Investigation of micropylar endosperm expressed genes during Arabidopsis seed development and germination

Thesis report

Yunmeng Zhang

Supervisor: Bas Dekkers
PPH80436

Table of Content

Summary

Arabidopsis thaliana is a popular model plant for seed science and reveal insights of physiological processes for study economically important crops. The dormancy of *Arabidopsis thaliana* is coat induced or coat enhanced, and evidence of micropylar endosperm region functioning in the seed dormancy and germination has been emerged. In this study we exploited the insight in micropylar endosperm region specific transcriptional factor genes. Firstly, transcription factors from microarray data of tissue specific expression at micropylar end endosperm during seed germination were identified and their expression confirmed using publicly available data set of seed development. Secondly, we started to make target genes's promoter::reporter gene contructs for localization. Finally, we ordered the T-DNA lines of these genes to study the functions of the genes. From the study, *WRKY8* was investigated that it is involved in the ABA pathway during seed germination by doing germination assay under ABA stress. Under osmotic and salt stress, *WRKY8* loss-of-function also influenced germination performance of the mutants. However, more replications are needed in future to confirm the result.

Introduction

1. Arabidopsis seed development

Seeds are the most complicated means of propagation resulted from plant evolution. They are not only starting materials of plants but also indispensable as food sources for human society (Joosen, et al. 2010). Higher plant seed development can be divided into two major phases, which initiate development of embryo and endosperm followed by seed maturation (Gutierrez, et al. 2007). Seed formation starts by a double fertilization event in angiosperms, the haploid egg cell and the diploid central cell of the female gametophyte within the ovule each fuse with one sperm cell from the pollen tube to form the zygote and the endosperm cell, respectively. The fertilized central cell undergoes a series of nuclear divisions without cytokinesis, resulting in filling the whole seed sac (Holdsworth, et al. 2008). The initial endosperm development and enlargement may affect seed size by enforcing a space limitation (Sun, et al. 2010). Accordingly, endosperm may initiate a signal to regulate the subsequent embryo development or the initial endosperm cavity may simply impose a physical restriction for subsequent embryo enlargement (Sun, et al. 2010). The zygote undergoes a series of differentiation events, resulting in the formation of the mature embryo, by consuming endosperm as a nutrient supply. The endosperm is either a transient or persistent structure within the seed. Mature Arabidopsis seeds contain only a single layer of endosperm cells, whereas the endosperm makes up most of the mass of cereal seeds. The diversity of seed structures implies that distinct genetic scheme underlie the development of each seed component (Bethke, et al. 2007). Seed development is completed by accumulation of storage compounds, a decrease in water content as well as rise of ABA levels. During the last stage desiccation tolerance and primary dormancy are induced (Holdsworth, et al. 2008).

2. Arabidopsis seed dormancy and germination

The completion of seed germination is a key stage in the life cycle of higher plant as well as a significant ecological and agronomical trait (Holdsworth, et al. 2008). Finch-Savage & Leubner-Metzger (2006) defined that germination is the emergence of the radicle through surrounding structures. Seed dormancy is the incapacity of a viable seed to germinate under

favourable conditions (Finch-Savage and Leubner-Metzger 2006). The dormancy of *Arabidopsis thaliana* is coat induced or coat enhanced (Penfield, et al. 2006). The germination process consists of two temporally distinct stages, which are testa rupture and endosperm rupture (Holdsworth, et al. 2008). Dormancy breaking and regulation of germination are determined by both internal and external cues (Holdsworth, et al. 2008).

The evolution of seed structure and seed dormancy has been reported by Martin (Martin 1946), and he defined seed types with distinct embryo to endosperm ratios, arranged them in a seed phylogenetic tree and proposed evolutionary seed trends, one of which is a general evolutionary trend with a decrease in relative endosperm size (Finch-Savage and Leubner-Metzger 2006). There seem to be evolutionarily conserved molecular mechanisms as well as species-specific adaptations for endosperm weakening and/or dormancy release. It can be hypothesized that the function of endosperm in higher plant seeds therefore evolved from providing nutrition towards regulating seed dormancy and germination.

3. Regulators of Arabidopsis seed dormancy and germination

The degree of dormancy and the capacity of the embryo to overcome dormancy are very complex traits which are dependent on hormone and environmental factors such as light, temperature and time of seed dry storage as well as by tissue specific factors (Koornneef, et al. 2002). These factors play either inhibiting or promoting roles in seed dormancy and germination (Figure 1).

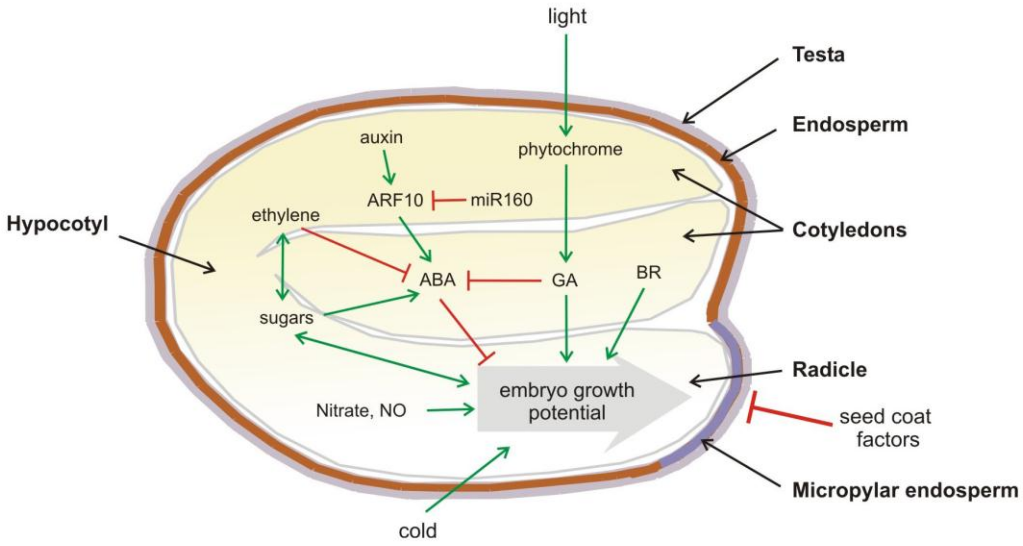


Figure 1. A schematic graph presents the processes controlling seed dormancy and germination of an Arabidopsis seed (Bentsink and Koornneef, 2008). Green arrows indicate germination promoting factors and red arrows represent germination inhibiting factors.

a. Hormonal and environmental regulators

Signal transduction pathways, mediated by environmental and hormonal signals, are necessary for Arabidopsis dormancy release and germination (Holdsworth, et al. 2008). Plant hormone abscisic acid (ABA) is the positive regulator of dormancy, whereas gibberellins (GA) release dormancy and promote the completion of germination, counteracting the effects of ABA (Kucera, et al. 2005). ABA mediates osmotic stress responses in plants as hypothesized by Zhu (2002), mainly due to salt and drought stress enhance ABA accumulation in plants and exogenous application of ABA can have similar effects as osmotic stress. Salt and drought stress signal transduction consists of ionic and osmotic homeostasis signalling pathways, detoxification response pathways, and pathways for growth regulation (Zhu 2002). Both ABA-dependent and –independent osmotic stress signalling exist in modifying expression of transcription factors as well as following downstream stress tolerance genes (Zhu 2002).

b. Role of different seed tissues in role of germination control

Seed dormancy release and germination of species with coat dormancy is determined by the balance of forces between the growth potential of the embryo and the constraint exerted by the covering layers (Kucera, et al. 2005). The role of this single endosperm layer as a mechanical barrier in dormancy and the need to weaken this cell layer for germination has recently been shown in several species including tomato, tobacco and lettuce (Bewley 1997; Kucera, et al. 2005). In Arabidopsis, the endosperm has been proposed to be the primary determinant of seed dormancy (Bethke, et al. 2007). Bethke's study showed that seeds remained dormant when the testa layer was removed from imbibed freshly harvested seeds of the dormant accession C24. The endosperm has been shown to control germination by secreting cell wall remodelling enzymes (CWREs) that weaken the mechanical resistance of the micropylar endosperm cap to radical protrusion (Bewley 1997; Kucera, et al. 2005). Analysis of endosperm specific transcriptome data sets can provide valuable information about regulation of CWREs expression (Penfield, et al. 2006). In Penfield's study, the genes showing enhanced expression in specific tissues provided clues to target CWREs for further functional characterization. In addition, the transcription factor *ABI5* was shown to be expressed in the micropylar endosperm, suggesting that this positive regulator of ABA signalling maybe involve in the ABA-regulated endosperm function (Penfield, et al. 2006).

To understand role of endosperm region in control of Arabidopsis seed germination, several screens have been performed to identify endosperm-expressed genes during seed germination (Dubreucq, et al. 2000; Liu, et al. 2005a). By screening 10 000 Arabidopsis transgenic lines carrying a gene-trap (GUS) construct as identifying markers, an extensin-like gene (*AtEPR1*) was shown to be expressed at the micropylar end of the germinating seeds (Dubreucq, et al. 2000). *AtEPR1* could involve in modifying the cell-wall structure specifically during seed germination, and its expression was induced by the application of exogenous GA (Dubreucq, et al. 2000). Using enhancer trap technologies (Liu, et al. 2005a) enabled the identification of seed germination associated genes. Blue Micropylar End 3 (*BME3*), a GATA zinc finger transcription factor) was identified as a positive regulator of germination (Liu, et al. 2005b). The gene is expressed in the embryo at the radicle tip, could function in

facilitating endosperm rupture since loss of function resulted in seeds lacking endosperm rupture which could be restored by exogenous GA treatment (Liu, et al. 2005b).

