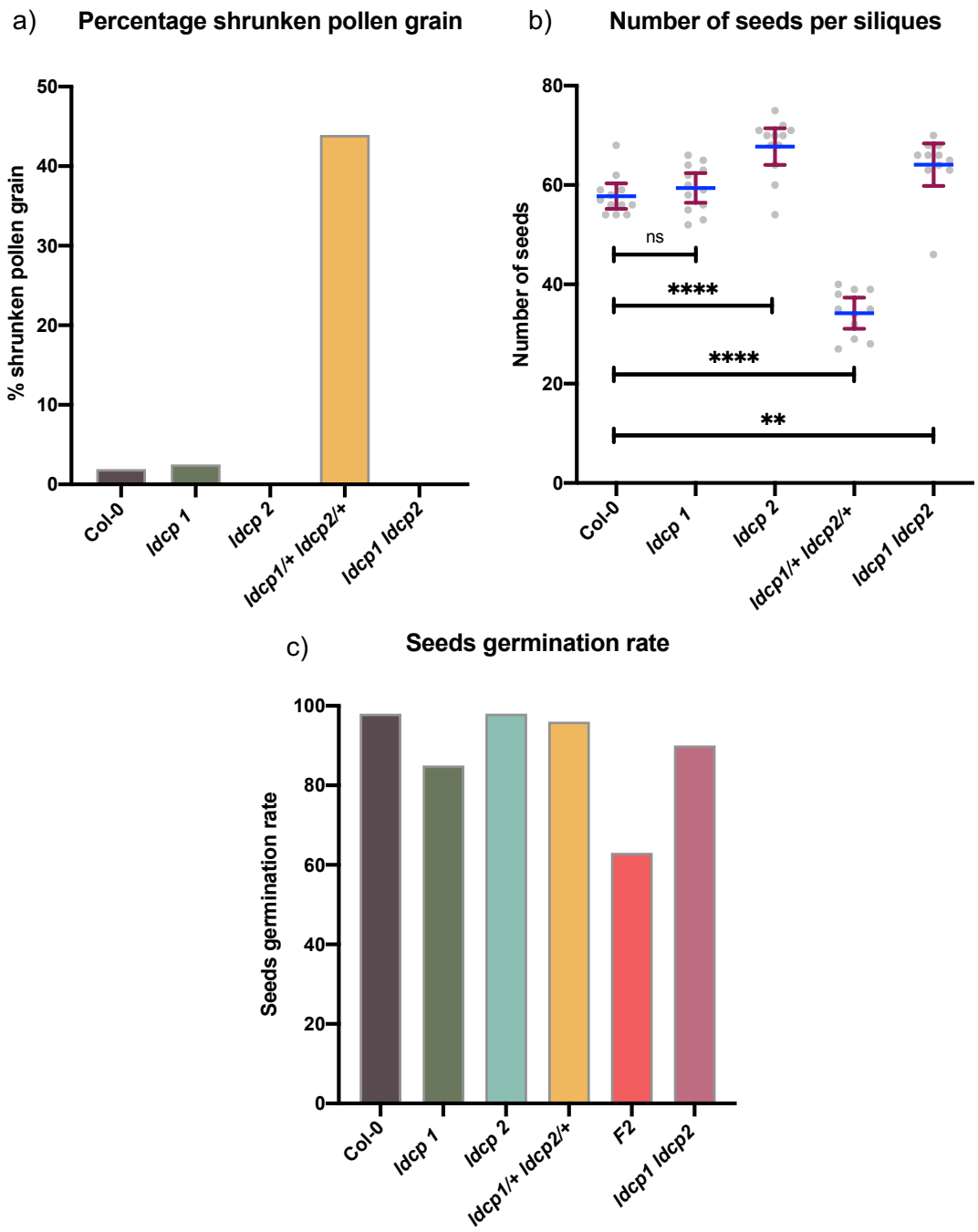


Figure 7 – Phenotypic quantification of T-DNA line mutants a) Percentage shrunken pollen grain of *A.thaliana* wild type and the mutants. b) Graph of the number of seeds per siliques of *A.thaliana* wild type and the mutants. The blue line represents the mean, the purple line represents the 95% confidence interval. Significant difference code: ns = non-significant, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$. c) Graph of seeds germination rate of *A.thaliana* wild type and the mutants. The graph and the statistical calculation were performed using Prism9 software.

Finally, the number of seeds per silique was also explored and statistical analysis indicated that the double heterozygous line displayed a clear reduction of nearly 50% in seeds per silique compared to WT (see figure 7b). The seeds obtained from the self-pollination of ldcp1/+ ldcp2/+ line were collected and sown to investigate the rate of germination and for further characterization of the *f2* generation. The

f2 seeds show a defect in the seed germination assay (see supplemental figure 2). On average, over 30% of the *f2* seeds did not germinate (see figure 7b). This result is contrasted by the large range of results obtained, from 10% to 100% germination rate. These results should be considered preliminary. Nevertheless, one of the siliques yielded 100% germination rate and was therefore a good candidate for full genotypic characterization. Based on genotyping analysis of every plant originated from this single silique, it seems that the inheritance is similar to a normal Mendelian distribution (data not shown). From the set of plants, we could identify 3 homozygous double mutants by genotyping. Single mutant lines and double mutants were also characterized using RT-PCR to confirm the PCR genotyping data and verify the transcript deletion (see figure 5b). The extracted RNA from unopened flower buds was sent for processing with the Smart-Seq2 protocol at the IGC Genomics Unit to obtain cDNAs and the PCR was performed. Ubiquitin 21 (UBC21) was used as control to verify the quality of the cDNA sample. *ldcp1* and one of the previously characterized double homozygous *ldcp1 ldcp2* were found to be true homozygous mutants for the gene *LDCP1*. Mutant *ldcp2* showed amplification in the RT-PCR for the identification of LDCP1, similar to Col-0 cDNA. Unfortunately, the genotyping for LDCP2 gene was troublesome, and we could not obtain information from this reverse PCR reaction. The only phenotype recorded for the apparently true double homozygous *ldcp1 ldcp2* was a significantly higher number of seeds per silique, quite similar to the phenotype recorded in *ldcp2* (see figure 7b).

