Sustainable Biotechnology M.Sc.
Master thesis

Selection of *Arabidopsis thaliana* plant lines for the study of Single locus proteomics

Author
Nicholas Bolsi

In collaboration with University of Copenhagen
Copenhagen Plant Science Center, Bülowsvej 21a

Supervisors
Mette Lübeck – Aalborg University
Kateřina Adamusová– University of Copenhagen
Sebastian Marquardt – University of Copenhagen

Copenhagen, Denmark, 1st June 2023

Contents

INTRODUCTION

Arabidopsis thaliana

Arabidopsis thaliana, commonly known as thale cress, is a small flowering plant in the family of *Brassicaceae,* native to Europe, Asia, and Africa. It is widely used as a model organism in plant research due to its characteristics. With its compact size, Arabidopsis thaliana is easily cultivated and handled in laboratory settings. It exhibits a high-speed growth cycle and the ability to self-pollinate [1], allowing for efficient experiments and observation of plant development. The plant possess a relatively small genome size (approximately 120 Mb) [2]. This simplifies genetic analysis. Furthermore, its capacity for abundant seed production enables large-scale studies. Furthermore, Arabidopsis thaliana is genetically manipulable, facilitating the introduction of specific genetic modifications. These combined features make Arabidopsis thaliana a tool for studying various aspects of plant biology, including genomics, plant development and responses to environmental factors.[2]

Figure 1.1: plants of Arabidopsis thaliana in soil after eight weeks of growing.

Transcription

Transcription is the process by which a sequence of ribonucleic acid (RNA) is synthesized using the template of deoxyribonucleic acid (DNA) in the genome. RNAs serve various functions, such as forming ribosomes (ribosomal RNA or rRNA) or acting as a template for protein synthesis (messenger RNA or mRNA). The transcription process is catalyzed by a family of enzymes called RNA polymerases, which can read the DNA sequences and match each deoxyribonucleotide with a complementary ribonucleotide. While prokaryotes possess a single RNA polymerase, eukaryotic organisms have evolved multiple specialized polymerases for producing different types of RNAs. RNA polymerase I is responsible for rRNA synthesis, RNA polymerase III is involved in the production of RNA transfer (tRNA) and smaller rRNA, and RNA polymerase II synthesizes mRNA and long non coding RNA (lncRNA). Additionally, plants possess two modified forms of RNA polymerase II, known as IV and V, which play a role in gene silencing.[3] My thesis focuses on the transcription by RNA polymerase II. An important characteristic of the transcription is that it is composed of four distinct phases, each characterized by a different chromatin profile, as it is possible to see in the next paragraph.

Histones PTM

The genome, which is located within the nucleus of a cell, spans a considerable length if fully unwrapped, so it needs a structure to make it compact enough to fit in a nucleus. In order to do so, eukaryotic organisms have evolved a structure known as chromatin. The main building blocks of chromatin are nucleosomes. Each of these is composed of two copies of histones H2A, H2B, H3, and H4, around which the DNA is wrapped, covering a length of approximately 140 base pairs per nucleosome.

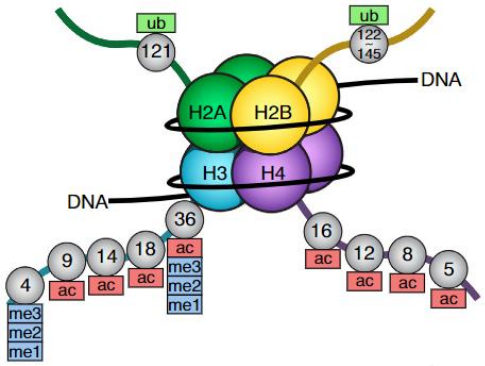

Figure 1.2: representation of a nucleosome and its modifications[4]

Histones play a crucial role in chromatin organization and gene expression. This happens through various "post-translational modifications" (histone PTMs), including methylation (me) and acetylation (ac)[5]. These modifications can influence the overall structure of chromatin. The structure of the chromatin can then impact the accessibility of DNA, affecting the activation or repression of genes. Thus, histone PTMs serve as a mechanism for modulating chromatin dynamics and controlling gene activity.[5]

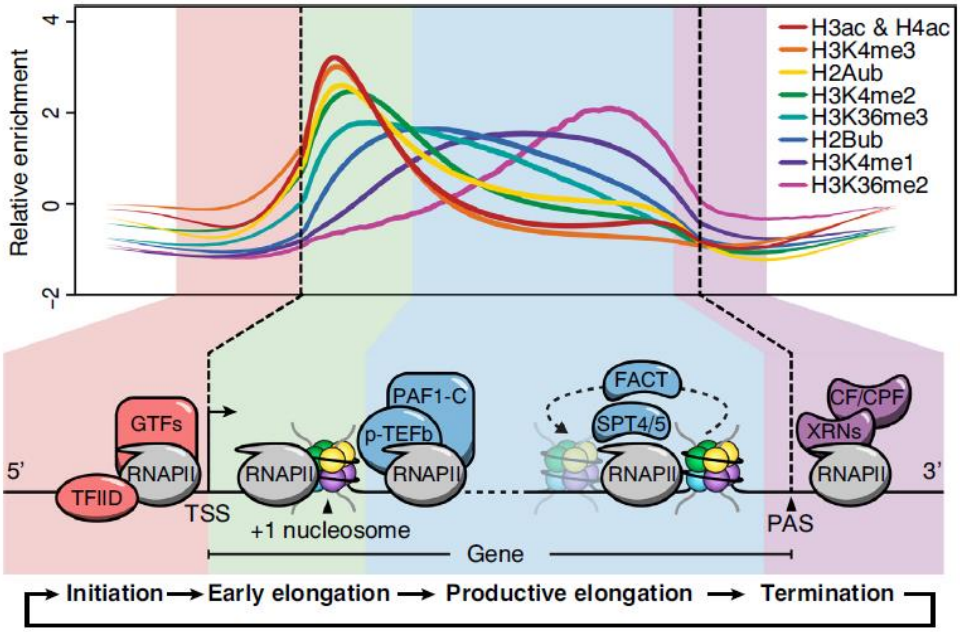

Figure 1.3: distribution of histon PTMs along a gene[4]

Tandem transcriptional interference

Transcriptional interference (TI) refers to the inhibitory effect on RNAPII transcription caused by the transcription of one transcription unit overlaying a second transcription unit. This interference occurs when a sequence of long non-coding DNA is transcribed, and its transcription invades the mRNA production of a downstream protein-coding gene. An example for tandem Transcriptional Interference in Arabidopsis thaliana is a T-DNA inserted in the promoter region of the Quasimodo (QUA1) gene, which encodes a glycosyltransferase essential for pectin production and cell adhesion in plants. The T-DNA insertion (*qua1-1*) causes a recessive loss of function mutation of QUA1 in *qua1-1* leads to a dwarfed phenotype compared to the wild type, and the cells exhibit poor adhesion to each other, due to the lack of pectin in the middle lamella: the mucilages rich substance that glues together the plant cells[6]. As an effect, the plant cells can be penetrated more easily by the ruthenium red dye, giving a positivity to its test.[7]

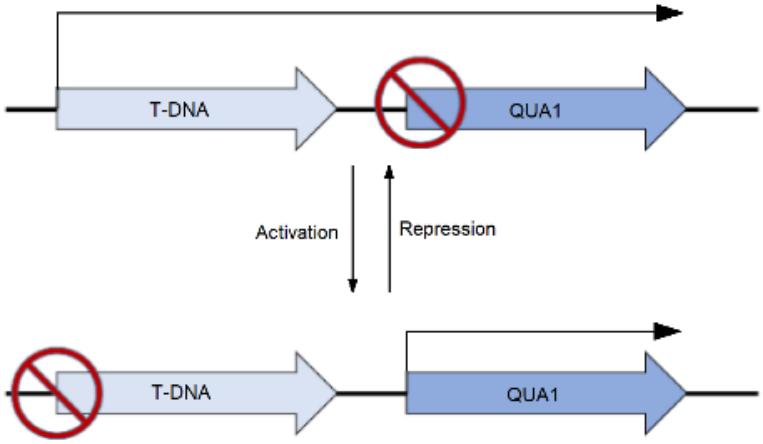

Figure 1.4: the mechanism of tandem Transcriptional Interference

The mechanism of tTI is activated when the T-DNA is transcribed, with transcriptional elongation continuing beyond the insertion site and extending to the end of the QUA1 gene. This results in the production of a nonfunctional transcript, inhibiting the proper initiation ofthe QUA1 mRNA . The region between the T-DNA insertion site and the gene body, referred to as the tTI region, undergoes a change in its transcriptional context. In wild type, this region acts as a promoter driving QUA1 initiation. However, in the *qua1-1* background, the tTI region

becomes moves in the middle of a longer transcript region elongating of transcription takes place and represses initiatio by tTI.

To summarize, the region between the T-DNA insertion site and the gene body experiences a shift in its transcriptional context. In the wild type line, it functions as a promoter driving QUA1 initiation, while in the *qua1-1* background, it becomes a mid-transcript region undergoing elongating transcription influenced by tTI.

Figure 1.5: A) schematic representation of the qua1-1 locus.[8] B) Ruthenium red staining of wild type (WS) and *qua1-1* hypocotyls.[8]

Single locus proteomics

To study how the chromatin influences transcription, it is needed to discover which proteins are present in certain loci of the DNA. PTMs in the genome it is widely used the Chromatin Immunoprecipitation.[9]With this technique, through antibodies, it is possible to purify DNA associated with the studied transcription factor, and study its presence in the genome by sequencing the purified DNA. However, since ChIP uses factor-specific antibodies, and it is necessary to have prior knowledge of the studied factor, it is not a suitable method for the identification of unknown factors. Such factors could be revealed by targeted proteomics of purified DNA loci. To examine the proteomics of specific loci, they need to be localized and isolated

To gain a comprehensive understanding of transcription processes like initiation, termination, and tTI, the proteins located around the specific loci can be isolated and purified, and their composition can be analysed through mass spectrometry. This approach allows for a targeted and detailed exploration of localized genomic events and their associated chromatin

components. Using this procedure comes with some drawbacks. One of these is the possibility of not being able to extract enough proteins from a sample, leading to a low sample output. This poses a problem because in locus-specific isolation methods, since all the identified proteins are compared to a background composed of unspecifically co-precipitated proteins.[10] To distinguish the proteins of interest bound to the single loci from the cellular background, it is necessary to purify effectively the immunoprecipitated proteins, so it is possible to to overcome the signal-to-noise ratio.