4. Scope of the study and research goals

Arabidopsis thaliana is a popular model plant for seed science and provides insights in physiological processes which can be translated to economically important crops (Joosen, et al. 2010). The availability of mutants, ecotypes, inbred populations and sequence information enables the molecular-genetic analysis of many germination-related traits (Joosen, et al. 2010). Arabidopsis possesses seed dormancy, as is the case for many other plant species.

Considering the emerging evidence of micropylar endosperm region functioning on the seed dormancy and germination, the current project exploits the recent insight in micropylar endosperm region specific transcriptional factor genes. Transcription factors are known to play a pivotal role in regulation of gene expression in plants at transcriptional level. We hypothesized that "the tissue specific transcription factor genes in the micropylar endosperm region during seed development and germination may regulate the seed germination behavior and dormancy release".

For this study we used gene expression data set that combined both spatial and temporal sampling during seed imbibition and germination. In total 11 time points were sampled during the germination time curve and up to four different seed tissues (micropylar and chalazal endosperm, peripheral endosperm, cotyledon and radicle). Transcription factors from data of tissue specific expression at micropylar endosperm and the chalazal endosperm were mined as basis of this thesis study. Six transcription factors (*WOX9/STIMPY, BS1/ TMO7/ ATBS1, LBD18/ASL20, WRKY8, WRKY28,* and *WRKY75*) were chosen from this dataset since they had at least three folds higher expression level in the micropylar and chalazal endosperm compared with peripheral endosperm at begin of seed germination (at three hour after sowing).

In the *in silico* analysis, expression of these genes at the end phase of seed development from public dataset was checked in order to confirm whether they were also highly expressed in the late seed development stage prior to dormancy/ germination. In

addition to the transcriptomics approaches, T-DNA lines of the target genes were ordered to study the function of the genes in regulation of seed germination. Moreover, in order to confirm the expression of genes in the seed tissue, we started making promoter::GUS gene constructs.

Material and Methods

1. *In silico* analysis of transcriptome data

Transcription factors from the microarray data sets of tissue specific expression in the beginning of seed germination at micropylar endosperm and the chalazal endosperm were mined as basis of this thesis study. In order to identify the expression level of selected genes during the last phase of seed development, the extensive amount of seed microarray data which is publicly available on the Bio Array Resource (BAR website: www.bar.utoronto.ca) (Winter, et al. 2007) was made use of. A comparison between the two data sets gives an indication whether the selected transcription factors have specific higher expression levels in micropylar end during the first phase of germination also specifically express during the last phase of seed development. Some other transcription factor genes found from BAR dataset which highly expressed in micropylar end at the last seed developmental stage were compared in the microarray data set. The result from the data comparison helped to target the interesting genes for further study.

2. Confirmation of tissue specific expression using promoter-GUS reporter constructs

The three Arabidopsis WRKY genes (*ATWRKY8*, *ATWRKY23* and *ATWRKY28*) were chosen from the eight transcription factor genes reported in previous section. Another three cell wall related genes (*AGP2, SKS9* and *AGP21*) were included based on gene expression differences at testa rupture. These genes were highly expressed in the micropylar endosperm of ruptured seeds compared to non-ruptured seeds. In total six genes were chosen for confirming tissue specific expression study (Appendix 1).

The reporter gene used here was *Escherichia coli* β-glucuronidase (GUS) (Jefferson, et al. 1987). The technology employed was Gateway cloning system. For the accurate profiling of expression patterns, primers were designed that amplified approximately 1.5 kb to 2.5 kb of the promoter region (Table 2 in Appendix 2-1). Another reporter construct, containing a strong promoter, in this case the cauliflower mosaic virus (CaMV) 35S, was used as control.

The Gateway cloning procedure consisted of several steps: DNA amplification by high fidelity polymerase (PFX) (Appendix 2-2); BP reaction (Appendix 2-3); Colony PCR check

(Appendix 2-4). The clones were checked by sequencing. Here four samples were send to EZ-Seq (Macrogen). Due to time limitation of this experiment, the LR reaction and transformation were not performed.

GUS expression pattern test was done with 35S::GUS line and activation tagging lines for endosperm region (from study of Liu et al, (2005)) (Appendix 2-5). The aim of this test was to optimize the protocol for endosperm GUS expression pattern analysis for future study. According to GUS staining protocol in Appendix 2-5, optimization was set into three treatments for the second step of the protocol, which were not to fix imbibed seeds in acetone, or fix the seeds for 1 minute or 20 minutes. Tissues observation from different treatments was set to be the 5^{th} and 26^{th} hour after imbibition.

3. Identification of gene function using T-DNA lines

a. T-DNA lines ordering and isolation of homozygous lines

T-DNA lines of targeting genes (from the previous two sections) were ordered from Nottingham Arabidopsis Stock Centre (NASC, http://arabidopsis.info/). The list of genes with available T-DNA lines was shown in Appendix 3-1. Plant materials including wild type (Columbia) and T-DNA lines were grown on the rock wool in a climate chamber (20°C day, 18°C night) with 16h of light (35 $W \cdot m^{-2}$) at a relative humidity of 70%. Seeds were bulk harvested and stored at 20°C under ambient relative humidity (around 40%).

For detection of T-DNA insertions in the genes of interest, we used two PCR reactions followed by analysis on a 1% agarose gel (Figure 2). The first PCR reactions contained a combination of both gene specific primers and T-DNA primer. In this way we were able to detect lines which carried the T-DNA insertions in the genes of interest (see Figure 2). Since insertion fragments were smaller than the gene specific ones, logically gene specific fragments were less efficient to detect and therefore it could be miss indicating with only this PCR reaction. For this reason, we performed a second PCR reaction which contained only the gene specific primers. The second PCR reaction confirms whether the lines detected is indeed homozygous (in case no wildtype product is detected).

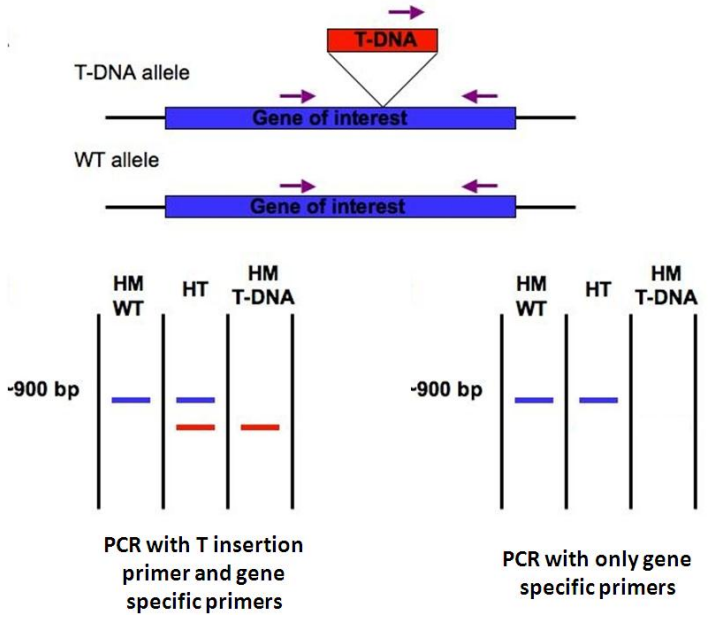


Figure 2. PCR principle for isolating homozygous T-DNA lines (Adjusted from figures on Website of Warwick School of Life Science, 2008). Upper graph shows the two possible alleles and binding position of screening primers present in T-DNA lines which need to be distinguished. Graphs below show expected PCR products on a gel amplified by two different primer sets for each potential genotype, homozygous wild type (HM WT), heterozygous (HT) and homozygous T-DNA knockout (HM T-DNA).

PCR pipetting scheme is in Appendix 3-2. Homozygous T-DNA lines and a wild type control were multiplied in 12 plants from their seeds for genotyping. Whereas for ordered heterozygous plants, 36 plants were growing for genotyping in order to select homozygous ones. All homozygous lines were confirmed by a second DNA extraction and PCR analysis so we were sure about the homozygous state of the T-DNA line used in further experiments.

b. Targeting gene expression analysis in homozygous T-DNA lines

After isolation of homozygous T-DNA lines, target gene expression levels were analysed by using RT-qPCR. Three T-DNA lines (GK-625G10, SALK_107668C and SALK_006855C) and a wild type control were analysed in this study. For RNA isolation, it was according to the protocol described by Dekkers et al. (2012). A second time of extraction was done with same sample. The whole seeds were frozen in liquid nitrogen at the third hour during seed germination phase because the genes specifically expressed at this time point. Seeds were imbibed in the petri dish with wetted filter paper at 22 °C in a climate chamber with continuous light.

RNA integrity of all samples was assessed by analysis on 1% agarose gel. Clear bands for all samples are visible, indicating that RNA was intact. Further, OD 260/280 ratios were determined using a Nanodrop ND-1000, and were close to 2.0 for all samples in this experiment. Genomic DNA was removed using a DNase treatment, and afterwards cDNA samples and RNA samples were compared to check the absence of DNA (minus RT control). Signals detected in minus RT control should above 5 Cq (quantification cycle) higher than they were in the cDNA samples. For RT-qPCR experiment, primers were designed preferably in the 3' part of the transcript. When possible the primer or primer pair was designed for spanning intron-exon border. The melting temperature of the primers was between 59 and 62 °C. In the end, data were analyzed. The selection of reference genes for RT-qPCR is a critical issue. Therefore we used recently identified seed specific reference genes (At2g20000, At2g43770 and At4g02080) from study of Dekkers et al. (2012) that were stable expressed in both endosperm and radicle during 3 hours after germination.

c. Assessment of Phenotypic traits

Phenotypic traits of homozygous T-DNA lines were assessed to study the regulation of a target transcription factor. Three T-DNA lines (GK-625G10, SALK_107668C and SALK_006855C) and a wild type control were studied under different environmental conditions. The seeds were bulk harvested and stored at 20°C under ambient relative humidity (around 40%).

i. Germination assay for dormancy release

The germination assay was performed using the Germinator software (Joosen, et al., (2010). To test dormancy release of the T-DNA lines, the seeds were analysed by germination assay at week 1, week 2 and week 3 after harvest, accordingly.

ii. Germination test under different stress conditions

Germination test of T-DNA lines under different stress conditions were according similarly to the protocol described by Joosen et al, (2010). ABA stress was set in six different concentrations (0, 0.025 µM, 0.1 µM, 0.5 µM, 1 µM and 5µM). For salt and osmotic stress tests seeds were germinated on 125 mmol NaCl and 350 mmol mannitol, respectively.