Successful implementation of a CRISPR-CAS9 protocol

Recently, T-DNA lines have shown their limits (chromosome rearrangements, limited number of mutants available, additional insertions in the same line). Therefore, in parallel to T-DNA mutant characterisation, the CRISPR-Cas9 approach was developed. Following the reception of the plasmids GEE, the plasmid concentration was checked and GEE was transformed into competent *E.coli* alpha DH5 cells. The cells containing the GEE plasmid were grown, purified and digested for size verification (see figure 8a). The size of the plasmid was as expected (around 16kb) and the digestion revealed no missing fragments. The sgRNA candidates (CLR3, CLR5, CLR7, CLR9, CLR13, CLR39) were individually inserted, by heat-shock, into the plasmid and the resulting plasmids re-transformed into competent cells. The transformation was verified via colony PCR (see figure 8c). The positive candidates were selected via antibiotic resistance. The ability of sgRNA candidates to cleave the genes of interest was examined by *In vitro* CRISPR-Cas9 cleaving assay. First, the sgRNA candidates were amplified and transcribed from the plasmid. The sgRNA candidates then formed a complex with commercially available Cas9 enzyme. The complex was exposed to the gene of interest *LDCP1* and *LDCP2,* previously amplified by PCR. If the Cas9 protein, guided by the sgRNA, recognised the gene of interest, it then cleaves the fragment producing a particular pattern. The result of the cleavage was evaluated via gel electrophoresis and from this preliminary assay, 3 candidates were kept for further experimental work (CLR3, CLR5 and CLR7). The cleaving efficiency was verified in a second round of *in vitro* CRISPR-Cas9 assays (see figure 8d). The plasmid was then inserted into *Agrobacterium tumefaciens* and PCR colony was performed to confirm the plasmid insertion (see figure 8e).

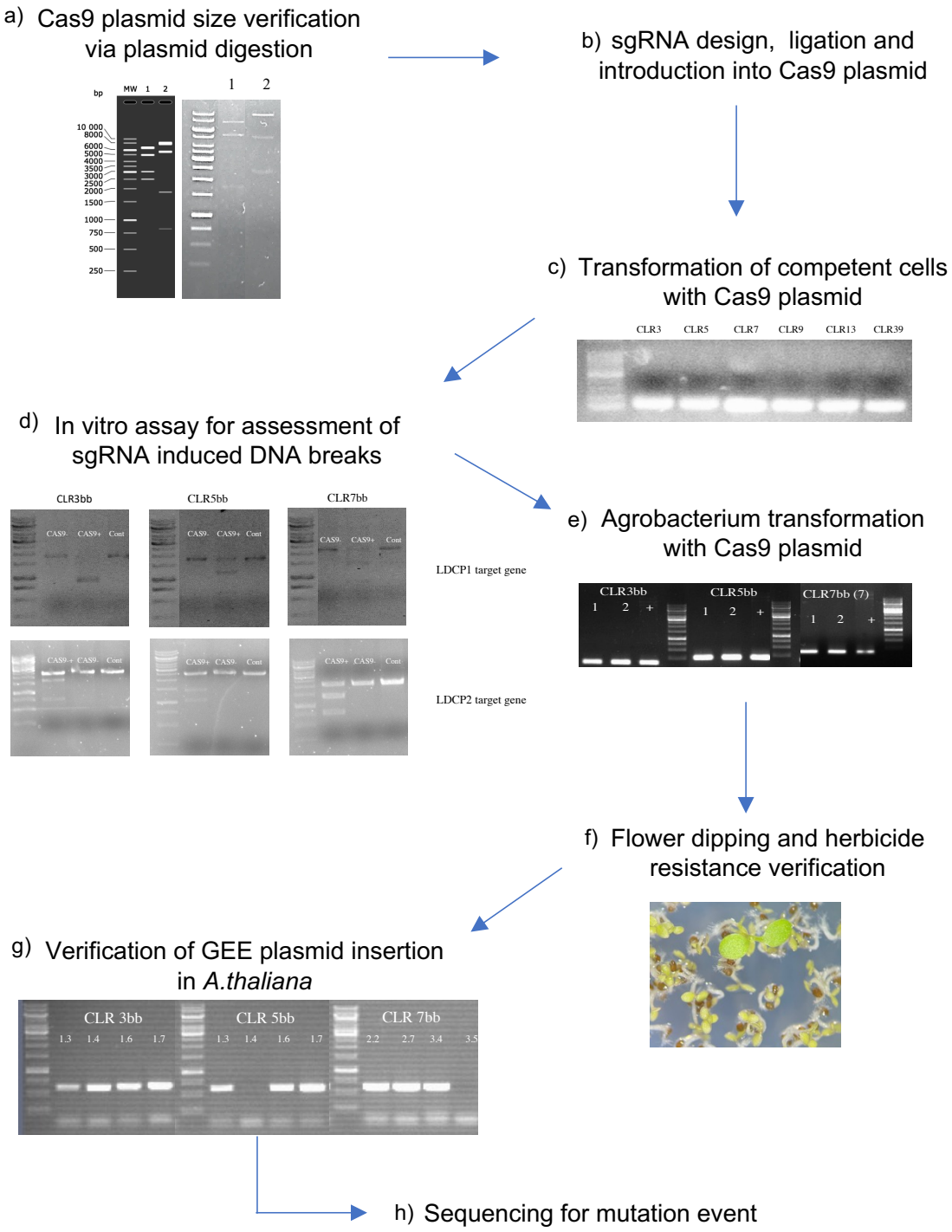


Figure 8– Schematic representation of CRISPR-Cas9 development. a) Size verification of plasmid (1) GEE and (2) ROXY using EcoRI | MluI and EcoRI | HpaI, respectively, including a 1kB DNA Ladder. Left panel includes the predicted digested size. b) sgRNA design and introduction was performed following the protocol provided by Dr. Corinna Thurow (see materials and methods for further information) c) Verification, via PCR, of Cas9 plasmid insertion into competent cells (expected product size – 300bp). d) *In vitro* CAS9 cleavage of *LDCP* 1 and *LDCP* 2 PCR products. e) PCR for the detection of the Cas9 plasmid, containing the specific sgRNA insertion, following a heat-shock agrobacterium transformation (expected product size – 130bp). f) Following the flower dipping steps, seeds were sown onto glyphosate containing media and resistant plant were moved onto soil. g) Verification of the Cas9 containing the specific sgRNA insertion into *A.thaliana* (expected product size – 620bp) h) Sequencing data was performed for 7 candidates containing both sgRNA insertion (data not shown).