The identification of genes regulated by the tandem Transcription Interference (tTI) mechanism presents certain challenges due to the nature of transcription events involved. Quantitative PCR (qPCR) is an example for a traditional method used to identify regulated genes. It cannot be used to study genes undergoing tTI since it only detects if a gene is transcribed, without distinguishing between interfering transcription and functional mRNA production. RNA-Seq is another potential method for identifying genes which undergo tTI. However, the main problem is that it requires stable RNA transcripts, and the interfering transcript is not guaranteed to have the necessary stability. More advanced methods like transcript isoform sequencing and nanopore direct RNA sequencing[11,12] can provide insights into transcript boundaries, but they still rely on comparing induced tTI conditions with unimpeded gene expression. Focusing on the identification of proteins specifically associated with tTI allows to overcome these limitations, and the screening of these proteins along the whole genome with the aid of antibodies enables the identification of loci affected by tTI.

The successful development of a purification method targeting specific chromatin modifications associated with Single Nucleotide Polymorphisms (SNPs) can indeed provide valuable insights into understanding human diseases caused by SNPs. SNPs are variations in a single nucleotide within the DNA sequence, and they can have significant effects on chromatin structure and function. By specifically targeting and purifying chromatin regions affected by SNPs, researchers can gain a better understanding of the underlying causes of these diseases.

Methods for single locus targeting

Zinc fingers

Zinc finger proteins (ZFPs) are a large class of DNA-binding proteins that utilize coordinated zinc ions to stabilize the typical ββα fold in their modular DNA-recognition domains. Each

zinc finger domain, comprising approximately 30 amino acids, possesses the ability to recognize a 3-4 base pair sequence in a sequence-specific manner.[13] Normally a zinc finger proteins contains three DNA binding domains, allowing to bind a more specific portion of DNA. It has been demonstrated that it is possible to produce a zinc finger protein capable of targeting a specific locus.[14]

For a long time, the design of zinc fingers domains was a rare way to engineer a tool for binding the DNA in specific loci with a wide range of applications, such as the transcriptional blockage of oncogenic genes[14] or the identification of histone modifications inducing gene repression. However, it has not been possible to create a 'recognition code' to design the DNA binding domain of zinc fingers free of constraints to any DNA sequence, due to the interaction of adjacent zinc finger proteins.[15] Furthermore, their specificity of the binding site can be influenced by the methylation of the cytosines in the target sequence, lowering their effectiveness.[14,16] More recently, they have been replaced by other DNA binding proteins where sequence-specific design is easier, such as TAL effectors and dCas9

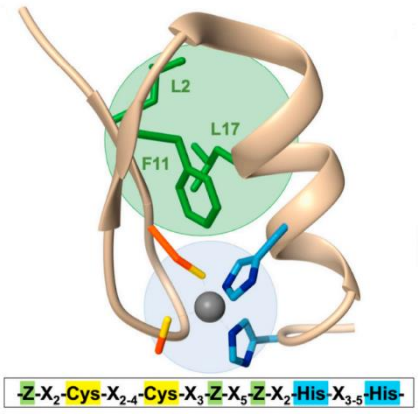

Figure 1.6: zinc finger representation. In green are indicated the amino acids deputed to the formation of the hydrophobic core

Transcription activator like proteins

Transcription activator like effectors (TALE) are another family of proteins capable of binding the DNA sequence. They originated from plant pathogenic bacteria, and when injected into plant cells, they are driven to the nucleus, where they can target specific DNA sequences.[17] Their DNA-binding domain is composed by a repetition of a sequence composed

of 34 aminoacids[18], with two polymorphisms in the positions 12 and 13 of each of these repetitions. These determine to which nucleotide the TAL protein binds. A combination in tandem of these repeated regions allows to create TALEs capable to bind any DNA sequence.[17]Studies about genome editing and genome expression modulation have been successfully carried out with the use of TALEs[19], as well as experiments involving single locus targeting.[20,21]

The use of TALEs has its weaknesses. It has been showed that it could band off target sequences, due to the protein context where the TAL is used. This kind of off-target is difficult to predict. Furthermore, as the zinc fingers, it can be affected by methyl groups attached to the cytosines in the target sequence, making their usage more problematic. It is also important to notice that the design of TALEs protein can be time consuming, since the amino acid repetitions must be designed and the binding efficiency must be tested.

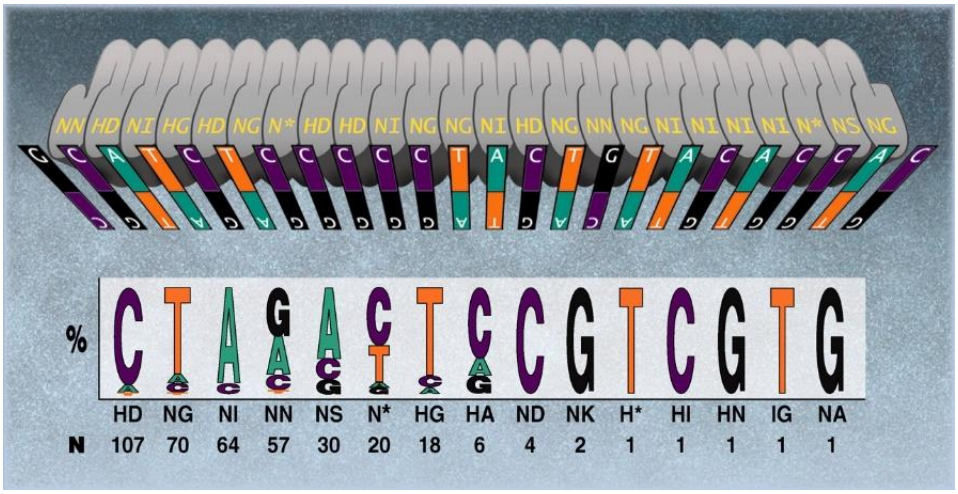

Figure 1.7: DNA-binding domain of the TAL proteins

dCas9

The CRISPR-Cas (Clustered Regularly Interspaced short palindromic repeats - CRISPR associated proteins) system is an adaptive immune system used by bacteria and archaea for defence against foreign DNA, such as bacteriophages or plasmids. [22] It involves the use of Cas proteins, with Cas9 being a well-known example. Cas9 is guided to the target DNA sequence by CRISPR RNA (crRNA) or trans-activating crRNA (tracrRNA [23]), which are stored in the organism's CRISPR loci after encountering foreign DNA.[24] Cas9 utilizes two nuclease domains cleave the foreign double strand DNA at multiple sites. In this way, the bacteria is able to

impede the proliferating of the foreign DNA inside its cell. In laboratory settings, when a cleavage of the DNA is not needed, a modified version called dCas9 (dead Cas9) is often used. This is paired with a single guide RNA: a hybrid of the two RNAs aforementioned [25]. dCas9 lacks the cleaving activity due to a deletion in its nuclease domains, and it can be guided to the target DNA sequence without producing any breaks in it.[23] This makes dCas9 a valuable tool for precise DNA-binding applications, including transcriptional blockage, gene expression modulation, epigenetic editing, and chromosomal region labelling for live cell imaging. [25"]

Even though it is a widely used technique, the use of dCas9 for locus targeting is not exempt from hurdles. In fact, off-target binding of the dCas9 protein has been observed when applied on a genome-wide scale.[26] The prevention of these effect has been achieved by reducing charge interaction between Cas9 and the non-targeted DNA[27], as well as by rationally redesigning the Cas9 to bind more specifically to PAMs.[28]

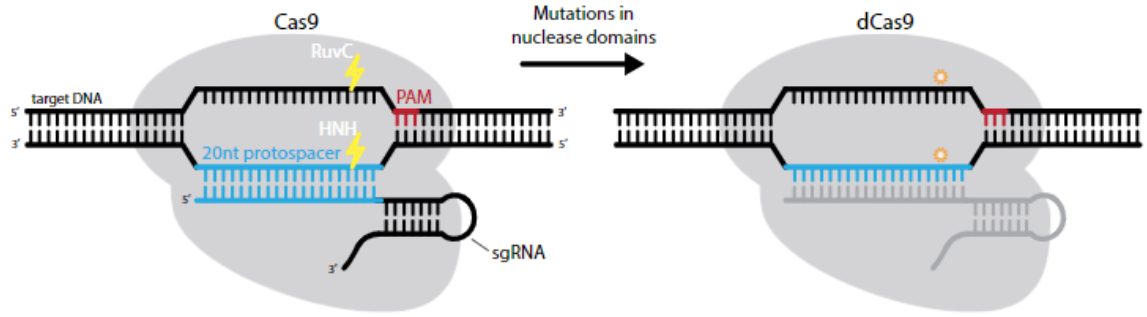

Figure 1.8: the action of the dCas9[25]

Workflow: from the wild type plants to the proteome analysis

To study the proteomics of single loci in plants, a series of steps are followed, starting from the wild type plant to the final analysis. The process has the following steps:

1. Protein Design: The first step involves designing the protein, such as zinc fingers or TAL proteins, which will specifically bind to the desired genomic loci. In the case of dCas9, it is necessary to design a guide RNA instead.

2. Gene Expression and Plasmid Preparation: The gene encoding the designed protein is inserted into a plasmid. This plasmid serves as a carrier for introducing the gene into

the plant cells. It needs to be amplified, a process typically carried out in *Escherichia coli.* The resulting organism produces biologically functional replicons.[29]

3. Transformation into *Agrobacterium tumefaciens*: After plasmid amplification, it is transferred to a strain of *Agrobacterium tumefaciens*, a bacterium commonly used for plant transformation. This step allows the introduction of the plasmid carrying the target gene into plant cells.[30]

4. Plant Transformation: The transformed Agrobacterium strain is used to infect the plant of interest. Through this process, the foreign gene (DNA binding protein) is integrated into the plant's genome, and the plants become capable of expressing the designed protein.

5. Selection of Homozygous Plants: To ensure that the plants express the DNA binding protein, a selection process is performed. An antibiotic resistance gene present in the plasmid is used for this purpose. First, the plants that can survive in a medium containing the antibiotic are selected. After propagating the selected lines, the plant lines that are homozygous are then chosen. During this selection process, it's important to note that heterozygous plant lines exhibit a 75% survival rate in the presence of the antibiotic, whereas all the homozygous plant lines consistently germinate correctly.

6. Genotyping and Western Blot: After propagation, the selected plant lines are tested to determine if they are homozygous for the DNA binding protein gene. Genotyping tests are carried out to confirm the presence of the target gene. Subsequently, a western blot analysis is performed to verify the expression of the protein. A tag can be used to facilitate the identification of the protein during the western blot.

7. Immunoprecipitation: Once it is confirmed that the plants are homozygous and expressing the DNA binding protein, immunoprecipitation is performed. This technique is used to isolate the proteome around the specific genomic loci targeted by the DNA binding protein.