Results and discussion

1. *In silico* analysis of transcriptome data

Experimental microarray data available in the lab was mined to find the transcription factors which have higher expression in the micropylar end endosperm in comparison to the peripheral endosperm during the seed germination (three hours after imbibition). Six transcription factors (*WOX9/STIMPY, BS1/ TMO7/ ATBS1, LBD18/ASL20, WRKY8, WRKY28,* and *WRKY75*) were found that showed at least three folds higher expression in the micropylar end endosperm comparison to the peripheral endosperm (Figure 3 A, B, C, F, G, and H). We hypothesize that these transcription factors may play a role in regulating seed development and/or germination. First we confirmed the six gene expression profiles by comparison at publicly available dataset (Figure 3-I, J, K, N, O, and P). From the figure 3 we can see that the genes with higher expression in the micropylar endosperm region during first phase of seed germination also highly expressed during last phase of seed development.

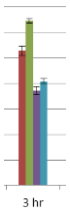
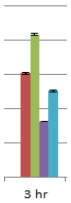
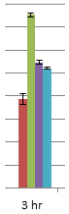
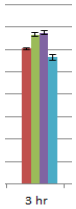
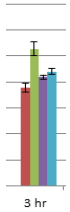
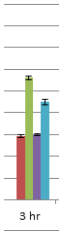
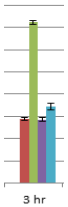
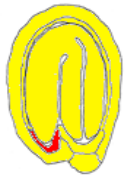
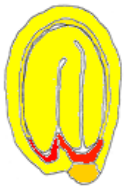
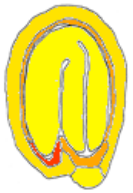
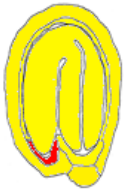
After confirmation of expression, these five candidate transcription factors were targeted for further study. From previous studies, these transcription factor genes were reported with various functions. *WOX9* (WUSCHEL-related homeobox gene) is involved in cell cycle control (Wu, *et, al.* 2005). *TMO7* (target of MONOPTEROS) encodes a basic helix–loop–helix (bHLH) transcription factor is required and partially sufficient for MP-dependent root initiation (Schlereth*et al.*2010) as well as regulating brassinosteroid (BR) signalling (Wang*et al.*2009). *LBD18* gene (lateral organ boundaries domain gene) referred to lateral organ boundaries, may function in initiation and emergence of lateral root formation (Lee, *et al.* 2009).

Three WRKY transcription factor family members were included, *WRKY8, WRKY28,* and *WRKY75*. WRKY transcription factors are one of the largest families of transcriptional regulators in plants. It has been reported that members of the family play roles in essential parts of signalling webs that modulate many plant processes (Rushton, et al. 2010). To be more specific, *WRKY8* is a defence gene localized in the nucleus. It is a negative regulator of basal resistance to *P. syringae* and positive regulator to *B. cinerea*. *WRKY28* is integrated in a model together with *WRKY46* for regulation of Salicylic acid (SA) biosynthesis (van Verk, et

al. 2011). *WRKY75* plays a critical role in regulating adaptive mechanisms against Phosphate (Pi) deficiency as well as root development (Devaiah, et al. 2007).

In order to reveal more information about WRKY family members from the gene expression datasets in this study, other WRKY transcription factor genes were also mined in the BAR website of gene expression during seed development. *WRKY23* and *WRKY25* were found to be also highly expressed in ME region during last stage of seed development (Figure 3-L and M), which indicated that they may function in micropular endosperm tissue during seed germination. To find out more about this question, microarray data were mined for these two genes (Figure 3-D and E). During the first phase of seed germination, *WRKY23* and *WRKY25* also highly expressed in the micropylar endosperm region. Therefore, they were also included in the targeted genes for further study. From previous studies, it has been reported that *WRKY23* expressed during early stages of nematode infection, initiated by auxin-independent signals (Grunewald, et al. 2008). *WRKY25* is involved in response to various abiotic stresses, especially salt stress (Jiang and Deyholos, 2009). Together with *WRKY26* and *WRKY33, WRKY25* positively regulated responses to heat stress, and these three genes interact synergistically and function overlapping in plant thermotolerance (Li, et al. 2011).

Figure 3. Target genes expression data comparison between first stage of seed germination (Microarray data of target gene expression level in different tissues during at 3rd hour after sowing) and last stage of seed development (target gene expression level from public dataset at maturation stage of seed development). Green colour of the bar in microarray dataset represents micropylar endosperm tissue while light blue bar represents lateral endosperm. Colour of the tissues from public dataset towards yellow means low level of target gene expression, while red means high level of target gene expression.

Gene name	*WOX9/STIMPY*	*BS1/ TMO7/ ATBS1*	*WRKY8*	*WRKY23*	*WRKY25*	*LBD18/ASL20*	*WRKY75*	*WRKY28*
First stage of seed germination	A. 	B. 	C. 	D. 	E. 	F. 	G. 	H. 
Last stage of seed development	I.  Expression level: 1200	J.  Expression level: 53.07	K.  Expression level: 476.32	L.  Expression level: 45.28	M.  Expression level: 40.22	N.  Expression level: 405.45	O.  Expression level: 8.42	P.  Expression level: 236.44

2. Confirm tissue specific expression

Promoter-reporter gene constructs are often employed as a powerful tool to obtain detailed expression profiles (Mitsuda and Ohme-Takagi 2009). Here the promoter region of each target gene with GUS reporter gene constructs were made to localize the gene expression during early seed germination. A proofreading polymerase (PFX) PCR was employed to amplify target promoter region. Figure 4 represented amplification the genes *WRKY8* and *SKS9,* as an example to show how we identify promoter region from targeted transcription factor genes. The promoter regions were inserted in pDONR207 vector using Gateway cloning and transformed to *E. coli*. Colony PCR was used to select colonies with correct insertions. Figure 5 showed seven colonies (a to g) which were tested by colony PCR. Three colonies likely contained right-size-insertion (b, f and g), and the PCR products were sequenced for further confirmation. Figure in Appendix 4-1 showed one of the gene (*SKS9*) sequencing results aligning with the wild type Colombia sequence. In the second line, although the errors in the end of the sequence can be ignored for analysis, two unclear reads appeared in the middle of the sequence. Therefore this colony can not be used for further study and another colony with same insertion should be tested in sequencing. Figure in Appendix 4-2 showed a sequence alignment (*WRKY28*) with only two errors in the end, which can be ignored. Therefore this colony can be used for LR reaction in further study.

In total, promoter regions of *WRKY8, WRKY23, WRKY28, AGP2, AGP21* and *SKS9* were amplified and transferred to pDONR207 vector. The promoters of *WRKY8, WRKY23, WRKY28* and *SKS9* were sequenced. The WRKY8 and WRKY28 gene promoters were tested with correct sequences in the colonies and therefore ready to be proceeded with LR reactions. *WRKY23* with failed sequencing as well as SKS9, as described in Figure 6, needed to be re-sequenced with another colony.

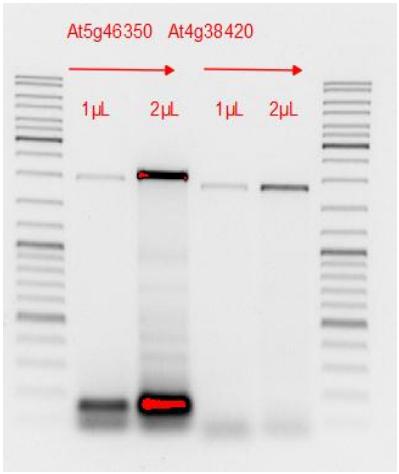


Figure 4. Amplification of target genes' promoter region by PFX PCR, followed by resolved in the agarose gel. First two bands were from promoter region of At5g46350 (*WRKY8*) gene with the right size in two different concentrations. The third and fouth bands indicated two different concentrations of At4g38420 (*SKS9*) gene's promoter region with correct size.

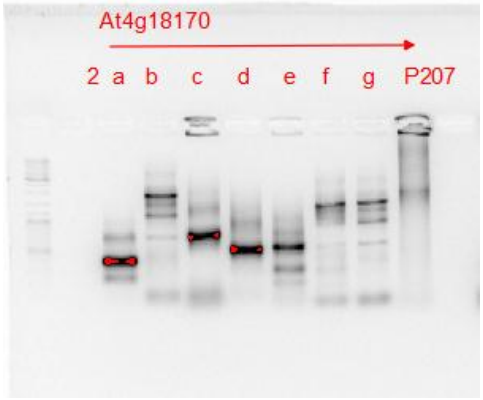
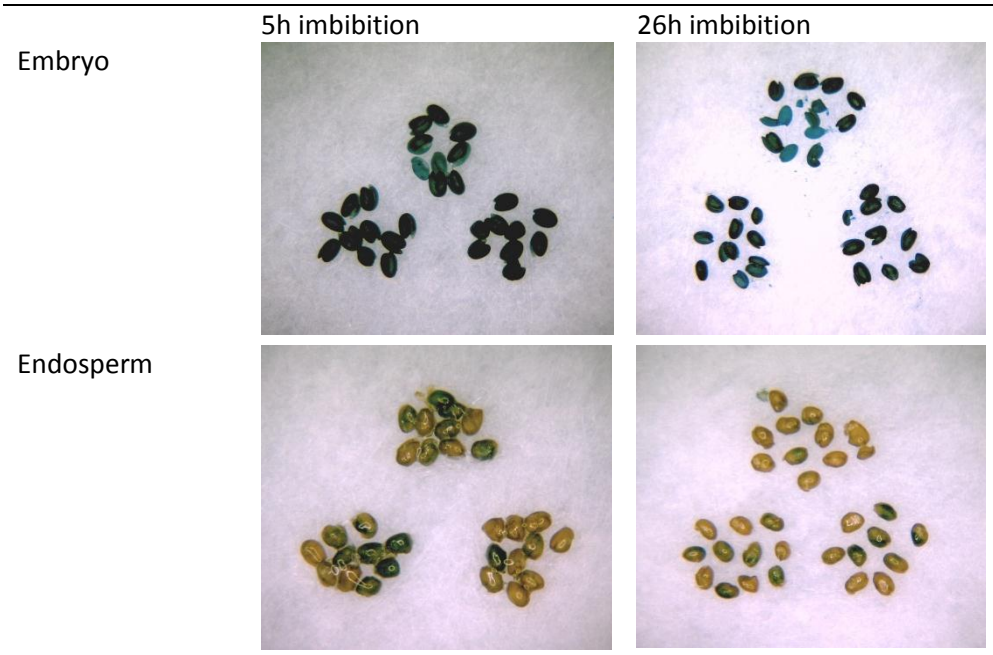