Agrobacterium positive candidates were grown in antibiotic containing media and extracted. The plasmid containing agrobacterium, of equal concentration, were mixed for 3 different transformation events (CLR3+CLR5, CLR3+CLR7 and CLR5+CLR7). The transformation by flower dipping was carried out using *Arabidopsis thaliana* Col-0, and the resulting seeds were then sown on glyphosate containing media and the resistant plants (see figure 8f) were transferred to soil. Over 30 resistant mutants were transferred and genotyped for either large deletions or for insertion of the plasmid (see figure 8g). No large deletions were found, but all the plants genotyped contained at least one plasmid insertion. We focused on 10 individual plants containing both insertions and to sequence the region of interest for possible CRISPR-CAS9 introduced mutation. No pattern of mutation was found in the plants containing both insertions.

Discussion

LDCPs are linked to pollen development defect

LDCPs have been quite heavily linked to gametophytic defects. Recently, in *Brassica campestris*, two novel proteins containing the same structural domains have been linked to similar pollen defect observed in our study (Xu et al, 2020b). The two genes (BRA015163 and BRA013181) encode a protein containing a CCCH zinc-finger domain, a LOTUS domain and an RNA recognition motif domain in the same order. The expression patterns of BRA015163 and BRA013181 are very similar to *LDCP1* and *LDCP2* as they are specifically expressed in inflorescences. Xu et al, 2020b created mutant lines for BRA015163 and BRA013181 that exhibited partial male sterility. They recorded a pollen abortion rate of 30% for the double homozygous mutants as compared to 4% for the control plants, including a 40% reduction in pollen viability in the mutants. They also observed that, *in vivo* and *in vitro*, pollen germination was reduced to half in the mutant plants. Finally, they observed the shape of the pollen and reported a shrunken and collapsed shape, with a lack of nuclei in abnormal pollen grains (Xu et al, 2020b), finding similar results to ours in *Arabidopsis* double heterozygous ldcp1/+ ldcp2/+. This suggests that the genes *LDCP1* and *LDCP2,* based on protein domain similarity, expression pattern and the phenotype, could be closely related to the genes *BRA015163* and *BRA013181*. *LDCP2,* which was found to be highly and specifically expressed during pollen development in *A. thaliana* (Xu et al, 2020a), was also investigated using over-expression mutant (Xu et al, 2022). A knock-out homozygous mutant was generated for the gene *LDCP2,* using the CRISPR-Cas9 technology, which revealed no developmental or sexual related phenotype, possibly resulting from gene redundancy. An over-expression mutant of *LDCP2* was generated, which led to pollen developmental defects. In the over-expression mutant, the pollen was shrunken, leading to male sterility. They also observed that the abnormal pollen development was always associated with the formation of a high number of granules, suggesting that an appropriate expression level of *LDCP2* is essential for proper pollen development in *A.thaliana* (Xu et al, 2022).

Pollen defect phenotype could also be explained by chromosome translocation

Unfortunately, the literature available regarding *LDCPs* does not hint at any phenotype being observed in the double heterozygous phase, which would disappear in the double homozygous. T-DNA insertion was suspected for a long time to cause chromosome re-arrangement (Castle et al, 1993) and this was recently heavily re-investigated due to the increased availability of chromosome analysis techniques. It has been shown that in *A. thaliana* GABI T-DNA lines, out of the 14 mutants being investigated, 6 displayed chromosomal rearrangement, mainly arm translocation (Pucker et al, 2021). So far, and to our knowledge, only one article has mentioned a heterozygous based phenotype disappearing in the homozygous lines. The Anti-Silencing Function (ASF1) is a gene linked with chromatin assembly during key stages such as transcription, replication, and repair (Adkins et al 2007). ASF1A and ASF1B are homologues in Arabidopsis and their respective single T-DNA lines do not show any gametophytic

defects but when cross-pollination occurs, the defects, quasi similar to ours, appear in the double heterozygous before disappearing in the double homozygous line. The phenotypes observed are showing roughly 50% of unfertilized ovules and reduced seed viability in the heterozygous line, whereas the homozygous line only shows an 8% ovule abortion rate (Min et al, 2020).

Limitation of the study

The Reverse-Transcriptase PCR to detect the presence of the transcript is key to any mutational studies as the insertion of T-DNA inside a gene does not always disrupt its transcription. Unfortunately, the RT-PCR for the detection of the *LDCP2* (At3G52980) transcript was very problematic. We used 3 techniques to produce cDNA, namely primers specific, oligo DT and SmartSeq2 amplification. We conducted the primer specific and the oligo DT assay in the lab, while the SmartSeq2 was sent to IGC´s Genomics Unit. In total, 10 different pairs of primers were used to span both predicted introns and to confirm the deletion of the transcript of the *LDCP2* gene. This gene contains 2 predicted introns but one of them was recently suspected to contain a stop-codon read-through event (Xu et al, 2020a). This read through event could explain some of the issues we have been experiencing regarding RT-PCR. Unfortunately, the genes *LDCP1* and *LDCP2* have a very high degree of identical DNA sequence, especially in the area spanning the intron of interest. Following optimization steps, we decided to choose an area for primer annealing with the highest number of mismatches. The forward primer, for gene *LDCP2*, was designed to have 6 mismatches and the reverse primer to have 8 mismatches with the gene *LDCP1*. The specificity of the pair of primers for the gene *LDCP2* was verified using the in-silico primer affinity tool Primer blast from NCBI. The mutant *ldcp1*, under the *LDCP2* RT-PCR set up, produced a fragment of similar size to Col-0 gDNA, which is quite surprising when considering no sign of gDNA contamination was present in the control reaction for Ubiquitin21. The mutant *ldcp2* amplified a fragment which seems to be slightly smaller than the Col-0 cDNA sample. Finally, the mutant *ldcp1 ldcp2*, which was genotyped by PCR to be a double homozygous mutant did not amplify a fragment using the primer for *LDCP2* identification, whereas its direct parental line *ldcp2* amplified a fragment, using the same PCR condition.