8. Mass Spectrometry Analysis: The isolated proteins obtained from immunoprecipitation are then subjected to mass spectrometry analysis. This method

allows for the identification and characterization of proteins[31] present in the proteome surrounding the targeted loci.

By following these steps, it is possible to analyze the proteome of a single genomic locus in plants and gain insights into the protein interactions and functions associated with that specific region.

This thesis focuses on steps 5 and 6 of this workflow, with the objective to select the most adapt plants for a future analysis of their loci.

RESULTS

Project 1: proteomics of tandem Transcription Interference

Objective

The primary objective of the first project in this thesis is to select suitable plants to analyze the proteomics of the promoter region of the Quasimodo (QUA1) gene under two conditions: 1.) In the mutant where the tandem Transcriptional Interference (tTI) mechanism is active, and 2.) in wild type conditions.

To study the single locus proteomics of promoters affected by tTI, a mutant is employed. This is *qua1-1*, which has a T-DNA insertion upstream of a promoter. This insertion is located 117 base pairs before the TSS of the QUA1 gene and inhibits the expression of the QUA1 gene by a read through transcription. In wild type plants, the transcription of the QUA1 gene occurs during its initiation phase, while in the *qua1-1* plant lines, the promoter DNA is marked with chromatin signals linked to elongation. Consequently, it is anticipated that there will be differences in the proteomics of the same sequence in these two types of plant lines. To target the promoter region, a Transcription Activator-Like (TAL) protein, designed to specifically bind the sequence AAAGTGCTCTCTCCCCAA, known as "TAL22," is employed. The gene encoding the TAL22 protein is inserted into the genome of *Arabidopsis thaliana* using the "floral dipping" transformation through *Agrobacterium tumefaciens*.

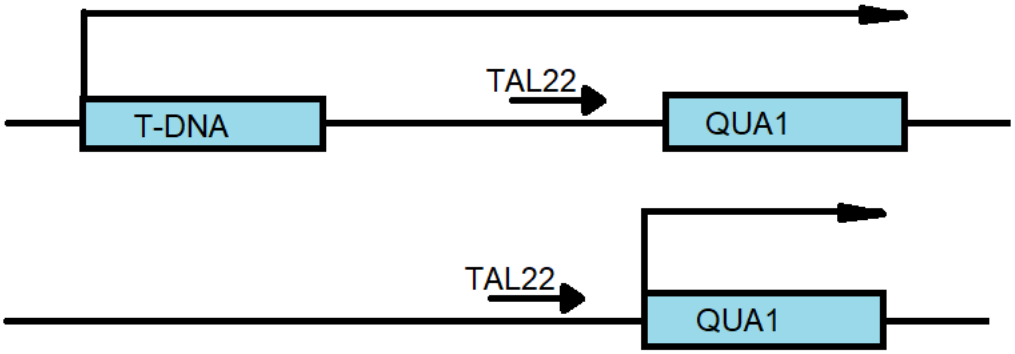

Figure 2.1: positioning of the TAL22 protein on the QUA1 promoter

Homozygous selection

For the purpose of this thesis, two *qua1-1* lines containing TAL22 and two WS wild type plant lines containing TAL22 were selected. Using the Marquardt lab plant database, the plant lines

which were already tested to be homozygous for the TAL22 gene were selected. This was tested by plating plants in a ½ MS medium with 30µg/ml hygromycin. (see Materials and methods). In the table 1 it is possible to see the name of the selected plant lines and their number in the CPSC database (see Materials and methods)

SMA number	Plant line	TAL22
6839	WS	Not transformed
5327	WS	A1
5328	WS	H4
5008	*qua1-1*	Not transformed
5340	*qua1-1*	C2
5342	*qua1-1*	E4

Table 2.1

DNA analysis: genotyping test

Two genotyping tests were performed to assess the presence of specific genes in the genome. The first test aimed to detect the presence of the TAL22 transgene, while the second test was conducted to identify the *qua1-1* T-DNA insertion.

For the TAL22 genotyping, primers were designed based on the plasmid used for plant transformation and were ordered from IDT (Integrated DNA Technologies). This decision was made after determining that the existing primers in the laboratory were not suitable for this purpose. Untransformed plant lines are used as negative control. The designed primers target a sequence in the construct carrying the TAL22 gene, which was inserted in a construct and transformed into Arabidopsis thaliana.

The *qua1-1* genotyping utilized the MLO 20, MLO 21, and MLO 22 primers. This primer set produces two types of bands: one around 450 base pairs, indicating the presence of the *qua1-1* insertion, and another approximately 800 base pairs in size, indicating its absence. This happens because both genotyping tests use the MLO20 as a direct primer, but the MLO21 binds to the T-DNA insertion, absent in the wild type plant lines. The MLO22 reverse primer binds after the T-DNA insertion, allowing to detect the wild type plants. These band patterns serve as indicators for identifying plants with or without the *qua1-1* T- insertion. A WS and a *qua1-1* untransformed and pre-tested plant lines are used as positive control of the genotyping.

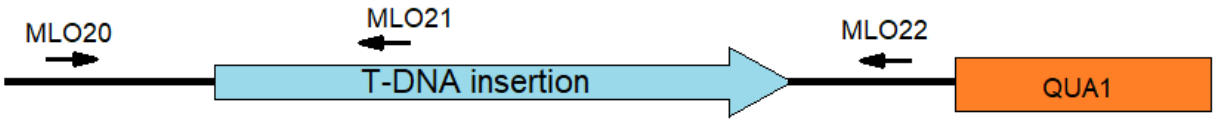

Figure 2.2: diagram of the primers used in the qua1-1 genotyping

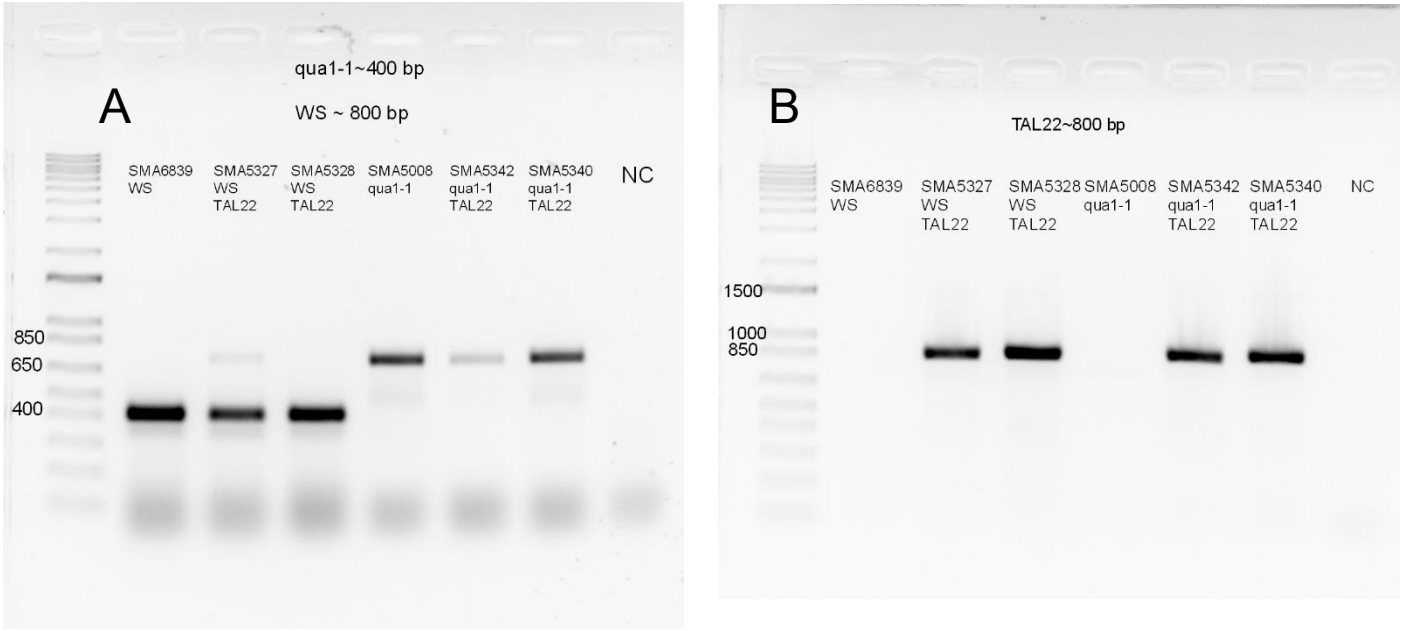

Figure 2.3: A) genotyping test for the qua1-1 T-DNA insertion. B) genotyping test for the TAL22 gene

The figure 2.3 shows the gels used for the genotyping tests for TAL22 and *qua1-1* genes after electrophoresis. In the *qua1-1* genotyping gel, it reveals that the SMA5327 (*qua1-1* TAL22 C2) plant line shows two bands (450 bp and 800 bp) indicating that it is heterozygous for the *qua1-1* T-DNA insertion. Therefore, this plant line cannot be directly used for the study's purpose and needs to be further selected. However, the SMA5328 (*qua1-1* TAL22 E4) plant line is homozygous for the *qua1-1* insertion, as is the *qua1-1* positive control. In contrast, all WS plant lines do not have a band specific for the *qua1-1* insertion, but they have a band in the wild type position. Furthermore, no bands are visible in the negative control.

Regarding the TAL22 genotyping gel, a distinct band is observed in all the transformed plant lines, indicating the presence of the TAL22 gene. On the other hand, no bands are visible in the untransformed plant lines and the negative control.

Protein analysis: western blot

To determine the expression of the FLAG tag attached to the TAL22 gene, the Western blot technique is utilized. This technique enables the detection and visualization of the 9x-FLAG tag through the use of an anti-FLAG antibody. The western blot also allows to determine the level of expression of the detected protein, signaled by the intensity of the band. In fact, the protein can have various levels of expression. This variability is caused by the causality of the construct containing the dCas9 gene position in the genome. When transformed, the inserted sequence can be positioned anywhere in the genome, and it could end up to be after a strong promoter, or in a part of the genome where the chromatin allows the attachment of the RNAPII more easily. In this case the protein will be well expressed, and the resulting signal will be higher than a protein expressed by a gene after a weak promoter. This is important to analyze in order to find plant lines which have a similar level of FLAG expression, which can help to minimize the error in the Mass Spectrometry analysis. The Western blot analysis provides crucial information about the expression of the FLAG tag, which is necessary for subsequent immunoprecipitation steps to analyze the proteome of the target sequence.