Figure 5. Colony PCR result for testing right size insertion of promoter region of At4g18170 (*WRKY28*). The band b, f and g indicated amplified regions with correct size, and a, c, d and e with smaller sizes of amplified regions indicated that these colonies contained wrong targeted gene region.

2.1 GUS expression protocol test

This test was aimed to optimize the protocol for GUS expression of target genes' study. The protocol contained an important step of fix tissues in acetone and it was 20 minutes in the original protocol. Since the endosperm tissue in Arabidopsis is only one cell layer, it was tested whether shorter periods of fixation would work better (Figure 6). Using 35S::GUS lines, most number of endosperm region can be observed at 5th hour of imbibition with fixation of 1 minute or not fixed in acetone (Figure 6).

Figure 6. GUS staining with 35S::GUS lines. According to GUS staining protocol in Appendix 2-5, optimization was set into three treatments for the second step of the protocol, which were not to fix imbibed seeds in acetone, or fix the seeds for 1 minute or 20 minutes (displayed as symbolized in the graph).



With the optimized protocol, two GUS activation tagging lines (from study of Liu et al, (2005)) were tested. These two lines were expected to be specifically expressed in the micropylar region. Examine of the performance of the two different lines can be used to predict the performance of the GUS constructs of targeted genes in future study. From Figure 7 we can see that the second line performed better with more endosperm region stained rather than the embryo. This line could be used as a positive control in later study of endosperm specific gene construct.

Figure 7. GUS staining test with two activation tagging lines in the endosperm micropylar region (displayed as symbolized in the graph).

3. Identification of gene function using T-DNA lines

Manipulation of target genes provides the possibilities to induce phenotype changes, therefore identification of biological function of target transcription factors. Two strategies are usually applied, which are “gain of function” and “loss of function” (Krysan, et al. 1999). T-DNA insertion lines, in which the large T-DNA fragment is inserted into genes, are available and provide a very good resources for loss-of-function analysis (Mitsuda and Ohme-Takagi 2009).

The identification of knockout mutants is the first step towards describing the function of a gene. After the isolation of a mutant line, plants homozygous for the mutation must be identified that only one T-DNA insertion is present. With a confirmed mutant in hand, the next step is to determine the consequences of the mutation on growth and development relative to the wild type. However, it has become apparent that many knockout mutants have no readily identifiable phenotype, unless grown under specific condition (Krysan, et al. 1999).

3.1 Genotyping of T-DNA lines by PCR, 2 times for confirmation

Polymerase chain reaction (PCR) methods have been developed that enable one to easily isolate individual plants that contain a particular T-DNA mutation of interest (Krysan, et al. 1999). We used two PCR reactions as described in materials and methods (see also Figure 2). Expected PCR products of homozygous T-DNA knockout on a gel amplified by T insertion primer and gene specific primers were one or two clear bands with correct T insertion size. While PCR products amplified by only gene specific primers were expected to be blank for homozygous T-DNA knockout on a gel. In total three lines were checked to be homozygous lines for further study of phenotyping, GABI-625G10 (*WRKY8*), SALK_107668 (*WRKY8*) and SALK_006855 (*WRKY23*) (Figure 8).

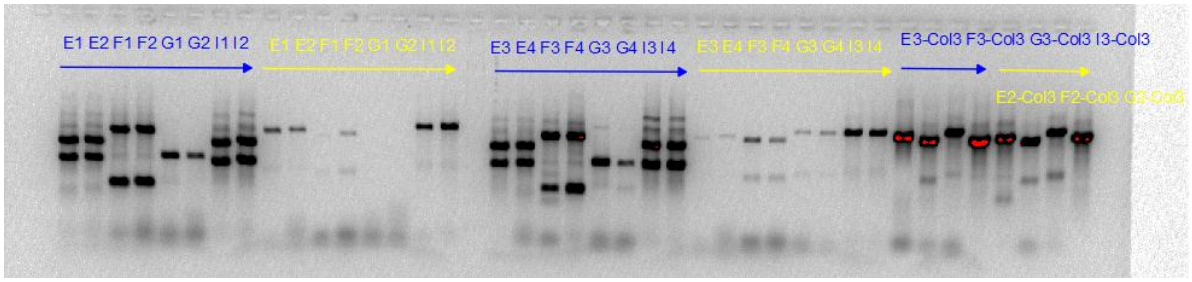


Figure 8. One of the gel photos for genotyping the T-DNA lines. *GABI-625G10 (WRKY8), SALK_107668 (WRKY8) and SALK_006855 (WRKY23)* were screened for individuals containing a homozygous T-DNA insertion. Bands with blue mark indicate mutants amplified by gene specific primer and border primers, while bands with yellow mark indicate mutants amplified by only gene specific primer. The last eight bands are wild type Arabidopsis DNA amplified by both primer sets for positive control. Those mutant numbers showing one or two clear lower bands (at around 400 bp) with blue mark and blank with yellow mark were the expected homozygous T-DNA lines for further study.

3.2 Gene expression analysis

We identified homozygous plants for three different T-DNA lines (section 3.1). Next we tested the expression levels of the mutated genes in these T-DNA lines in order to make sure the T-DNA lines were “loss of function”. From Figure 9 we can see that expression level of *WRKY8* was significantly reduced in the mutants. Therefore, phenotyping of these lines was reliable. However, *WRKY23* was even significantly higher expressed in its mutants compared in the wild type (Figure 10). The reason could be that the T-insertion of the line SALK_006855 was on one of its alternative splicings, and the gene could still express through other splicings (Appendix 4-3).