Future prospect

T-DNA mutants have been used for the last two decades for reverse genetics approach with great success, but recently limitations have been pointed out. The main disadvantages are possible chromosome rearrangements (Tamura et al, 2016; Min et al, 2020), the limited number of mutants available for one gene and the potential additional insertions in the same line. On the other hand, CRISPR technology appears as a potential candidate to overcome these limitations. The study we carried out is a good example of the limitations of using T-DNA lines to study processes involving several genes. It also shows the need for mutational study to include mixed approaches and complementation to answer a complex biological question. Whether the pollen defect phenotype observed in the double heterozygous mutant line was caused by T-DNA related chromosome rearrangement or currently unknown DNA maintenance mechanism, CRISPR-Cas9 was designed to overcome the limitation encountered. The implementation, for the first time in our lab, of the CRISPR-Ca9 technology was a

success. We could insert our sgRNA candidates into the Cas9 plasmid, to check the ability to cleave our gene of interest and finally to transform our plants and check for the insertion. We could also sequence some of the lines but found no DNA insertions or deletions of nucleotides in the targeted regions. Concerning the future prospect of the T-DNA lines, preliminary assays to understand whether the heterozygosity of either *LDCP1* or *LDCP2* is responsible for the heterozygous-only phenotypes can be performed. Crossing both lines with the wild type Col-0 can be a quick and easy option. T-DNA junction mapping can also help to identify possible chromosome translocation events (Valente et al, 2018). The future prospect of this project would be to further characterise the rest of the lines and the second generation of CRISPR lines. This should allow for the identification of lines where Cas9 could create an insertion or deletion, or even better, a whole segment deletion.

References

Adkins, M.W., Carson, J.J., English, C.M., Ramey, C.J. and Tyler, J.K., 2007. The histone chaperone anti-silencing function 1 stimulates the acetylation of newly synthesized histone H3 in S-phase. Journal of Biological Chemistry, 282(2), pp.1334-1340.

Alkan, F., Wenzel, A., Anthon, C., Havgaard, J.H. and Gorodkin, J., 2018. CRISPR-Cas9 off-targeting assessment with nucleic acid duplex energy parameters. Genome biology, 19(1), pp.1-13.

Anantharaman, Vivek, Dapeng Zhang, and L. Aravind. "OST-HTH: a novel predicted RNA-binding domain." Biology direct 5.1 (2010): 1-8.Bennetzen JL. 2000. Transposable element contributions to plant gene and genome evolution. Plant Mol Biol. 42:251–269.

Bente, H., Mittelsten Scheid, O. and Donà, M., 2020. Versatile in vitro assay to recognize Cas9-induced mutations. Plant Direct, 4(9), p.e00269.

Brown, R. C. & Lemmon, B. E. Spores before sporophytes: hypothesizing the origin of sporogenesis at the algal–plant transition. New Phytol. 190, 875–881 (2011).cells of Arabidopsis and rice. Proceedings of the National Academy of Sciences, USA Clark KA, Krysan PJ. Chromosomal translocations are a common phenomenon in Arabidopsis thaliana T-DNA insertion lines. Plant J. 2010;64(6):990–1001.

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.

Ding, D., Wei, C., Dong, K., Liu, J., Stanton, A., Xu, C., Min, J., Hu, J. and Chen, C., 2020. LOTUS domain is a novel class of G-rich and G-quadruplex RNA binding domain. Nucleic acids research, 48(16), pp.9262-9272.

Ding,D., Liu,J., Midic,U., Wu,Y., Dong,K., Melnick,A., Latham,K.E. and Chen,C. (2018) TDRD5 binds piRNA precursors and selectively enhances pachytene piRNA processing in mice. Nat. Commun., 9, 127.

Edlund AF, Swanson R, Preuss D (2004) Pollen and Stigma Structure and Function: The Role of Diversity in Pollination. Plant Cell 16: 84–97

Edwards, K., Johnstone, C. and Thompson, C., 1991. A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic acids research, 19(6), p.1349.

European Environment Agency, published on the 15th of December 2021, available at : https://www.eea.europa.eu/ims/greenhouse-gas-emissions-from-agriculture

European Parliamentary Research Service. (May 2020) Author: James McEldowney Members' Research Service PE 651.922

Fan, L.M., Wang, Y.F., Wang, H. and Wu, W.H., 2001. In vitro Arabidopsis pollen germination and characterization of the inward potassium currents in Arabidopsis pollen grain protoplasts. Journal of Experimental Botany, 52(361), pp.1603-1614.

Fang L, Leliaert F, Zhang ZH, Penny D, Zhong BJ (2017) Evolution of the Chlorophyta: insights from chloroplast phylogenomic analy- ses. J Syst Evol 55(4):322–332. https://doi.org/10.1111/jse.12248

FAO. 2020. Emissions due to agriculture. Global, regional and country trends 2000–2018. FAOSTAT Analytical Brief Series No 18. Rome

Fedoroff, N.V., 2012. Transposable elements, epigenetics, and genome evolution. Science, 338(6108), pp.758-767.

Flores-Tornero, M., Sprunck, S. and Dresselhaus, T., 2021. Identification and characterization of reproductive mutations in Arabidopsis. In Arabidopsis Protocols (pp. 371-390). Humana, New York, NY.

Galindo-González, Leonardo, et al. "LTR-retrotransposons in plants: Engines of evolution." Gene 626 (2017): 14-25.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Pnas 109, E2579–E2586.

Gasser CS, Skinner DJ (2019) Development and evolution of the unique ovules of flowering plants. Current topics in develop- mental biology, vol 131. Elsevier.

Gehring, M., 2019. Epigenetic dynamics during flowering plant reproduction: evidence for reprogramming?. New Phytologist, 224(1), pp.91-96.