The extracted proteins are loaded in a pre-casted SDS-PAGE gel and undergo electrophoresis to divide them by their molecular weight. Subsequently, the proteins have been transferred onto the membrane, this is incubated with the anti-FLAG antibody. Following the primary antibody incubation, the rabbit anti-mouse polyclonal immunoglobulins are used. The secondary antibody recognizes and binds to the primary antibody, allowing for the detection and visualization of the FLAG tag on the membrane. To control that the loading of the proteins has been carried out correctly, the picture of the stain free gel is showed next to each membrane.

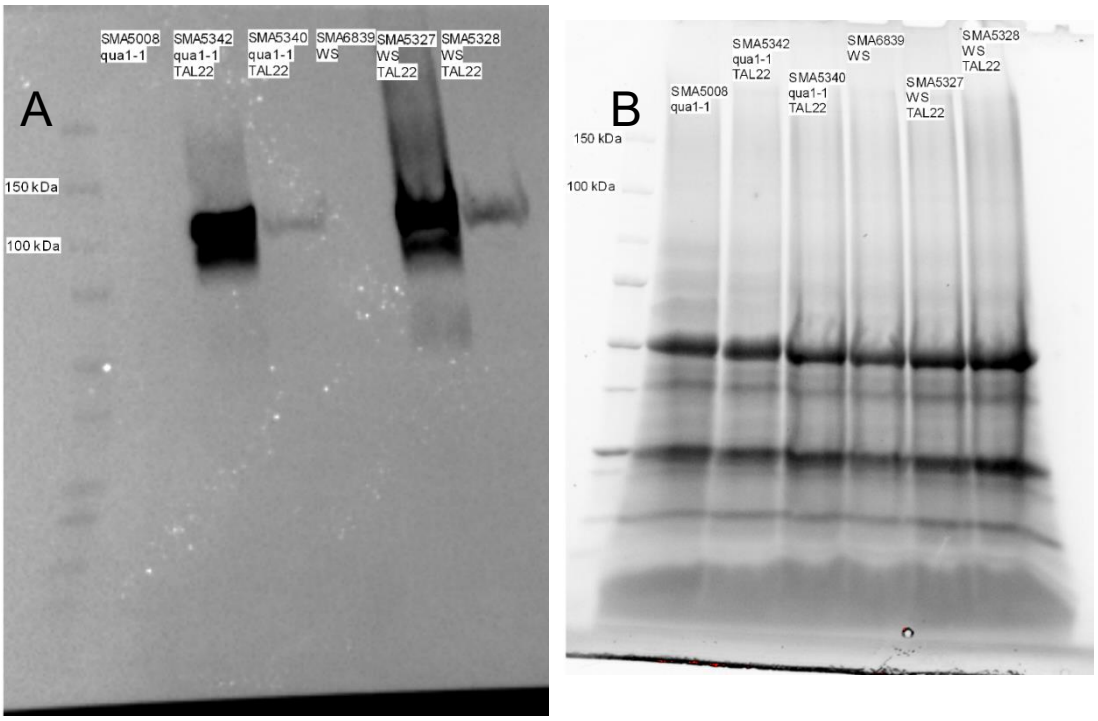

Figure 2.4 A) Western blot to detect the dCas9-FLAG complex. B) Stain free gel for the loading control

The figure 2.4 shows the blotted membrane after both incubations. It is noticeable a band around the 110 kDa in all the transformed plant lines. This indicates that the primary antibody has correctly bound the 9xFLAG-TAL22 complex. We can also notice that the SMA5340 and SMA5328 plant lines have a weak band, and this indicates that the FLAG is expressed in a smaller, but similar quantity. This makes these plant lines an ideal couple of candidates for the analysis of their proteome.

Phenotypic analysis: hypocotyl length evaluation

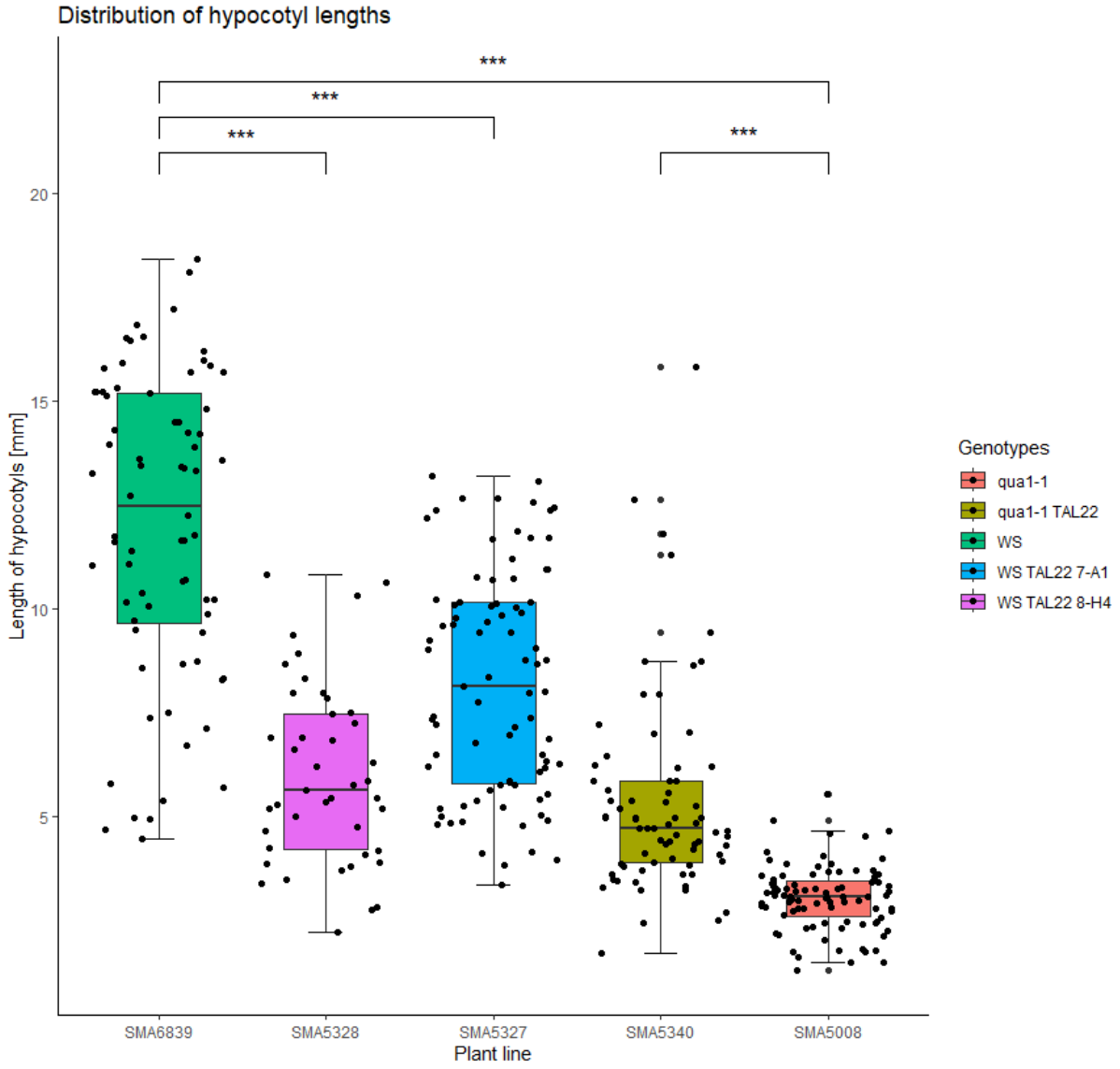

Figure 2.5 length of the hypocotyls put in boxplots.

The boxplots show the length of the hypocotyls of the different plant lines. On the top are represented the Student's t-test confronting the length of the hypocotyls of two distinct plant lines. First, we can see that the WS (SMA6839) and the *qua1-1* (SMA5008) lines have a significant difference in length, confirmed by the statistical test. Next, the difference between the two *qua1-1* plant lines is observable: the *qua1-1* TAL22 plant line appears to be longer. This partial restoration of the wild type phenotype could be an effect of the TAL22 positioned in the promoter region interrupting the tTI and allowing the QUA1 transcription to take place. Lastly, it is also possible to see differences in length between the WS and the WS TAL22 plant lines. The transformed plant lines appear to be shorter than the untransformed ones. This could

be explained with the TAL22 positioning in the promoter region of the QUA1 gene, competing with the RNAPII for the binding of the promoter.

Phenotypic analysis: Ruthenium red staining test

Another method to check if the wild type phenotype has been restored in the *qua1-1* plant lines is to test how they react to the ruthenium red staining test. This is a test that involves the submersion of hypocotyls in ruthenium red: a cationic dye that binds to polyanionic molecules, with pectins as the main example in plants.[32] When the QUA1 gene is inactivated, the lack of pectins in the lamella promotes an insufficient cell adhesion, allowing the ruthenium red to penetrate more easily in the plant cells, giving the hypocotyls a red color. [7]

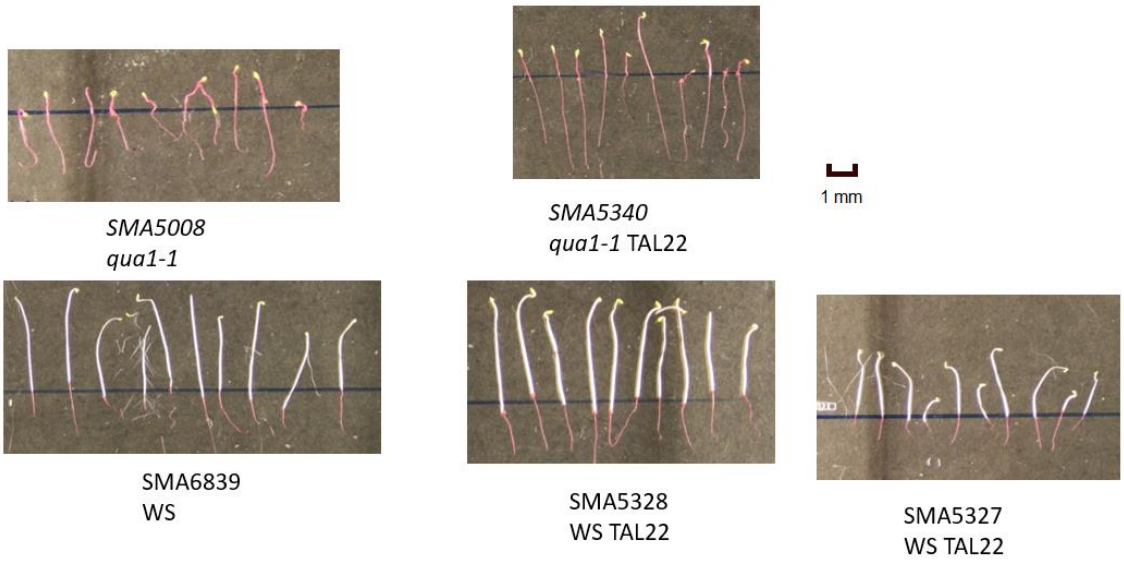

Figure 2.6 Hypocotyls after staining with ruthenium red

In the case of the *qua1-1* TAL22 plant line, the positive staining indicates that the tTI mechanism prevents the proper transcription of the QUA1 gene. On the other hand, in the WS TAL22 plant lines, it can be observed that the presence of TAL22 in the promoter region does not impede the transcription of the QUA1 gene.