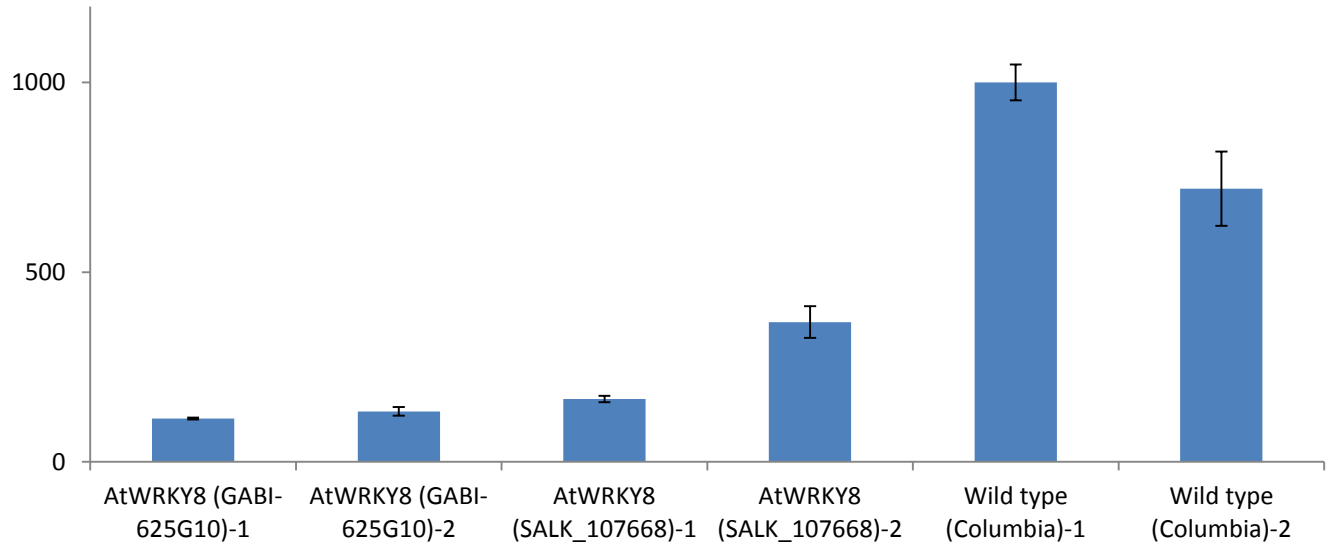


Figure 9. *WRKY8* gene expression level in its mutants compared to wild type

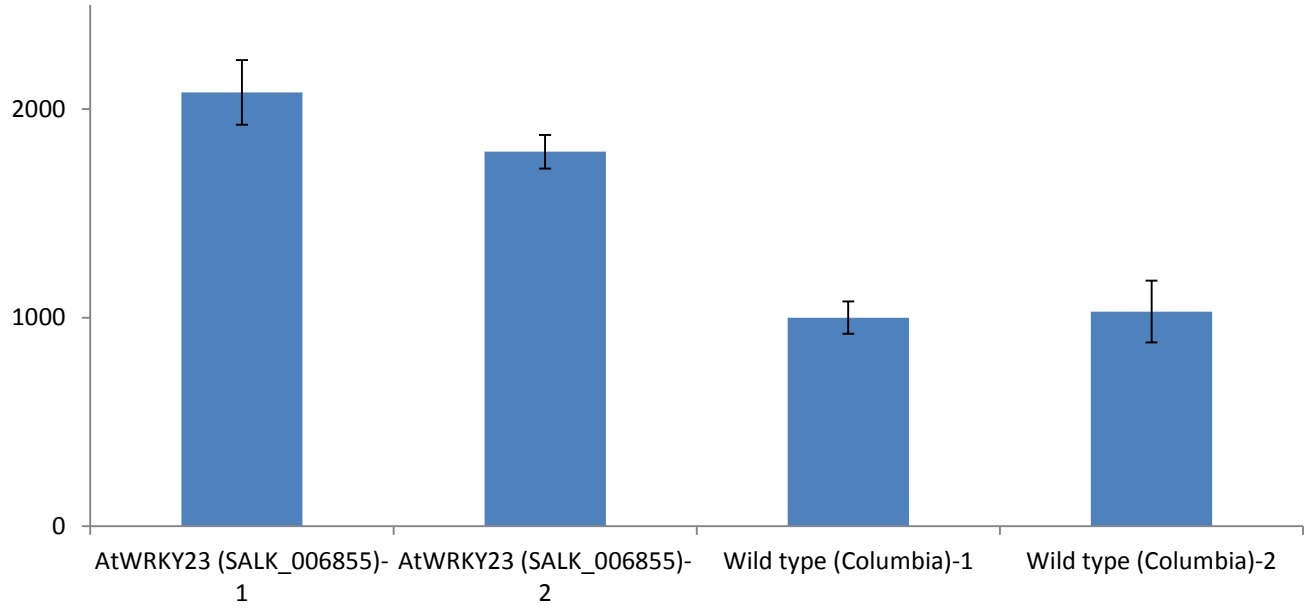


Figure 10. *WRKY23* gene expression level in its mutants compared to wild type

3.3 Phenotyping of T-DNA lines

Phenotypic analysis of transcription factors includes the identification of a phenotype that is regulated by a target transcription factor (Mitsuda and Ohme-Takagi 2009). Phenotypes regulated by target genes include not only visible morphological changes, but also not significant changes which are only visible under certain conditions (Mitsuda and Ohme-Takagi 2009). To induce the visible changes, several different stress conditions were applied including a concentration range of ABA, salt stress, osmotic stress and sub-optimal temperatures during germination.

3.3.1 Germination assay for dormancy release

Compare germination rates of seeds sowing in week 1, 2 and 3 after seed harvest, the three T-DNA lines and the wild type had an increase of germination percentage. And for each genotype, germination percentage in week 1 was significantly less than it was in week 2 and week 3. The difference of the increase rates between line GK-625G10 (*WRKY8*) and SALK_006855C (*WRKY23*) was significant, but all the three mutants did not significantly differ from the wild type (Columbia) (Figure 11). Therefore this experiment, do not prove a role for the transcription factors *WRKY8* and *WRKY23* in regulation of dormancy release under normal conditions.

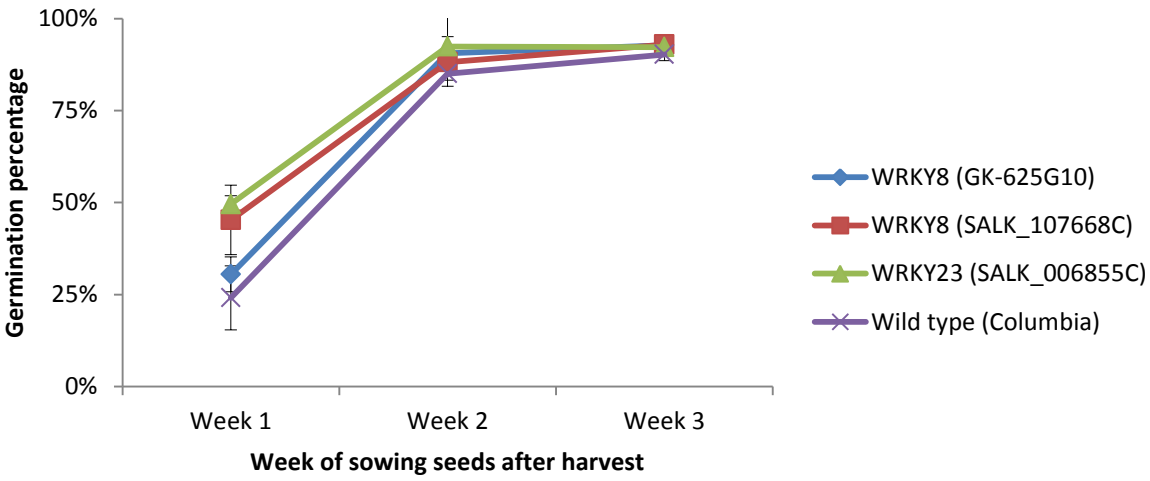


Figure 11. Dormancy release of three T-DNA lines and the wild type control.

3.3.2 Germination assay under ABA stress

The table showed that under different ABA concentrations the genotypes had different germination performances. With an increase of ABA concentration, the germination percentage generally decreased. The germination percentage of different lines was not influenced until 0.1 µmol, except that the mutant SALK_107668C (*WRKY8*) had a slightly lower germination percentage than the others at concentration of 0.025 µM (Figure 12). When the concentration is higher than 0.5 µmol, the germination decreased dramatically The germination of the mutants decreased more dramatically compared to the wild type since the mutants had the significantly lower of germination percentage compared with wild type under the concentration of 0.5 µmol and 1 µmol (Table 1). At very high level concentration of ABA (5 µmol), the germination of different genotypes were similarly low (Table 1). From this study we can find that mutants with knock down of gene *WRKY8* were more sensitive than the wild type, therefore *WRKY8* could play a role in the ABA pathway during seed germination. Since the gene *WRKY23* was not knocked down in the T-DNA line SALK_006855C as examined by RT-qPCR in previous experiment (Figure 9), the changing of sensitivity towards different ABA concentrations can not approve that *WRKY23* is involved in the ABA pathway.

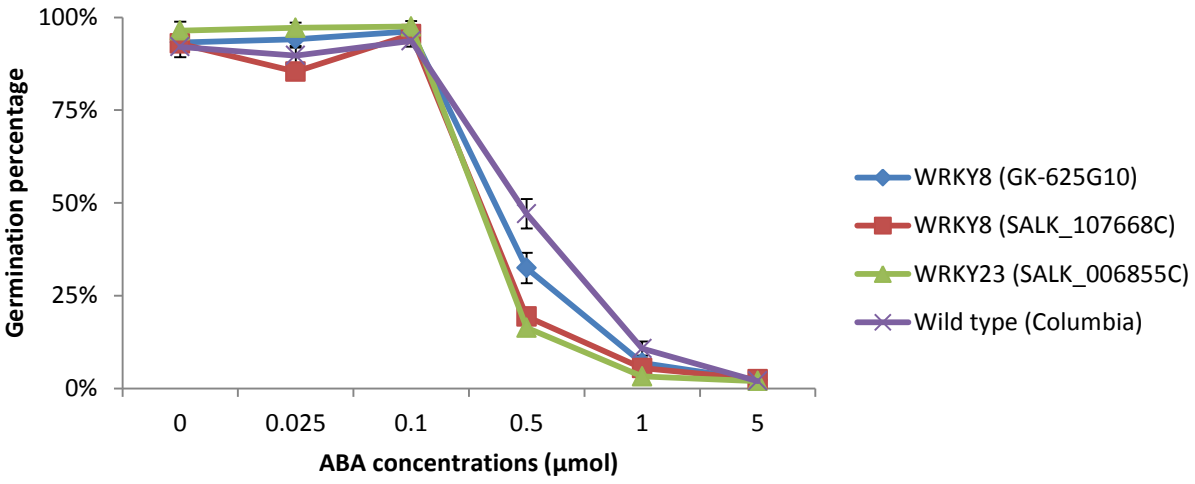


Figure 12. Germination percentage of each T-DNA lines and wild type under different ABA concentration.

Table 1. Germination percentage of each T-DNA lines and wild type under different ABA concentration (different small case of letters near the percentage number indicates significant difference between the number).

ABA concentration (µmol)	*WRKY8* (GK-625G10)	*WRKY8* (SALK_107668C)	*WRKY23* (SALK_006855C)	Wild type (Columbia)
0	93% j	93% j	96% j	92% ij
0.025	94% j	85% i	97% j	90% ij
0.1	96% j	95% j	98% j	94% j
0.5	32% g	19% f	16% ef	47% h
1	7% cd	6% bc	3% ab	11% de
5	2% a	2% a	2% a	2% a

3.3.3 Germination assay under osmotic and salt stresses

Under salt and osmotic stress conditions, germination percentage and speed differ from normal conditions. And the differences varied in different T-DNA lines and wild type. All three T-DNA lines seemed to have higher percentage of seed germination compared with wild type under salt stress as well as osmotic stress and the germination percentage of line GK-625G10 (*WRKY8*) was significantly higher than the wild type under both stress conditions (Figure 13 and Table 2). It could indicate that *WRKY8* gene loss of function resulted the line GK-625G10 has less sensitivity under salt and osmotic stresses compared with wild type. As described by Zhu (2002), both salt and drought stress enhance ABA accumulation in plants. The pathways of these different stresses are related. *WRKY8* is involved in the ABA pathway, as studied in previous experiment, therefore loss of function also influence T-DNA line germination under osmotic and salt stress.