Hall, T.M.T. Multiple modes of RNA recognition by zinc finger proteins. Curr. Opin. Struct. Biol. 2005, 15, 367–373.

He ZH, Dong HT, Dong JX, Li DB, Ronald PC. 2000. The rice Rim2 transcript accumulates in response to Magnaporthe grisea and its predicted pro- tein product shares similarity with TNP2-like proteins encoded by CACT A transposons. Mol Gen Genet. 264:2–10.

Hedhly, A., Hormaza, J.I. and Herrero, M., 2009. Global warming and sexual plant reproduction. Trends in plant science, 14(1), pp.30-36.

Ingouff, M., Selles, B., Michaud, C., Vu, T.M., Berger, F., Schorn, A.J., Autran, D., Van Durme, M., Nowack, M.K., Martienssen, R.A. and Grimanelli, D., 2017. Live-cell analysis of DNA methylation during sexual reproduction in Arabidopsis reveals context and sex-specific dynamics controlled by noncanonical RdDM. Genes & development, 31(1), pp.72-83.

IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.

Jeske, Mandy, et al. "The crystal structure of the Drosophila germline inducer Oskar identifies two domains with distinct Vasa helicase-and RNA-binding activities." Cell reports 12.4 (2015): 587-598.

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.

Jupe, F., Rivkin, A.C., Michael, T.P., Zander, M., Motley, S.T., Sandoval, J.P., Slotkin, R.K., Chen, H., Castanon, R., Nery, J.R. and Ecker, J.R., 2019. The complex architecture and epigenomic impact of plant T-DNA insertions. PLoS genetics, 15(1), p.e1007819.Lacal I, Ventura R. Epigenetic inheritance: concepts, mechanisms and perspectives. Frontiers in molecular neuroscience. (2018) Sep 28;11:292.Laity, J.H.; Lee, B.M.; Wright, P.E. Zinc finger proteins: New insights into structural and functional diversity. Curr. Opin. Struct. Biol. 2001, 11, 39–46.

Lee, L.Y. and Gelvin, S.B., 2008. T-DNA binary vectors and systems. Plant physiology, 146(2), pp.325-332.

Leonelli, S., 2007. Arabidopsis, the botanical Drosophila: from mouse cress to model organism. Endeavour, 31(1), pp.34-38.

Lexa M, Kejnovský E, Steflová P et al (2014a) Quadruplex- forming sequences occupy discrete regions inside plant LTR retrotransposons. Nucleic Acids Res 42:968–978

Liang, Z., Zhang, K., Chen, K. and Gao, C., 2014. Targeted mutagenesis in Zea mays using TALENs and the CRISPR/Cas system. Journal of Genetics and Genomics, 41(2), pp.63-68.Mashiguchi, K., Hisano, H., Takeda-Kamiya, N., Takebayashi, Y., Ariizumi, T., Gao, Y., Ezura, H., Sato, K., Zhao, Y., Hayashi, K.I. and Kasahara, H., 2019. Agrobacterium tumefaciens enhances biosynthesis of two distinct auxins in the formation of crown galls. Plant and Cell Physiology, 60(1), pp.29-37.

Meissner, Scott. "Plant Sexual Reproduction: Perhaps the current plant two-sex model should be replaced with three-and four-sex models?." (2020).

Min, Y., Frost, J.M. and Choi, Y., 2020. Gametophytic Abortion in Heterozygotes but Not in Homozygotes: Implied Chromosome Rearrangement during T-DNA ertion at the ASF1 Locus in Arabidopsis. Molecules and cells, 43(5), p.448.

Mojica, F.J., Díez-Villaseñor, C., García-Martínez, J. and Soria, E., 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. Journal of molecular evolution, 60(2), pp.174-182.

Nelson, G. C., Valin, H., Sands, R. D., Havlík, P., Ahammad, H., Deryng, D., ... & Willenbockel, D. (2014). Climate change effects on agriculture: Economic responses to biophysical shocks. Proceedings of the National Academy of Sciences, 111(9), 3274-3279.

Oliver, Keith R., Jen A. McComb, and Wayne K. Greene. "Transposable elements: powerful contributors to angiosperm evolution and diversity." Genome biology and evolution 5.10 (2013): 1886-1901.

Orgel LE, Crick FH. 1980. Selfish DNA: the ultimate parasite. Nature 284: Park K, Kim MY, Vickers M, Park JS, Hyun Y, Okamoto T, Zilberman D, Fischer Patil, V.S., Anand, A., Chakrabarti, A. and Kai, T., 2014. The Tudor domain protein Tapas, a homolog of the vertebrate Tdrd7, functions in the piRNA pathway to regulate retrotransposons in germline of Drosophila melanogaster. BMC biology, 12(1), pp.1-15.

Patil, V.S., Anand, A., Chakrabarti, A. and Kai, T., 2014. The Tudor domain protein Tapas, a homolog of the vertebrate Tdrd7, functions in the piRNA pathway to regulate retrotransposons in germline of Drosophila melanogaster. BMC biology, 12(1), pp.1-15.

Pauwels, L., De Clercq, R., Goossens, J., Iñigo, S., Williams, C., Ron, M., Britt, A. and Goossens, A., 2018. A dual sgRNA approach for functional genomics in Arabidopsis thaliana. G3: Genes, Genomes, Genetics, 8(8), pp.2603-2615.

Provart, N.J., Alonso, J., Assmann, S.M., Bergmann, D., Brady, S.M., Brkljacic, J., Browse, J., Chapple, C., Colot, V., Cutler, S. and Dangl, J., 2016. 50 years of Arabidopsis research: highlights and future directions. New Phytologist, 209(3), pp.921-944.Pucker, B., Kleinbölting, N. and Weisshaar, B., 2021. Large scale genomic rearrangements in selected Arabidopsis thaliana T-DNA lines are caused by T-DNA insertion mutagenesis. BMC genomics, 22(1), pp.1-21.

Pucker, B., Kleinbölting, N. and Weisshaar, B., 2021. Large scale genomic rearrangements in selected Arabidopsis thaliana T-DNA lines are caused by T-DNA insertion mutagenesis. BMC genomics, 22(1), pp.1-21.