Furthermore, the test also provides an initial indication of the difference in hypocotyl length. The transformed *qua1-1* plant lines exhibit longer hypocotyls compared to their untransformed counterparts. This elongation could be a consequence of partial transcription of the QUA1 gene. Conversely, in the WS plant lines, the SMA5327 plant line appears shorter than the wild type plant line. This, along with the white color of the hypocotyls, suggests that the presence

of TAL22 partially inhibits the transcription of the QUA1 gene, but it remains sufficiently active to promote cellular adhesion, resulting in a negative test outcome.

Project 2: Proteomics of initiation and elongation transcription sites

Objective

The second aim of my project is to investigate the proteomic differences between a DNA sequence located in the transcription initiation region of the Quasimodo gene and a DNA sequence found in the gene body of the same gene, where transcription is in its elongation phase.

To target these specific sequences, the protein dCas9 is employed, along with two different guide RNAs. The guide RNA QT3 is used to target the promoter region, while the guide RNA QG3 is utilized to target the gene body. The distance between the guide RNAs is of 1517 base pairs (figure 2.7). Additionally, plants that possess the gene for dCas9 but do not express any guide RNA were also chosen. These plants serve as a negative control in the experiment. The purpose of including this negative control is to have a plant line that is genetically identical to the others but lacks any specific sequence to bind to. This helps to distinguish the effects of the guide RNAs from any non-specific effects related to the presence of dCas9 alone. The table 2 shows the plant lines that were selected to undergo the experiment.

SMA number	Target sequence	Name
6839	N/A	WS
6407	QT3	1A
4586	QT3	1D
4935	QT3	F3
4646	QG3	C5-11A
4648	QG3	C5-11C
4672	QG3	2B-4C
4673	QG3	2B-4D
4675	QG3	2B-4F
4945	No gRNA	#2
4947	NogRNA	#3

Table 2.2

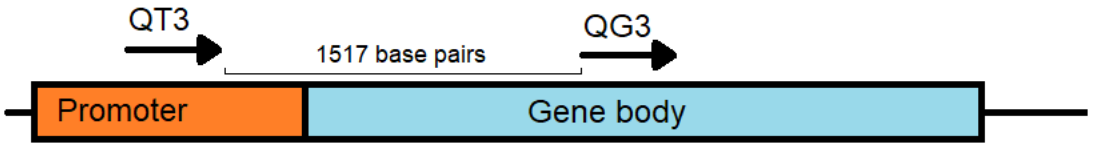

Figure 2.7: diagram of the guide RNAs in the QUA1 gene

DNA analysis: genotyping test

To determine the presence of the gene encoding dCas9 in the plant lines, genotyping tests were performed. The primers MLO1806 and MLO1809 were utilized for this purpose.

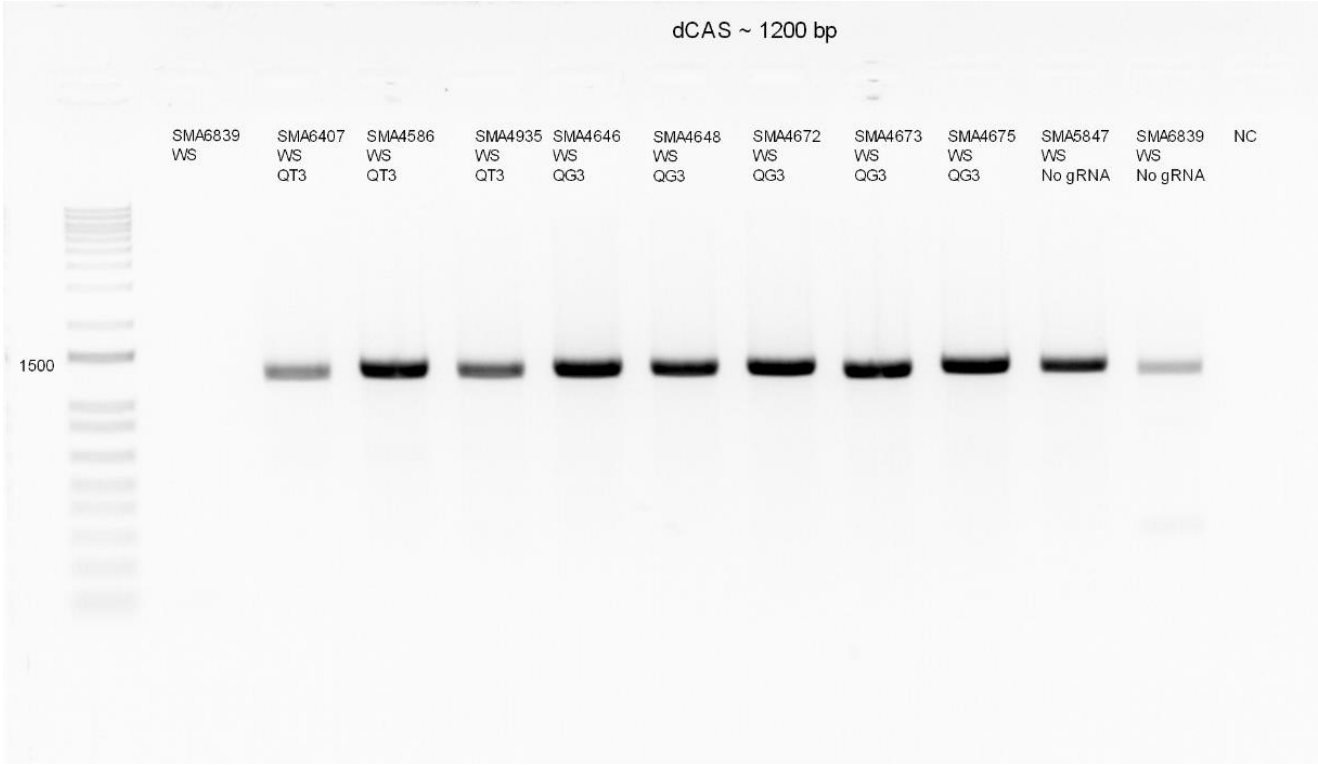

Figure 2.8 Genotyping test for the dCas9

The presence of a clear band at approximately 1200 base pairs in all of the transformed plant lines, while the negative controls (water and untransformed plant line) do not show any bands, indicates that all of the selected plant lines have successfully undergone the desired transformation. This positive result suggests that these plant lines can proceed to the next step of the experimental process.

Protein analysis: western blot

Once the presence of the dCas9 gene has been confirmed, it becomes essential to evaluate whether the FLAG tag attached to it is being expressed. This assessment is crucial because the FLAG tag serves as the target for immunoprecipitation, which is a necessary step in analysing the proteomics of the target sequences. To detect the expression of the FLAG tag, the Western blot technique is employed. This technique enables the identification of the FLAG tag using an anti-FLAG antibody, which specifically binds to the FLAG. The western blot is conducted as described in the first project: following the primary antibody incubation, the secondary antibody (rabbit anti-mouse) recognizes and binds to the primary antibody, allowing the visualization of the FLAG tag on the membrane.

In the figure 2.9 it is possible to notice that there is not a single band for all of the samples. This is an indication that the proteins have been degraded during the extraction. It is possible that the seedlings have not been correctly freezed after their collection, or that the extracted proteins were heated for a longer time. Nevertheless, it is possible to clearly see the bands on the membrane, which is a sign that the primary antibody has correctly bound the FLAG tag.

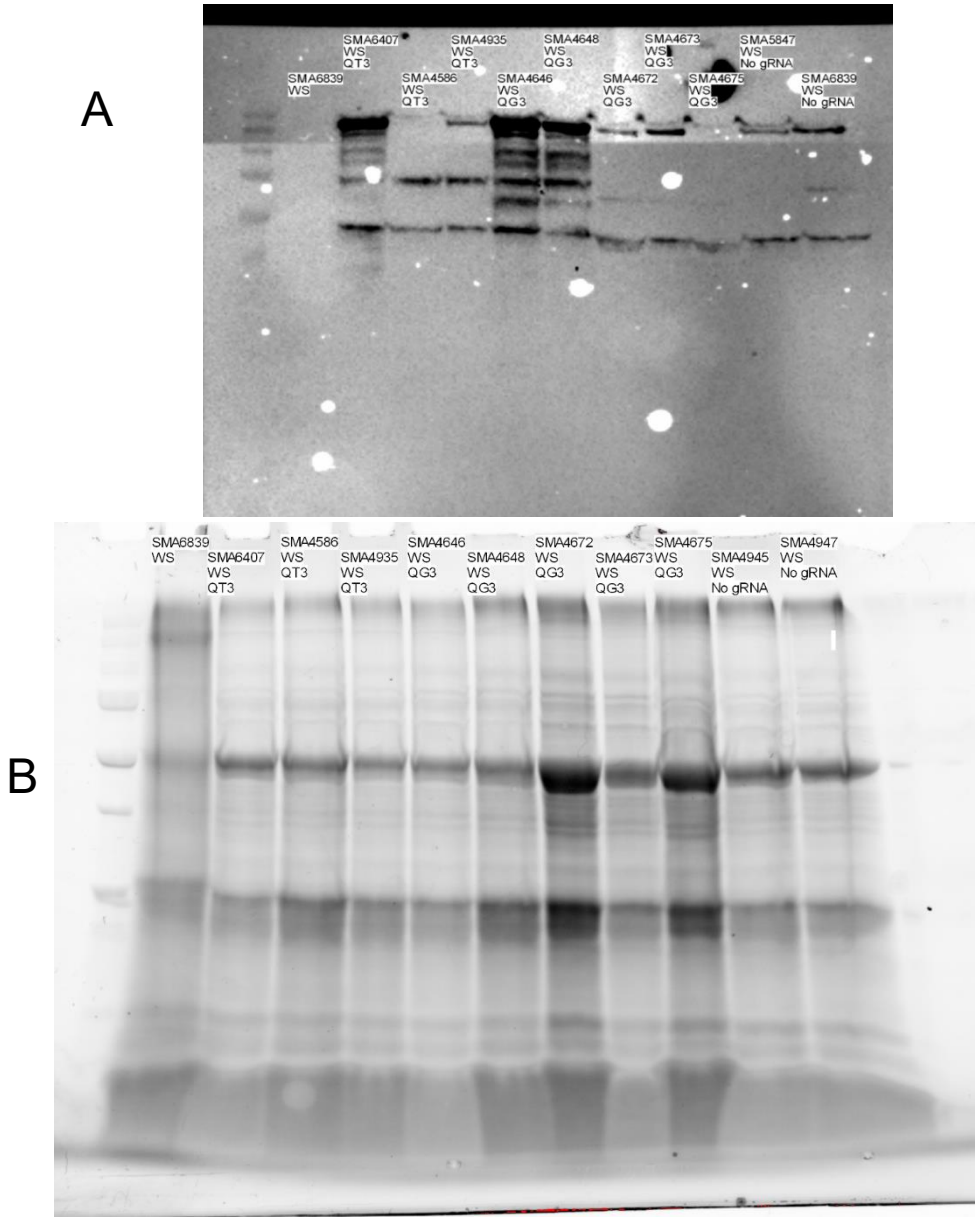

Figure 2.9: A) Western blot to detect the presence of dCas9. B) Gel after protein electrophoresis to assess the loading of proteins

Homozygous selection

The plant lines used in this project were not already been tested to be homozygous for the dCas9. This was tested by making plants grow in ½ MS medium where hygromycin was diluted at 30µL.