Exogenous application of ABA can have similar effects as osmotic stress (Zhu 2002). In this study, ABA, osmotic and salt stress all lead a decreasing germination percentage in mutants and wildtype compared in normal condition. However, sensitivity change of line GK-625G10 in this study under ABA stress (higher sensitivity) was different from under salt and osmotic stress (lower sensitivity). The reason could be that germination is a very complex process which may be regulated by many genes. Therefore only one gene loss of function did not had obvious morphological difference, with other genes also play roles in stress conditions. Further study is needed and different concentrations of salt and osmotic stresses are recommended to apply.

Sensitivity of germination percentage of line SALK_107668C (*WRKY 8*) did not change significantly compared with wild type although *WRKY 8* was also lower expressed in the mutants. To explore why germination performance of line GK-625G10 (*WRKY8*) and SALK_107668C (*WRKY8*) differed under stress conditions, the position of two T-DNA insertions were checked in the TAIR website. As shown in Figure in Appendix 4-4, T-DNA insertion of line SALK_107668C was in the intron of *WRKY8* while T-DNA insertion of line GK-625G10 was in the extron *WRKY8*. This could explain why *WRKY8* did not influence the germination performance of line SALK_107668C under ABA, salt and osmotic stresses.

Time to take for 50% of the seeds to germinate is an indicator of germination speed. Each T-DNA lines under salt and osmotic stresses did not significantly differ in germination speed compared with wild type therefore it did not seem to be an useful indicator in this study (Figure 14). Since the gene *WRKY23* was not knocked down in the T-DNA line SALK_006855C as examined by RT-qPCR in previous experiment (Figure 9), the changing of sensitivity under salt and osmotic stresses can not prove that *WRKY 23* is involved.

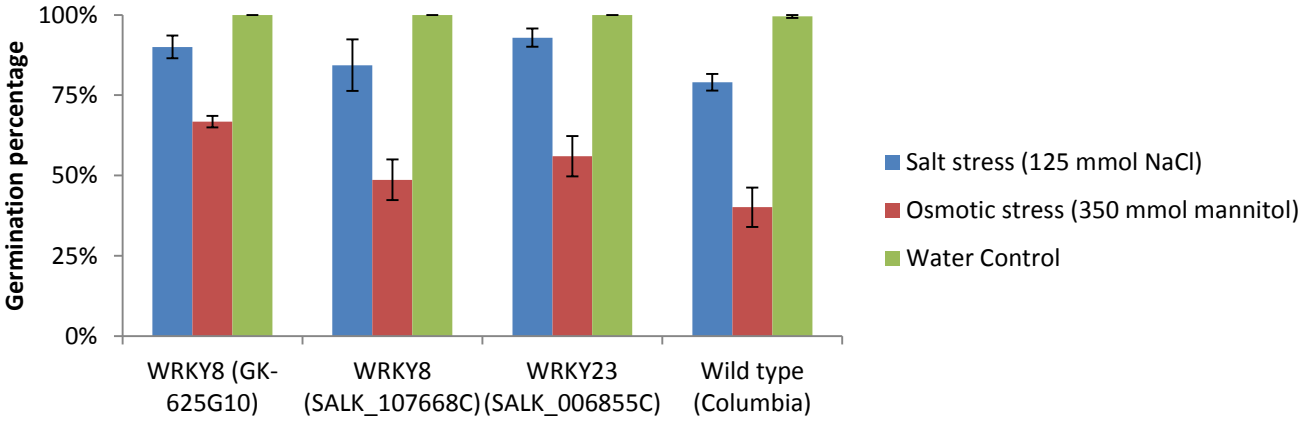


Figure 13. Germination percentage of each T-DNA lines and wild type under salt and osmotic stresses.

Table 2. Germination percentage of each T-DNA lines and wild type under salt and osmotic stresses (different small case of letters near the percentage number indicates significant difference between the number).

	Salt stress (125 mmol NaCl)	Osmotic stress (350 mmol mannitol)	Water Control
WRKY8 (GK-625G10)	90% e	67% bd	100% f
WRKY8 (SALK_107668C)	84% cdef	49% a	100% f
WRKY23 (SALK_006855C)	93% e	56% ab	100% ef
Wild type (Columbia)	79% c	40% a	100% f

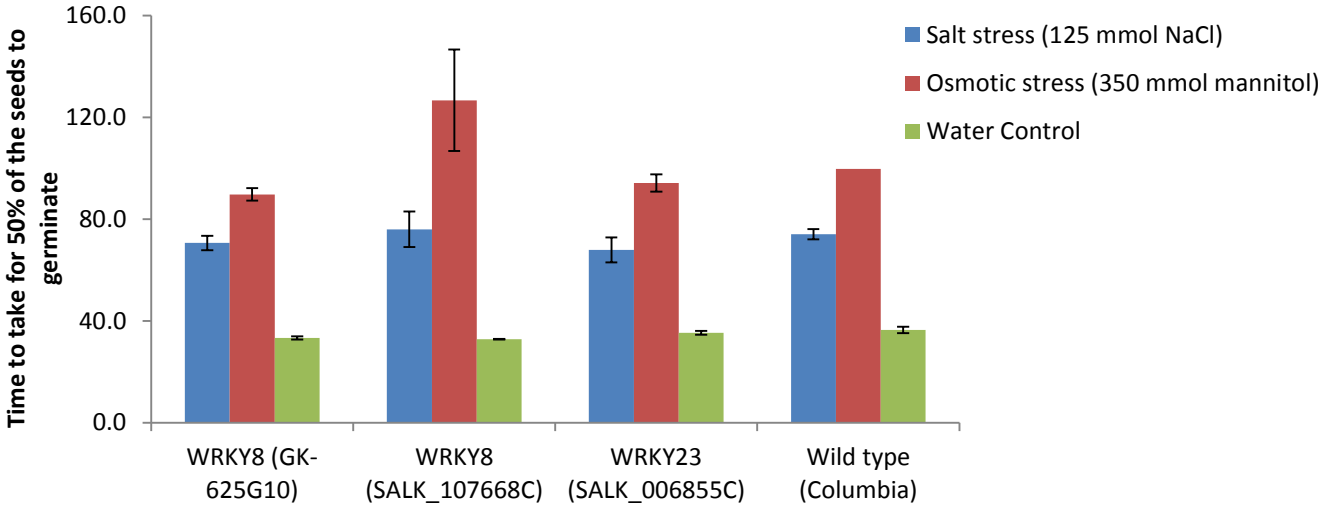


Figure 14. Time to take for 50% of the seeds to germinate of each T-DNA lines and wild type under salt and osmotic stresses.

3.3.4 Germination assay under different temperature

Germination assay under 10 and 30 ℃ aimed to find whether the target genes have involved in temperature regulated pathway. However, as illustrated in Figure 15 and 16, germination percentage as well as germinate speed of T-DNA lines did not significantly differ from wild type under different temperatures. Therefore, *WRKY8* could not be involved in the temperature regulated pathways. *WRKY25* was reported to be responded to heat stress (Li, et al. 2011), but no homozygous T-DNA lines were selected for *WRKY25* in this study. In the future study, the heterozygous T-DNA lines can be further grown and selected for phenotyping.

Another reason for lack of phenotype with loss of function mutants could be that high plants have the ability to adapt their physiology to stresses and constraints without undergoing morphological changes. Our inability to detect slight morphological alterations and/or weak reductions in fitness is also one of the reason. Furthermore, the gene could be patial or functional redundantly during the complex regulation. Therefore, phenotyping must be integrated with other approach such as transcriptome analysis to reveal gene function.

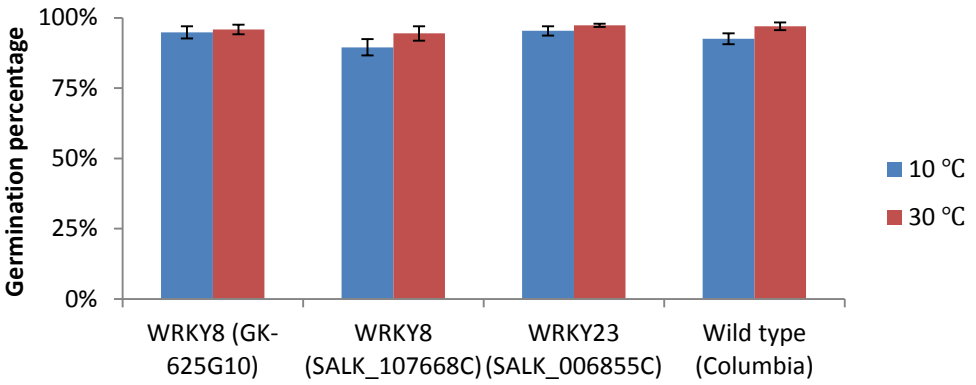


Figure 15. Germination percentage of each T-DNA lines and wild type under 10 ℃ and 30 ℃.

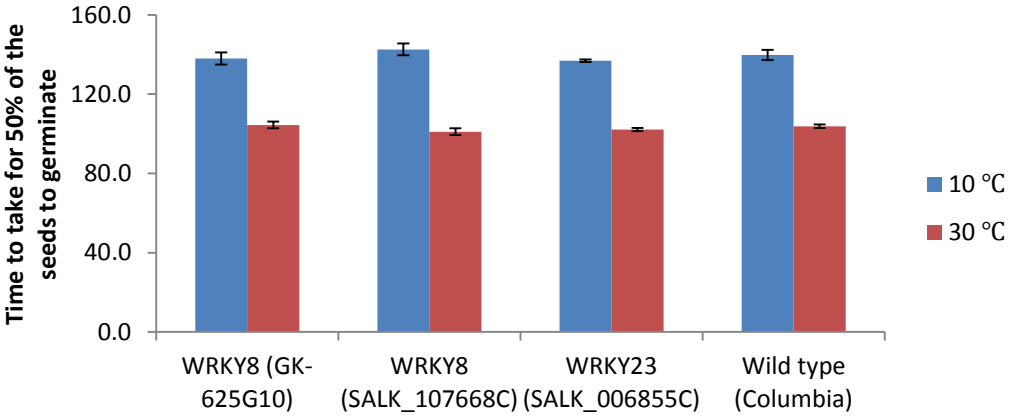


Figure 16. Time to take for 50% of the seeds to germinate of each T-DNA lines and wild type under 10 ℃ and 30 ℃.