Purugganan, M.D., 2019. Evolutionary insights into the nature of plant domestication. Current Biology, 29(14), pp.R705-R714.

Ramakrishnan, M., Satish, L., Kalendar, R., Narayanan, M., Kandasamy, S., Sharma, A., Emamverdian, A., Wei, Q. and Zhou, M., 2021. The dynamism of transposon methylation for plant development and stress adaptation. International Journal of Molecular Sciences, 22(21), p.11387.

Ricroch, A., 2019, August. Global developments of genome editing in agriculture. In Transgenic research (Vol. 28, No. 2, pp. 45-52). Springer International Publishing.RL, Feng X, Choi Y et al. 2016. DNA demethylation is initiated in the central Scopus.com. 2022. Welcome. [Accessed 2 June 2022]. Available at: https://tinyurl.com/4rw3smj9

Scott, J.C., 2017. Against the grain: A deep history of the earliest states. Yale University Press.

Sessions, A., Burke, E., Presting, G., Aux, G., McElver, J., Patton, D., Dietrich, B., Ho, P., Bacwaden, J., Ko, C. and Clarke, J.D., 2002. A high-throughput Arabidopsis reverse genetics system. The Plant Cell, 14(12), pp.2985-2994.

Shimon and Horwitz. "The impact of Arabidopsis research on plant biotechnology." Biotechnology advances 13.3 (1995): 403-414.Shivanna, K. R. (2003). Pollen biology and biotechnology. Plymouth: Science Publ. 2003Sigman, M.J., Panda, K., Kirchner, R., McLain, L.L., Payne, H., Peasari, J.R., Husbands, A.Y., Slotkin, R.K. and McCue, A.D., 2021. An siRNA-guided ARGONAUTE protein directs RNA polymerase V to initiate DNA methylation. Nature plants, 7(11), pp.1461-1474.

Singh, J., Mishra, V., Wang, F., Huang, H.Y. and Pikaard, C.S., 2019. Reaction mechanisms of Pol IV, RDR2, and DCL3 drive RNA channeling in the siRNA-directed DNA methylation pathway. Molecular cell, 75(3), pp.576-589.

Skea, J., Shukla, P. and Kılkış, Ş., 2022. Climate Change 2022: Mitigation of Climate Change.Slotkin, R.K., Vaughn, M., Borges, F., Tanurdzic, M., Becker, J.D., Feijo´, J.A., and Martienssen, R.A. (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell 136, 461–472

Slotkin, R.K., Vaughn, M., Borges, F., Tanurdžić, M., Becker, J.D., Feijó, J.A. and Martienssen, R.A., 2009. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell, 136(3), pp.461-472.

Smith, P.; Davis, S.J.; Creutzig, F.; Fuss, S.; Minx, J.C.; Gabrielle, B.; Kato, E.; Jackson, R.B.; Cowie, A.; Kriegler, E.; et al. Biophysical and economic limits to negative CO2 emissions. Nat. Clim. Chang. 2015, 6, 42–50.Stoll, B. and Binder, S., 2016. Two NYN domain containing putative nucleases are involved in transcript maturation in Arabidopsis mitochondria. The Plant Journal, 85(2), pp.278-288.

Sturme, M.H., van der Berg, J.P., Bouwman, L.M., De Schrijver, A., de Maagd, R.A., Kleter, G.A. and Battaglia-de Wilde, E., 2022. Occurrence and Nature of Off-Target Modifications by CRISPR-Cas Genome Editing in Plants. ACS Agricultural Science & Technology, 2(2), pp.192-201.

Su, You-Qiang, et al. "MARF1 regulates essential oogenic processes in mice." Science 335.6075 (2012): 1496-1499.

Su, You-Qiang, Fengyun Sun, Mary Ann Handel, John C. Schimenti, and John J. Eppig. "Meiosis arrest female 1 (MARF1) has nuage-like function in mammalian oocytes." Proceedings of the National Academy of Sciences 109, no. 46 (2012): 18653-18660.

Tamura, K., Kawabayashi, T., Shikanai, T. and Hara-Nishimura, I., 2016. Decreased expression of a gene caused by a T-DNA insertion in an adjacent gene in arabidopsis. Plos one, 11(2), p.e0147911.

Umen J, Coelho S (2019) Algal sex determination and the evolution of anisogamy. Annu Rev Microbiol 73:267–291. https://doi.org/ 10.1146/annurev-micro-020518-120011

Ünal, M., Vardar, F. and Aytürk, Ö., 2013. Callose in plant sexual reproduction. In Current progress in biological research. IntechOpen.

Valente, A.S., Tutone, M., Brodie, E., Peper, H. and Pillitteri, L.J., 2018. T-DNA associated reciprocal translocation reveals differential survival of male and female gametes. Plant Gene, 15, pp.37-43.

Vora, D.S., Dhanjal, J.K. and Sundar, D., 2020. Engineering of Cas9 for improved functionality. In Genome Engineering via CRISPR-Cas9 System (pp. 111-122). Academic Press.

Wang, Y., Meng, Z., Liang, C., Meng, Z., Wang, Y., Sun, G., Zhu, T., Cai, Y., Guo, S., Zhang, R. and Lin, Y., 2017. Increased lateral root formation by CRISPR/Cas9-mediated editing of arginase genes in cotton. Science China. Life Sciences, 60(5), p.524.

Wieczorek, A. M. & Wright, M. G. (2012) History of Agricultural Biotechnology: How Crop Development has Evolved. Nature Education Knowledge 3(10):9

Xu, L., Liu, T., Xiong, X., Liu, W., Yu, Y., & Cao, J. (2020a). AtC3H18L is a stop-codon read-through gene and encodes a novel non-tandem CCCH zinc-finger protein that can form cytoplasmic foci similar to mRNP granules. Biochemical and Biophysical Research Communications.

Xu, L., Liu, T., Xiong, X., Liu, W., Yu, Y. and Cao, J., (2020c). Overexpression of two CCCH-type zinc-finger protein genes leads to pollen abortion in Brassica campestris ssp. chinensis. Genes, 11(11), p.1287.