SMA number	Name	dCas9
6839	WS	WT
6407	1A	homozygous
4586	1D	homozygous
4935	F3	homozygous
4646	C5-11A	homozygous
4648	C5-11C	homozygous
4672	2B-4C	heterozygous
4673	2B-4D	heterozygous
4675	2B-4F	heterozygous
4945	#2	homozygous
4947	#3	homozygous

Table 2.3: homozygous selection of the plant lines

Three of the QG3 plant lines (SMA4672, 4673 and 4675) resulted to be heterozygous, therefore they cannot be used to study the single locus proteomics.

Selection of a negative control

During the study of the proteomics in the tTI loci, a deletion in the DNA promoter region of the *qua1-1* plant lines has been discovered. This deletion involves also the locus where the QT3 guide RNA should bind. This opens the possibility of using these plant lines as negative control in the proteomics of initiation and elongation transcriptional sites. The use of these plant lines will add another control to the experiment, where we can see what effects can be caused by the addition of guide RNA.

DNA analysis: Genotyping test

The genotyping test has been executed to check if the plant lines actually derived from a *qua1-1* plant line. This was done to ensure that the plant lines have the deletion in the QT3 position that impedes the guiding of dCas9. In this case the primers 20, 21 and 22 were used. To genotype the dCas9 gene, instead, is important to ensure that the negative control has identical characteristics to the analysed plant lines. For the dCas9 genotyping, the MLO1806 and 1809 primers were used.

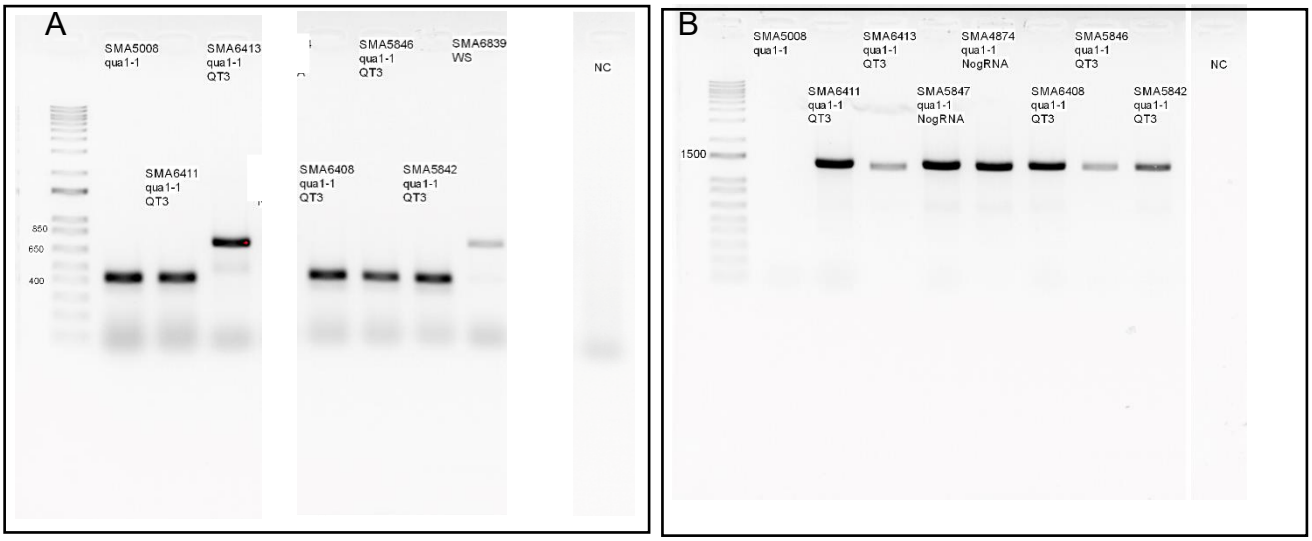

Figure 2.10 A) Genotyping of qua1-1. B) Genotyping of dCas9

The *qua1-1* genotyping shows that the plant lines SMA6411, 6408, 5846 and 5842 have clear bands around 400 bp. This confirms their homozygosity for the *qua1-1* insertion and the fact that the plant lines have been correctly obtained by propagating homozygous *qua1-1* plant lines. The plant line SMA6413, instead, shows a band in the length characteristic of the wild type plants, meaning that the dCas9 would bind to it due to the absence of the deletion. The plant lines SMA5847 and 4874 cannot be considered to the purpose of this selection since they do not produce a guide RNA.

The dCas9 genotyping, instead, shows clear bands around 1200 base pairs in all of the samples. The dCas9 is present in all of the analysed samples

Protein analysis: western blot

As in the other studies, the western blot is executed to ensure the expression of the FLAG-dCas9 complex.

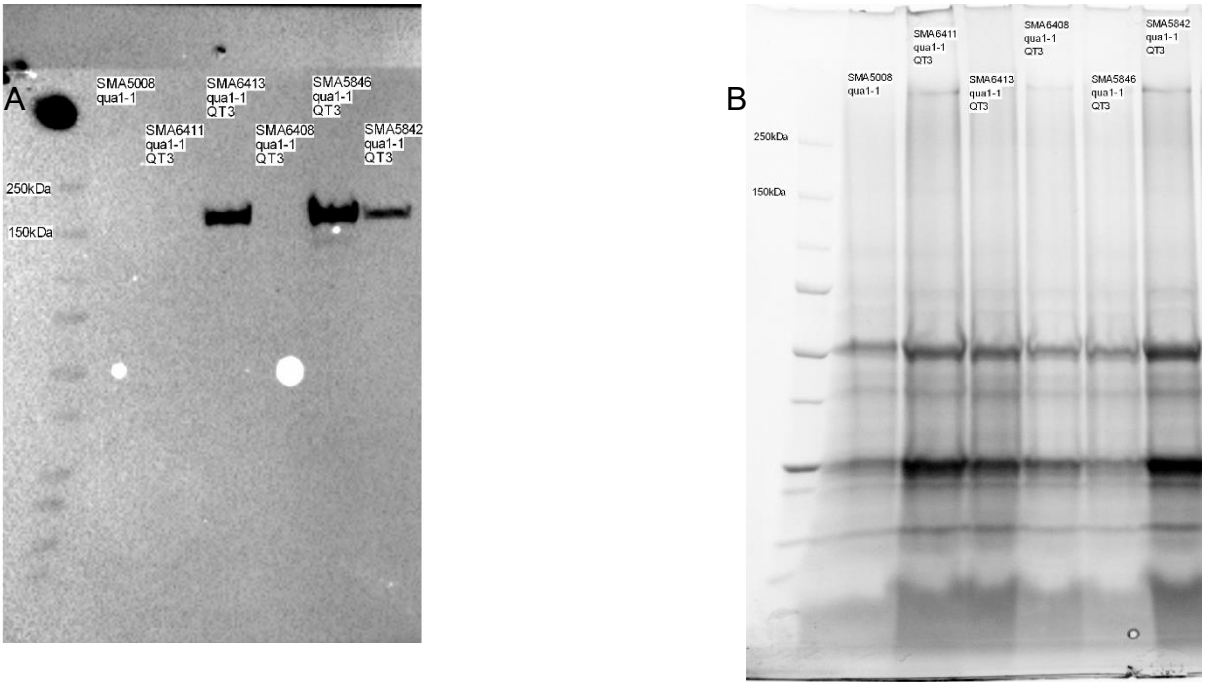

Figure 2.11 A) Western blot to detect the 3xFLAG. B) Loading control

From the figure 2.10 it appears that the plant lines SMA6413, 5846 AND 5842 are the only ones that express the FLAG tag, while the SMA6411 and 6408 plant lines do not have any detectable expression of it.

DISCUSSION

To summarize, two projects have been carried out: in the first, it is studied the proteomics of tandem Transcriptional Interference, by selecting plants that express the TAL22 that wil bind to the promoter of the QUA1 gene. In the second project, instead, the proteomics of two different loci in the same plant is analyzed, so it is possible to notice differences in the protein profiles between the initiation and the elongation transcriptional sites. In this second project, it was possible to have an additional negative control, represented by a *qua1-1* QT3 plant line. This was possible due to a deletion that involves the QT3 position. In these ploant lines, the dCas9 cannot be guided anywhere to the genome

Proteomics of tandem Transcriptional Interference

In the project studying the proteomics of tandem Transcriptional Interference, the study of the genotype has revealed one of the plant lines which was not suitable for the study since it is heterozygous (SMA5327). The observation of the FLAG expression allowed pairing each WS plant lines to a *qua1-1* line with a corresponding expression level. However, since the plant line SMA5327 has been excluded for the heterozygous phenotype, only one of the couples can be used. These are the WS line 5328 and the *qua1-1* line 5340.

During the study of the phenotype it has been found out that the TAL22 insertion between the *qua1-1* T-DNA insertion and the gene body can partially restore the wild type phenotype in the *qua1-1* plant lines. The restoring of the wild type phenotype indicates that the QUA1 gene is transcribed, so the TAL22 can inhibit the mechanism of tandem Transcription Interference. This could be a main drawback in the study of proteomics of tTI loci using this kind of proteins, since the restoring of the phenotype could mean the reversion of the loci's protein profile from an elongation site to the one characteristic of an initiation transcription site.

With the collected data, it was possible to extrapolate a possible couple of plant lines to confront the proteomics of *qua1-1* and wild type plant lines. These are the SMA5340 for the *qua1-1* and SMA5328 for the WS.

Proteomics of initiation and elongation sites

To study of the proteomics of the initiation and elongation transcription sites we utilized dCas9 lines targeting promoter and gene body of QUA1 gene.