Conclusion

Considering the emerging evidence of micropylar endosperm region functioning on the seed dormancy and germination, in this study we exploited the insight in micropylar endosperm region specific transcriptional factor genes. We hypothesized that “the tissue specific transcription factor genes in the micropylar endosperm region during seed development and germination may regulate the seed germination behavior and dormancy release”.

Firstly, we identified transcription factors from microarray data of tissue specific expression at micropylar end endosperm during seed germination and confirmed their expression using publicly available data set of seed development. Secondly, target genes’s promoter::reporter gene contructs were started to make for localization. Finally, the T-DNA lines of these genes were ordered to study the functions of the genes. *WRKY8* was investigated that it is involved in the ABA pathway during seed germination by germination assay under ABA stress. Under osmotic and salt stress, *WRKY8* loss-of-function also influenced germination performance of the mutants. However, more replications are needed in future to confirm the result.

Reference

Bethke, P. C., et al.
2007 The arabidopsis aleurone layer responds to nitric oxide, gibberellin, and abscisic acid and is sufficient and necessary for seed dormancy. Plant Physiol 143(3):1173-1188.

Bewley, J. D.
1997 Seed germination and dormancy. Plant Cell 9(7):1055-1066.

Devaiah, B. N., A. S. Karthikeyan, and K. G. Raghothama
2007 WRKY75 transcription factor is a modulator of phosphate acquisition and root development in Arabidopsis. Plant Physiol 143(4):1789-801.

Dubreucq, B., et al.
2000 The Arabidopsis AtEPR1 extensin-like gene is specifically expressed in endosperm during seed germination. Plant Journal 23(5):643-652.

Finch-Savage, W. E., and G. Leubner-Metzger
2006 Seed dormancy and the control of germination. New Phytol 171(3):501-23.

Grunewald, W., et al.
2008 A role for AtWRKY23 in feeding site establishment of plant-parasitic nematodes. Plant Physiol 148(1):358-68.

Gutierrez, L., et al.
2007 Combined networks regulating seed maturation. Trends Plant Sci 12(7):294-300.

Holdsworth, M. J., L. Bentsink, and W. J. Soppe
2008 Molecular networks regulating Arabidopsis seed maturation, after-ripening, dormancy and germination. New Phytol 179(1):33-54.

Jefferson, R. A., T. A. Kavanagh, and M. W. Bevan
1987 GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO Journal 6(13):3901-3907.

Joosen, R. V. L., et al.
2010 Germinator: A software package for high-throughput scoring and curve fitting of Arabidopsis seed germination. Plant Journal 62(1):148-159.

Koornneef, M., L. Bentsink, and H. Hilhorst
2002 Seed dormancy and germination. Curr Opin Plant Biol 5(1):33-36.

Krysan, P. J., J. C. Young, and M. R. Sussman
1999 T-DNA as an insertional mutagen in Arabidopsis. Plant Cell 11(12):2283-2290.

Kucera, Birgit, Marc Alan Cohn, and Gerhard Leubner-Metzger
2005 Plant hormone interactions during seed dormancy release and germination. Seed Science Research 15(4):281-307.

Li, S., et al.
2011 Arabidopsis thaliana WRKY25, WRKY26, and WRKY33 coordinate induction of plant thermotolerance. Planta 233(6):1237-1252.

Liu, P. P., et al.
2005a Large-scale screening of Arabidopsis enhancer-trap lines for seed germination-associated genes. Plant J 41(6):936-44.

Liu, P. P., et al.

2005b The BME3 (Blue Micropylar End 3) GATA zinc finger transcription factor is a positive regulator of Arabidopsis seed germination. Plant J 44(6):960-71.

Martin, A. C.

1946 The Comparative Internal Morphology of Seeds. American Midland Naturalist 36(3):513-660.

Mitsuda, N., and M. Ohme-Takagi

2009 Functional analysis of transcription factors in arabidopsis. Plant and Cell Physiology 50(7):1232-1248.

Penfield, S., et al.

2006 Arabidopsis ABA INSENSITIVE4 regulates lipid mobilization in the embryo and reveals repression of seed germination by the endosperm. Plant Cell 18(8):1887-1899.

Rushton, P. J., et al.

2010 WRKY transcription factors. Trends Plant Sci 15(5):247-58.

Sun, X., et al.

2010 Transcriptional and hormonal signaling control of Arabidopsis seed development. Curr Opin Plant Biol 13(5):611-20.

van Verk, M. C., J. F. Bol, and H. J. Linthorst

2011 WRKY transcription factors involved in activation of SA biosynthesis genes. BMC Plant Biol 11:89.

Winter, D., et al.

2007 An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. PloS one 2(1).

Zhu, J. K.

2002 Salt and drought stress signal transduction in plants. Annu Rev Plant Biol 53:247-73.

Appendix

1. Gene list in each study

Gene name	At number	WP1	WP2	WP3
WOX9/STIMPY	At2g33880	Choose from transcriptional experiment during first phase seed germination, and confirmed with public dataset (during last phase of seed development)		No available (homozygous) T-DNA lines
BS1, TMO7, ATBS1	At1g74500			No effect at transcriptome level (descript in paper)
AtWRKY8	At5g46350			With available (homozygous) T-DNA lines
AtWRKY28	At4g18170			
AtWRKY75	AT5g13080			
LBD18	AT2g45420	Choose from public dataset, gene express specially high in ME during last phase of seed development	Already available	No available (homozygous) T-DNA lines
AtWRKY23	At2g47260			With available (homozygous) T-DNA lines
AtWRKY25	AT2g30250			No available (homozygous) T-DNA lines
AGP2	At2g22470		Choose from transcription experiment, which genes express in high level at certain point (25h) during germination in ME of ruptured seeds, compared to non-ruptured seeds.	
SKS9	At4g38420			
AGP21	At1g55330			

2. Confirm tissue specific expression

2-1 Getaway primer design

Gene name	Forward primer	Forward Tm	Reverse primer	Reverse Tm
At1g55330	GGGGACAAGTTTGTACAAAAAAGCAGGCTAAATTCTTGGCTATAAGTATTTCTGTC	60.28	GGGGACCACTTTGTACAAGAAAGCTGGGTTAATTCTCCTTTGATGTCTTGAAATAAG	62.74
At2g22470	GGGGACAAGTTTGTACAAAAAAGCAGGCTGAGTGATAGATTGGAGATAAATCATAC	59.38	GGGGACCACTTTGTACAAGAAAGCTGGGTTTTAGATTATCTCTTCTTTGTCTGAATCT	61.17
At2g47260	GGGGACAAGTTTGTACAAAAAAGCAGGCTCAAGGGAAGGAGGGAGTAAC	61.9	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTCTTCTCTCGCTGAGCTTTAC	62.92
At4g18170	GGGGACAAGTTTGTACAAAAAAGCAGGCTAAGTAACGATCCACTCCACATAG	61.28	GGGGACCACTTTGTACAAGAAAGCTGGGTAAAGAAGACAAAGAAGAAGAGATGAG	61.62
At5g46350	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCTGAGTTGCTTTAATTTC	65.34	GGGGACCACTTTGTACAAGAAAGCTGGGTGACGAAGAACAAAGAGAAAAAACTTAAACC	66.8
At4g38420	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCTCAAGAAGAAACCCGAAGATG	65.74	GGGGACCACTTTGTACAAGAAAGCTGGGTACTCGCTTTGACAAGAAGAAGA	62.46
attB1 site for: GGGGACAAGTTTGTACAAAAAAGCAGGCT attB2 site rev: GGGGACCACTTTGTACAAGAAAGCTGGGT				

2-2 DNA amplification by high fertility polymerase (PFX)

10X AccuPrime PFX Reaction mix	2.5 µl
Primer Forward	0.75µl
Primer Reverse	0.75µl
Template DNA	1µl
AccuPrime PFX DNA Polymerase	0.5µl
Autoclaved, distilled water	19.5µl
Total	25µl

2-3 BP reaction and transforming competent cells

Reaction:

PCR product (40ng)	2µl
pDONR 207 vector (150 ng/µl)	2µl
BP clonase and buffer	1µl
Total	5µl

After put the tube at 25°C for 1 hour, add Poteinase 0.5µl and put the tube in the 37°C oven for 10 min.

Medium preparation:
Prepare 50 mL LB (Luria-Bertani) medium with 1.5% agar containing 1000 times dilution of Gentamysin stocks solution for selection. LB medium contains per liter: tryptone 10g, yeast extract 5g and Nacl 10g. PH value of the LB medium was adjusted to 7.0.
Transforming competent cells:
50µl competent cells with 1µl BP reaction was treated with electrophoresis at 2.5 kV for several seconds. Afterwards, 450µl of SOC medium was added immediately and the cells were put at 37 °C for 1 hour.
75µl of the competent cells after treated were transfer to the medium for low efficiency treatment, and the other left cells were votexed and centrifuged and then transferred for high efficiency treatment. After store the plates containing competent cells on the medium for one night at 37 °C, growing colonies can be observed.