Xu, L., Liu, T., Xiong, X., Shen, X., Huang, L., Yu, Y. and Cao, J., (2022). Highly Overexpressed AtC3H18 Impairs Microgametogenesis via Promoting the Continuous Assembly of mRNP Granules. Frontiers in Plant Science, p.2195.

Xu, L., Xiong, X., Liu, W., Liu, T., Yu, Y. and Cao, J., (2020b) BcMF30a and BcMF30c, Two Novel Non-Tandem CCCH Zinc-Finger Proteins, Function in Pollen Development and Pollen Germination in Brassica campestris ssp. chinensis. International journal of molecular sciences, 21(17), p.6428.

Yabuta, Y., Ohta, H., Abe, T., Kurimoto, K., Chuma, S., & Saitou, M. (2011). TDRD5 is required for retrotransposon silencing, chromatoid body assembly, and spermiogenesis in mice. Journal of Cell Biology, 192(5), 781-795.

Yang, X., Cheema, J., Zhang, Y., Deng, H., Duncan, S., Umar, M.I., Zhao, J., Liu, Q., Cao, X., Kwok, C.K. and Ding, Y., 2020. RNA G-quadruplex structures exist and function in vivo in plants. Genome biology, 21(1), pp.1-23.

Yang, Y., Zhu, K., Li, H., Han, S., Meng, Q., Khan, S.U., Fan, C., Xie, K. and Zhou, Y., 2018. Precise editing of CLAVATA genes in Brassica napus L. regulates multilocular silique development. Plant biotechnology journal, 16(7), pp.1322-1335.Zhang F (2019). Development of CRISPR-Cas systems for genome editing and beyond. Quarterly Reviews of Biophysics 52, e6, 1–31. https://doi.org/10.1017/ S0033583519000052

Zhang, James Z., Robert A. Creelman, and Jian-Kang Zhu. "From laboratory to field. Using information from Arabidopsis to engineer salt, cold, and drought tolerance in crops." Plant physiology 135.2 (2004): 615-621.

Supplemental material

Supplemental table 1 - List of reagents used for PCR reaction with DreamTaq DNA Polymerase (Thermo Scientific).

Reagents	Quantity
Dream taq Buffer	10x
dNTP	50x
Forward primer	10x
Reverse Primer	10x
Dream taq Polymerase	250x
DNA sample	200x
H2O	Fill-up

Supplemental table 2 - List of reagents used for the GEE plasmid digestion using BsaI (NEB)

Reagent	Volume
CutSmart buffer	4ul
pBCsGFPEE ("GEE plasmid")(500 ng/ul)	34ul
BsaI (NEB)	2ul
Total	*40ul*
Leave at 37 C for o/n digestion	

Supplemental table 3 - List of reagents used for the ligation between sgRNA and plasmid

Reagent	**Volume**
pGEE digested with BsaI (500 ng of digested plasmid)	17 ul
Annealed oligos	0.5ul
T4 DNA ligase Ligation buffer (ThermoScientific)	1ul
T4 ligase – (5 U/ul) (ThermoScientific)	0.5ul
Total	***20ul***
Incubate o/n at room temperature	

Supplemental table 4 - List of primers used to identify the insertion of the GEE plasmid into the competent cells via colony PCR.

PCR colony primers (only forward; reverse is always the same L236: (GAGCCGGAAGCATAAAGTGTAAAGC)
Expected product size – around 650 bp. Annealing temperturature 57˚C

Target	Primer ID	sequence(5'-3')
CLR3bb	CLR4 (oligo BOTTOM)	AAACGCTTTGCCATGTCTTTGTGA C
CLR5bb	CLR6 (oligo BOTTOM)	AAACGCAGAGAGTAGTTTCACTGA C
CLR7bb	CLR8 (oligo BOTTOM)	AAACTTGTTGAAGTAATGGCAGAT C
CLR9bb	CLR10 (oligo BOTTOM)	AAACCATGGACAACATTCAGTTAT C
CLR13bb	CLR14 (oligo BOTTOM)	AAACCATGGACAACATTCAGTTAT C
CLR39bb	CLR40(oligo BOTTOM)	AAACTCAACTTGAACCAGAGAATG C

Supplemental table 5 - List of primers used for the amplification of the sgRNA template for *in vitro* transcription of the sgRNA

In vitro transcription primer (only forward; reverse is always the same CLR22: AAAAAAAGCACCGACTCGGTGCC)
Expected product size – around 130 bp. Annealing temperturature 63˚C

Target	Primer ID	sequence (5´-3´)
CLR3bb	CLR21	GTTTCTTTAATACGACTCACTATAGGTCACAAAGACATGGCAAAGCG
CLR5bb	CLR23	GTTTCTTTAATACGACTCACTATAGGTCAGTGAAACTACTCTCTGCGTTTTAG
CLR7bb	CLR25	GTTTCTTTAATACGACTCACTATAGG ATCTGCCATTACTTCAACAA (GTTTTAGAGC)
CLR9bb	CLR27	GTTTCTTTAATACGACTCACTATAGG ATAACTGAATGTTGtCCATG (GTTTTAGAGC)
CLR13bb	CLR31	GTTTCTTTAATACGACTCACTATAGGATAACTGAATGTTGtCCATG (GTTTTAGAGC)
CLR39bb	CLR41	GTTTCTTTAATACGACTCACTATAGGCATTCTCTGGTTCAAGTTGA (GTTTTAGAGC)

Supplemental table 6 - List of reagents used for the *in vitro* transcription of the sgRNA