A particularity is that it was possible to use the *qua1-1* plant lines to have a negative control due to lack of sequence which could be targeted by QT3. These plant lines have been genotyped for the dCas9 gene and for the *qua1-1* insertion, and their expression of the FLAG tag has been checked. These tests were used to identify *qua1-1* plant lines with identical characteristics to the WS plant lines, with the only difference that the same guide RNA would not bind anywhere in the genome.

Growing plants in homozygous selection allowed doing a first selection of the QG3 plant lines. It found two homozygous plant lines among a pool of 14. Subsequently, these plant lines were genotyped for the dCas9 along with three QT3 plant lines and two lines without guide RNA.

All of them have the dCas9 in their genome. The western blot for the FLAG-dCas9 expression, instead, showed the inability to bind the FLAG in one of the QT3 lines: the SMA4586. Furthermore, it was possible to pair the SMA6407 QT3 plant line with the SMA4646 and SMA4648 QG3 lines.

With the collected data, it was possible to select the plant lines SMA6407 and 4935 to determine the protein profile of the transcriptional initiation region, while the SMA4646 and 4648 can be used to study the elongation sites. As negative control, are present two plant lines without guide RNA: SMA4945 and 4947, along with three plant lines which have a QT3 guide RNA, but do not target the genome due to a deletion of the target sequence: SMA5846 and 5842.

Conclusions and future steps

During this study, it has been possible to execute tests that allowed for an accurate selection of plant lines which can express DNA binding proteins. These can be used directly to immunoprercipitate the studied loci and analyze their protein, or they can undergo further studies before using them.

This plant selection could have an extra step before the actual immunoprecipitation. This is the execution of a CUT&RUN to isolate the DNA around the sequences bound by the dCas9/TAL22 and sequence it. This would give information about how specifically the proteins are binding the target sequence in the genome. The CUT&RUN should be conducted targeting the FLAG tag: the same tag targeted during the immunoprecipitation. If from the sequencing it results that the DNA has been targeted correctly, it is possible to confirm that immunoprecipitating the FLAG-TAL22 or the FLAG-dCas9 would give us information about the proteomics of solely the DNA region of interest. Successfully, it would be possible to immunoprecipitate the loci and analyze their proteomics, producing results for the study of single locus proteomics.

MATERIAL AND METHODS

CPSC plant database

The location where the studies have been conducted is the Copenhagen Plant Science Center, a research center in the University of Copenhagen where studies on plants are conducted. Over the year, the plants are propagated in large amounts and undergo different transformations. In order to keep track of every plant line and transformation, it is used an online database. In this, every time a plant line is generated, propagated or transformed, a number is assigned to it.

By knowing the number of the plant line in use, it is possible to retrieve different information about that specific plant line in the laboratory. These include the generation of the plant line, how it has been labelled for the experiments outside the database, its genotype and from which plant line it has been generated. This last information is particularly important when having a faulty plant line, since it allows to go back to the precedent generation and verify its genotype or propagate it again, so the new generation can be studied for the experiment.

For the *Arabidopsis thaliana*, the SMA (Sebastian Marquardt Arabidopsis) database has been used. *A. thaliana* is not the only organism whose information are stored, but there are databases for storing information about lines of *Solanum lycopersicum, Agrobacterium tumefaciens* and all the PCR primers that have been used in the laboratory.

Plant selection

With the purpose of studying the proteomics of certain loci, the plants have been transformed with *Agrobacterium tumefaciens* to express different DNA binding proteins (TAL and dCas9). The TAL proteins are designed to bind the DNA without any additional insertion, but the dCas9 needs the expression of a guide RNA to target a sequence in the genome.

In the table 1 are indicated the name of the plant lines transformations, and to each of these names are paired with their characteristic genotype and it is indicated which sequence they bind in the genome.

Name	genotype	Binding sequence
WS	wild type	Absent
qua1-1	*qua1-1* T-DNA insertion	Absent
QT3	dCas9; guide RNA on the QT3 position	TTCAATTATAAGCTATTTAA
QG3	dCas9; guide RNA on the QG3 position	TAATCCCAAAGCATGTGCGTGGG
No gRNA	Presence of dCas9 gene; absence of guide RNA	Absent
TAL22	Presence of the TAL protein 22	AAAGTGCTCTCTCCCCAA

Table 4.1: List of the plant transformations used and their binding DNA sequence

Seeds sterilization and growing

Equipment

- 70% ethanol + 0,05% Triton-x
- 96% ethanol
- ½ Murashige and Skoog medium
- Sterile hood
- Sterile filter paper
- P1000 pipette and sterile pipette tips
- Sterile toothpicks
- Petri plates

The sterilization of the *Arabidopsis thaliana* seeds is an important step to avoid the contamination of the analysis and the correct growing of the plants. Being the ½ MS medium rich in nutrients, the bacteria and fungi present on the seeds can grow uncontrollably, altering the growth of the studied plant.

In order to sterilize the seeds, they are put in an Eppendorf tube, together with a solution of 70% ethanol and 0.05% Triton-X. These are sterilised for 10 minutes. During this step it is important to mix by inversion the solution, so the seeds on the bottom of the tube can be reached as well. After removing the first solution, the seeds are washed with 96% ethanol for one minute, and then they are put on sterile filter paper to dry. The seeds are then put in plates. Sterilised toothpicks are used for a precise seed placement.

Homozygous selection

Equipment

- ½ Murashige and Skoog medium
- Hygromycin
- Petri plates
- Growth chamber

The genome of Arabidopsis thaliana is modified by inserting the dCas9 and TAL22 genes, along with a hygromycin resistance gene. If the plants possess homozygous hygromycin resistance, they will also be homozygous for the DNA binding protein. To confirm this, the plants undergo sterilization and are placed on petri plates containing ½ MS and 30µg/ml hygromycin. After a period of two days in darkness followed by fourteen days in light, the plants are ready for evaluation. Homozygosity of the hygromycin resistance gene is indicated by a 100% growth rate in the presence of hygromycin medium. However, if there are seeds that germinated but did not grow properly, it can be inferred that the plant line is heterozygous.

Hypocotyls measurements

- Petri plates
- sterile seeds
- ½ Murashige and Skoog medium
- Sterile toothpicks
- Aluminium foil
- Growth chamber

After the sterilization of the seeds, these are ready to be put in plates. To study the hypocotyls measurements, it is important to have a well distanced seeds sowing. To achieve this, the single seeds are placed in petri plates filled with 50 ml of ½ Murashige and Skoog medium with the use of a sterilized toothpick. The seeds are positioned along four rows, with 25-30 seeds per row. After sowing the seeds, the plates are closed with micropore tape, to ensure the air flow and impede the entrance of microorganisms. At this point, the hypocotyls are grown in the dark at 4 °C for two nights, then those are transferred in a growth chamber and exposed to light for 6 hours. Lastly, the plates are covered again with aluminium foil and rotated to make the

hypocotyls grow parallel to the medium layer for five days. After this time, the hypocotyls are scanned and their length is measured using the programme Imagej. This allows to set a certain distance in pixel as a scale (in this case, 1 cm), and count the number of pixel occupied by each hypocotyls.

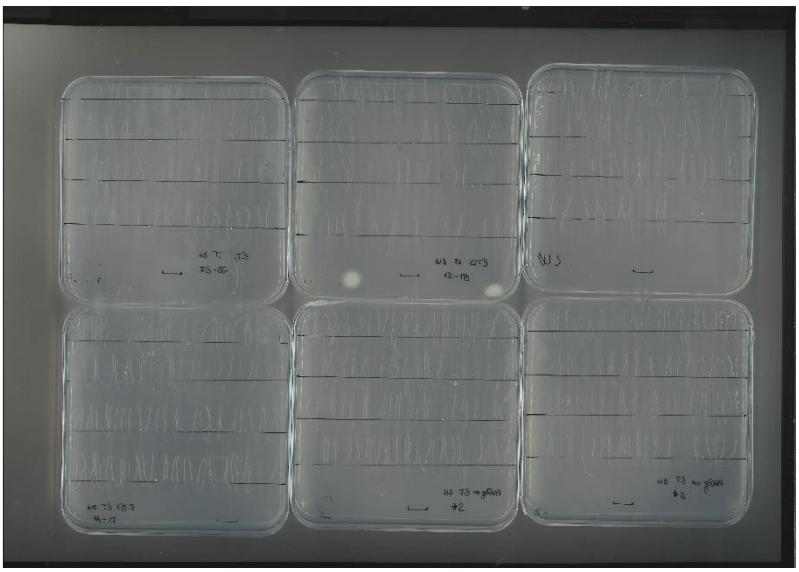

Figure 3.1 disposition of the seedlings for the hypocotyls analysis

Ruthenium red staining test

- Ruthenium red dye
- Milliq water
- Petri dishes
- Camera

After growing in dark, 10 to 12 seedlings are collected to test the penetration of the ruthenium red dye in their cells. The seedlings are collected in a petri dish, where they are submerged in the ruthenium red for two minutes. After this time, the ruthenium red is removed and the seedlings are washed two times with milliq water. The hypocotyls are now ready to be visually analysed. To keep a record of the test results, a camera is used, allowing to visualize the stained hypocotyls.

DNA extraction

Equipment

- Qiagen TissueLyzer
- Seedlings of *Arabidopsis thaliana*
- DNA extraction buffer (Tris-HCl 0.2M, NaCl 0.25M, EDTA pH 8 25mM, SDS 0.5%)
- Isopropanol 100%
- 70% ethanol
- Vent
- Sterile water

The seedlings are freezed in liquid nitrogen and milled in a Qiagen TissueLyzer for two minutes at 30 Hz. After centrifuging, the DNA extraction buffer is added, and the solution mixed by inversion. After a centrifugation, part of the supernatant is transferred in a fresh Eppendorf tube, together with Isopropanol. The mixture is incubated overnight at -80°C, during which the pellet will precipitate, and it will be centrifuged the next day. The isopropanol is removed and the pellet is suspended in Milliq water after a wash with ethanol.

PCR for genotyping

Equipment

- *Arabidopsis thaliana* DNA
- PCR primers
- dNTP mixture
- 10x homemade PCR buffer
- 7x taq DNA polymerase
- Thermocycler

To genotype plant lines, it was performed PCR with different primers: for the TAL22, a direct and a reverse primer were designed with IDT. For dCas9, the primers MLO1806 and MLO1809 were used, and for the *qua1-1* T-DNA insertion, MLO20, MLO21 and MLO22 were used.