2-4 Colony PCR

Autoclaved, distilled water	9 µl
$MgCl_2$	1.25µl
Buffer	1.25µl
M13 Forward	0.25µl
Gene specific primer Reverse	0.25µl
dNTP	0.25µl
Firepoll Polymerase	0.25µl
Template DNA	Use wood stick to dip the target colony into the reaction
Total	12.5µl

2-5 GUS expression pattern test

Solutions:
Rinse solution: 50 mM NaPO4, pH 7.2 (68% Na2HPO4 en 31.6% NaH2PO4); 0.5 mMK3Fe(CN)6; 0.5 mMK4Fe(CN)6
X-gluc staining solution: 50 mM NaPO4, pH 7.2 (68.4% Na2HPO4 en 31.6%; NaH2PO4); 0.5 mMK3Fe(CN)6; 0.5 mMK4Fe(CN)6; 2 mM X-gluc

Experiment:
Day 1:

- For each treatment: put the seedlings in a well of a 24 wells-plate filled with 1 ml ice-cold acetone 90%
- Incubate in acetone for 20 minutes.
- Remove the acetone (try not to damage the seedlings!) and rinse with 2x 1 ml rinse solution.
- Remove the rinse solution.
- Add 500 µl GUS staining solution (X-gluc) making sure all seedlings are completely covered by liquid
- Vacuum infiltrate twice for 30 seconds.
- Wrap the plates in aluminium-foil and incubate at 37^{O}C for 24 hours.

Day 2: Scoring GUS activity

- Remove the X-gluc staining solution
- Rinse twice with water
- Add 1 ml 30% ethanol to fix the tissue and to start with removing chlorophyll
- Incubate for 10 min. and remove solution
- Add 1 ml 70% ethanol
- Incubate for 10 min. and remove solution.
- Add 1 ml 96% ethanol
- Incubate for 10 min. and remove solution.
- add some water to untangle the seedlings for easy viewing.

3. Identification of gene function using T-DNA lines

a) T-DNA lines ordering list

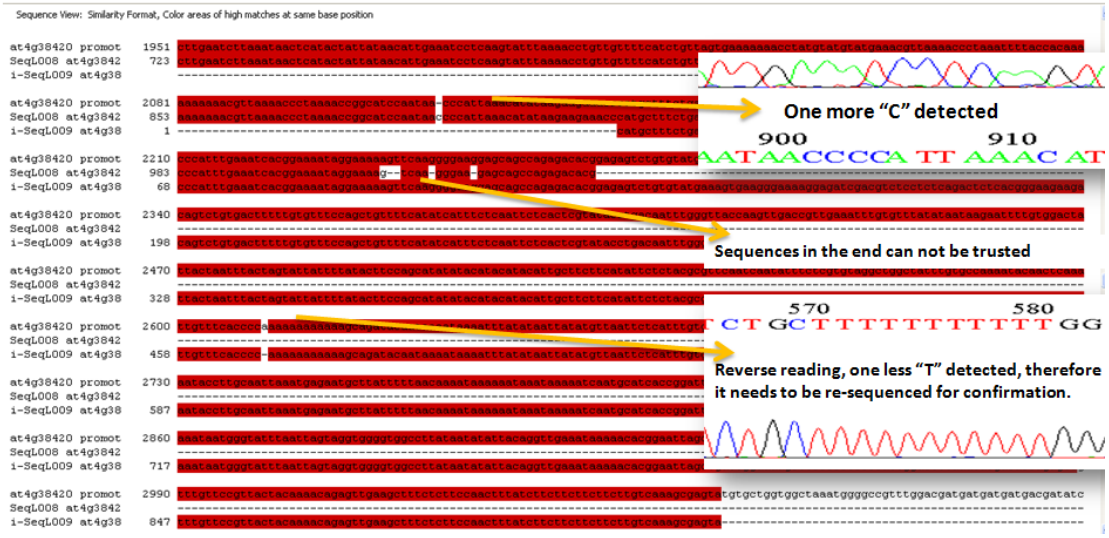
	Gene name	At nr.	T-DNA lines	NASC ID	Background	Availability	Segregation status
A	BS1, TMO7, ATBS1	At1g74500	SALK_058700 (K)	N558700	Col-0	yes	segregating
B	BS1, TMO7, ATBS1	At1g74500	SALK_071814 (Z) (AN)	N571814	Col-0	yes	segregating
C	BS1, TMO7, ATBS1	At1g74500	SALK_080003 (BM)	N580003	Col-0	yes	segregating
D	AtWRKY8	At5g46350	GK-625G10	N313365	Col-0	yes and 12 more	all unknown
E	AtWRKY8	At5g46350	SALK_107668C	N666984	Col-0	yes	homozygous
F	AtWRKY8	At5g46350	SALK_064419C	N686131	Col-0	yes	homozygous
G	AtWRKY23	At2g47260	SALK_006855C	N685132	Col-1	yes	homozygous
H	AtWRKY23	At2g47260	GK-125C01	N317435	Col	yes and 12 more	all unknown
I	AtWRKY75	AT5g13080	SALK_042481C	N674673	Col	yes	homozygous

b) PCR pipetting scheme

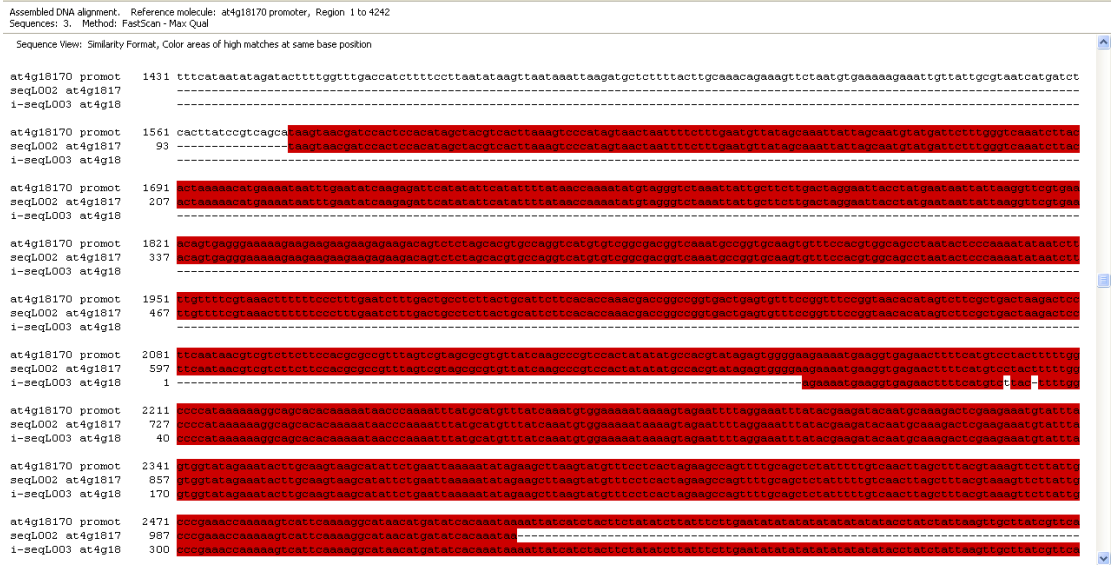
Autoclaved, distilled water	7 µl
$MgCl_2$	1.25µl
Buffer	1.25µl
M13 Forward	0.25µl
Gene specific primer Reverse	0.25µl
dNTP	0.25µl
Firepoll Polymerase	0.25µl
Template DNA	2 µl
Total	12.5µl

4. Result and discussion

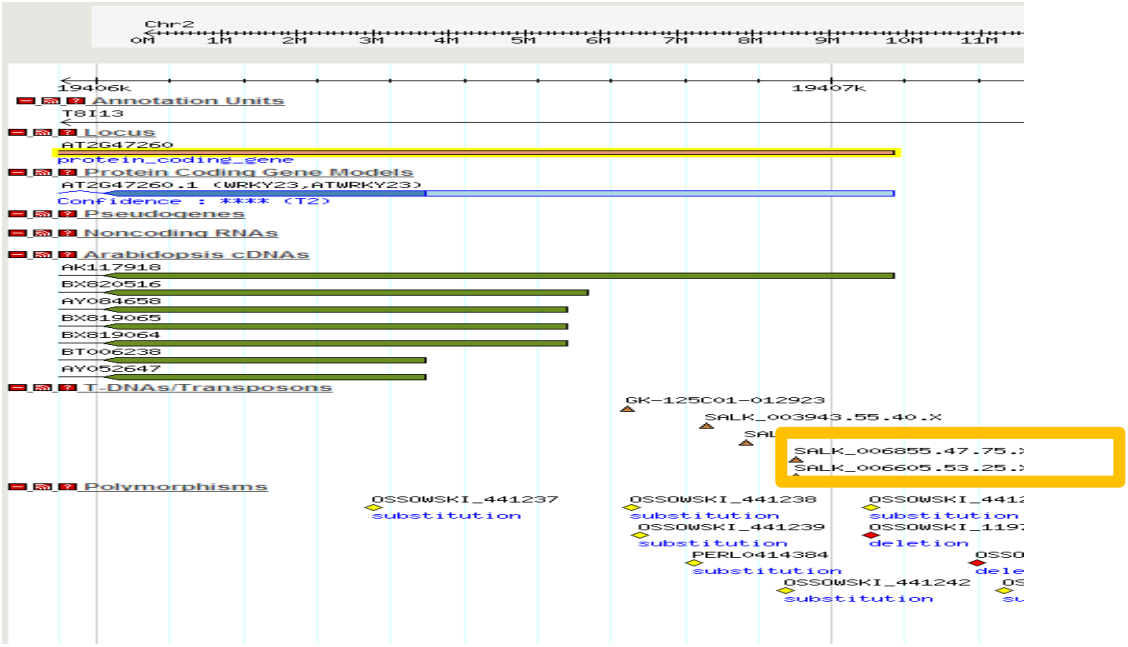
4-1 Gene sequence result of At4g38420 (*SKS9*) promoter region, aligning with the wild type Colombia sequence. At around position 905, reads from the machine (4“C”s) did not match with the peak (3“C”s). Misdetection at the end of a sequence cannot be trusted. At around position 570, one more “T” was detected.



4-2 Gene sequence result of At4g18170 (28) promoter region, aligning with the wild type Colombia sequence, with only two errors at the end of the sequence, which could be ignored.



4-3 Partial *WRKY23* gene structure and position of the T insertions (pointed in yellow square) (TAIR, 2011)



4-4 Partial *WRKY 8* gene structure and position of the T insertions (pointed in yellow square) (TAIR, 2011)