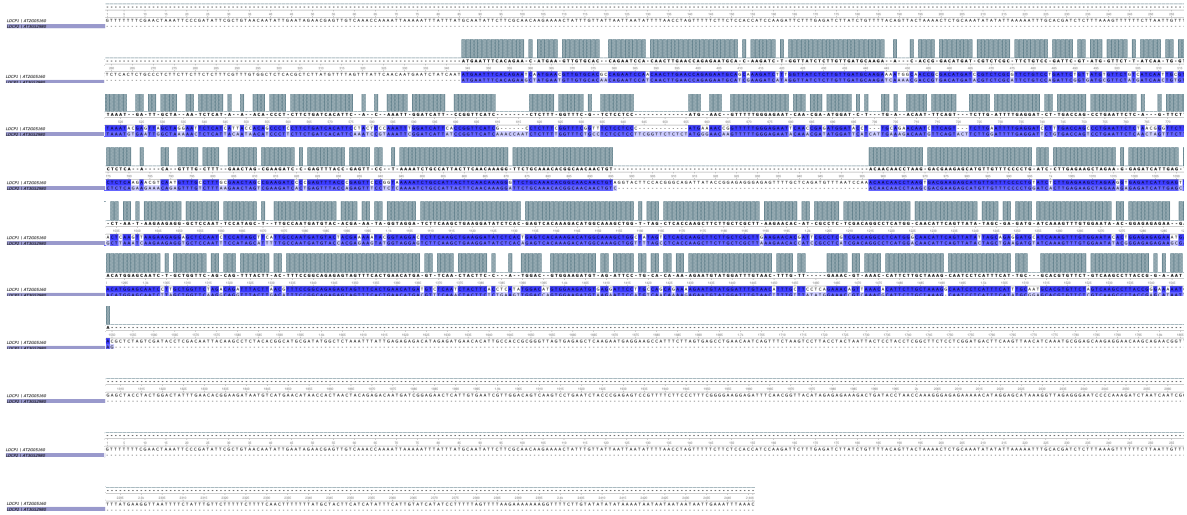
Reagent	T7 +	T7 -
H2O	4 ul	4 ul
T7 transcription buffer (ThermoScientific) (5x)	4 ul	4 ul
riboNTPs (80 mM mix, 20 mM each)	2 ul	2 ul
DNA template (50 ng/ul)	10 ul	10 ul
T7 RNA polymerase (ThermoScientific) (20 U/ul)	0.7 ul	-
RNAse Inhibitor (Riboguard, 40U/ul)	0.2ul in total	
Total	***20 ul***	***20 ul***
	Incubate at 37 C o/n ;	

Supplemental table 7 - List of reagents used for the *in vitro* Cas9 cleavage of PCR products

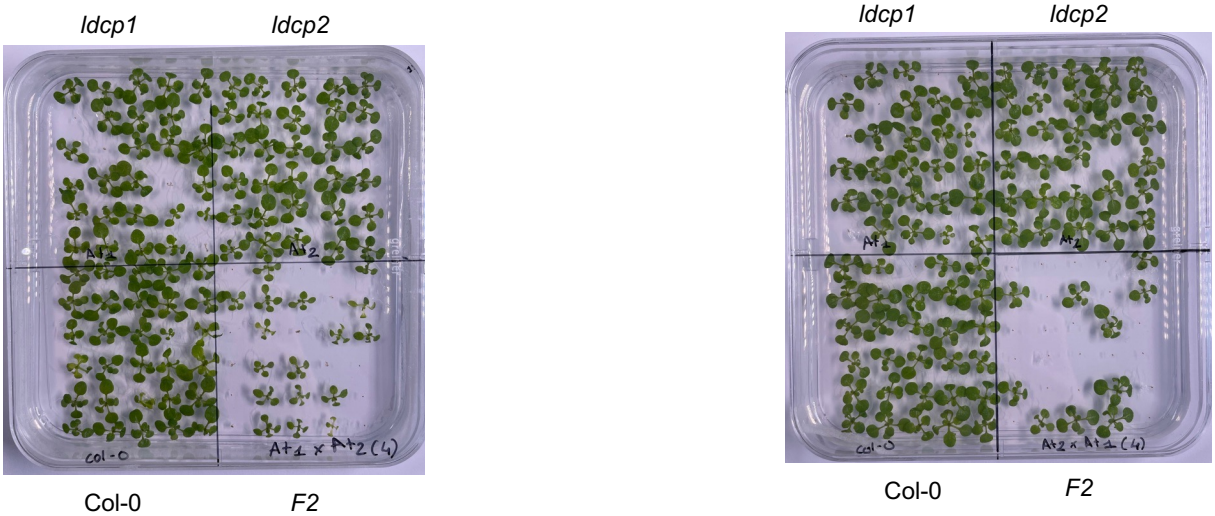
Reagent	Cas9 +	Cas9 -
H2O	28 ul	29 ul
NEB 3.1 buffer (10x)	4 ul	4 ul
sgRNA (20 nM final; around 0.8 ng/ul final)	1.6 ul	1.6 ul
Cas9-NLS (20 nM final; 1000 nM initial; NEB)	0.8 ul	-
SubTotal	***34 ul***	***34 ul***
Incubate 22 C, 10 minutes to form RNP complex;		
DNA cleavage template (2 nM; approx. 120 ng of 2.5 kbp product; 20 ng/ul initial)	6 ul	6 ul
Total	***40 ul***	***40 ul***

Supplemental table 8 - List of primers used for the amplification of DNA template for *in vitro* Cas cleavage

	LDCP1 forward	LDCP1 reverse	LDCP2 forward	LDCP2 reverse
	Sequence (5'-3')		Sequence (5'-3')	
CLR3	GGGTTCATGAGTGAGTATCACTC	AGCTCAACCGTTCTGCTTGTTC	GAACCAGAGAATGCATCGAAG	ATTGTCCGCGGTAATCCG
CLR5	GGGTTCATGAGTGAGTATCACTC	AGCTCAACCGTTCTGCTTGTTC	GAACCAGAGAATGCATCGAAG	ATTGTCCGCGGTAATCCG
CLR7	GGGTTCATGAGTGAGTATCACTC	AGCTCAACCGTTCTGCTTGTTC	CGCCGATGATCACGATAATAAG	GAGAACACGTGCTCCGC



Supplemental figure 1 – Gene alignment of *LDCP1* and *LDCP2*. The alignment shows a 82% similarity. This alignment was performed using ClustalW provided by UNIPRO Ugene.



Supplemental figure 2 – Seed germination assay. Germination was performed using AGM media. Homozygous single mutant *ldcp1*, *ldcp2* homozygous single mutant and the *f2* segregation progeny were sown and the germination rate was calculated using Col-0 as a control.