MLO number	Genotyping test	Sequence
20	*qua1-1*	CGTTCTGATATCCGAAATAAACATGTCTATTCA
21	*qua1-1*	CTGATACCAGACGTTGCCCGCATAA
22	*qua1-1*	GTTGGATATGATCTGTTCGGAGAGAT
1806	dcas9	ACCGCCAAGTACTTCTTCTAC
1809	dcas9	GAACGATCGGGGAAATTCGC
N/A	TAL22 direct	TAGCGAGAGCCTGACCTATT
N/A	TAL22 reverse	CGGTCGGCATCTACTCTATTTC

In order to execute the PCR, µL of extracted DNA solution was loaded in a PCR master mix composed of

- 2 µL of 10x homemade Taq buffer
- 0.8 µL of a solution of deoxynucleotides
- 0.7 µL of x7 polymerase
- 0.8 µL of each of the needed primers at a concentration of 10 mM (three primers are used in the *qua1-1* genotyping)
- Milliq water to reach the final volume of 19 µL (13 µL in the *qua1-1* genotyping; 13.8 µL in the TAL22 and dCas9 genotyping.

After loading the DNA in the mixture, it is put in a thermocycler where the DNA is amplified.

The PCR conditions used for each experiments are the following:

dCas9

The DNA is first denaturated at 98°C for two minutes, then the PCR cycle starts. The DNA is denaturated at 98°C for 30 seconds, the primers are annealed at 57°C for 30 seconds and lastly the DNA polymerase enlongs the DNA at 65°C for 2 minutes. The cycle is repeated for 34 times. After the last cycle, the temperature stays at 65°C for five more minutes, in order to ensure the a complete elongation of the dna, and lastly the temperature is dropped to 4 °C to conserve the amplified DNA.

TAL22

The DNA is denaturated at 98°C for two minutes. The DNA is denaturated at 98°C for 30 seconds, the primers are annealed at 55°C for 30 seconds and lastly the DNA polymerase enlongs the DNA at 65°C for 50 seconds. After repeating the cycle for 34 times, the temperature stays at 65°C for five more minutes, and then it drops to 4 °C to conserve the amplified DNA.

qua1-1

The DNA is denaturated at 98°C for two minutes. The cycle is composed of a denaturation temperature holded for two minutes, an annealing temperature of 60°C for 30 seconds and an elongation temperature of 65°C for 1 minute. The cycle is repeated for 34 times and the temperature reaches 72°C for five minutes before dropping to 4°C.

Gel electrophoresis for genotyping test

Equipment

- Gel template and comb
- 1% agarose gel in TAE buffer
- Gel electrophoresis apparatus
- GelRed

Once the DNA has been amplified, it can be loaded on a 1% agarose gel and undergo gel electrophoresis in order to assess the length of the PCR amplified DNA. The electrophoresis is conducted at 120 Volt for 30 minutes. These conditions ensure a uniform distribution of the ladder without it escaping the gel.

To detect the amplified DNA, GelRed is used. It is a fluorescent nucleic acid dye, which can detect the DNA in the agarose gel with great sensitivity.[33]It binds the DNA by intercalation, and when exposed to ultra-violect light, it fluoresces with an orange colour, allowing the detection of DNA.

Protein extraction for Western blots

Equipment

- Arabidopsis thaliana seedlings

- 5x loading dye (Tris 0.25 M; SDS 10%, Glycerol 50%, Bromophenolblue 0.25%, DTT 0.1M)
- Qiagen TissueLyzer
- Heating block
- centrifuge

To extract proteins, the seedlings were grinded in a Qiagen TissueLyzer for two minutes at the frequency of 30 hz. The plants are then weighted and dissolved in the triple of their weight of 5x loading dye and boiled at 98°C for 5 minutes. It is then centrifuged at 17 G for 5 minutes and, lastly, the supernatant is transferred to a fresh Eppendorf tube. The proteins are now ready for the western blot.

Western blot

Equipment

- Bio-Rad's 12% Criterion™ TGX Stain-Free™ Precast Gels
- Low fat milk
- PBS buffer
- Anti-FLAG monoclonal antibody
- Rabbit anti-mouse polyclonal antibody
- Super Signal West Dura Extended Duration Substrate
- PVDF membrane
- ChemiDoc^{TM} MP Imaging System
- Trans-Blot ®TurboTM Transfer

In this experiment, 20 µl of proteins were loaded onto Bio-Rad's 12% Criterion™ TGX Stain-Free™ Precast Gels. The gel was subjected to electrophoresis at 180V for a duration of 35-45 minutes. To ensure proper protein loading and running, the protein loading control was visualized using the A ChemiDocTM MP Imaging System, using the "free stain gel" setting. Next, the proteins were transferred (blotted) onto a PVDF membrane using the Trans-Blot® TurboTM Transfer system. The membrane was initially blocked in a solution containing 5% low-fat milk powder in PBS buffer for one and a half hours at room temperature. Subsequently, it was immersed in another solution consisting of a primary antibody for the FLAG tag and 3%

skim milk powder, diluted in PBS buffer, and incubated overnight at 4°C. After the primary antibody incubation, the membrane was washed and then incubated for one hour in PBS containing a secondary antibody (Polyclonal Rabbit anti-Mouse Immunoglobulins) and 1.4% low-fat milk powder. The membrane was washed twice for 10 minutes each with PBS, followed by another wash in PBS-T. Subsequently, it was incubated for 2 minutes in 1ml of Super Signal West Dura Extended Duration Substrate. Finally, the membrane was imaged using the BioRad ChemiDocTM imaging system.

Bibliography

1. Meyerowitz, E. M. *ARABIDOPSIS THALIANA*. www.annualreviews.org (1987).

2. Meinke, D. W., Cherry, J. M., Dean, C., Rounsley, S. D. & Koornneef, M. *Arabidopsis thaliana: A Model Plant for Genome Analysis*. http://www.kazusa.or.jp/arabi/ (2023).

3. Haag, J. R. & Pikaard, C. S. Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nat Rev Mol Cell Biol* **12**, 483–492 (2011).

4. Leng, X., Thomas, Q., Rasmussen, S. H. & Marquardt, S. A G(enomic)P(ositioning)S(ystem) for Plant RNAPII Transcription. *Trends in Plant Science* vol. 25 744–764 Preprint at https://doi.org/10.1016/j.tplants.2020.03.005 (2020).

5. Kouzarides, T. Chromatin Modifications and Their Function. *Cell* **128**, 693–705 (2007).

6. Zamil, M. S. & Geitmann, A. The middle lamella—more than a glue. *Phys Biol* **14**, 015004 (2017).

7. Kohorn, B. D. *et al.* Mutation of an Arabidopsis Golgi membrane protein ELMO1 reduces cell adhesion. *Development* **148**, (2021).

8. Nielsen, M. *et al.* Transcription-driven chromatin repression of Intragenic transcription start sites. *PLoS Genet* **15**, (2019).

9. Carey, M. F., Peterson, C. L. & Smale, S. T. Chromatin Immunoprecipitation (ChIP). *Cold Spring Harb Protoc* **2009**, pdb.prot5279 (2009).

10. Gauchier, M., van Mierlo, G., Vermeulen, M. & Déjardin, J. Purification and enrichment of specific chromatin loci. *Nat Methods* **17**, 380–389 (2020).

11. Pelechano, V., Wei, W., Jakob, P. & Steinmetz, L. M. Genome-wide identification of transcript start and end sites by transcript isoform sequencing. *Nat Protoc* **9**, 1740–1759 (2014).

12. Parker, M. T. *et al.* Nanopore direct RNA sequencing maps the complexity of Arabidopsis mRNA processing and m6A modification. *Elife* **9**, (2020).

13. Pavletich, N. P. & Pabo, C. O. Zinc Finger-DNA Recognition: Crystal Structure of a Zif268-DNA Complex at 2.1 Å. *Science (1979)* **252**, 809–817 (1991).

14. Choo, Y., Sánchez-García, I. & Klug, A. In vivo repression by a site-specific DNA-binding protein designed against an oncogenic sequence. *Nature* **372**, 642–645 (1994).

15. Wolfe, S. A., Grant, R. A., Elrod-Erickson, M. & Pabo, C. O. Beyond the "Recognition Code". *Structure* **9**, 717–723 (2001).

16. Vandevenne, M. *et al.* New Insights into DNA Recognition by Zinc Fingers Revealed by Structural Analysis of the Oncoprotein ZNF217. *Journal of Biological Chemistry* **288**, 10616–10627 (2013).

17. Becker, S. & Boch, J. TALE and TALEN genome editing technologies. *Gene and Genome Editing* **2**, 100007 (2021).

18. Kay, S., Hahn, S., Marois, E., Hause, G. & Bonas, U. A Bacterial Effector Acts as a Plant Transcription Factor and Induces a Cell Size Regulator. *Science (1979)* **318**, 648–651 (2007).

19. Sanjana, N. E. *et al.* A transcription activator-like effector toolbox for genome engineering. *Nat Protoc* **7**, 171–192 (2012).

20. Maeder, M. L. *et al.* Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat Biotechnol* **31**, 1137–1142 (2013).

21. Zhang, F. *et al.* Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol* **29**, 149–153 (2011).

22. Bhaya, D., Davison, M. & Barrangou, R. CRISPR-cas systems in bacteria and archaea: Versatile small RNAs for adaptive defense and regulation. *Annu Rev Genet* **45**, 273–297 (2011).

23. Jinek, M. *et al. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity*. https://www.science.org.

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

25. Brocken, D. J. W., Tark-Dame, M. & Dame, R. T. dCas9: A Versatile Tool for Epigenome Editing. *Curr Issues Mol Biol* 15–32 (2018) doi:10.21775/cimb.026.015.

26. Kuscu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat Biotechnol* **32**, 677–683 (2014).

27. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science (1979)* **351**, 84–88 (2016).

28. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–485 (2015).

29. Cohen, S. N., Chang, A. C. Y., Boyer, H. W. & Helling, R. B. Construction of Biologically Functional Bacterial Plasmids *In Vitro*. *Proceedings of the National Academy of Sciences* **70**, 3240–3244 (1973).

30. Lloyd, A. M. *et al.* Transformation of *Arabidopsis thaliana* with *Agrobacterium tumefaciens*. *Science (1979)* **234**, 464–466 (1986).

31. Glish, G. L. & Vachet, R. W. The basics of mass spectrometry in the twenty-first century. *Nature Reviews Drug Discovery* vol. 2 140–150 Preprint at https://doi.org/10.1038/nrd1011 (2003).

32. Luft, J. H. Ruthenium red and violet. I. Chemistry, purification, methods of use for electron microscopy and mechanism of action. *Anat Rec* **171**, 347–368 (1971).

33. Baum, L. & Fu, W.-L. *Simple and Practical Staining of DNA with GelRed in Agarose Gel Electrophoresis*. https://www.researchgate.net/publication/44605778.